A Case-Based Reasoning Framework Augmented with Causal Graph Bayesian Networks for Multi-Hazard Assessment of Earthquake Impacts

Yiding Dou[1,†], Jiaming Zhang[1,†], Yuxin Li[1], Ruyi Qi[1], Zimeng Yuan[1], Yanbing Bai[1,*], Erick Mas[2] and Shunichi Koshimura[2]

[1]Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing BJ 100872, China
[2]International Research Institute of Disaster Science, Tohoku University, Sendai 980-8572, Japan

Abstract

Earthquakes often lead to significant secondary hazards such as landslides, liquefaction, and aftershocks, which in turn cause great damage to buildings and seriously jeopardize socio-ecological welfare. The prevailing models for post-earthquake damage assessment predominantly utilize deep learning methods and InSAR-based Damage Proxy Maps. However, these approaches require data of high quality, with both multi-temporal and spatiotemporal resolution, and are heavily reliant on supervised learning, limiting their applicability on a broader scale. This paper presents a Case-Based Reasoning Framework Augmented with Causal Graph Bayesian Networks for Multi-Hazard Impact Assessment. This method demonstrates strong adaptability to noisy data, making it an innovative tool in the field of earthquake damage estimation. We applied this framework to analyze the catastrophic earthquakes that struck Turkey and Japan in 2023 and 2024, respectively, using them as bases for our case-based reasoning process. For the Turkey case, our model achieved a precision of 99.9%, a recall of 40.2%, and an F1 score of 57.4% in detecting landslides—significantly surpassing the performance of the USGS a priori model. In detecting liquefaction, the model showed a recall of 95.9% and an F1 score of 70.6%, both substantial improvements over the preliminary model. For the 2024 Noto Peninsula earthquake, our method enhanced the Area Under the Curve (AUC) index from 0.73 to 0.77, further validating the effectiveness of our approach. This study offers a highly precise, scalable, and unsupervised learning method for estimating earthquake disaster damage, providing a valuable asset for optimizing post-disaster resource allocation, reducing economic losses and accurately repairing the environment.

Keywords

Case-Based Reasoning, Causal Graph Bayesian Networks, Multi-Hazard Assessment, Earthquake

1. Introduction

Earthquakes are an unpredictable and devastating disaster that causes significant damage, economic loss, and risk to life, posing a great challenge to socioeconomic welfare [1]. Diverse secondary hazards such as landslides, liquefaction, and aftershocks also cause huge building damage, and responding to this disaster places high demands on rescue and reconstruction capabilities [2]. For example, the devastating Kahramanmaras earthquake sequence occurred in February 2023 in the Turkey-Syria seismic belt, affecting 11 cities in Turkey and shaking an area of about 90,000 square kilometers. The earthquake triggered massive landslides and liquefaction [3] and building damage [4]. Therefore, accurate and efficient assessment of various disasters caused by earthquakes is very important for post-disaster relief resource provision and social reconstruction. With the release and popularization of open-source satellite data, InSAR-based damage mapping [5, 6, 7, 8] and machine learning-driven damage mapping approaches [9, 10, 11, 12, 13, 14, 15] have played an important role in estimating building damages. However, there are three main problems. Firstly,the models mostly rely on supervised or semi-supervised learning [16] based on ground truth data and are difficult to realize fast response.

ICCBR AI Track'24: Special Track on AI for Socio-Ecological Welfare at ICCBR2024, July 1, 2024, Mérida, Mexico
*Corresponding author.
†These authors contributed equally.
douyiding@ruc.edu.cn (Y. Dou); zhangjiaming2002@ruc.edu.cn (J. Zhang); liyuxin1129@ruc.edu.cn (Y. Li); 2021201582@ruc.edu.cn (R. Qi); 2022201542@ruc.edu.cn (Z. Yuan); ybbai@ruc.edu.cn (Y. Bai); mas@irides.tohoku.ac.jp (E. Mas); koshimura@irides.tohoku.ac.jp (S. Koshimura)

Secondly, existing models usually use the traditional supervised learning mode for model training, i.e., they need a sufficient amount of ground truth data in the affected area as labels. The unavailability of ground truth data immediately after an earthquake creates a serious obstacle to the fast response of the relevant models. Thirdly, previous approaches were overly dependent on high-quality or multi-temporal data [17, 18]. When applying these methods to noisy data in disaster scenarios, they lack generalizability or exhibit significantly reduced accuracy; Despite this, most existing post-earthquake damage assessment models are not interpretable as they ignore multiple hazards and the complex causal relationships between impact processes. The latest research progress shows that causal Bayesian networks are very effective for estimating cascading disasters [19, 20, 21].

To address the above three problems in the estimation of newly occurring large earthquakes, we introduce the Case-Based Reasoning (CBR) [22, 23, 24] approach. Case-Based Reasoning (CBR) is an artificial intelligence [25] technique that solves current problems by utilizing past cases. In the retrieval session, we retrieved a large number of cases of using causal Bayesian networks to study the complex correlation of nodes in various fields, especially disaster prediction, and cases of using high-precision satellite images to study earthquake hazards, with the expectation of finding the most similar cases to the current earthquake prediction for reuse. Finally, we utilized a causal Bayesian network model [2] learned from Damage Proxy Maps and a priori data generated based on high-resolution satellite imagery to quantitatively model the causal dependencies behind typical secondary hazard geologic processes and various a priori information, to model the rapid estimation of post-earthquake building damage as well as landslides and liquefaction. In this paper, we reuse the causal Bayesian network model to the 2023 Turkey earthquake sequence and the 2024 Noto Peninsula earthquake, and the parameters of the Bayesian network were adjusted and optimized in context, realizing the revision of the original case on a new earthquake prediction problem. Our causal Bayesian network can quantify the causal effect of parent nodes on child nodes, output the causal coefficients between secondary hazards and seismic damages, and approximate the posterior probabilities such as building damage using variational inference.

The model has three main stages:

Firstly, we collected a large amount of raw data from relevant earthquakes on open-source websites and uniformly converted them into raster data that can be accurately input into a causal Bayesian network for comprehensive validation.

Secondly, we defined conditional probabilities, for each pair of parent and child nodes in the causal graph, defining the conditional probability relationship between them, and subsequently investigated the introduction of a variational inference algorithm into the Bayesian network to estimate unobserved intermediate seismic hazards, building damages, and quantitative causal dependencies between them. In addition, the system can flexibly integrate available building footprint information provided by OpenStreetMap or Microsoft to further improve the inference performance of the whole graph [2].

