Computer Modeling in Engineering & Sciences

Tech Science Press

Doi:10.32604/cmes.2025.074768

REVIEW

The Transparency Revolution in Geohazard Science: A Systematic Review and Research Roadmap for Explainable Artificial Intelligence

Moein Tosan[1,*], Vahid Nourani[2,3], Ozgur Kisi[4,5,6], Yongqiang Zhang[7], Sameh A. Kantoush[8], Mekonnen Gebremichael[9], Ruhollah Taghizadeh-Mehrjardi[10] and Jinhui Jeanne Huang[11]

[1]Department of Irrigation and Reclamation Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, 31587-77871, Iran

[2]Center of Excellence in Hydroinformatics, Faculty of Civil Engineering, University of Tabriz, 29 Bahman Ave, Tabriz, 51666-16471, Iran

[3]Department of Civil Engineering, Faculty of Engineering and Architecture, World Peace University, Sht. Kemal Ali Omer St. No:22, Yenisehir, Nicosia/TRNC, Mersin 10, Türkiye

[4]Department of Civil Engineering, Lübeck University of Applied Sciences, Lübeck, 23562, Germany

[5]Department of Civil Engineering, Ilia State University, Tbilisi, 0162, Georgia

[6]School of Civil, Environmental and Architectural Engineering, Korea University, Seoul, 02841, Republic of Korea

[7]Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China

[8]Disaster Prevention Research Institute (DPRI), Kyoto University, Kyoto, 611-0011, Japan

[9]Department of Civil and Environmental Engineering, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

[10]Department of Soil Science and Geomorphology, University of Tübingen, Tübingen, 72076, Germany

[11]College of Environmental Science and Engineering/Sino-Canada Joint R&D Centre for Water and Environmental Safety, Nankai University, Tianjin, 300071, China

*Corresponding Author: Moein Tosan. Email: moein69tosan@alumni.ut.ac.ir

Received: 17 October 2025; Accepted: 16 December 2025; Published: 29 January 2026

ABSTRACT: The integration of machine learning (ML) into geohazard assessment has successfully instigated a paradigm shift, leading to the production of models that possess a level of predictive accuracy previously considered unattainable. However, the black-box nature of these systems presents a significant barrier, hindering their operational adoption, regulatory approval, and full scientific validation. This paper provides a systematic review and synthesis of the emerging field of explainable artificial intelligence (XAI) as applied to geohazard science (GeoXAI), a domain that aims to resolve the long-standing trade-off between model performance and interpretability. A rigorous synthesis of 87 foundational studies is used to map the intellectual and methodological contours of this rapidly expanding field. The analysis reveals that current research efforts are concentrated predominantly on landslide and flood assessment. Methodologically, tree-based ensembles and deep learning models dominate the literature, with SHapley Additive exPlanations (SHAP) frequently adopted as the principal post-hoc explanation technique. More importantly, the review further documents how the role of XAI has shifted: rather than being used solely as a tool for interpreting models after training, it is increasingly integrated into the modeling cycle itself. Recent applications include its use in feature selection, adaptive sampling strategies, and model evaluation. The evidence also shows that GeoXAI extends beyond producing feature rankings. It reveals nonlinear thresholds and interaction effects that generate deeper mechanistic insights into hazard processes and mechanisms. Nevertheless, several key challenges remain unresolved within the field. These persistent issues are especially pronounced when considering the crucial necessity for interpretation stability, the demanding scholarly task of reliably distinguishing correlation from causation, and the development of appropriate methods for the treatment of complex spatio-temporal dynamics.

KEYWORDS: Explainable artificial intelligence (XAI); geohazard assessment; machine learning; SHAP; trustworthy AI; model interpretability

1 Introduction

Geohazards, spanning from landslides and floods to wildfires and earthquakes, represent a mounting threat to human populations and critical infrastructure. This complex problem is being significantly intensified by the dual pressures of global climate change and rapid urbanization [1]. Consequently, the capacity to accurately predict where and when these destructive events might occur is not just beneficial but critical for establishing effective risk management strategies [2], informing strategic land-use planning decisions [3], and operating reliable early warning systems [4]. For decades, the field has tackled this challenge using two main approaches; mechanistic models and data-driven statistical models [5]. Physics-based models, built on the principles of geomechanics and hydrology, give valuable insight into how hazards form and their underlying processes [6]. However, these models often face significant limitations when applied at a regional scale. Specifically, such regional modeling necessitates immense demands on computational power and requires a high quality and depth of detailed geotechnical input data which is rarely available in practice.

To circumvent these traditional limitations, data-driven methodologies, particularly those based on machine learning (ML), have emerged as a major paradigm shift in the field of geohazard susceptibility mapping and forecasting [7,8]. By effectively utilizing vast geospatial datasets, these ML models have provided a considerable advancement in predictive capability [9]. This progress encompasses a diverse array of techniques, ranging from powerful ensemble models—such as random forest (RF) [10] and eXtreme gradient boosting (XGBoost) [11–13]—to more sophisticated deep learning (DL) architectures, including convolutional neural networks (CNNs) [14,15], long short-term memory (LSTM) networks [16,17], and graph neural networks (GNNs) [18]. However, this undeniable success is accompanied by a fundamental challenge: the well-known black-box problem [19,20]. The very complexity that grants these models their predictive strength simultaneously obstructs an understanding of their underlying mechanisms and internal decision-making processes. This lack of transparency remains as a major obstacle to their practical implementation, especially in high-stakes, mission-critical scenarios [21]. When it comes to high-stakes decisions, such as those made by policymakers, stakeholders, or civil protection agencies, simply receiving a prediction isn't sufficient; they must be able to comprehend, validate and ultimately justify the rationale behind the model's recommendation [22].

Building on this critical need for justified decision-making, explainable AI (XAI) has emerged precisely to address the transparency challenge in geohazard prediction systems [23]. Essentially, XAI provides researchers and practitioners with an extensive toolkit of methods, all specifically engineered to take the complex predictions and often-opaque internal logic of advanced predictive models and make them intelligible to human domain experts [24,25]. In the context of earth sciences, the utilization of geohazard explainable artificial intelligence (GeoXAI) has recently seen a rapid expansion [26,27]. Its function has matured well beyond simple, passive post-hoc interpretation—such as merely ranking the importance of input features [28]. Instead, GeoXAI now actively influences the modeling process itself [29]. Currently, researchers employ GeoXAI for strategic tasks like intelligent feature selection [30], targeted model simplification [31], and rigorously validating a model's underlying logic against established physical principles [32]. At the vanguard of this field, XAI is even being employed as a tool for scientific discovery, helping scholars uncover latent physical parameters [33], pinpoint novel hazard precursors, and fundamentally connect the physics-based and data-driven modeling domains.

While GeoXAI has advanced rapidly, a genuinely comprehensive synthesis of the field is still absent in the current literature. Most existing review papers focus narrowly on a few particular hazards or methodological tools (see Table 1). Therefore, there is a recognized and critical need for the community to consolidate this dispersed knowledge more holistically and, crucially, to establish a clear, unified agenda for future research efforts. This systematic review is designed to address that gap. First, it maps out the intellectual and thematic structure of GeoXAI by using quantitative and thematic analyses. Building from there, the review synthesizes the mechanistic insights into hazard processes that XAI has produced, such as identifying thresholds, interactions, and the spatio-temporal dynamics of hazards. This paper then traces the evolution of XAI, following its development from a passive post-hoc explanation tool into a more active component within the modeling lifecycle. Finally, the review outlines the frontier challenges—which include uncertainty quantification, causal inference, and merging physics-based with data-driven approaches—to propose a research roadmap aimed at developing robust, transparent, and operational geospatial artificial intelligence (GeoAI) systems. By bringing together these state-of-the-art advances, this work aims to give researchers and practitioners a solid foundation and a forward-looking perspective for advancing trustworthy AI in geohazard science and risk management. Accordingly, this study aims to provide a comprehensive and systematic synthesis of explainable artificial intelligence (XAI) applications in geohazard science, identify their methodological evolution, and establish a research roadmap to guide future GeoXAI development.

Table 1: Comparison of existing review papers with the present study

Reference	Thematic focus	Methodological scope	Key limitations & distinction from this review
[34]	Landslide susceptibility	Statistical and ML models	Seminal review on best practices for susceptibility assessment, but pre-dates the widespread adoption of modern XAI tools (e.g., SHAP).
[35]	Landslide susceptibility	Comparative ML performance	Focuses strictly on predictive performance metrics (AUC/accuracy) of algorithms; lacks systematic discussion on model transparency.
[36]	General geohazards	AI applications (general)	Provides a high-level narrative overview of AI trends; discusses XAI only briefly as an emerging challenge without a systematic synthesis.
[37]	Flood mapping	Deep learning architectures	Comprehensive review of CNNs/DL models, but treats explainability as a minor subsection rather than a core evaluation criterion.
[38]	Disaster risk management	XAI in disaster management	Covers a broad spectrum of disaster phases (response, recovery); lacks specific focus on the physical modeling mechanisms of hazard generation.
This study (2026)	GeoXAI (Multi-hazard)	Systematic review of XAI and physical insights	First systematic synthesis of XAI methods specifically for understanding geohazard mechanisms, spatiotemporal dynamics, and addressing the correlation-causation gap.

2 Methodology

To ensure a comprehensive and unbiased synthesis, the review followed the PRISMA 2020 protocol, providing a transparent and standardized approach for systematic literature analysis. This ensures that the final set of selected articles accurately reflects the contemporary research landscape and satisfies the core objectives of this review. The methodology was guided by closely following the Preferred Reporting Items for systematic reviews and meta-analyses (PRISMA) framework [39,40]. This approach was explicitly engineered to be transparent, fully reproducible, and rigorously thorough, thereby ensuring that the final set of selected articles accurately reflects the contemporary research landscape and completely satisfies the core objectives of this review.

2.1 Search Strategy and Database Selection

The web of science (WoS) core collection was chosen as the primary database for this work. The selection of this database was based on several critical factors: its stringent curation process, complete citation tracking capabilities [41,42], and its strong reputation for consistently indexing high-impact, peer-reviewed articles [43,44]. Consequently, this choice aligns strongly with the study's primary objective: to effectively synthesize the most influential and methodologically sound research specifically within the GeoXAI field. To ensure the successful retrieval of a body of studies that was simultaneously comprehensive and precisely focused, a detailed search string was subsequently developed and meticulously fine-tuned by the researchers. The final query was structured around three essential core pillars: the geohazard context, the underlying ML methodology, and the explainability component. A typical search string is shown below:

TS = ("landslide" OR "flood" OR "wildfire" OR "earthquake" OR "drought" OR "geohazard*") AND ("machine learning" OR "artificial intelligence" OR "deep learning" OR "ensemble learning") AND ("explainable AI" OR "XAI" OR "interpretability" OR "SHAP" OR "LIME" OR "Grad-CAM" OR "causal inference" OR "physics-informed").

The search was deliberately restricted to articles published between 01 January 2021, and September 2025. This precise temporal constraint constitutes a strategic choice, implemented to capture the modern paradigm of algorithmic explainability effectively. Before the main review, a preliminary scoping analysis decisively revealed that the first significant wave of studies applying contemporary XAI frameworks (e.g., SHAP) to geohazard assessment began to proliferate quite notably in 2021. Consequently, this focus ensures that the review analyzes a coherent body of cutting-edge literature that is actively shaping the future direction of the field, thereby avoiding a broad overview of older, conceptually distinct research focused merely on general model interpretation.

2.2 Screening and Selection Process

To ensure a systematic and unbiased review, the article selection process rigorously followed the four-stage PRISMA protocol (Fig. 1).

Identification: Executing the optimized search string on the WoS database initially yielded 482 articles, from which 34 records were removed after initial filters.

Screening: Following identification, the titles and abstracts of all 448 articles were independently screened by two separate reviewers. The reviewers' role was to meticulously assess relevance against the established core inclusion criteria. This step proved critical: articles were primarily excluded if they clearly fell outside the thematic scope. Examples of such exclusions include studies where XAI was applied only to medical imaging, where ML was used solely for mineral exploration, or, importantly, geohazard studies that completely lacked an XAI component. This initial screening successfully resulted in a substantial reduction of the pool, moving us toward the set of potentially eligible papers.

Eligibility: The full text of all remaining articles (108) was subsequently retrieved for a detailed eligibility assessment. During this more rigorous review phase, articles were systematically excluded if they failed to meet the necessary methodological depth (for instance, if the application of XAI was superficial or limited solely to a standard feature importance plot without any deeper analysis), were not original peer-reviewed research (e.g., conference papers, existing reviews, or book chapters), or failed to satisfy other specific inclusion criteria.

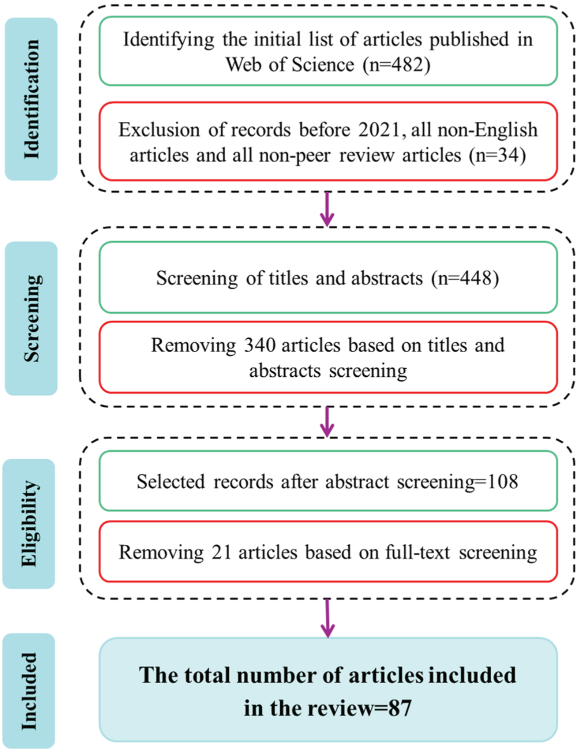

Identification
Identifying the initial list of articles published in Web of Science (n=482)
Exclusion of records before 2021, all non-English articles and all non-peer review articles (n=34)
Screening
Screening of titles and abstracts (n=448)
Removing 340 articles based on titles and abstracts screening
Eligibility
Selected records after abstract screening=108
Removing 21 articles based on full-text screening
Included
The total number of articles included in the review=87

Figure 1: PRISMA-based workflow showing the four-step process (identification, screening, eligibility, and inclusion) used for selecting the 87 studies analyzed in this GeoXAI systematic review. This ensures transparency and reproducibility of the literature selection process

Inclusion: Following the full-text review, a final consensus was reached on 87 articles that fully met all established criteria. This highly curated collection of studies serves as the foundation for quantitative mapping, thematic synthesis, and critical analysis presented in the subsequent sections of this paper. Fig. 1 summarizes this entire process using the PRISMA flow diagram. With the final corpus of 87 studies established, the following section synthesizes their collective findings—mapping how XAI has been adopted, adapted, and advanced within geohazard modeling over the past five years.