Finally, by inputting the collated data from the Turkey and Japan earthquakes and using the expectation maximization (EM) algorithm to optimize the model parameters in the causal map, including the causal coefficients and the a posteriori probabilities of the unobserved variables, and by adjusting the hyper-parameters in the network with the characteristics of the actual seismic region, the relevant causal coefficients computed based on the sequence of this earthquake can be obtained and more accurate landslide, liquefaction, and house damage probability.

The main contributions of this paper are as follows:

Theoretical level 1. This study is based on the causal Bayesian network model [2], introduces case-based reasoning theory, and constructs an unsupervised causal graph Bayesian network case-based reasoning earthquake multihazard assessment framework. This method is crucial for early earthquake disaster grasp and rapid post-disaster response.

2. This study selected hot cases such as the 2023 Turkey Earthquake and the 2024 Japan Noto Peninsula Earthquake to conduct model robustness and transferability analysis, verifying the effectiveness of the method.

Socio-ecological welfare level 1. Our rapid prediction model can optimize resource allocation, reduce economic damage, and safeguard life safety. Rapid and accurate post-disaster damage prediction can improve the efficiency of disaster avoidance and timely relocation of people and valuable belongings. It can also help rescue teams prioritize resource allocation to save more lives and reduce injuries.
2. By accurately mapping the impacts of second disasters, the model allows environmental agencies to better assess the extent of ecological damage and plan the necessary environmental remediation measures, such as stabilizing soils, to facilitate ecosystem recovery. Meanwhile, the model profoundly reveals the process of geological changes in earthquakes, contributing to further ecological research.
3. Model-assisted timely rescue and recovery can maintain social stability. The detailed damage reports provided by the system can help organizations to formulate more effective recovery and reconstruction plans, so as to quickly restore social order and public confidence and reduce social unrest.

2. Related Work

Current methods for assessing earthquake damage predominantly encompass two strategies. The first leverages multi-temporal radar interferometry to create damage proxy maps (DPMs), which are instrumental in detecting and annotating changes within remote sensing images of areas affected by disasters. This approach offers a rapid estimation of earthquake impacts, exemplified by Yun & Sang-Ho et al.(2015) [5], who utilized the Interferometric Synthetic Aperture Radar (InSAR) coherence model to assess the 2015 Nepal earthquake's aftermath. The second strategy involves assessing building damage through deep learning models applied to remote sensing data, as demonstrated by Bai et al.(2018,2022)in evaluating the building damage from the 2011 Great East Japan Earthquake using post-disaster TerraSAR data [10] and Xview2 Challenge [26]. However, these conventional methods demand high-quality, multi-temporal, and spatiotemporal resolution data, with the latter largely dependent on supervised learning from extensively annotated datasets, which restricts their widespread adoption.

In pursuit of a more universally applicable approach for large-scale disaster damage assessment, Xu [2] integrated causal inference with remote sensing to refine the accuracy of predicting earthquake-triggered landslides, liquefaction, and structural damages. This method inputs DPMs alongside pre-existing data on landslides, liquefaction, and building damage footprints to model causal relationships among these hazards, thereby facilitating efficient seismic event assessments across various countries and regions without the need for ground truth data. Despite its promising results, the robustness and transfer ability of this method await further validation.

Case-based reasoning (CBR) methods, especially those grounded in geographic space, have recently made significant strides in disaster damage assessment [27, 28, 29, 30, 31]. For instance, Zhao et al.(2021)capitalized on spatial proximity features to develop a spatial CBR approach for regional landslide risk evaluation [27], while Wang et al.(2022)introduced a CBR framework to forecast the spatial distribution of economic losses from typhoon storm surges [28]. The resilience and adaptability of these CBR-based assessment methods still require confirmation.

The catastrophic earthquakes in Turkey and Japan, occurring in 2023 and 2024 respectively, present a unique opportunity to refine and apply the space-based CBR method for a comprehensive assessment of complex earthquake hazards. Drawing inspiration from this research trajectory, we aim to enhance Xu's model [2] and adapt it to the recent earthquakes in Turkey and Japan, to estimate a broader spectrum of earthquake-induced hazards.

3. Application of CBR Model

CBR (Case-Based Reasoning) is a problem-solving methodology that leverages past experiences or cases to solve new problems. It operates on the principle that similar problems tend to have similar solutions.

3.1. Unsolved Case Representation

3.1.1. Dataset

We mainly selected the 7.8 magnitude earthquake that occurred 23 kilometers east of Nodagi, Gaziantep Province, Turkey in 2023 and the 7.6 magnitude earthquake that occurred in Noto Peninsula, Japan in 2024 for our research and experiments. Building footprint, DPM, Landslides, and Liquefaction, are required to construct the causal Bayesian network. On the open web data platform, we collected a quantity of raw data and constructed it as an original dataset.

DPMs DPMs(Damage Proxy Maps) are tools developed by the NASA Advanced Rapid Imaging and Analysis (ARIA) team for rapidly assessing the extent of damage after disasters. They utilize satellite remote sensing data, particularly radar interferometry data, and optical imagery, to identify and quantify surface damage caused by earthquakes. We collected DPMs of the relevant earthquakes on the official ARIA website. Among them, the images of the Turkey earthquake cover part of the seismic region (35.51°N - 38.97°N, 35.58°E - 38.80°E) and are generated at a resolution of 2.8E-4 (latitude/longitude) resolution was generated. Seismic images of the Noto Peninsula in Japan covering part of the area (36.31°N-37.57°N, 136,24°E-137.36°E) were generated at the same resolution.

Landslides and Liquefaction Landslide Maps and Liquefaction Maps are raster data generated by the USGS using known geographic features and ground shaking data, which consisted on Bayesian networks as prior data. We collected landslide and liquefaction image data of the related earthquakes on the official website of USGS (earthquake.usgs.gov). In Turkey, Landslide Maps and Liquefaction Maps were generated for some areas (35.06°N -37.07°N:34.47°E -37.15°E) after the earthquake with an accuracy of 16.0 km width of the real area covered by each square pixel element. In Japan, Landslide Maps were generated for some areas (34.75°N - 40.13°N:133.88°E - 140.80°E) with a resolution of 0.002083333, respectively, Landslide maps and liquefaction maps were generated for some areas (34.75°N - 40.13°N, 133.88°E - 140.80°E) in the Noto peninsula region at a resolution of 4.2E-4 (latitude/longitude).

Building Footprints Building footprints refer to the outlines or shapes of buildings projected onto the ground. We obtained building footprint images for part of the affected area from the HDX website (data.humdata.org), whose data comes from OSM.

Ground Truth We have collected real earthquake data as test comparisons from the Japan Institute of Land and Geography, GitHub, and other online platforms, including real liquefaction maps of the ground, landslide maps, and building damage maps in the aftermath of the earthquake.

3.1.2. Processing

The building footprint, DPM, Landslides, and Liquefaction image data inputted into the Bayesian network must have the same spatial coverage area and precision. Of the raw data, Building Footprints images are in shapefile format and the rest are raster images in tif format. Only when each pixel of the four image datasets is aligned, can the corresponding information on buildings, landslides, liquefaction, etc., be accurately fed into the causal Bayesian network for prediction. Therefore, we primarily conducted a three-step preprocessing on the original dataset:

Building Footprints Labeling The original dataset's building footprint image data was in shapefile format, requiring conversion to raster data for input into the Bayesian network. Firstly, we used ArcGIS to convert the building footprint image to raster data. Then, we assigned values to each pixel in the data, converting it into a binary variable ranging from 0 to 1. Pixels with buildings were marked as 1, while those without were marked as 0.

Image Cropping The spatial ranges of different image data in the original dataset varied. The over-

lapping spatial portion was deemed the effective study area. Thus, we utilized ArcGIS for cropping to ensure that the input building footprint, DPM, Landslides, and Liquefaction image data were of consistent size and spatial overlap.

Precision Alignment (Resampling) The input image data required consistent precision, yet the original dataset's precision differed among building footprints, landslides, and liquefaction images compared to DPMs, which were coarser. Using the pixel size of DPMs as the reference, we divided the pixel granularity of the other three image datasets to match DPMs' pixel size. Finally, the precision of the four image datasets was unified to 0.00027777778 degrees latitude/longitude. Thus, we obtained the final dataset, as shown in Table 1. The four image datasets used for model training are unlabeled. Building footprints are binary variables ranging from 0 to 1, while the remaining three datasets are continuous variables. As the degree of damage/liquefaction/landslide increases, the corresponding value approaches 1.

3.2. Case retrieval

In our problem, Earthquakes can directly trigger secondary disasters such as landslides and liquefaction, which may lead to structural damage and casualties. Concurrently, these secondary disasters, including landslides and liquefaction, can also cause injury to buildings. The complexity is further exacerbated by environmental noise and varying geographical conditions, which complicate the causal relationships between different events. In such intricate scenarios, the Bayesian causal network demonstrates its formidable problem-solving capabilities by parsing the intricate causal relationships among a multitude of variables within a vast dataset.

The Causal Bayesian Network [32, 33, 34] is a directed acyclic graphical model of probabilities, where each node represents a variable, and the network is utilized to delineate the causal relationships among these variables. Currently, Causal Bayesian Inference is extensively applied across various domains. Lijing Wang et al.(2022) explored the application of Bayesian Network (BN) structure learning algorithms in the integration of machine learning and causal knowledge and compared the differences in graphical structure and data interpretation capabilities between purely machine-learned BNs and knowledge-based BNs [35]. Anthony C et al.(2018) proposed the CausalGNN method, which combines graph embedding and causal modules, utilizing graph-based nonlinear transformations to learn spatiotemporal embeddings, and providing epidemiological context through ordinary differential equations, thereby surpassing various baseline models in predicting daily new cases of COVID-19 [36]. Meghamala Sinha et al.(2021) introduced a method named Kg2Causal, which leverages a large-scale, general-purpose biomedical knowledge graph as prior knowledge to enhance the performance of data-driven causal network learning [37].

3.3. Case reuse

We used a model based on the one put forward by Xu et al.(2022) [2]. The model is a causal Bayesian network tailored for seismic hazard analysis, aiming to understand and predict the impacts of earthquakes by considering various factors and their interdependencies. In our context, integrating DPM information for the assessment of secondary seismic disasters, particularly building damage, is analogous to disease diagnosis and prognosis, making the application of Bayesian Causal Networks highly appropriate here.

In our research, We engage with a multitude of stochastic variables, focusing on discerning the causal relationships between the existing coarse-grained landslide and liquefaction probability distributions, which are rich in information and environmental noise, and the joint secondary disasters of landslides, liquefaction, and building damage. This is achieved by leveraging Damage Proxy Maps (DPMs) and local geographical context, thereby synthesizing a joint probability distribution for secondary disasters. Our model incorporates landslide and liquefaction probability maps from the USGS team, DPMs from the ARIA team, and building footprints from OpenStreetMap as inputs. The model then synthesizes and outputs posterior probability maps for landslides, liquefaction, and building damage, as well as causal