2.3 Inclusion and Exclusion Criteria

To effectively maintain both the required focus and requisite academic rigor of this review, the following explicit criteria were consistently applied throughout the entire screening process:

2.3.1 Inclusion Criteria

- The studies selected had to explicitly apply one or more XAI techniques to either the modeling or assessment of a geohazard (e.g., landslides, floods, wildfires, erosion, earthquakes, subsidence).
- Importantly, the paper needed to implement a recognized post-hoc XAI framework (such as SHAP, LIME, or Grad-CAM) or utilize an intrinsically interpretable (glass-box) model [45]. Studies reporting only standard, model-internal feature importance metrics were generally considered insufficient for inclusion. Such papers were ruled out unless they demonstrated a genuine engagement with the broader XAI literature and provided analysis beyond these basic measures. This criterion was central to distinguishing genuine GeoXAI research from broader ML applications.
- The publication must be a peer-reviewed, original research article in a scientific journal.

- The article must be published in English between 01 January 2021, and September 2025.

2.3.2 Exclusion Criteria

- Studies using ML for geohazards but lacking any explicit explainability component.
- Studies focused on XAI applications in non-geospatial or non-environmental domains.
- Review articles, meta-analyses, book chapters, conference proceedings, and pre-prints.
- Methodological papers that develop new XAI techniques without applying them to a concrete—geohazard case study.

3 Results

3.1 The Emerging Landscape: A Quantitative and Thematic Analysis

This section provides a comprehensive statistical and thematic overview of the GeoXAI research landscape, mapping its main application areas, dominant methodologies, and key geographical research centers. The field where XAI meets geohazard assessment; a domain termed GeoXAI; is no longer just a theoretical idea but has become a rapidly expanding and vibrant area of scientific research [46,47]. A statistical analysis of the papers foundational to this review reveals a clear and accelerating growth pattern. The field has seen an exponential jump in publications, especially since 2022, a surge that points to a critical shift within the geoscientific community (Fig. 2). There is a clear move away from simply accepting ML models as high-performing black-boxes and a growing demand for transparency, trustworthiness, and deeper mechanistic insight.

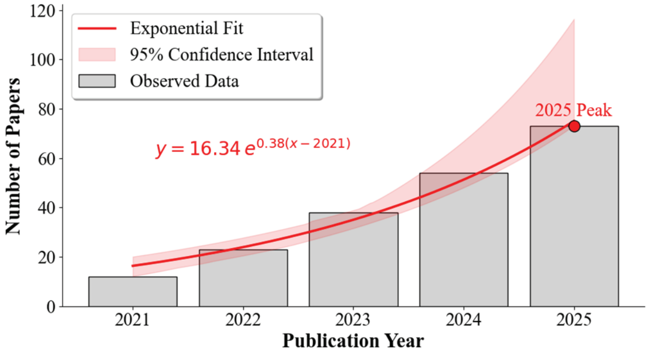


Figure 2: Temporal distribution of the reviewed GeoXAI publications (2021–2025), illustrating the exponential growth of the research field

As summarized in Fig. 3, the thematic landscape of GeoXAI shows a field with both focused depth and expanding breadth. Landslides and various forms of flooding (e.g., pluvial, fluvial, flash) have become the dominant testbeds for this research, accounting for a large majority of applications. There is a logical reason for this focus. These hazards are typically backed by large historical inventories and well-defined conditioning factors, which offer a solid basis for building and testing complex models. At the same time, the field is quickly broadening to cover other geohazards, now including wildfires, various forms of erosion (gully, soil), and earthquake-related risks. The growing flexibility and adoption of this methodology across the Earth sciences is highlighted by more specialized applications, such as studies on sinkhole susceptibility [48], land subsidence [49], and even failures in geotechnical infrastructure [50].

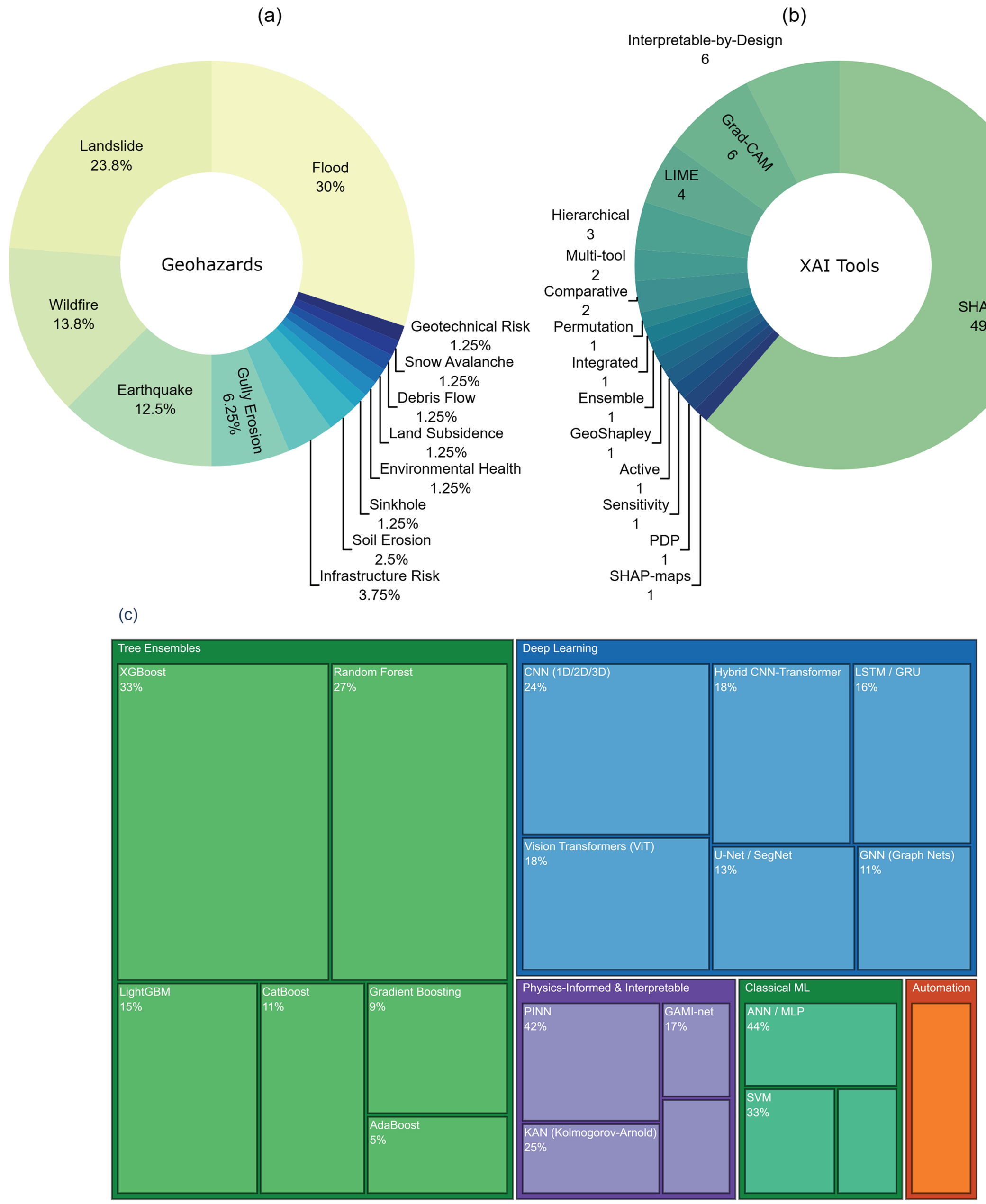


Figure 3: Thematic landscape of the reviewed GeoXAI studies showing (**a**) distribution of research by hazard type, (**b**) frequency of explainability methods, and (**c**) hierarchy of ML/DL models. The figure highlights landslides and floods as dominant applications and SHAP as the prevailing interpretation tool

The methodological core of the entire GeoXAI toolkit rests decisively on the performance ceiling of its chosen ML algorithms. In this area, tree-based ensemble methods are clearly the primary methods. Algorithms like RF, XGBoost, and Light Gradient Boosting Machine (LightGBM) are deployed extensively, mainly because of their well-established reputation for exceptional robustness and consistent high predictive accuracy across various geohazard datasets. Moving past these standard ensembles, researchers are increasingly leveraging sophisticated DL architectures for more advanced predictive modeling, particularly when intricate data types are involved [12]. For instance, CNNs are remarkably effective when it comes to processing gridded spatial information, while researchers typically utilize LSTM networks for detailed time-series data analysis. Additionally, the field's advanced toolkit has recently expanded its capacity with GNNs, a class of models uniquely engineered to capture and learn directly from the intrinsic spatial relationships embedded within complex geographic data structures [51]. A significant development parallel to these modeling advancements is the rise of automated ML (AutoML). The synthesis reveals that AutoML is shifting from a mere efficiency tool to a critical component of the GeoXAI workflow. Frameworks such as AutoGluon [52], tree-based pipeline optimization tool (TPOT) [53], and Bayesian optimization engines like Optuna [54,55] are now being employed to democratize high-performance modeling. For instance, [52] demonstrated that an AutoML framework could generate a state-of-the-art landslide susceptibility model in just 156 s, a task that typically requires weeks of manual tuning. Crucially, within the GeoXAI paradigm, AutoML does not just automate prediction; it enhances process transparency. Studies like [53] utilize the visualization tools inherent in these frameworks (e.g., hyperparameter importance plots) to explain how the model reached its optimal architecture, thereby adding a layer of interpretability to the optimization process itself.

In the domain of interpretability, the research community has shown an overwhelming consensus, effectively converging on SHAP (SHapley Additive exPlanations) as the *de facto* standard for post-hoc explanation. This consensus exists because of the method's rigorous game-theoretic foundation, which has firmly established SHAP as the premier analytical tool available. This robust theoretical underpinning is what enables researchers to meticulously dissect and precisely quantify the individual, specific contribution of every conditioning factor to the final prediction outcome. Other methods, however, play important complementary roles. local interpretable model-agnostic explanations (LIME) is often employed for instance-specific diagnoses [56,57], while gradient-weighted class activation mapping (Grad-CAM) is the preferred technique for visualizing the spatial focus of CNNs in image-based tasks [58,59]. The increasing adoption of a multi-tool approach—where the insights derived from several distinct XAI techniques are strategically fused—is a compelling indicator of the growing methodological sophistication and maturity now characterizing the geohazard modeling community [60,61].

Finally, the geographic distribution of GeoXAI research highlights its global relevance (Fig. 4). While China is a leader in producing a high volume of impactful studies, research groups from Europe, the United States, Iran, and Türkiye are also making significant contributions. The widespread nature of this work points to a shared international agreement: tackling the common challenges of geohazards requires transparent and trustworthy AI. In essence, this quantitative overview shows a field defined by its rapid growth. It has a diversifying range of applications, an increasingly sophisticated toolkit, and a truly global research community.

To provide a systematic and more detailed understanding of the methodological toolkit that underpins GeoXAI, we comparatively summarize the principal explainability frameworks found in the literature. Table 2 offers a structured overview of these different methods. The comparison itself is structured around several key aspects for each framework: its core principles, how it's typically implemented, the predictive models it's often paired with, and specific examples of its application to geohazards.

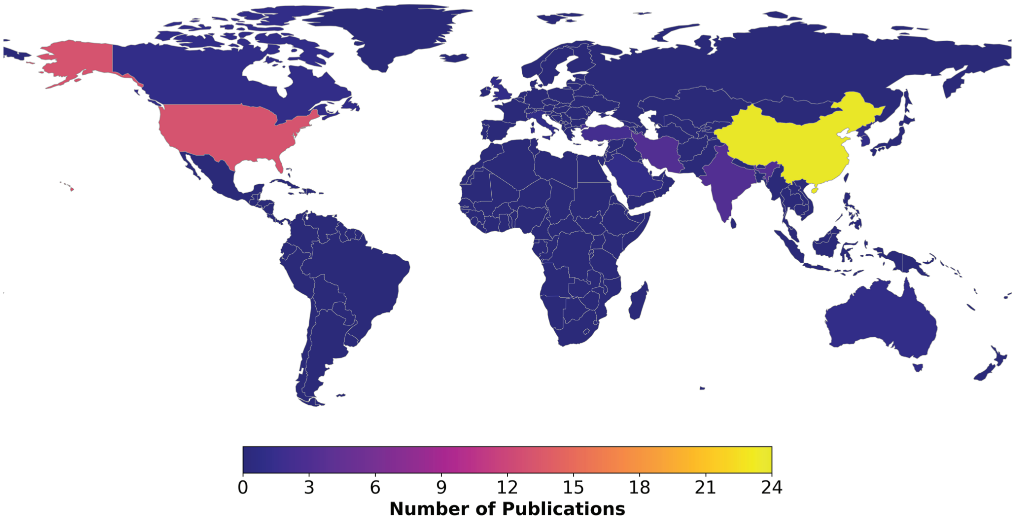


Figure 4: Geographic distribution of GeoXAI research based on the affiliations of the primary authors, highlighting the global research hubs and collaborative landscape

Table 2: A comparative overview of the primary XAI methodologies applied in geohazard assessment, detailing their core principles, typical applications, advantages, limitations, and representative studies from the reviewed literature

XAI method	Core principle /mechanism	Typical applications and context	Advantages	Limitations	Representative studies
SHAP	A game-theoretic method that fairly attributes prediction output to each input feature based on its marginal contribution across all possible feature coalitions.	Domains: Landslides, floods, wildfires, etc. Models: RF, XGBoost, CNN, LSTM, GNN	• Strong theoretical foundation. • Provides both global and local explanation consistency. • Quantifies feature interaction effects.	• Computationally intensive. • Can be misinterpreted with correlated features. • Sensitive to the background dataset.	• Choubin et al. [62]: Exemplifies a comprehensive end-to-end workflow. • Yao et al. [63]: Exemplifies deep spatio-temporal analysis. • Lv et al. [64]: Exemplifies comparative model interpretation.
LIME	Approximates a complex model with a simple, interpretable surrogate (e.g., linear model) in the local vicinity of a single prediction.	Domains: Flood segmentation, image classification, localized diagnosis. Models: Any black-box model, especially CNNs.	• Highly intuitive for individual cases. • Model-agnostic. • Comput ationally efficient for single explanations.	• Local explanations only; no global view. • Can be unstablze depending on neighborhood definition. • May not be faithful if local boundary is highly non-linear.	• Purohit and Pujari [56]: Exemplifies use as a diagnostic tool for error analysis. • Gao and Wang [65]: Exemplifies personalized risk explanation.