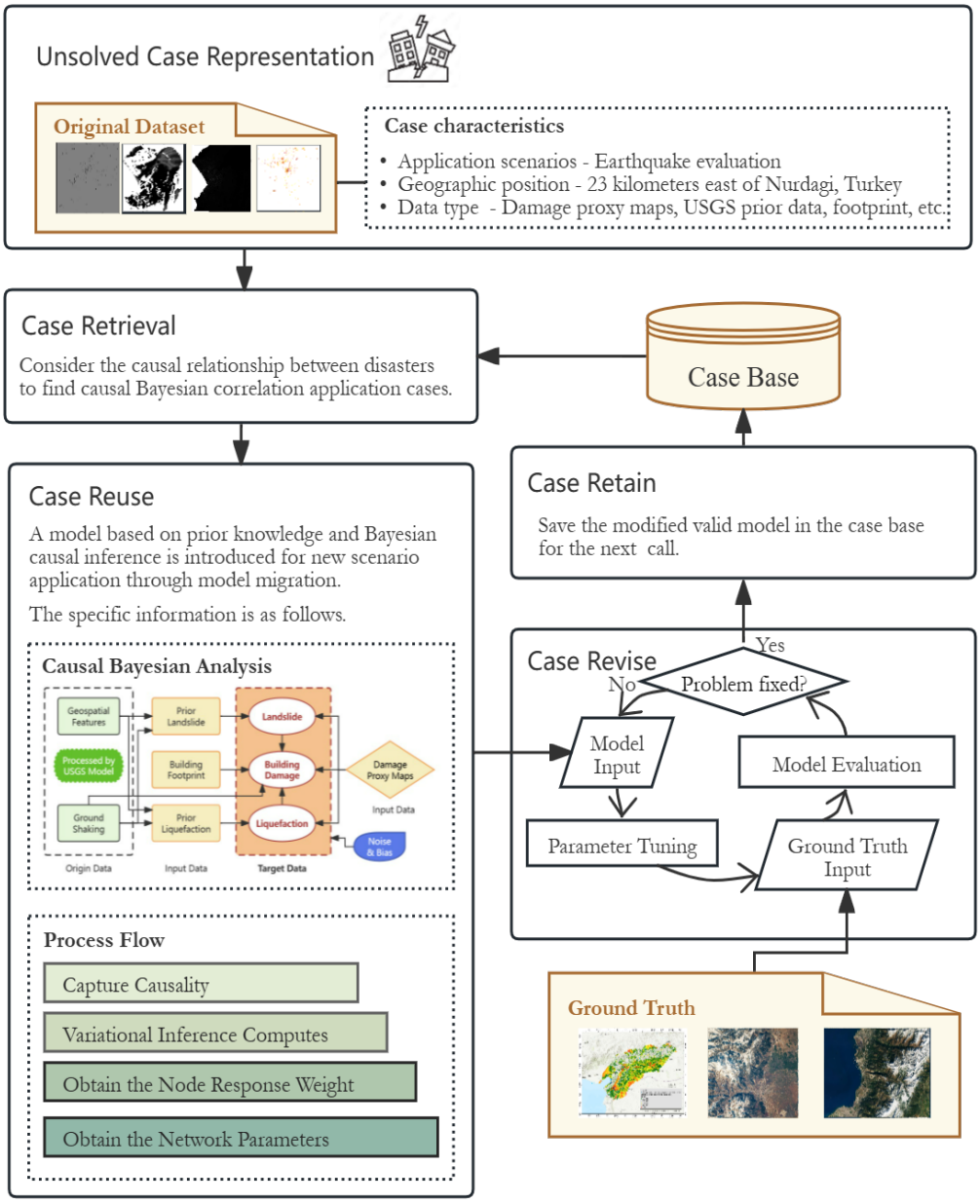

Figure 1: The overview of the case training process in CBR. After case input, the process of case retrieval, case reuse (causal Bayesian variational inference is used here), case revise, and case retain outputs the results and enriches the case base.

Table 1
Final Data Description.

Data Name	Data Type	Spatial Range		Resolution (degrees)
		Turkey	Japan	
Building footprints	Binary (0 or 1)			
DPMs	Continuous (0-1)	35.77°N—36.56°N,	37.37°N-37.41°N,	0.0000278
Landslides	Continuous (0-1)	35.68°E—36.50°E	136.77°E-136.94°E	
Liquefaction	Continuous (0-1)			

coefficients between them. The model framework is shown in Figure 2. Building upon this foundation, our model is capable of elucidating complex causal relationships to demystify the mechanisms underlying the occurrence of secondary disasters accompanying earthquakes. It efficiently assesses the primary causes of building damage in specific seismic events and facilitates rapid predictions, all predicated on a rich dataset.

In the specific expression of the formula, We let i denote the impact nodes, which include DPMs (Damage Proxy Maps), LS (Landslides), LF (Liquefaction), and BD (Building Damage). We denote x_i as the binary variable representing the activity and inactivity of the unobserved ground damage and impact nodes, where i can represent LS, LF, BD here, $i \in \{1, 2, 3\}$and is a binary variable with $x_i \in \{0, 1\}$. Outside of the aforementioned definition, a bias node x_0 is defined as $x_0 = 1$, indicating that it is always active regardless of the activity of its parent nodes, to address certain anomalies, such as instances where building damage occurs despite the inactivity of parent nodes like landslides and liquefaction. The parent nodes k of the child node i are denoted as w_{ki}, where DPMs are denoted separately as y. With the random ϵ_y disturbance term following a normal distribution, then $y|p(i)$ follows a log-normal distribution. When examining the causal influence relationship between parent and child nodes, the causal influence coefficient of parent node k on child node i is defined as $p(i)$, where satisfies [2]:

$$w_{ki} = \begin{cases} \theta_k & \text{if } i \rightarrow LS, \\ \mu_k & \text{if } i \rightarrow LF, \\ \phi_k & \text{if } i \rightarrow BD, \\ \eta_k & \text{if } i \rightarrow DPMs \end{cases} \tag{1}$$

Based on the above expressions, the Bayesian causal network model is formulated as:

$$p(X_i = 1|X_{p(i)}, E) = W_{oi} + W_{e,i} + \sum_{k \in P(i)} W_{ky} X_k \tag{2}$$

$$log(Y_i) = W_{ey} + W_{oy} + \sum_{k \in P(i)} W_{ik} X_i \tag{3}$$

3.4. Case revise

To achieve the adaptive application of the model in different cases, reasonable parameter estimation and optimization adjustments are required. We iterate and optimize the model parameters by applying the model to existing earthquake cases and comparing them with real data. Furthermore, in the practical application of the model, various adjustments and solutions have been implemented to address specific challenges. As previously mentioned, the intricate causal inferences involve a multitude of factors, including secondary disasters, environmental conditions, and geographical elements. Relying solely on traditional Bayesian Causal Network inference to predict the posterior distributions of secondary disasters is insufficient, as these algorithms may struggle to capture and represent the complexity, particularly when nonlinear relationships among variables are present. Furthermore, the model's Bayesian Causal Network encompasses a variety of variable types, such as binary, log-normal, and logit-normal, which poses additional challenges for traditional Bayesian algorithms. Moreover, the traditional Bayesian network operates under the assumption of conditional independence among variables, an assumption that is violated in our scenario where secondary disasters exhibit complex and interdependent causal relationships. To address these issues, the model employs a variational inference algorithm that estimates the posterior distributions by maximizing the marginal likelihood of observed variables, a technique proven effective for complex probabilistic models with numerous unobserved variables. By maximizing the variational lower bound, the algorithm optimizes the posterior distributions of location-specific multi-hazard and impact variables along with their causal dependencies, providing a theoretical guarantee for the optimality of joint posterior inference. Specifically, for each

location , we $l$define a variational distribution $q(x^l)$ to further decompose the unobserved variables:

$$q(x^l) = \prod_i q(x_i^l) = \prod_i (q_i^l)^{x_i^l}(1-q_i^l)^{1-x_i^l} \tag{4}$$

By maximizing the lower bound, we express the optimal combination of expected posterior estimation and causal relevance estimation. Define a nonlinear function:

$$\begin{aligned} F(.) = & \sum_{k\in\, p(i)} \{(-1)^z(w_{0i} + w_{ki} + \frac{w_{\epsilon_i}^2}{2}) + (-1)^z(1-q_k)f((-1)^z w_{0i} + \frac{w_{\epsilon_i}^2}{2})\} \\ & + \sum_{\substack{j\in\, c(i) \\ i\in\, \{1,2,3\}}} \{(-1)^z q^j f((-1)^z(w_{0i}+w_{ij}) + \frac{w_{\epsilon_j}^2}{2}) + (-1)^z(1-q^j f((-1)^z w_{0i} + \frac{w_{\epsilon_j}^2}{2})\} \\ & - \frac{2\sum_{m\in\,(i,y)} w_{iy}w_{my}q_m}{2w_{\epsilon_y}^2} + \frac{(2logy - 2w_{0y} - w_{iy}^2)w_(iy)}{2w_{\epsilon_y}^2} \end{aligned} \tag{5}$$

where $f(x)$,v , and p(i) represent the parent nodes, c(i) represents the child nodes, and s(i) represents the set of spouse nodes sharing the same child nodes. Based on this, we provide the optimal posterior estimation [2].