(Continued)

Table 2 (continued)

XAI method	Core principle /mechanism	Typical applications and context	Advantages	Limitations	Representative studies
Permutation feature importance (PFI)	Measures feature importance by calculating the decrease in model performance (e.g., accuracy, RMSE) when a single feature's values are randomly shuffled.	Domains: Flood forecasting, gully erosion. Models: RNN, GRU, LSTM (time-series), RF.	• Model-agnostic and intuitive. • Directly measures impact on generalization error. • Effective for capturing temporal dependency in time-series when applied to lag features.	• Computationally expensive for large datasets. • Can be biased if features are highly correlated (creates unrealistic data combinations). • Provides global importance only (no local instance explanation).	• Gholami et al. [61]: Exemplifies use with RNN/GRU for temporal susceptibility. • Slater et al. [66]: Exemplifies sensitivity analysis in dynamic flood modeling.
Grad-CAM	A visualization technique that produces a heatmap highlighting important input regions for a prediction by using the gradients of the target class.	Domains: Flood segmentation, image classification. Models: CNNs, U-Net, vision transformers.	• Visually intuitive (where the model looks). • Computa tionally efficient. • Excellent for validating physical plausibility of spatial reasoning.	• Explanations can be coarse/low-resolution. • Explains where but not necessarily 'why'. • Limited to models with convolutional-like structures.	• Sanderson et al. [58]: Exemplifies validation of multi-modal data fusion. • Wang et al. [67]: Exemplifies validation of a model's adaptive reasoning.
Interpretable by design/Physics-informed	A class of glass-box models where transparency is an intrinsic architectural feature, often by embedding physical laws (e.g., PDEs) into the model's structure or loss function.	Domains: Landslide mechanics, wildfire spread, flood forecasting. Models: PINNs, KANs, custom architectures.	• Produces physically plausible predictions. • Can infer unobserved physical parameters. • Provides the highest level of transparency and trust.	• Requires well-defined governing equations or strong domain knowledge. • Can be complex to design and train. • Applicability may be limited to problems with known physics.	• Dahal and Lombardo [68]: Exemplifies inferring latent geotechnical parameters. • Vogiatzoglou et al. [69]: Exemplifies parameter learning for physical models. • Taghizadeh et al. [70]: Exemplifies novel interpretable architecture (KAN).

To address the need for a quantitative synthesis beyond thematic mapping, we analyzed the reported performance metrics and XAI efficacy across the reviewed studies (summarized in Table 3). This synthesis reveals distinct methodological patterns tailored to specific hazard types. For Landslide susceptibility, tree-based ensembles (RF, XGBoost) paired with SHAP are the dominant standard, consistently delivering high predictive performance (Mean AUC ≈ 0.92) and robust feature attributions. In contrast, Flood assessment exhibits a bifurcation: susceptibility mapping relies heavily on tree-based SHAP (similar to landslides), while flood inundation and forecasting increasingly employ deep learning architectures (CNNs, LSTMs) interpreted via Grad-CAM and time-dependent SHAP, respectively. While a direct statistical meta-analysis is constrained by the heterogeneity of validation metrics (e.g., AUC vs. RMSE vs. NSE), the aggregated data indicates that hybrid models (integrating XAI in the loop) generally yield a performance improvement of 2%–5% over baseline black-box models by facilitating the removal of spurious features.

Table 3: Quantitative synthesis of dominant ML architectures, XAI methods, and reported performance ranges across key geohazards

Geohazard type	Dominant ML models	Primary XAI methods	Reported performance range (typical)	Key physical insights uncovered by XAI
Landslides	Tree ensembles (RF, XGBoost, LightGBM)	SHAP (TreeExplainer), PDP	AUC: 0.85–0.98 Acc: 85%–93%	Identification of slope-rainfall thresholds; distinguishing conditioning vs. triggering factors.
Floods (susceptibility)	RF, SVM, ensemble stacking	SHAP, permutation importance	AUC: 0.88–0.96	Identifying topographic controls (TWI, distance to river) vs. urban factors (imperviousness).
Floods (forecasting /mapping)	CNN, LSTM, GNN	Grad-CAM, deep SHAP, IG	NSE: 0.70–0.93 IoU: 0.60–0.85	Visualizing spatial attention on river networks; quantifying time-lag effects of upstream runoff.
Wildfires	RF, MaxEnt, DNN	SHAP, interaction plots	AUC: 0.82–0.94	Synergistic effects of temperature and anthropogenic proximity; seasonal fuel preconditioning.
Gully/soil erosion	RF, XGBoost	SHAP, spatial maps	AUC: 0.90–0.96	Thresholds for slope-area relationships; impact of land-use change on erosion initiation.
Earthquakes	Hybrid (CNN-XGBoost), PINN	SHAP, physics-based extraction	Acc: 85%–99% R^2: 0.97 (magnitude)	Validating seismic precursors; distinguishing geophysical signals from noise in deformation data.

3.2 *Unpacking the Black-Box: Core Insights from XAI Applications*

The true value of integrating XAI into geohazard modeling lies in its capacity to transform a model's outputs from opaque predictions into a authentic source of new scientific knowledge and actionable intelligence [71]. By systematically unpacking the black-box, researchers are empowered to move decisively beyond simply knowing what a model predicts, allowing them instead to fully understand why it arrives at certain decisions [49]. Critically, this essential process of inquiry unfolds across two distinct, yet complementary, scales: the global and the local.

3.2.1 Global vs. Local Interpretability: From General Drivers to Site-Specific Diagnosis

A fundamental duality exists in model interpretation. Global interpretability aims to elucidate the model's behavior in its entirety, specifically by identifying the average influence of each conditioning factor across the whole study area [72,73]. Essentially, it answers the strategic question: What are the most important drivers of this hazard in this region? Local interpretability, in sharp contrast, involves dissecting a single, specific prediction for an individual data point, such as an isolated pixel, a distinct slope unit, or a single structure. It answers the tactical question: Why was this specific location assigned to this particular risk level? While both global and local interpretations are essential, the synthesis reveals that it is the power of local interpretation that most effectively bridges the gap between high-performance predictive modeling and crucial operational decision-making.

At the global scale, interpretation provides the crucial first step of understanding broad hazard mechanisms. It provides a crucial initial step, often achieved through analyzing SHAP summary plots or reviewing feature importance rankings. These global interpretation methods are instrumental in two primary areas: first, they confirm established domain knowledge, and second, they reliably pinpoint the dominant predisposing and triggering factors across a macro scale. Such global interpretations, for example, have been successfully used in various studies to identify key drivers. Reference [74] found the Fire Weather Index to be the dominant factor driving wildfire risk throughout Italy. Similarly, these techniques consistently validate that elements like distance to streams and topographic wetness are critical in assessing flood susceptibility [62]. Moreover, research has clearly demonstrated that specific seismological parameters are paramount for accurately assessing earthquake probability [75]. Notably, this level of analysis effectively provides the broad, essential scientific context required for effective regional planning and informed policy development. However, integrating spatiotemporal analysis into these large-scale predictions remains a frontier challenge. In continental-scale assessments, treating space and time as static features can obscure critical regional variances. Recent applications address this by integrating dynamic inputs directly into the XAI framework. For instance, in a pan-European wildfire study, [76] utilized Spatio-temporal SHAP Maps to demonstrate that the drivers of fire risk shift from solar radiation in Southern Europe to precipitation deficits in Northern Europe, and that these drivers have distinct seasonal time-lags. This demonstrates that for large-scale modeling to be operationally valid, XAI must move beyond global averages to reveal the heterogeneous spatiotemporal mechanisms driving the hazard.

Crucially, local interpretations do more than just illustrate global trends on a smaller scale; they often reveal that site-specific mechanisms can deviate significantly from the regional average. A compelling example is found in Teke and Kavzoglu [57]. A global landslide model identified slope and elevation as the dominant regional drivers. However, when researchers conducted a local analysis of three distinct landslides in that same area, they found that a completely different factor was the primary cause for each one. This finding is profound because it shows that a one-size-fits-all mitigation strategy based only on global drivers would be fundamentally flawed. As noted in Ibrahim et al. [50], the local context—whether it's a unique geological feature or a human influence—can often override the general trend.

Ultimately, the most mature GeoXAI frameworks leverage a synergistic workflow between the two scales. Peng et al. [60] offers a clear example of this synergy. The study first used global SHAP to understand network-wide trends in flood-induced pavement damage. It then employed local LIME to diagnose the specific impact of a flood on an individual road segment. This ability to shift from the general trend (what matters everywhere) to the specific cause (why it mattered here) is the hallmark of an effective explainable system. While global interpretability provides the scientific foundation, it is the local explanation that makes AI operational. The capacity to deliver a specific diagnosis for a single hillslope, building, or asset is what turns a prediction into actionable knowledge and builds the trust required for real-world adoption.

Despite the utility of global and local interpretations, it is crucial to distinguish between explaining a model's internal logic and establishing true physical causality. This review highlights that post-hoc tools like SHAP explain the model, not necessarily the physical reality. A critical comparative study by [64] demonstrated that different high-performance models (e.g., XGBoost vs. DenseNet) can yield contradictory feature attributions for the same landslide inventory, revealing that explanations are often model-dependent rather than physically absolute. Furthermore, standard XAI frameworks face significant challenges with time-variant features. As noted by [75], algorithms like SHAP were primarily designed for static tabular data; consequently, applying them to dynamic, time-series geohazard data often requires simplified aggregations that may obscure the temporal evolution of risk triggers, such as the changing lag-effects of rainfall.

3.2.2 Beyond Linear Importance: Uncovering Thresholds and Complex Interactions

Global feature importance rankings provide a valuable overview, but the most significant scientific insights from XAI are found when the analysis moves beyond simple linear attribution. The real value comes from exploring the non-linear, conditional, and interdependent relationships that drive complex Earth systems. High-performance ML models are particularly good at learning these nuanced patterns from data. In turn, XAI tools like SHAP dependence plots and interaction analyses provide a way to make these learned mechanisms transparent and open to interpretation. The synthesis of the literature conducted effectively reveals two critical capabilities within this domain: (1) The identification of quantitative, physically meaningful thresholds that translate a model's continuous predictions into actionable rules, and (2) The discovery of complex interaction effects that reveal the synergistic recipes for geohazards.

A key function of XAI is to demystify the complicated decision boundaries hidden inside any black-box model. Tools like SHAP dependence plots, for example, are highly effective at visualizing the intricate relationship between a specific feature's value and its isolated, marginal impact on the final model prediction. Significantly, this unique capability successfully reveals critical tipping points where the inherent geohazard susceptibility dramatically changes. This specific capability has been leveraged powerfully and successfully to accelerate the operationalization of predictive models, demonstrating particular utility for deployment within critical early warning systems [77]. For example, several specific studies [78,79] successfully utilized their XAI frameworks to directly extract the precise numerical parameters needed to trigger alerts. This led to the identification of clear thresholds—such as daily rainfall (>5 mm), antecedent moisture (>15 mm), and reservoir water level (~145 m)—which definitively signify a sharp increase in landslide risk.

This approach extends beyond dynamic triggers to static predisposing conditions, providing an evidence-based foundation for land management and zoning policies. The work on sinkhole susceptibility in Bilgilioğlu et al. [48] serves as a clear practical example. Instead of producing just a complex risk map, the model's internal logic was translated into a set of simple, testable rules for planners. For instance, the analysis established a straightforward geological rule: sinkhole risk increases substantially within 2000 m of a fault line. It also produced a novel hydrochemical rule, showing that risk also goes up when the seasonal bicarbonate difference exceeds 0.68. This ability to convert the complex reasoning of an AI model into a short list of practical, quantitative rules is what bridges the gap between an abstract prediction and an actionable, on-the-ground risk management strategy. As Fig. 5 illustrates, advanced AI can be translated into practical, on-the-ground risk management strategies.

Geohazards rarely stem from a single cause; instead, they typically arise from the complex interaction of multiple factors. A significant strength of XAI is its ability to look beyond the main effects of individual variables and quantify how they work together, whether they amplify (synergy) or reduce (antagonism) risk. The literature reviewed here reveals several classes of these complex interactions, deepening the mechanistic understanding of hazard processes (Fig. 6).

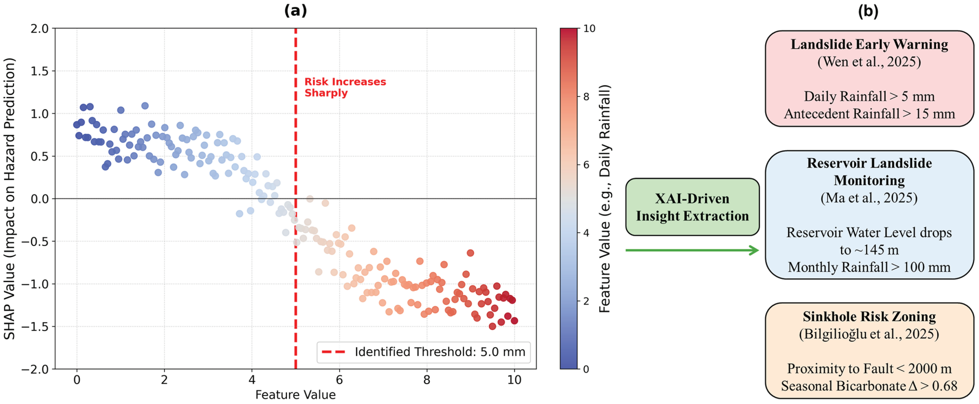
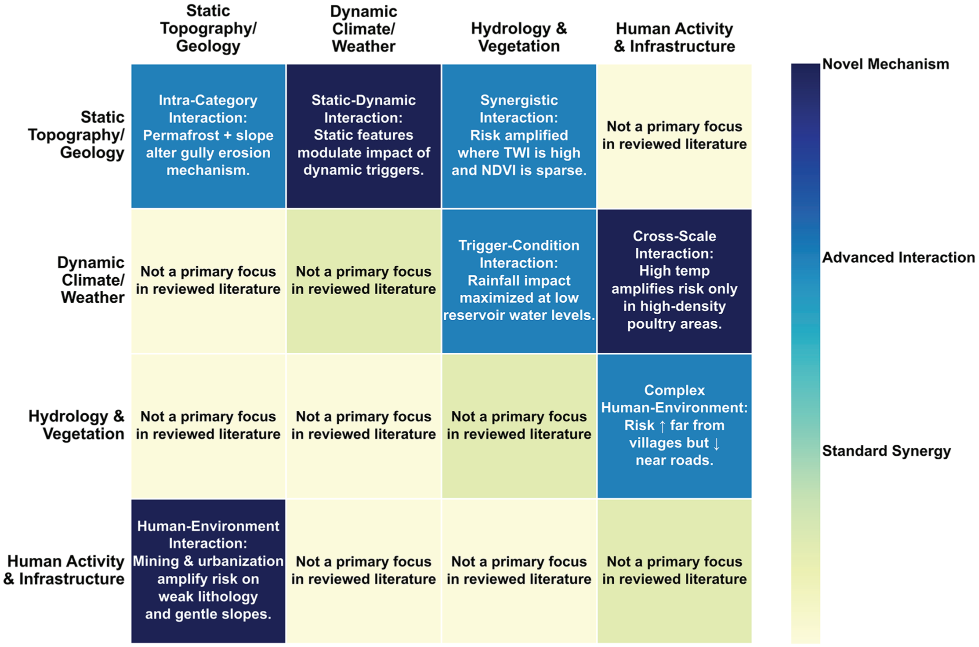