Recognizing that complex causal relationships do not significantly manifest in all regions, the model leverages this insight to enhance computational efficiency, particularly valuable for post-earthquake loss assessment where time is of the essence. In areas devoid of building footprints or where the prior probability of a single disaster is low, local pruning strategies can be employed to designate certain nodes as inactive, effectively setting their posterior probabilities to zero. This approach streamlines the impact estimation process. To strike a balance between model complexity and predictive accuracy, it is imperative to ensure that the model's accuracy remains above a performance benchmark. This is achieved through gradual pruning and iterative model adjustments.

3.5. Case retain

By applying these methods, the model can be extended to various types of disasters, offering significant practical value that could be proved in the results. With a rich case library formed from a large number of earthquake cases and the adaptability training of the model to different earthquakes, we can effectively obtain probability estimates for various specific seismic losses in each location. The model's adaptability and efficiency, coupled with its ability to maintain high predictive accuracy, make it a robust tool for disaster risk analysis across diverse seismic environments.

4. Result

4.1. Model performance

Our model has output three probability maps for landslides, liquefaction, and building damage, visually presenting the likelihood of these damages occurring at different locations with varying shades of color depth. We compared these probability maps with the actual damage maps obtained after the earthquake to assess the performance of the model. On the other hand, we demonstrated the refinement of our model's results by comparing them with the landslide and liquefaction probability maps from the USGS as prior information in the input data. From both evaluations, we achieved favorable outcomes.

It is important to evaluate the accuracy of this multi-hazard fast response model for damage prediction of newly occurring earthquakes, as well as to compare the performance improvement of the prior model. We summarize the performance of our system and prior methods using the model's Precision, Recall, F1 score, and ROC curves, respectively. Table 2 shows the results of our test of causal Bayesian networks on the 2023 Turkey earthquake dataset.

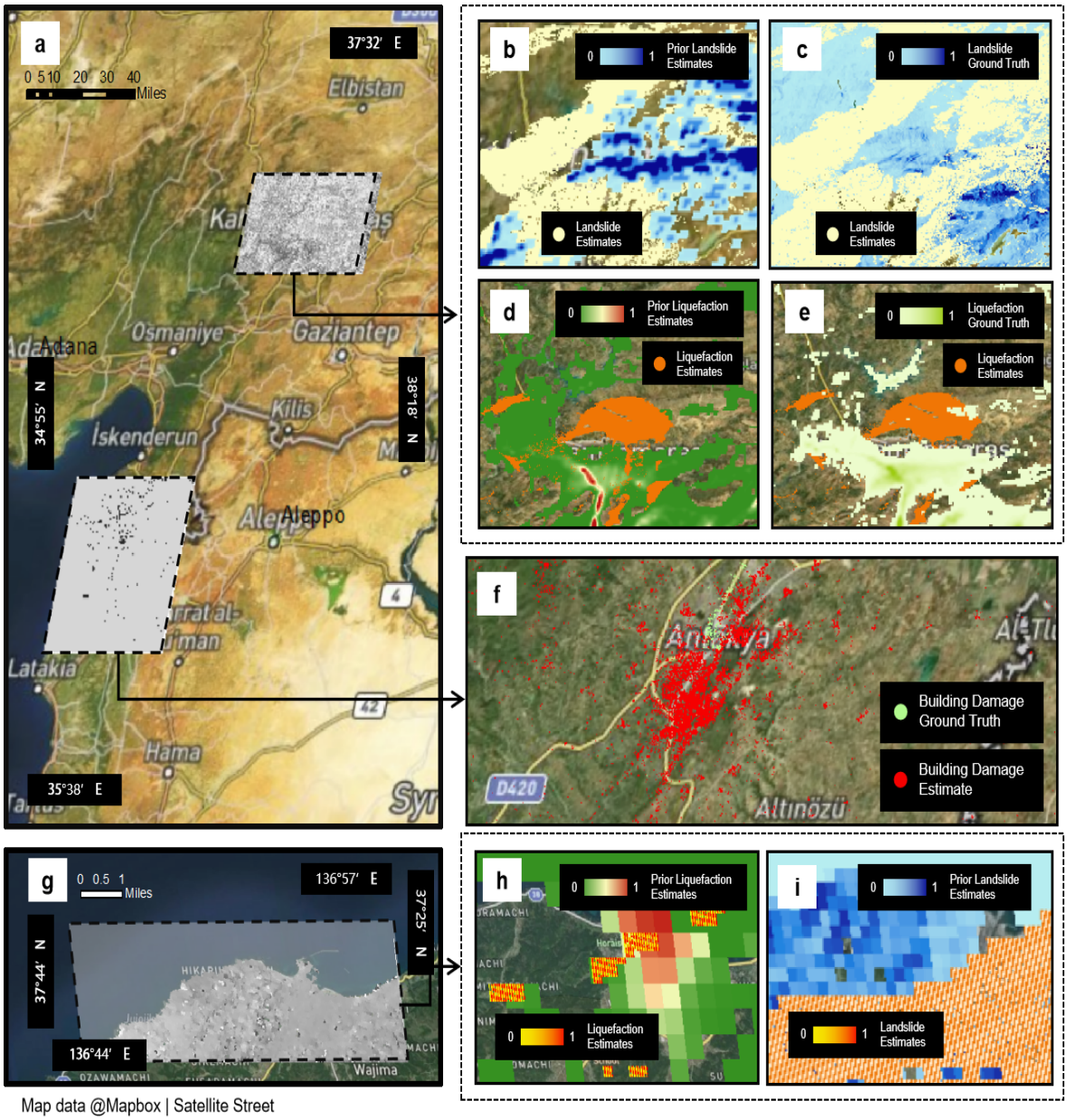

Figure 2: Geospatial prior and posterior estimation models. a, h Damage proxy map with parallelogrammic extents of each focused area. **b, i** Prior landslide estimates (in blue) with our landslide estimates (in yellow). **c** Our landslide estimates (in yellow) with ground truth (in blue). **d, h** Prior Liquefaction estimates with our landslide estimates. **e** Our liquefaction estimates with ground truth. **f** Our building damage estimates (in red) with ground truth (in green).