Figure 5: Conceptual workflow illustrating the translation of black-box model logic into actionable knowledge. This process is exemplified with studies that used XAI to derive quantitative thresholds for landslide early warning [78], reservoir monitoring [79], and sinkhole risk zoning [48]



	Static Topography/ Geology	Dynamic Climate/ Weather	Hydrology & Vegetation	Human Activity & Infrastructure
Static Topography/ Geology	Intra-Category Interaction: Permafrost + slope alter gully erosion mechanism.	Static-Dynamic Interaction: Static features modulate impact of dynamic triggers.	Synergistic Interaction: Risk amplified where TWI is high and NDVI is sparse.	Not a primary focus in reviewed literature
Dynamic Climate/ Weather	Not a primary focus in reviewed literature	Not a primary focus in reviewed literature	Trigger-Condition Interaction: Rainfall impact maximized at low reservoir water levels.	Cross-Scale Interaction: High temp amplifies risk only in high-density poultry areas.
Hydrology & Vegetation	Not a primary focus in reviewed literature	Not a primary focus in reviewed literature	Not a primary focus in reviewed literature	Complex Human-Environment: Risk ↑ far from villages but ↓ near roads.
Human Activity & Infrastructure	Human-Environment Interaction: Mining & urbanization amplify risk on weak lithology and gentle slopes.	Not a primary focus in reviewed literature	Not a primary focus in reviewed literature	Not a primary focus in reviewed literature

Figure 6: Synthesis of multi-factor interaction classes uncovered by XAI, exemplified by studies on static-dynamic coupling (Wang et al. [80], Slater and Coxon [66]), cross-scale environmental interactions (Wang et al. [81], Choubin et al. [62], Zhang et al. [82]), and the interplay between natural triggers and human activity (Ma et al. [79], Fang et al. [83], Nam et al. [53], Iban Aksu [84])

One important class is the cross-scale interaction, where a large-scale context alters the impact of a local-scale trigger. For instance, a study on avian influenza [81] found that high temperatures (>30°C)—normally considered low-risk—become a significant risk amplifier when they occur within a macro-scale context of high poultry density. Another critical category is the static-dynamic interaction, where a dynamic event changes the importance of a static landscape feature. Research in Wang et al. [80] found that the dynamic forces of a typhoon significantly amplified the role of static features like vegetation. This proves that the vulnerability of a landscape is not fixed but changes based on interacting events.

Crucially, XAI's utility extends beyond mere model validation; it has also proven essential for accurately quantifying complex human-environment interactions. For example, a notable investigation utilized an interpretable-by-design model to find that the combined effect of local mining activities and rainfall was a considerably more potent predictor of landslides in a Karst region than if those factors were assessed individually [83]. The key takeaway is this: when a predisposing anthropogenic stressor intersects with a natural trigger, the resultant risk of failure is powerfully elevated, often to a disproportionately high degree.

This ability to dissect a model's logic elevates XAI from a simple interpretation tool to a computational laboratory. It allows researchers to probe the inner workings of a trained model to generate new, data-driven, and often non-intuitive hypotheses about the complex, conditional, and synergistic mechanisms that drive geohazards. While these interpretive analyses deepen scientific understanding, the next step in GeoXAI's evolution is its integration into the modeling workflow itself—where explainability no longer follows modeling, but actively guides it.

3.3 XAI as an Active Agent: Enhancing the Geohazard Modeling Workflow

A significant paradigm shift has occurred in how XAI is used in geohazard science. It began as a post-hoc tool, something used simply to interpret a model after it was already built. Now, however, it is emerging as an active component that is integrated directly into the modeling workflow itself. This new paradigm is built around a core iterative feedback loop: Model → Explain → Refine → Final Model. In this in-the-loop approach, an explanation is not treated as a final report card after the fact. It becomes an active blueprint for making the model better. Researchers use these insights to guide and refine the entire modeling route—from how the data is prepared and features are engineered, all the way to the final optimization. The systems that come out of this process are more than just transparent; they are measurably more robust, efficient, and physically plausible.

3.3.1 Data-Driven Refinement: XAI for Feature Engineering and Sampling Strategies

The most notable application where XAI functions as an active agent involves the data-driven optimization of the complete modeling pipeline. This all-encompassing procedure spans every stage, from the critical initial steps of feature selection and engineering right through to the systematic generation of the necessary training data itself (Fig. 7).

One fundamental and practical aim of the in-the-loop methodology is the development of models that are both simpler and more efficient, a goal typically accomplished through strategic feature selection followed by thorough model simplification. XAI makes this possible by providing a quantitative ranking of predictor importance. This allows researchers to create a more focused (parsimonious) model by keeping only the most impactful features. A case study on post-fire gully erosion [85] shows this process in action. An initial, complex model used 21 different factors. After a consensus-based SHAP analysis across four different models, the researchers identified the eight most dominant factors. They subsequently trained a new, highly simplified model utilizing only those eight identified predictors. The outcome was striking: the much simpler model performed almost identically to its complex predecessor. The performance trade-off was minimal; the model's area under the curve (AUC) only dropped from 0.989 to 0.973. This provides powerful evidence that XAI can be used to reduce a model's complexity without a major sacrifice in predictive performance. This same technique—using SHAP to distill a complex research model into a more practical tool with fewer inputs—has also been effectively applied to ice-jam flooding [86]. Several other studies have similarly confirmed the tangible value of XAI when applied to rigorous feature selection [61,80,92].

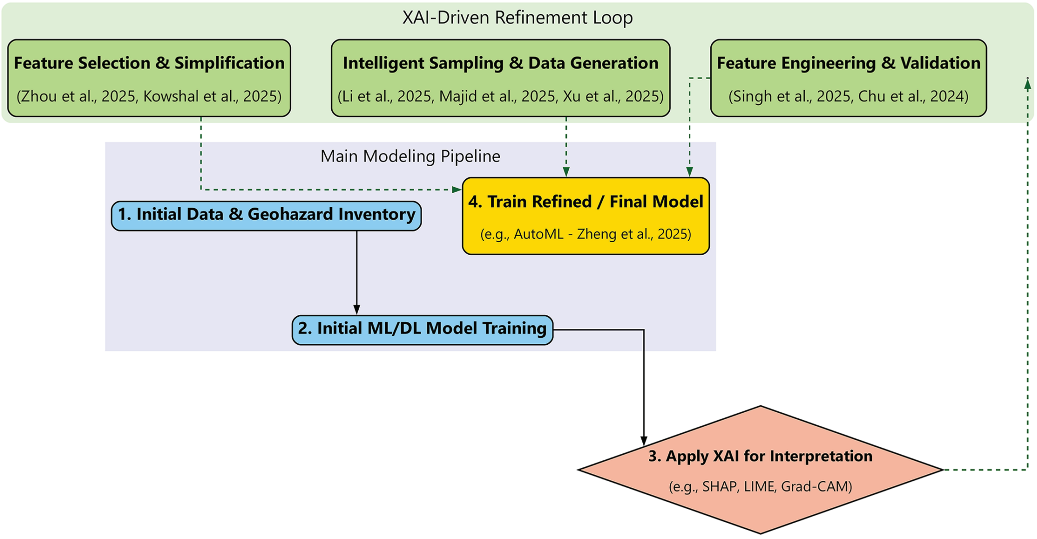

XAI-Driven Refinement Loop
Feature Selection & Simplification
(Zhou et al., 2025, Kowshal et al., 2025)
Intelligent Sampling & Data Generation
(Li et al., 2025, Majid et al., 2025, Xu et al., 2025)
Feature Engineering & Validation
(Singh et al., 2025, Chu et al., 2024)
Main Modeling Pipeline
1. Initial Data & Geohazard Inventory
4. Train Refined / Final Model
(e.g., AutoML - Zheng et al., 2025)
2. Initial ML/DL Model Training
3. Apply XAI for Interpretation
(e.g., SHAP, LIME, Grad-CAM)

Figure 7: The GeoXAI active modeling workflow, illustrating the use of XAI as an in-loop agent for model refinement. This is exemplified by studies that leverage XAI insights for feature selection and simplification (Zhou et al. [85], Kowshal et al. [86]), intelligent sampling and data generation (Li and Tian [87], Majid et al. [88], Xu et al. [89]), and feature engineering and validation (Singh and Roy [90], Chu et al. [91])

More innovatively, XAI is being used to guide intelligent data generation and sampling. A groundbreaking workflow developed by Li and Tian (2025) [87] demonstrates a powerful feedback loop for landslide susceptibility: (1) an initial model is trained; (2) SHAP is used to identify a rule for landscape stability (NDVI > 0.8); and (3) this rule is then used to filter the original random negative samples and replace them with a more physically plausible, higher-quality set. The implementation of this specific XAI-driven sampling methodology led to a dramatic enhancement in model performance, with the AUC soaring from 0.914 to an impressive 0.986. The principle of using data-driven insights to guide data generation, as exemplified by the XAI-driven workflow, is also central to other sophisticated methods such as hybrid modeling. These frameworks frequently leverage physics-based models (such as RIVICE, RUSLE, or TRIGRS) to effectively create large, synthetic, and physically consistent training datasets—a critical technique where observational data are naturally scarce [86,88,89]. Furthermore, utilizing unsupervised clustering strategically to engineer a more robust target variable out of raw historical data constitutes yet another sophisticated technique within the realm of data refinement [93].

XAI also serves as a critical validation tool for physics-informed feature engineering. In this workflow, domain knowledge is used to create new, powerful predictors, and XAI is then used to confirm their efficacy. For earthquake magnitude prediction [90], new features were engineered based on seismological principles; a subsequent SHAP analysis provided definitive proof of their value by showing that the new energy variable had become the single most dominant predictor in the model. A similar approach was used to validate engineered spatiotemporal features for urban flood forecasting, where SHAP confirmed the new features contributed significantly (approx. 14%) to the model's output [91].

The most advanced expression of this active refinement paradigm is AutoML, where the AI system itself takes on the role of refinement. An AutoML framework built for landslide susceptibility Zheng et al. [52] offers a clear example. It automates the entire modeling pipeline, from algorithm selection and hyperparameter tuning to the final ensembling. The system was able to produce a state-of-the-art model (AUC = 0.90) in only 156 s. This result showcases the immense potential of AutoML to accelerate the development of objective, high-performance, and reproducible geohazard models.

3.3.2 Building Confidence: XAI for Model Comparison and Validation

While continuous refinement efforts generate high-performing models, it is crucial to acknowledge that strong metric scores don't inherently equate to trustworthiness [94,95]. In fact, conventional validation—which often relies solely on a single measure, such as the AUC—is frequently a poor indicator of genuine reliability in real-world scenarios [96]. A model might achieve outstanding performance benchmarks yet rely on spurious correlations or physically implausible reasoning. Such hidden dependencies make the system brittle and untrustworthy, especially in mission-critical contexts [97]. The literature reviewed for this paper strongly suggests, therefore, that XAI offers the essential framework for a much deeper, multi-faceted validation. As vividly shown in Fig. 8, this XAI-driven methodology pushes beyond simple performance statistics, focusing instead on establishing confidence in a model's internal logic, its underlying robustness, and its concrete practical utility.

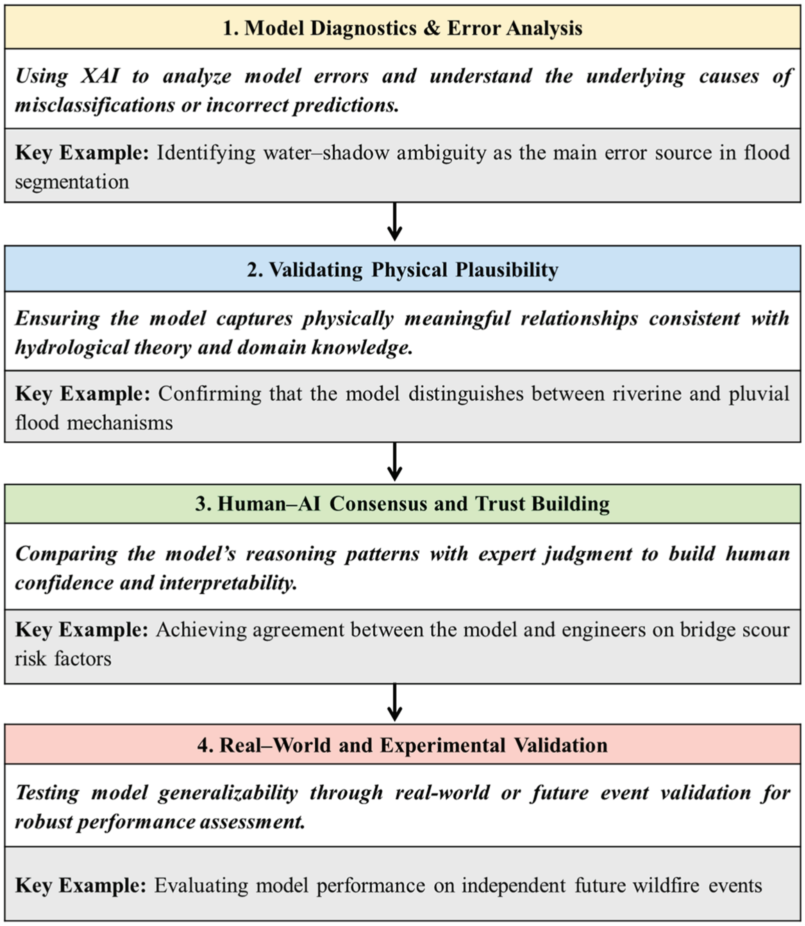

1. Model Diagnostics & Error Analysis
Using XAI to analyze model errors and understand the underlying causes of misclassifications or incorrect predictions.
Key Example: Identifying water–shadow ambiguity as the main error source in flood segmentation
2. Validating Physical Plausibility
Ensuring the model captures physically meaningful relationships consistent with hydrological theory and domain knowledge.
Key Example: Confirming that the model distinguishes between riverine and pluvial flood mechanisms
3. Human–AI Consensus and Trust Building
Comparing the model's reasoning patterns with expert judgment to build human confidence and interpretability.
Key Example: Achieving agreement between the model and engineers on bridge scour risk factors
4. Real–World and Experimental Validation
Testing model generalizability through real-world or future event validation for robust performance assessment.
Key Example: Evaluating model performance on independent future wildfire events

Figure 8: A spectrum of XAI-driven validation methods that build model trustworthiness, exemplified by studies on: (1) diagnostics and error analysis [56], (2) validation of physical plausibility [67], (3) establishing human-AI consensus [98], and (4) validation against independent, real-world events [99]

One of the most valuable and readily available applications of XAI is its use in model diagnostics and detailed error analysis. It's vital to recognize that the actual utility of XAI goes significantly beyond simply spotting that a model produced an error; critically, it offers researchers the unique ability to precisely pinpoint and fully understand the specific mechanism that caused the mistake [100,101]. A premier illustration of this diagnostic capability, for example, is the utilization of LIME for analyzing a state-of-the-art flood segmentation model [56]. XAI essentially offers a post-mortem analysis, allowing us to diagnose exactly why a model fails, even when its overall quantitative metrics are otherwise excellent. Take, for instance, a flood model that initially posted high overall accuracy scores: researchers utilized LIME heatmaps to carefully investigate the specific regions it had misclassified. The resulting heatmaps proved highly instrumental, revealing that the model's core failure mode was rooted in water-shadow ambiguity. To be specific, the model consistently mistook dark shadows for actual floodwater, a clear problem arising from their similar visual textures. This kind of specific, diagnostic insight is critical; crucially, you can't get it solely from an aggregate accuracy score. Pinpointing the exact underlying reason for the failure provides a clear, direct pathway for targeted improvements—a necessary effort that forms a fundamental step toward developing any truly robust predictive system. This same potent diagnostic capability, notably, was instrumental in explaining the catastrophic failure of an optical satellite model that struggled with cloudy conditions. The failure stemmed from the fact that the system had erroneously concentrated its analysis on the clouds themselves, entirely neglecting the underlying ground features it was initially engineered to observe [58].