We evaluate the predictions of our system using ground-truth data collected by a post-event reconnaissance team and compare the prediction performance with existing USGS earthquake ground damage models (i.e., prior models for landslides, and liquefaction). For landslide predictions, our model predicted a Precision of 99.9%, Recall of 40.2%, and F1 score of 57.4%, which is a significant increase compared to the USGS a priori model (Nowicki Jessee, 2018), and we predicted landslides very accurately while predicting more imminent areas of landslides. In the prediction of liquefaction, our model has a recall of 95.9% and an F1 score of 70.6%, both of which are much improved over the a priori model, but the accuracy still needs to be improved. Overall, we accurately predicted most of the areas where liquefaction is imminent.

Also, we summarize the performance of our system and a priori method for building damage prediction using a subject operating characteristic (ROC) curve which visualizes the variation of true positive rate

Table 2
Performance from our tested model on the Turkey data. Prior model refers to the current earthquake prediction model used by USGS.

	Landslide Evaluation			Liquefaction Evaluation		
Type	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Prior Model	0.99973	0.24926	0.39903	0.89326	0.49025	0.63306
New Model	0.99887	0.40221	0.57350	0.55894	0.95868	0.70616

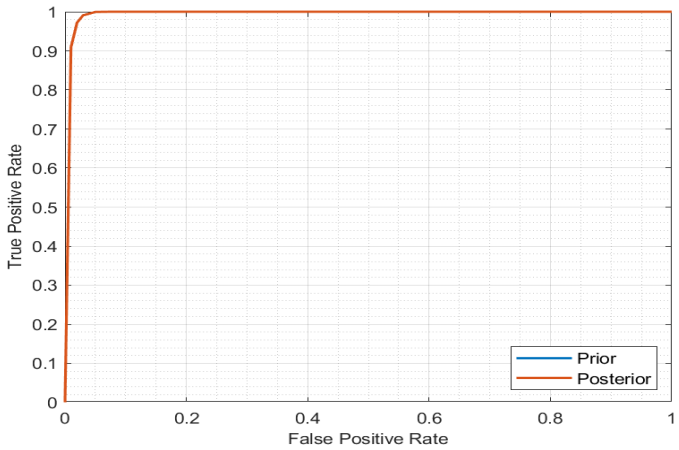

Figure 3: Comparisons of receiver operating characteristic (ROC) curves. The figure shows the ROC curves of our model and the prior model for house damage in the Turkey earthquake.

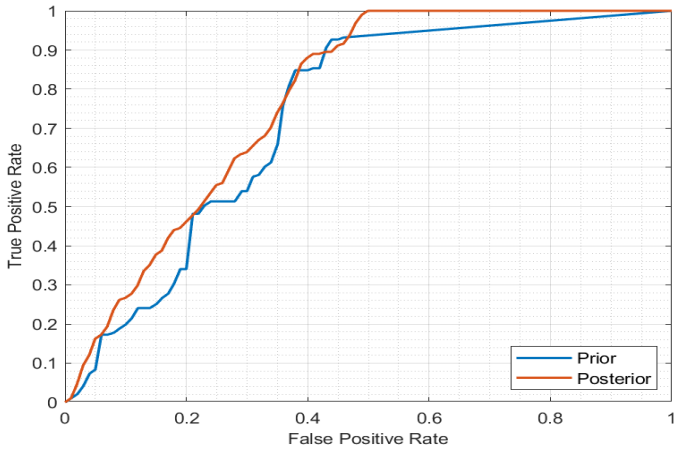

Figure 4: Comparisons of receiver operating characteristic (ROC) curves. The figure shows the ROC curves of our model and the prior model for landslides in the Japan earthquake.

versus false positive rate. Figure 4 gives the performance evaluation results for building damage in the Turkey earthquake, and it can be seen that we have high accuracy in predicting building damage in the Turkey earthquake. Figure 5 gives the performance evaluation results for landslide prediction in the Noto Peninsula earthquake in Japan, and the AUC of our model is 0.77, which improves the performance of our model by 5.36% compared to the USGS model. In summary, our model effectively integrates information from building footprints, a priori models, and DPMs, which significantly improves the performance of ground fault estimation compared to traditional geospatial models and models based solely on DPMs.

4.2. Estimation of the selected earthquake

4.2.1. 2023 Turkey earthquake

We analyzed the causal dependencies between ground shaking, landslide, liquefaction, and building damage using a weight matrix. We found that in the southwestern region, building damage is mainly caused by liquefaction rather than landslides. In earthquakes in Turkey, the causal coefficient from liquefaction to building damage in the southwest is 3.4078, significantly higher than the coefficient from landslides to building damage, which is 0.0301. In the central-northern region, building damage is caused by the combined effects of landslides and liquefaction, with coefficients for liquefaction and landslides at 1.9050 and 1.9021 respectively.

Soil liquefaction refers to the significant reduction in strength and stiffness of saturated or partially saturated soil lacking cohesion in response to ground motion during earthquakes, essentially making the soil behave like a liquid. Therefore, liquefaction occurring beneath buildings and other structures can lead to significant damage, including severe tilting of buildings, ground subsidence, and lateral flow of soil during intense seismic events. In the 7.8 magnitude earthquake in Turkey, extensive liquefaction phenomena were observed in Gölbaşı (Adıyaman) by the lakeshore, the İskenderun port area, and along the Aşu River near Antakya. These areas are located in the southwestern part of Turkey, consistent with our model results. In the central-northern regions of the study area, which are mostly mountainous, landslides also become significant factors affecting building damage, consistent with the model results.

4.2.2. 2024 Japan earthquake

In the Japanese earthquake of January 2024, we found that building damage was caused by both landslides and liquefaction, with a causal coefficient of 0.7125 from liquefaction to building damage and 0.7105 from landslides to building damage. Since there was no significant difference between the two coefficients, we conclude that building damage in the Japanese earthquake resulted from the combined effects of these two factors.

Around 1:10 AM on January 1, 2024, a magnitude 7 earthquake centered in Suzu City, Prefecture, struck, causing severe devastation across the Noto Peninsula. Consequently, even in the central regions, there was significant slope degradation. This demonstrates consistency between landslide and liquefaction-induced building damage and the model results.

4.3. Comparison with State-of-the-Art

The United States Geological Survey (USGS) has developed a series of global earthquake monitoring and rapid assessment systems, including ShakeMap, Did You Feel It? (DYFI), Earthquake Early Warning (EEW), and Prompt Assessment of Global Earthquakes for Response (PAGER), all of which have reached the state-of-the-art (SOTA) level, but each has certain drawbacks. ShakeMap provides earthquake intensity maps to assist decision-makers and emergency response agencies in quickly understanding the extent of earthquake impact in different areas; however, its accuracy and reliability are constrained by the seismic monitoring network, and it only provides earthquake intensity assessments rather than comprehensive earthquake disaster assessments. The DYFI system relies on subjective reports from the public, which may introduce subjectivity and uncertainty, and may lack sufficient quantity and quality of reports in some areas, affecting the accurate assessment of earthquake impact. The EEW system provides limited warning time, possibly only a few seconds to tens of seconds, and is susceptible to false alarms and missed alerts, reducing people's trust in the system. The PAGER system aims to provide rapid assessment after earthquakes, but still faces certain delays and accuracy issues, especially in the early stages following an earthquake event.

Our model, leveraging Damage Proxy Maps (DPM), offers several advantages over existing earthquake monitoring and assessment systems. By utilizing DPM, we have achieved higher-resolution disaster predictions with enhanced accuracy and speed of response. Here are the key benefits of our model:

1. Improved Resolution: Our model provides higher-resolution predictions compared to existing systems, allowing for more detailed and localized assessments of earthquake impacts. This finer granularity enables decision-makers and emergency response agencies to better allocate resources and prioritize areas for assistance.
2. Enhanced Accuracy: Leveraging advanced probabilistic modeling techniques through DPM, our model delivers more accurate assessments of earthquake intensity and potential damage. By incorporating a wider range of data sources and advanced algorithms, we minimize the limitations associated with traditional seismic monitoring networks, resulting in more reliable predictions.
3. Rapid Response: Our model enables swift response to earthquake events by providing timely and actionable information to relevant stakeholders. The speed of our system's predictions, facilitated by DPM, allows for quicker mobilization of resources and implementation of emergency measures, potentially reducing the impact of earthquakes on affected communities.

5. Discussion

Our method, which takes into account the complex interplay of multiple factors post-earthquake, provides a joint estimation of multiple seismic disaster losses. Quantifying the various triggering factors that cause building damage offers significant assistance for actual post-disaster response scenarios. Notably, the application of variational inference and local pruning ensures computational efficiency and scalability, effectively utilizing information from DPMs without waiting for ground truth data, thus accelerating the process of obtaining results. This enables the model to rapidly analyze and estimate losses immediately after an earthquake, making it suitable for various real-world scenarios. In earthquake disaster events, ground motion has always been a critical factor, either directly or indirectly, leading to building damage. However, there are also complex causal relationships between the various post-earthquake events triggered by ground motion. In the February 2023 Turkey earthquakes, the results indicate that the most significant factor affecting building damage was landslides. This observation is consistent with the extensive landslide phenomena noted in the severely impacted regions within the vicinity of our study area. Similarly, In the January 2024 earthquake in Japan, the model results indicate that the building damage caused by both landslides and liquefaction is consistent with the real data. The unique circumstances present in different earthquakes can greatly affect the pattern of loss formation following the seismic event. Rapidly analyzing and estimating these conditions is the central aim of this research. Based on a multi-layer Bayesian causal network, this loss estimation method can be extended to loss estimation for a broader range of catastrophic events. By transforming heterogeneous factors such as disasters and their induced secondary hazards and impacts into nodes within a Bayesian physical causality network, this approach integrates complex information into a holistic model connected by causal relationships.

Acknowledgments

This work was jointly supported by National Natural Science Foundation of China (NSFC) under grants 62206301; JSPS KAKENHI (Grants-in-Aid for Scientific Research, 21H05001); Public Health & Disease Control and Prevention, Fund for Building World-Class Universities (Disciplines) of Renmin University of China. Project No. 2024PDPC; the Major Project of the MOE (China) National Key Research Bases for Humanities and Social Sciences (22JJD910003); and this research was supported by Public Computing Cloud, Renmin University of China. We sincerely thank Dr. Zuo Zhenpeng of Boston University for providing data processing support.

References

[1] Y. Zhang, Z. Ruohan, L. Shang, H. Zeng, Z. Yue, D. Wang, On optimizing model generality in ai-based disaster damage assessment: A subjective logic-driven crowd-ai hybrid learning approach,

in: International Joint Conferences on Artificial Intelligence Organization, 2023.

[2] S. Xu, A. Dimakopoulos, D. J. Wald, H. Y. Noh, Seismic multi-hazard and impact estimation via causal inference from satellite imagery, Nature Communications 13 (2022) 1395. URL: http://creativecommons.org/licenses/by/4.0/, available under a CC BY 4.0 License. Code adapted with changes.

[3] T. Görüm, H. Tanyas, F. Karabacak, A. Yılmaz, S. Girgin, K. E. Allstadt, M. L. Süzen, P. Burgi, Preliminary documentation of coseismic ground failure triggered by the february 6, 2023 türkiye earthquake sequence, Engineering Geology 327 (2023) 107315. doi:10.1016/j.enggeo.2023.107315.

[4] T. Wang, J. Chen, Y. Zhou, et al., Preliminary investigation of building damage in hatay under february 6, 2023 turkey earthquakes, Earthquake Engineering and Engineering Vibration 22 (2023) 853–866. doi:10.1007/s11803-023-2201-0.

[5] S. H. Yun, K. Hudnut, S. Owen, F. Webb, M. Simons, P. Sacco, A. Coletta, Rapid damage mapping for the 2015 mw 7.8 gorkha earthquake using synthetic aperture radar data from cosmo–skymed and alos-2 satellites, Seismological Research Letters 86 (2015) 1549–1555.

[6] O. L. Stephenson, T. Köhne, E. Zhan, B. E. Cahill, S.-H. Yun, Z. E. Ross, M. Simons, Deep learning-based damage mapping with insar coherence time series, IEEE Transactions on Geoscience and Remote Sensing 60 (2021) 1–17.

[7] C. Wang, Y. Liu, X. Zhang, X. Li, V. Paramygin, A. Subgranon, P. Sheng, X. Zhao, S. Xu, Causality-informed rapid post-hurricane building damage detection in large scale from insar imagery, in: Proceedings of the 8th ACM SIGSPATIAL International Workshop on Security Response using GIS, 2023, pp. 7–12.

[8] S.-H. Yun, E. Fielding, M. Simons, P. Agram, P. Rosen, S. Owen, F. Webb, Rapid and reliable damage proxy map from insar coherence, in: IEEE GeoScience and Remote Sensing Society (IGARSS 2012), 2012.