Significantly, XAI moves beyond mere error diagnosis; it is, in fact, vital for validating the physical plausibility of a model's underlying reasoning [102]. This concept fundamentally involves using explainability tools as a necessary sanity check, which ensures the model has, in fact, learned scientifically sensible relationships instead of spurious ones [103]. A compelling illustration of this comes from research on a hydrology-aware DL model: validating it with Grad-CAM clearly demonstrated that the model had, in fact, acquired a sophisticated, adaptive reasoning strategy. Specifically, for riverine flood prediction, the model correctly focused its attention on the river network; however, it then smartly shifted its focus to local topography when dealing with pluvial floods occurring further away from the main channel [67]. Similarly, SHAP analysis was effectively used to confirm that a complex forecasting model had learned a physically sound heuristic by primarily relying on the most recent downstream data. This finding built confidence that the model's high accuracy was not just a result of spurious or non-physical patterns [104]. Ultimately, when an XAI-derived explanation aligns perfectly with known physical principles, such as the confirmed impact of a large dam on river dynamics [105], it provides powerful validation of the model's learned representation of the entire system.

Perhaps the most advanced form of validation is building trust through human-AI consensus. A deeper form of validation goes beyond data and physics to the epistemological level, where a model's logic is compared directly to the reasoning of human domain experts. A study on bridge scour risk [98] provides a powerful example. Confidence in the model was established not just through a high AUC, but by showing that both the model and a group of 26 field engineers came to the same conclusion: they were in unanimous agreement on the most important risk factor. This consensus between the AI's feature ranking and the experts' collective judgment is a far more convincing validation of the model's core logic than any statistical metric alone. A similar approach was used to validate a landslide model. In that case, the SHAP-based explanation for a specific event was shown to align with the findings of an independent, *in-situ* geotechnical investigation [106].

XAI is a key component of more rigorous validation frameworks for both experimental and real-world settings. The ultimate test of a model's generalizability is its performance on an entirely independent, future event. This gold standard approach was employed in one study by validating a wildfire damage model on the major 2025 Southern California wildfires. This test provided exceptional confidence in the model's actual utility for real-world scenarios [99]. Furthermore, moving past simple accuracy metrics, XAI enables a deep comparative validation process, allowing researchers to diagnose not merely whether one model outperforms another, but fundamentally why. As an example, a SHAP analysis was successfully used to reveal that a comparatively weaker support vector machine (SVM) model had learned relationships for key variables that were both physically implausible and often contradictory, particularly when assessed against a much more accurate RF model. This analytic step thus validated the superior physical logic and logical consistency inherent in the better-performing algorithm [107]. Such advanced validation strategies, fully empowered by XAI, are essential. They are actively helping the field transition from simply developing models that are accurate toward engineering systems that are demonstrably robust, physically plausible, and entirely trustworthy.

3.4 The Spatiotemporal Frontier: Challenges and Opportunities in GeoXAI

Despite the significant progress in applying XAI to geohazard assessment, a critical review of the literature points to a fundamental limitation [108]. The majority of current applications, while analyzing geographic phenomena, use ML and interpretation methods that are inherently aspatial and static [109]. In practice, most frameworks treat geographic data as a simple feature-based representation [75]. This approach ignores cornerstone principles of geography, such as spatial autocorrelation (the idea that nearby things are more related than distant things), spatial heterogeneity (the fact that relationships can change across a landscape), and the fundamental concept of scale [110]. This aspatial assumption—that a single, global model can capture the processes driving a hazard uniformly across a diverse landscape—is often invalid and can lead to models that are not only less accurate but whose explanations are incomplete or even misleading [111].

Adopting this spatial viewpoint immediately brings the major challenge of spatial heterogeneity (or non-stationarity) into focus. This describes the core phenomenon where the drivers of a geohazard, along with their relative levels of importance, change fundamentally from one location to another [112]. Given this fundamental challenge, a considerable portion of the research community is now deliberately transitioning away from traditional global, one-size-fits-all modeling strategies [113]. Therefore, the overarching objective has shifted toward creating far more sophisticated frameworks: those designed specifically to effectively capture and provide clear explanations for these critical spatial variations within the drivers of geohazards [114]. Broadly, these emerging solutions for effectively handling spatial heterogeneity can be sorted into three clear, distinct methodological approaches (Fig. 9). The first approach is defined as knowledge-driven stratification: this involves researchers applying deep domain expertise to logically partition the larger study area into smaller, more homogeneous and physically meaningful sub-regions. This necessary step occurs prior to initiating the primary modeling process. By building and interpreting separate models for distinct geomorphological zones or areas with different triggering mechanisms, studies have quantitatively proven that the dominant drivers of landslides are fundamentally context-dependent and vary significantly between these zones [64,115].

A second, more sophisticated approach employs geo-algorithmic models engineered specifically for inherent spatial awareness. A key example is the geographical RF (GRF), a technique that operates by iteratively fitting localized models within a moving spatial window. This localized adaptation is critical because it allows the GRF to account for particular regional conditions, consequently yielding significantly more realistic risk patterns. This methodology has been empirically shown to drastically outperform conventional aspatial models that ignore geographic context [116]. In a similar vein, the recent adoption of GNNs is noteworthy because these networks are fundamentally designed to learn from the topological relationships established among spatial units. The result is the production of geographically much more coherent and plausible regionalization outcomes [51]. The third, and arguably the cutting-edge, strategy is the development of GeoXAI-native frameworks. This strategy ensures that both the core predictive model and its accompanying interpretation methodology are developed from first principles to inherently handle spatial data intrinsically. From this area, key innovations are emerging, including GeoMLR (which explicitly incorporates geographic coordinates directly as modeling features) and the novel GeoShapley interpretation method. These tools are explicitly designed to quantitatively measure and, isolate purely spatial effects from other concurrent environmental factors [28]. A key output from these frameworks is the SHAP map, which visualizes how a feature's impact is distributed across space. This turns the abstract idea of feature importance into a tangible geographic pattern, offering a unique view into the model's spatial reasoning [117].

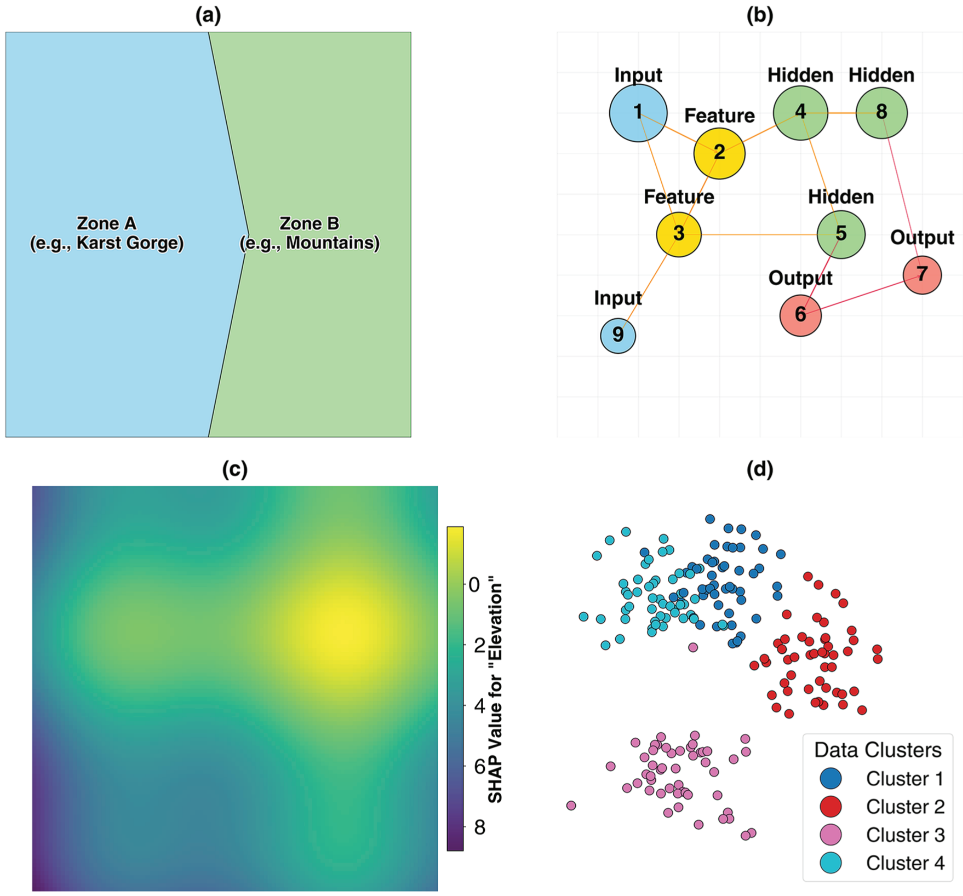


Figure 9: A comparison of four methodological approaches for analyzing spatial heterogeneity in GeoXAI. (**a**) Knowledge-driven Stratification, where the study area is pre-divided into homogeneous zones. (**b**) Geo-algorithmic approach, using inherently spatial models like GNNs. (**c**) GeoXAI-Native Interpretation, using spatially explicit explanation techniques like SHAP maps. (**d**) Data-driven discovery, using unsupervised clustering to automatically identify zones of similar behavior

Beyond the spatial frontier lies the temporal frontier. The vast majority of GeoXAI studies produce static susceptibility maps, which represent a long-term average probability but ignore the dynamic of geohazards. However, geohazards are dynamic processes [118,119]. A growing body of work is now using XAI to dissect these temporal dynamics (Fig. 10). This includes analyzing long-term dynamics and time-lag effects, moving beyond immediate triggers to understand how hazards are preconditioned over time. A landmark study using an LSTM-SHAP framework on wildfire data discovered that the risk of a summer fire was significantly influenced not just by summer weather, but also by the climate conditions of the preceding spring, which governed the growth of fuel [76]. Other research is tackling temporal non-stationarity by systematically evaluating how model interpretations change when trained on different chronological datasets [90].

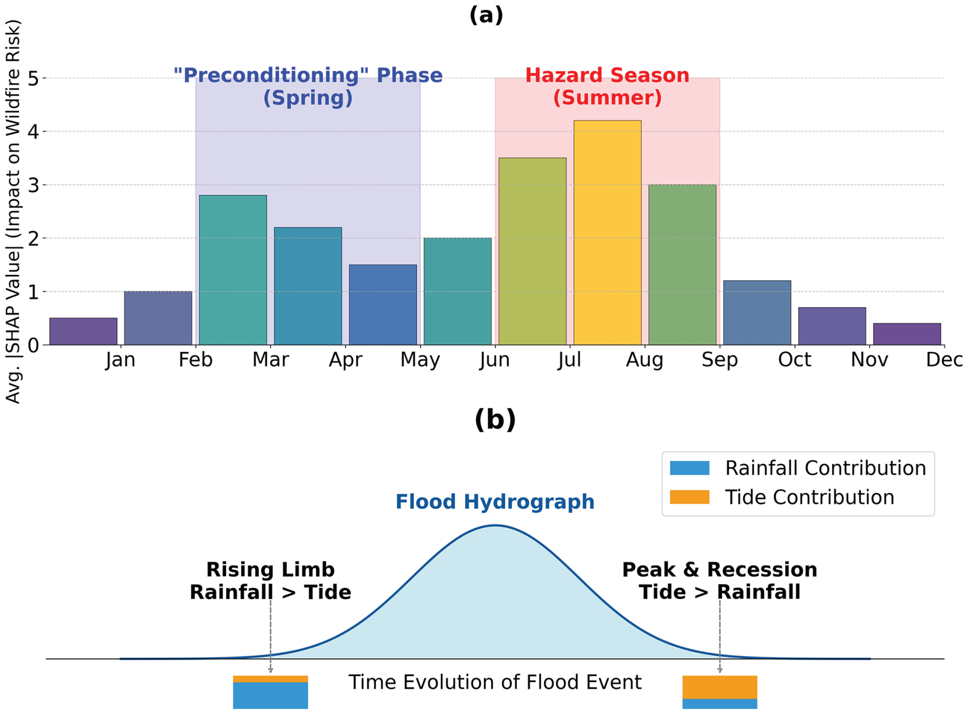


Figure 10: Visualization of two distinct types of spatio-temporal dynamics analyzed with XAI. (**a**) Uncovering long-term, time-lag effects, where antecedent conditions (the preconditioning phase such as spring climate) have a delayed but critical impact on a future hazard season (summer wildfire risk) [76]. (**b**) Dissecting intra-event dynamics, showing how the relative importance of different triggering factors (rainfall vs. tide) evolves during the course of a single compound flood event [120]

Analyzing the intra-event spatio-temporal dynamics is perhaps the most sophisticated application currently available. This methodology is unique because it specifically reveals how the drivers of a singular hazard event evolve across both spatial extent and temporal duration. Consider, for example, the dynamics of compound flooding: a time-dependent SHAP analysis distinctly illustrated the shifting hierarchy among the influential drivers throughout the event. Rainfall was found to be the most critical factor during the initial onset, yet high tides ultimately gained the dominant influence precisely when the flood reached its maximum peak [120]. Similarly, a parallel methodological approach was successfully employed to dissect the complex, internal kinematics of an active landslide. This specific analysis demonstrated how the relative impact of rainfall compared to reservoir levels fluctuated, not only across distinct segments of the landslide mass but also notably varied over the total duration of the event [63]. These pioneering investigations strongly indicate the trajectory for future GeoXAI work: a notable movement toward formulating and employing genuinely spatio-temporal models. Such models must be capable of capturing, and subsequently explaining, the complete, dynamic complexity intrinsically tied to geohazard processes.