[9] P. Ge, H. Gokon, K. Meguro, A review on synthetic aperture radar-based building damage assessment in disasters, Remote Sensing of Environment 240 (2020) 111693.

[10] Y. Bai, et al., A framework of rapid regional tsunami damage recognition from post-event terrasar-x imagery using deep neural networks, IEEE Geoscience and Remote Sensing Letters 15 (2018) 43–47. doi:10.1109/LGRS.2017.2772349.

[11] B. Adriano, N. Yokoya, J. Xia, H. Miura, W. Liu, M. Matsuoka, S. Koshimura, Learning from multimodal and multitemporal earth observation data for building damage mapping, ISPRS Journal of Photogrammetry and Remote Sensing 175 (2021) 132–143.

[12] R. Anniballe, F. Noto, T. Scalia, C. Bignami, S. Stramondo, M. Chini, N. Pierdicca, Earthquake damage mapping: An overall assessment of ground surveys and vhr image change detection after l'aquila 2009 earthquake, Remote sensing of environment 210 (2018) 166–178.

[13] H. Mueller, A. Groeger, J. Hersh, A. Matranga, J. Serrat, Monitoring war destruction from space using machine learning, Proceedings of the national academy of sciences 118 (2021) e2025400118.

[14] S. Wiguna, B. Adriano, E. Mas, S. Koshimura, Evaluation of deep learning models for building damage mapping in emergency response settings, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 17 (2024) 5651–5667.

[15] A. Rao, J. Jung, V. Silva, G. Molinario, S.-H. Yun, Earthquake building damage detection based on synthetic-aperture-radar imagery and machine learning, Natural Hazards and Earth System Sciences 23 (2023) 789–807.

[16] S. Lee, Application of logistic regression model and its validation for landslide susceptibility mapping using gis and remote sensing data, International Journal of Remote Sensing 26 (2005) 1477–1491.

[17] E. Weber, H. Kané, Building disaster damage assessment in satellite imagery with multi-temporal fusion, arXiv preprint arXiv:2004.05525 (2020).

[18] H. Liu, C. Song, Z. Li, Z. Liu, L. Ta, X. Zhang, B. Chen, B. Han, J. Peng, A new method for the identification of earthquake-damaged buildings using sentinel-1 multi-temporal coherence optimized by homogeneous sar pixels and histogram matching, IEEE Journal of Selected Topics in

Applied Earth Observations and Remote Sensing (2024).
[19] X. Li, P. M. Bürgi, W. Ma, H. Y. Noh, D. J. Wald, S. Xu, Disasternet: Causal bayesian networks with normalizing flows for cascading hazards estimation from satellite imagery, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023, pp. 4391–4403.
[20] C. Wang, Y. Liu, X. Zhang, X. Li, V. Paramygin, P. Sheng, X. Zhao, S. Xu, Scalable and rapid building damage detection after hurricane ian using causal bayesian networks and insar imagery, International Journal of Disaster Risk Reduction (2024) 104371.
[21] L. Huang, T. Chen, Q. Deng, Y. Zhou, Reasoning disaster chains with bayesian network estimated under expert prior knowledge, International Journal of Disaster Risk Science 14 (2023) 1011–1028.
[22] A. Schultheis, C. Zeyen, R. Bergmann, An overview and comparison of case-based reasoning frameworks, in: Case-Based Reasoning Research and Development. ICCBR 2023, volume 14141 of *Lecture Notes in Computer Science*, Springer, Cham, 2023. doi:10.1007/978-3-031-40177-0_21.
[23] S. Slade, Case-based reasoning: A research paradigm, AI magazine 12 (1991) 42–42.
[24] D. Leake, X. Ye, D. J. Crandall, Supporting case-based reasoning with neural networks: An illustration for case adaptation., in: AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering, volume 2, 2021.
[25] T. N. Wolf, F. Bongratz, A.-M. Rickmann, S. Pölsterl, C. Wachinger, Keep the faith: Faithful explanations in convolutional neural networks for case-based reasoning, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, 2024, pp. 5921–5929.
[26] Z. Xia, Z. Li, Y. Bai, et al., Self-supervised learning for building damage assessment from large-scale xbd satellite imagery benchmark datasets, in: International Conference on Database and Expert Systems Applications, Springer International Publishing, Cham, 2022, pp. 373–386.
[27] Z. Zhao, J. Chen, K. Xu, et al., A spatial case-based reasoning method for regional landslide risk assessment, International Journal of Applied Earth Observation and Geoinformation 102 (2021) 102381.
[28] K. Wang, Y. Yang, G. Reniers, et al., Predicting the spatial distribution of direct economic losses from typhoon storm surge disasters using case-based reasoning, International Journal of Disaster Risk Reduction 68 (2022) 102704.
[29] F. Yu, X. Li, Case-based reasoning for disaster management: Structure design for cascading disasters case base, in: Disaster Risk Reduction for Resilience: Disaster Risk Management Strategies, Springer, 2022, pp. 295–314.
[30] S. Möhrle, Case-Based Decision Support for Disaster Management, KIT Scientific Publishing, 2020.
[31] J. Chen, K. Xu, Z. Zhao, X. Gan, H. Xie, A cellular automaton integrating spatial case-based reasoning for predicting local landslide hazards, International Journal of Geographical Information Science 38 (2024) 100–127.
[32] C. Cundy, A. Grover, S. Ermon, Bcd nets: Scalable variational approaches for bayesian causal discovery, Advances in Neural Information Processing Systems 34 (2021) 7095–7110.
[33] K. Chobtham, A. C. Constantinou, Bayesian network structure learning with causal effects in the presence of latent variables, in: International Conference on Probabilistic Graphical Models, PMLR, 2020, pp. 101–112.
[34] V. Aglietti, N. Dhir, J. González, T. Damoulas, Dynamic causal bayesian optimization, Advances in Neural Information Processing Systems 34 (2021) 10549–10560.
[35] L. Wang, A. Adiga, J. Chen, A. Sadilek, S. Venkatramanan, M. Marathe, Causalgnn: Causal-based graph neural networks for spatio-temporal epidemic forecasting, Proceedings of the AAAI Conference on Artificial Intelligence 36 (2022) 12191–12199.
[36] A. C. Constantinou, N. Fenton, M. Neil, How do some bayesian network machine learned graphs compare to causal knowledge, arXiv 2101.10461 (2021).
[37] M. Sinha, S. A. Ramsey, Using a general prior knowledge graph to improve data-driven causal network learning, in: AAAI Spring Symposium Combining Machine Learning with Knowledge Engineering, 2021.