To capture this dynamic nature, recent studies have moved beyond static mappings to employ specialized spatiotemporal architectures. The review highlights the adoption of LSTM networks for capturing temporal dependencies in flood runoff [105] and hybrid architectures like LSTM-MHA (multi-head attention) for modeling compound flood dynamics [120]. Furthermore, GNNs, such as the HydroGraphNet employed by [70], represent a significant leap, allowing for the modeling of flood propagation on unstructured meshes. However, these architectures present unique XAI challenges. Standard SHAP implementations often struggle to attribute importance to specific time-steps in recurrent layers. To address this, recent works have utilized time-dependent SHAP visualizations to track how feature influence evolves during an event [76], revealing that drivers like antecedent soil moisture may be critical only in specific pre-event windows.

This final synthesis, which spans from the initial thematic mappings to the dissection of complex spatio-temporal dynamics, lays out a clear evolutionary trajectory for GeoXAI [87,121]. The function of explainability has fundamentally transitioned [122]; it has moved beyond its foundational role as a post-hoc tool for model transparency and has become a sophisticated framework for generating new scientific knowledge [123]. To capture this evolution, we propose a conceptual model—the hierarchy of GeoXAI Insights—which organizes the state-of-the-art into a four-tiered hierarchy (Fig. 11). The framework's foundation is Model Transparency (Section 3.2.1), which is essential for building the trust required for real-world use. Building upon this, the next level is Mechanistic Insight (Section 3.2.2), where XAI is used to find non-linear thresholds and complex interactions. These insights then enable Active Model Refinement and Validation (Section 3.3), where XAI is integrated straightway within the modeling workflow to improve robustness and physical plausibility. The hierarchy culminates in the frontier of scientific discovery and causal inference (Section 3.4), where advanced GeoXAI is used not just to predict but to infer physical parameters, but also to discover novel hazard precursors, and dissect complex causal chains. This framework serves a dual function: it acts as a structured summary of the field's current capabilities and as a conceptual bridge connecting this review of the state-of-the-art to the future challenges in developing trustworthy GeoAI. Having now established the methodological and spatio-temporal frontiers of GeoXAI, the subsequent section will critically examine the unresolved challenges that must be addressed to ensure the reliability, robustness, and real-world deployment of these emerging systems.

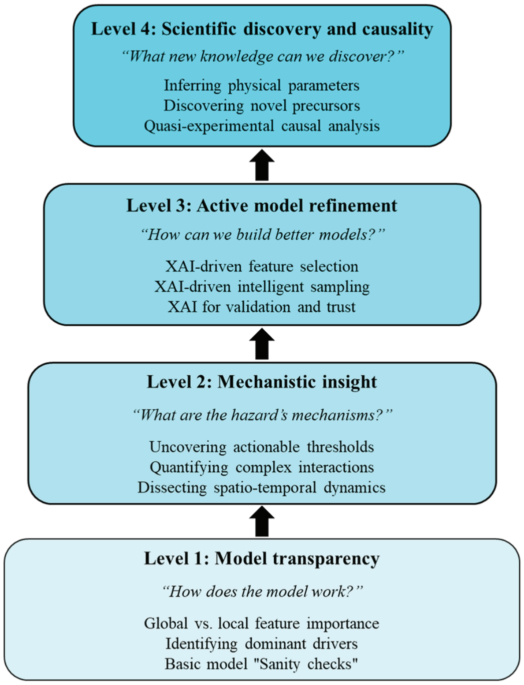

Level 4: Scientific discovery and causality
"What new knowledge can we discover?"
Inferring physical parameters
Discovering novel precursors
Quasi-experimental causal analysis
Level 3: Active model refinement
"How can we build better models?"
XAI-driven feature selection
XAI-driven intelligent sampling
XAI for validation and trust
Level 2: Mechanistic insight
"What are the hazard's mechanisms?"
Uncovering actionable thresholds
Quantifying complex interactions
Dissecting spatio-temporal dynamics
Level 1: Model transparency
"How does the model work?"
Global vs. local feature importance
Identifying dominant drivers
Basic model "Sanity checks"

Figure 11: The GeoXAI hierarchy of insights, a conceptual model illustrating the hierarchical progression of knowledge gained from XAI applications. The framework progresses from foundational model transparency (Matin and Pradhan [124], Pradhan et al. [113], Choubin et al. [62]), to deriving mechanistic insight (Wen et al. [78], Ma et al. [79], Wang et al. [81]), to enabling active model refinement (Zhou et al. [85], Li and Tian [87], Wang et al. [98]), and culminating in the research frontier of scientific discovery and causality (Dahal et al. [68], Graciosa et al. [125], Wei et al. [105])

4 Critical Challenges and Frontier Themes: Moving towards Trustworthy AI

The successful deployment of XAI across a varied spectrum of geohazards certainly represents a major stride forward for the field [126]. Yet, effectively moving these systems from promising laboratory results toward achieving truly reliable operational status fundamentally demands that the community successfully tackle a more complex array of interwoven challenges [127]. Interestingly, the very quest for transparency has brought a new issue into focus: even when the black-box is opened, the resulting explanations are not consistently stable, unique, or genuine reflections of the underlying physical reality [128]. Consequently, this section undertakes a critical review of the frontier themes poised to define the next ten years of GeoXAI research. The primary research themes center squarely on three critical areas: effectively managing the inherent uncertainty found in the explanations derived, successfully pushing the analysis past simple correlation toward genuine causation, and working toward a more rigorous, deeper integration of data-driven and physics-based models. To organize these key challenges and future-looking themes effectively, this review brings them together into a concise, forward-looking summary. Table 4 fulfills this specific function: it offers a systematic breakdown of the most vital research gaps identified, clearly clarifies the fundamental importance of these gaps for the broader discipline, and subsequently proposes prioritized trajectories for future investigation. This essential tabular framework, in turn, provides the direct basis for the comprehensive research roadmap detailed in the final section of this manuscript.

Table 4: Synthesis of the critical research gaps, their scientific relevance, and proposed directions for the next decade of GeoXAI research. The table consolidates evidence from 87 reviewed studies and structures the research roadmap into thematic priorities

Research gap	Problem statement and significance	Suggested research direction	Priority	Representative studies
Robustness and standardization	Interpretations are often unstable, varying with model choice [64] or data representation [129]. This undermines their scientific validity and trustworthiness.	• Develop quantitative metrics for interpretation stability. • Promote multi-model consensus analysis. • Quantify uncertainty in explanations.	High	• Yan et al. [55]: Exemplifies stability metrics (e.g., CV of feature ranks). • Pradhan et al. [129]: Exemplifies the impact of spatial units on interpretation.
GeoXAI-native algorithms	The common practice of applying aspatial XAI tools to spatial problems is a methodological mismatch that can yield physically implausible or incomplete explanations.	• Invest in inherently spatial algorithms (e.g., GNNs). • Develop glass-box models (e.g., PINNs, KANs). • Create spatially explicit explanation techniques (e.g., SHAP maps).	High	• Ke et al. [28]: Exemplifies a spatially-aware framework (GeoMLR). • Dahal and Lombardo [68]: Exemplifies an interpretable-by-design PINN.
Integration with causal reasoning	Models excel at identifying correlations but not causation, limiting their scientific value and the ability to design effective, targeted interventions.	• Employ XAI within quasi-experimental designs. • Use comparative XAI to dissect multi-stage process chains. • Integrate with formal causal inference frameworks.	High	• Wei et al. [105]: Exemplifies a quasi-experimental (before-and-after) design. • Liu et al. [130]: Exemplifies dissecting different process stages (initiation vs. development).
Operationalization and governance	A significant gap persists between developing interpretable research models and deploying them as robust, ethical, and practical decision-support tools for end-users.	Without a clear path to operationalization, the practical impact of GeoXAI research remains limited, and ethical issues like data bias go unaddressed.	High	• Develop MLOps frameworks for full lifecycle management. • Design human-AI collaborative systems. • Build actionable decision-support tools and ethical guidelines.

4.1 The Uncertainty of Explanations: How Robust Are Interpretations?

A core, often unstated, assumption in many XAI applications is that the generated explanation represents a singular, robust truth. The reality, however, is more complex [131]. An explanation is not a direct measurement of a natural process, but rather an artifact of a specific modeling pipeline [132]. As such, it is susceptible to significant uncertainty stemming from the choices made during that process (Fig. 12).

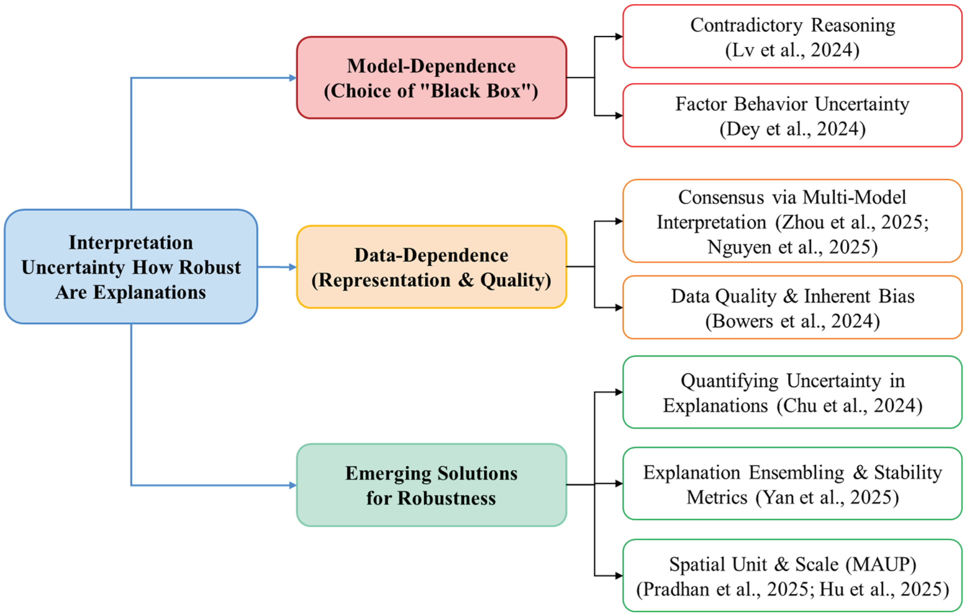

Interpretation Uncertainty How Robust Are Explanations
Model-Dependence (Choice of "Black Box")
Contradictory Reasoning (Lv et al., 2024)
Factor Behavior Uncertainty (Dey et al., 2024)
Data-Dependence (Representation & Quality)
Consensus via Multi-Model Interpretation (Zhou et al., 2025; Nguyen et al., 2025)
Data Quality & Inherent Bias (Bowers et al., 2024)
Emerging Solutions for Robustness
Quantifying Uncertainty in Explanations (Chu et al., 2024)
Explanation Ensembling & Stability Metrics (Yan et al., 2025)
Spatial Unit & Scale (MAUP) (Pradhan et al., 2025; Hu et al., 2025)

Figure 12: Conceptual mind map of the challenges and solutions related to the robustness of GeoXAI explanations. The analysis highlights key challenges such as model-dependence (Lv et al. [64], Dey, et al. [133]) and data-dependence (Pradhan et al. [129], Hu et al. [134], Bowers et al. [135]), along with emerging solutions like multi-model consensus (Zhou et al. [85], Nguyen Van et al. [136]), the use of stability metrics [55], and quantifying explanation uncertainty [91]

A primary source of this uncertainty comes from model dependence. This synthesis shows compelling evidence that different black-box models, even when they are equally accurate, can learn fundamentally different; and sometimes contradictory; mechanisms to solve the same problem. A critical investigation by Lv et al. [64], for instance, found that while both an XGBoost and a DenseNet model accurately predicted landslide susceptibility, their SHAP-based interpretations were at odds for key factors; XGBoost learned a negative correlation with relief, while the DenseNet model learned a positive one. This phenomenon, which Dey, Das and Roy [133] have termed factor behavior uncertainty, demonstrates that an explanation is a property of the model's learned solution, not necessarily a fixed law of the geohazard system. This is further backed up by studies on stacking ensembles. These show that individual base models might prioritize entirely different sets of features (e.g., topographic vs. climatic) but still contribute to a highly accurate final prediction [137]. The implication is that refering to the interpretation from a single, arbitrarily chosen model architecture can be a unreliable and potentially misleading approach.

A second, equally critical source of uncertainty is data-dependence. A model's interpretation of the world can be fundamentally altered by how geospatial data is represented, scaled, and sampled. The most profound challenge in this context is the modifiable areal unit problem (MAUP). A foundational study by [129] provided a stark example of this; they found that simply changing the spatial unit of analysis from slope units to hydrological response units didn't just change the model's accuracy—it completely inverted the feature importance hierarchy, swapping elevation and slope as the top-ranked landslide driver.

The choice of spatial scale has also been shown to have a significant impact on both model performance and interpretation stability [134]. In addition, the quality and inherent biases within the data itself can lead to robust explanations of an inaccurate reality. A crucial cautionary study by [135] represented how socioeconomic biases in flood insurance data led to a physically counter-intuitive explanation. It's a critical lesson: XAI will faithfully explain any bias a model learns from the data it is fed.

An efficient and highly effective strategy involves achieving a consensus-based interpretation by systematically applying XAI techniques across a diverse ensemble of models. Notably, research teams analyzing four [85] and six [136] distinct model architectures discovered a striking agreement on the most dominant geohazard drivers. To expand upon this encouraging conceptual basis, a more structured approach elevates the process: it involves actively combining the explanations themselves and then setting up specific quantitative metrics intended to measure their inherent stability. As an illustration, one pioneering work [55] put forward a novel hybrid framework specifically engineered to average the SHAP values derived from multiple models. Crucially, they introduced the coefficient of variation (CV) of feature ranks as a metric, proving that this ensemble explanation was substantially more stable than any single model's interpretation alone. The most advanced frontier, however, is the direct quantification of uncertainty in these explanations. By applying SHAP to probabilistic models, such as Bayesian neural networks (BNNs), researchers can generate a complete distribution of SHAP values for each feature. This capability then permits the clear calculation and visualization of confidence intervals (CI) surrounding an interpretation [91]. This clear evolution, which transitions from simply supplying an explanation, to rigorously ensuring its robustness, and ultimately to directly quantifying its inherent uncertainty, is indeed a truly critical step toward building genuinely trustworthy AI systems for comprehensive geohazard assessment.

4.2 From Correlation to Causation: The Next Frontier in Interpretability

A fundamental limitation currently affecting virtually all applications of ML within the geosciences stems from their inherently correlational nature. Indeed, even when leveraging the transparency provided by XAI, what is ultimately being explained remains the model's discovered association between patterns within the input data, rather than a definitively demonstrated causal mechanism [138]. A model may discover, for example, that Factor A is a robust predictor of Hazard B, yet it remains completely agnostic about whether A causes B, B causes A, or if some unobserved confounding factor C influences both. This crucial distinction isn't merely theoretical; it actually forms the bedrock of effective intervention. It follows, therefore, that attempting to mitigate a factor which is only correlated and lacks a true causal link proves to be a largely futile endeavor; however, focused intervention along a genuine causal pathway holds the authentic potential to genuinely avert catastrophic disasters [139]. Consequently, the single greatest challenge; and, simultaneously, the most significant opportunity; for the next decade of GeoXAI research lies in taking the difficult, yet essential, leap from merely explaining correlations to reliably inferring causality.

While the robust formal integration of causal inference frameworks remains in a unequivocally nascent state, this review nonetheless succeeds in highlighting key pioneering research already utilizing XAI to achieve substantially more causally informed insights. These researchers are moving past basic feature importance metrics to effectively disentangle the complex web of geohazard drivers. A key emerging strategy, for instance, uses XAI to cleanly differentiate predisposing conditions where from proximate triggers when. In short-term landslide forecasting, for example, a SHAP analysis showed that static factors like slope correlate strongly with the long-term spatial distribution of landslides, yet it is the dynamic trigger variables, specifically recent rainfall, that actually contain the predictive power for an imminent event [140]. This vital separation is inherently critical, as it truly represents the first major step required to achieve a causal understanding of how hazard initiation processes unfold.

A more sophisticated approach involves using comparative XAI to dissect the drivers of different stages within a causal chain. A novel framework for gully erosion [130] accomplished this by building and interpreting two separate models: one for the potential for gully formation and another for the risk of further development. The comparative SHAP analysis yielded a profound mechanistic insight: gully initiation was primarily driven by climatic factors, whereas the subsequent development was governed by a more complex interplay of topography, climate, and human activities. This demonstrates how XAI can be used to understand the evolving causal recipe of a multi-stage geomorphic process.

Perhaps the most powerful current approach for inferring causal impact involves using XAI to analyze quasi-experimental scenarios or natural experiments. A key study on riverine flooding [105] demonstrates this by using a before-and-after framework to quantify the impact of a massive intervention: the construction of the Three Gorges Dam. The comparative XAI analysis provided direct, compelling evidence that the system's causal structure had shifted. Before the dam's construction, downstream flooding was primarily driven by upstream runoff. After the dam was built, the analysis showed that the primary driver had changed to local precipitation. This use of XAI to illuminate the mechanistic changes resulting from a major intervention represents a significant step forward. It helps in the development of geohazard models that are more causally robust and scientifically powerful.

These studies are the vanguard of a critical new frontier. The future of GeoXAI will not be defined by simply building more accurate correlational models, but by integrating them with formal causal inference frameworks (e.g., structural causal models, do-calculus). In this new paradigm, XAI is used to generate testable causal hypotheses. The subsequent validation for these derived hypotheses would then need to be secured through other established methods, such as physics-based modeling, highly targeted field experiments, or rigorous quasi-experimental studies. This entire, crucial process marks a deliberate pivot away from mere simple pattern recognition and toward the dedicated discovery of underlying physical mechanisms. This necessary shift in focus is precisely what will change the primary function of GeoXAI; instead of remaining merely a powerful engineering tool for prediction, it successfully becomes an instrument for fundamentally advancing the scientific understanding of complex Earth systems.

4.3 Bridging Data and Physics: The Role of XAI in Hybrid and Physics-Informed Models

A fundamentally sound critique frequently directed at purely data-driven models centers on their inherent physics-agnostic nature. It is, in fact, quite true that when these computational frameworks operate entirely without the guidance of relevant physical constraints, even those models that attain excellent statistical accuracy run a substantial risk of learning spurious or inherently implausible correlations [141]. Consequently, this critical deficiency severely compromises their overall reliability, especially when they are tasked with critical operations such as extrapolation or, more broadly, with generating authentic scientific insight [142]. Consequently, this challenge has long created a visible great divide: highly flexible, though potentially error-prone, data-driven methods occupy one side, facing off against robust, yet frequently computationally intractable, physics-based models on the other [143]. The leading edge for designing truly trustworthy AI systems, especially those addressing geohazards, is now undeniably focused on successfully closing this methodological gap [144]. This unification is typically achieved by developing and deploying hybrid models, which are critical because they effectively merge the core strengths of both data-driven and physics-based paradigms [145]. This reviewed literature examined here, consequently, clearly charts an evolutionary trajectory pointing toward these more sophisticated, integrated hybrid architectures. Notably, as Fig. 13 vividly illustrates, XAI assumes an absolutely pivotal role in this progression, supplying the essential tools required to rigorously validate and illuminate the often-complex internal functions of these integrated systems.

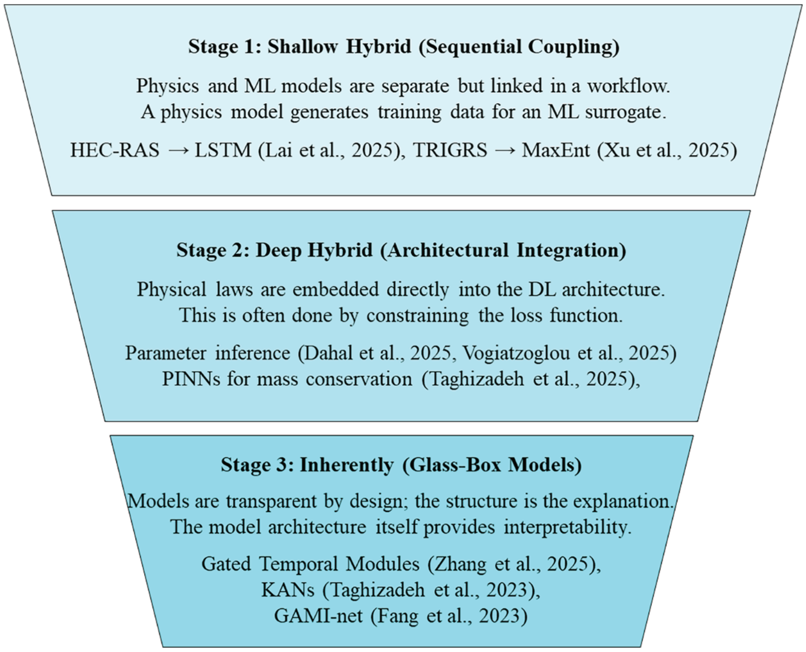

Stage 1: Shallow Hybrid (Sequential Coupling)
Physics and ML models are separate but linked in a workflow.
A physics model generates training data for an ML surrogate.
HEC-RAS → LSTM (Lai et al., 2025), TRIGRS → MaxEnt (Xu et al., 2025)
Stage 2: Deep Hybrid (Architectural Integration)
Physical laws are embedded directly into the DL architecture.
This is often done by constraining the loss function.
Parameter inference (Dahal et al., 2025, Vogiatzoglou et al., 2025)
PINNs for mass conservation (Taghizadeh et al., 2025),
Stage 3: Inherently (Glass-Box Models)
Models are transparent by design; the structure is the explanation.
The model architecture itself provides interpretability.
Gated Temporal Modules (Zhang et al., 2025),
KANs (Taghizadeh et al., 2023),
GAMI-net (Fang et al., 2023)

Figure 13: The evolutionary spectrum of data-physics integration in GeoAI, illustrating the progression from shallow hybrid models that sequentially couple physics and ML (Lai et al. [120], Xu et al. [89], Kowshal et al. [86]), to Deep Hybrid models with architectural integration of physical laws (Taghizadeh et al. [70], Dahal and Lombardo [68], Vogiatzoglou et al. [69]), and culminating in inherently interpretable glass-box systems (Zhang et al. [146], Fang et al. [83], Taghizadeh et al. [70])

The most established approach is shallow or sequential hybridization, where physics-based and ML models operate as distinct but connected components in a workflow. A dominant strategy is using physics to inform data generation, where a trusted, albeit computationally expensive, mechanistic model is used to create a large, physically plausible synthetic dataset. This dataset then serves to train a highly efficient ML surrogate model suitable for real-time applications. For the specific challenge of compound flooding, a hydrodynamic model (HEC-RAS) was utilized to train a DL surrogate; this model proved to be significantly faster—216 times faster, in fact. Furthermore, the crucial validation step involved SHAP analysis, which was instrumental in confirming that the surrogate had genuinely and successfully learned the correct underlying physical dynamics [120]. This approach has also been used to generate high-quality training data for landslide susceptibility by first identifying physically unstable slopes with a slope stability model (TRIGRS) [89].

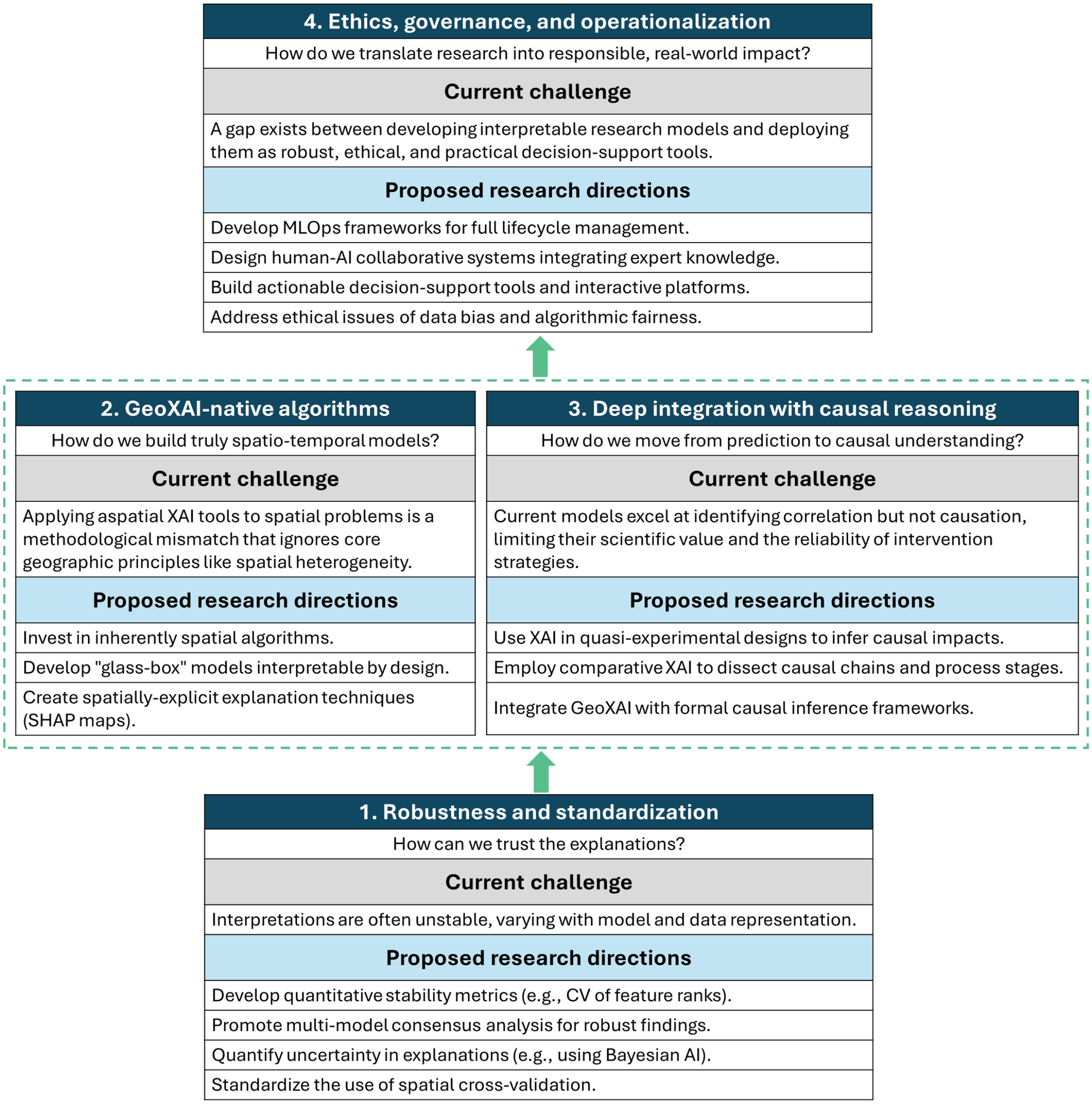
A more profound integration is found in deep hybridization, where physical principles are embedded directly into the architecture of a DL model. This is the domain of Physics-informed neural networks (PINNs). In this setup, the model's loss function is importantly augmented by including a term that directly penalizes any violation of the system's known governing equations; this could be, for example, the foundational laws of mass or energy conservation. This vital mechanism forces the network to find solutions that not only line up statistically with available observational data, but also inherently adhere to fundamental physics. Crucially, evidence consistently demonstrates that this approach notably lowers prediction errors in especially critical applications, such as in flood forecasting [70]. Furthermore, a genuinely transformative use of this framework appears when solving inverse problems: here, researchers have cleverly designed PINNs to infer previously unobserved, spatially-varying physical parameters—for instance, wildfire heat transfer coefficients [69] or certain geotechnical properties like the soil friction angle —simply by leveraging more accessible proxy environmental data [147]. Significantly, this successfully shifts the role of AI; it moves the system beyond being a mere passive predictor, establishing it instead as an active engine for scientific discovery.

The development of inherently interpretable, or glass-box models, clearly represents the logical culmination of this established evolutionary trend. In these advanced architectures, high predictive performance and essential transparency are not separated; rather, they are co-designed from the outset, entirely avoiding the pitfalls of clumsy or inefficient post-hoc retrofitting. This critical difference highlights a needed, fundamental shift: instead of just applying an explanatory tool like SHAP to an inherently opaque black-box, the next wave of models should be thoughtfully built using components that are intelligible by design. Several truly innovative architectures already demonstrate the viability of this movement. Several truly innovative architectures already illustrate this movement's direction. For instance, some models incorporate custom gated modules where the weights derived during the training process inherently function as the explanation [146]. Furthermore, important frameworks like GAMI-net are specifically built to isolate main effects from their corresponding interaction effects cleanly [83]. The emerging Kolmogorov-Arnold networks (KANs) likely represent the most promising development in this area. This methodology is potentially groundbreaking precisely because it employs activation functions that are both highly visualizable and, more importantly, remarkably learnable splines [70]. Importantly, these interpretable-by-design frameworks are frequently shown to reach state-of-the-art performance, effectively offering a genuine fusion of data-driven modeling power and physical domain understanding. Notably, they resolve the historical trade-off between accuracy and interpretability, clearly signaling a future where the most powerful geohazard models are also, by their very nature, the most transparent.

4.4 The Computational Tax of Interpretability

A critical, yet often underreported, barrier to the operational deployment of GeoXAI is the computational cost associated with generating explanations, particularly for high-dimensional geospatial datasets. This synthesis reveals a stark trade-off between model complexity and computational efficiency. While lightweight models like XGBoost can be trained in under a minute (e.g., 57.42 s in [148]), complex ensemble architectures coupled with SHAP calculations can increase processing time by nearly 600% (to 385.47 s). Similarly, spatially-explicit models impose heavy computational loads; [116] reported that a GRF required approximately 24 h for training due to the calculation of local estimators, compared to minutes for global models. Furthermore, physics-informed architectures (PINNs), while transparent, demand significant resources, with training times exceeding 11 h for wildfire spread parameterization [69]. Therefore, future research must balance the depth of explanation with the latency requirements of real-time early warning systems.

Fig. 14 synthesizes the evolution of GeoXAI into a conceptual model—the pyramid of GeoXAI Insights. This framework illustrates the progression from a simple tool for model transparency to a sophisticated method for scientific discovery, organizing the state-of-the-art into a four-tiered hierarchy. Building on the frontier challenges identified above, the following roadmap translates these conceptual issues into actionable research priorities and strategic directions for the coming decade of GeoXAI development.



4. Ethics, governance, and operationalization

How do we translate research into responsible, real-world impact?

Current challenge

A gap exists between developing interpretable research models and deploying them as robust, ethical, and practical decision-support tools.

Proposed research directions

- Develop MLOps frameworks for full lifecycle management.
- Design human-AI collaborative systems integrating expert knowledge.
- Build actionable decision-support tools and interactive platforms.
- Address ethical issues of data bias and algorithmic fairness.

2. GeoXAI-native algorithms

How do we build truly spatio-temporal models?

Current challenge

Applying aspatial XAI tools to spatial problems is a methodological mismatch that ignores core geographic principles like spatial heterogeneity.

Proposed research directions

- Invest in inherently spatial algorithms.
- Develop "glass-box" models interpretable by design.
- Create spatially-explicit explanation techniques (SHAP maps).

3. Deep integration with causal reasoning

How do we move from prediction to causal understanding?

Current challenge

Current models excel at identifying correlation but not causation, limiting their scientific value and the reliability of intervention strategies.

Proposed research directions

- Use XAI in quasi-experimental designs to infer causal impacts.
- Employ comparative XAI to dissect causal chains and process stages.
- Integrate GeoXAI with formal causal inference frameworks.

1. Robustness and standardization

How can we trust the explanations?

Current challenge

Interpretations are often unstable, varying with model and data representation.

Proposed research directions

- Develop quantitative stability metrics (e.g., CV of feature ranks).
- Promote multi-model consensus analysis for robust findings.
- Quantify uncertainty in explanations (e.g., using Bayesian AI).
- Standardize the use of spatial cross-validation.

Figure 14: A research roadmap for trustworthy GeoAI, structured around four foundational pillars exemplified by studies on robustness and standardization (Yan et al. [55], Zhou et al. [85], Chu et al. [91]), the development of GeoXAI-native algorithms (Li et al. [51], Dahal and Lombardo [68], Taghizadeh et al. [70]), the integration with causal reasoning (Wei et al. [105], Liu et al. [130]), and the challenges of operationalization and governance (Singh and Roy [90], Wang et al. [98], Wen et al. [78])

5 A Research Roadmap for the Next Decade

Although the incorporation of XAI has unquestionably yielded major advancements in geohazard assessment, this systematic synthesis brings to light several critical research gaps that must be immediately addressed to realize the full potential of this rapidly emerging domain. These pressing challenges aren't specific to one step; rather, they affect the entire modeling pipeline, ranging from the fundamental data utilized to fuel these models to the theoretical basis of the interpretations, and, crucially, their eventual, reliable operational deployment.

5.1 Data-Centric Challenges and Future Directions

Inventory quality and bias: The adage garbage in, garbage out is amplified in XAI; an explanation of a model trained on biased data is merely a transparent reflection of that bias. Geohazard inventories are often incomplete, spatially biased towards accessible areas, and lack consistent quality. Future research must move beyond simply using available data to actively developing frameworks for quantifying and propagating inventory uncertainty through the modeling pipeline. Methodologies like physics-informed data generation, data-driven sampling based on preliminary XAI insights, and the creation of community-driven, standardized benchmark datasets are critical for building models that learn a more faithful representation of reality.

Spatio-temporal data representation: Most current studies still rasterize complex geohazards into simple point or pixel representations, losing valuable information about the hazard's geometry, volume, and temporal evolution. The next generation of GeoXAI models must be built to handle richer data structures. This includes leveraging dynamic, multi-modal remote sensing time-series; provided that their fidelity is rigorously benchmarked against topographic biases [149]; integrating sensor network data, and utilizing graph-based representations that capture the topological relationships between spatial units, thus providing a more holistic input for both prediction and explanation.

Operationalizing models in data-scarce regions: A critical practical constraint in geohazard assessment is data scarcity, particularly for rare events like large-scale landslides or in ungauged catchments. This review demonstrates that XAI serves as a crucial validation layer when training data is limited. For instance, [50] successfully developed a rockbolt failure model using only 160 instances; here, SHAP was not just an explanation tool but a safety mechanism, confirming that the model did not overfit to noise but learned physically valid rules (e.g., roadway length influence). Similarly, in data-scarce urban flood settings, integrating active learning strategies with XAI has demonstrated the potential to maintain high segmentation accuracy with significantly fewer labeled samples by identifying and prioritizing ambiguous regions for labeling. Thus, XAI acts as a bridge, allowing researchers to deploy models in data-poor regions with a level of confidence that black-box metrics alone cannot provide.

5.2 Methodological and Interpretability Frontiers

The need for GeoXAI-native algorithms: The prevailing, dominant approach—applying ML and XAI algorithms that are fundamentally aspatial to problems inherently structured in space—represents a significant methodological compromise. A substantial research gap therefore emerges: the urgent need for GeoXAI-native algorithms. Within these necessary future architectures, core geographic principles—specifically spatial heterogeneity, scale, and autocorrelation—must be integrated directly as core architectural components, rather than being treated simply as subsequent adjustments or afterthoughts. The pioneering successes already demonstrated by intrinsically spatial models, such as GNNs and Geographically Weighted (GW) models, already offer a compelling, clear direction. When these are coupled with the strong potential of glass-box architectures like PINNs and the new KANs, the path forward becomes unambiguous. The future of GeoXAI absolutely lies in deliberately building models that are interpretable by design and inherently structured to effectively handle the unique, complex challenges posed by spatio-temporal geohazard data.

Technical pathways for interpretable spatial DL: Moving beyond general recommendations, this review identifies specific architectural strategies to reconcile the trade-off between the complexity of spatial models (like GNNs) and interpretability. The first pathway is intrinsic integration; rather than applying post-hoc tools to opaque GNN embeddings, researchers can embed interpretable layers directly into the network. A prime example is the integration of KANs within a GNN architecture, as demonstrated by [70], where learnable spline activation functions allow for the direct visualization of non-linear transformations at each graph node. The second pathway is spatially-adapted attribution; modifying standard XAI algorithms to explicitly treat geographic location as a variable. Study [28] introduced GeoShapley, which decomposes the standard SHAP value into intrinsic feature effects and location effects, thereby preventing the model's spatial biases from being conflated with physical feature importance. Future work should prioritize these hybrid architectures over the naive application of standard SHAP to spatial graph models.

The science of interpretation robustness: As this review has shown, explanations can be fragile. A critical frontier is the development of a formal science of interpretation that moves beyond generating a single explanation to assess its reliability. Future work must standardize the use of multi-model consensus analysis, create formal metrics to quantify interpretation stability across data perturbations, and develop methods to calculate and visualize the CI around feature attributions, for instance, through the application of XAI to probabilistic Bayesian models.

Defining quantitative metrics for trustworthiness: To transition trust from a subjective concept to an engineering requirement, future GeoXAI frameworks must adopt standardized quantitative metrics. This review identifies three promising candidates for community adoption:

1. **Stability metrics:** The CV of feature importance rankings across cross-validation folds, as demonstrated by [55], provides a scalar measure of explanation robustness against data perturbations.
2. **Uncertainty quantification:** The calculation of CI around SHAP values using probabilistic architectures (e.g., BNNs), as shown by [91], allows decision-makers to distinguish between certain and ambiguous risk drivers.
3. **Consistency indices:** The consistency index (C) proposed by [28], which quantifies the alignment between feature attribution and physical monitoring data.

5.3 From Explanation to Causation and Operation

Bridging correlation and causation: The most significant theoretical gap is that current models are primarily correlational. For these models to achieve genuine value for intervention and policy formulation, they must furnish clear insights into the underlying causal mechanisms. Consequently, a key, non-negotiable component of the future research agenda must involve formally integrating GeoXAI with established causal inference frameworks. This crucial integration will empower the community to effectively move beyond simple predictive modeling. Ultimately, this paradigm shift will allow us to use models reliably for several advanced purposes: rigorously testing causal hypotheses, systematically evaluating the potential impact of different mitigation strategies, and directly addressing crucial counterfactual questions concerning geohazard processes.

The last mile problem; operationalization and governance: A gap persists between producing an interpretable research model and deploying a trusted operational tool. Future research must focus on the full lifecycle of GeoAI systems, including MLOps frameworks for continuous monitoring and maintenance. Furthermore, a critical agenda is the development of human-AI collaborative systems that effectively integrate data-driven explanations with expert knowledge. This must be done alongside the creation of ethical frameworks to govern their use in high-stakes decision-making, ensuring that the transparency provided by XAI leads to fair, equitable, and effective risk management. Taken together, these data-centric, methodological, and operational priorities set the stage for a more transparent and causally informed generation of geohazard models, as summarized in the concluding section below.

6 Limitations and Future Directions

While this review follows a rigorous PRISMA workflow, several limitations should be acknowledged to contextualize the findings. First, the search strategy relied primarily on the WoS core collection. Although WoS is a premier database for high-impact natural science research, this choice may have excluded engineering-oriented studies indexed exclusively in Scopus, IEEE Xplore, or relevant grey literature. Second, restricting the search to English-language publications may overlook important regional GeoXAI contributions, particularly from non-English-speaking countries with high geohazard exposure. Third, given the fast-evolving nature of AI, the review reflects a specific snapshot of the XAI boom (2021–2025), and newer methodological advances may not yet be represented. Finally, publication bias—where journals tend to publish positive or methodologically successful outcomes—could influence the apparent effectiveness of XAI methods, as unsuccessful or inconclusive applications are rarely reported.

7 Conclusion

This systematic review charted the emergence and rapid maturation of XAI within geohazard science, a new paradigm appropriately designated here as GeoXAI. The collective synthesis of recent literature decisively reveals a fundamental and necessary shift in the field: researchers are now moving deliberately away from opaque black-box predictions and toward the development of models that are transparent, scientifically insightful, and ultimately trustworthy. A key contribution of GeoXAI lies in its inherent capacity to transform any predictive model into a computational laboratory, thereby facilitating a detailed dissection of underlying hazard mechanisms. This capability encompasses identifying critical, actionable thresholds for early warning systems, uncovering complex, synergistic interactions among conditioning factors, and providing a multi-faceted validation framework that, through rigorous error diagnosis, necessary checks for physical plausibility, and direct comparison with established human expertise, serves to bolster confidence. The shift from global to local diagnoses has represented a crucial step forward, making these powerful insights tangible for on-the-ground decision-makers. In parallel, the deliberate integration of these methodologies with physics-based principles is giving rise to a new generation of models that are both demonstrably more accurate and scientifically sound. Looking to the future, the continued advancement of this burgeoning frontier depends on a focused and strategic research effort. Future work should prioritize establishing new community standards for both model robustness and reproducibility—an imperative that includes the routine implementation of multi-model consensus analysis and the mandatory quantification of interpretation uncertainty. Alongside this, researchers must strategically invest in creating algorithms inherently designed for the unique properties of spatio-temporal data (GeoXAI-native), which marks a vital evolution from mere post-hoc explanation toward true glass-box transparency. The field must also deliberately shift its focus beyond correlational analysis and toward causal inference, an advancement that will empower models to not only predict a phenomenon but also to thoroughly explain the fundamental why behind it. Finally, focused attention on the critical last mile of operational deployment is essential, necessitating the creation of robust ethical governance frameworks and integrated human-AI collaborative systems to ensure these powerful technologies are deployed both effectively and equitably. The XAI frameworks synthesized in this review serve as key drivers of this progress by strengthening the physical interpretability of complex models and providing a transparent lens into their internal logic. By focusing on these pillars, the GeoXAI community can realize the full potential of this technology and foster a new generation of geohazard science that is more predictive, interpretable, and societally relevant.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Vahid Nourani; methodology, Vahid Nourani; software, Moein Tosan; validation, Sameh A. Kantoush and Mekonnen Gebremichael; formal analysis, Ruhollah Taghizadeh-Mehrjardi and Jinhui Jeanne Huang; investigation, Moein Tosan; resources, Vahid Nourani; data curation, Ozgur Kisi and Yongqiang Zhang; writing—original draft preparation, Moein Tosan; writing—review and editing, Vahid Nourani, Ozgur Kisi, Yongqiang Zhang, Sameh A. Kantoush, Mekonnen Gebremichael, Ruhollah Taghizadeh-Mehrjardi and Jinhui Jeanne Huang; visualization, Vahid Nourani; supervision, Vahid Nourani. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study, including the full list of reviewed articles and the detailed search query, are available from the Corresponding Author, Moein Tosan, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Wang X, Cao Y, Jiang S, Zhu D, Yang M, Su C. Assessing seasonal geohazard risk related to precipitation on traditional villages under climate change: a region-level analysis of southwest Yunnan. China Geomat Nat Hazards Risk. 2025;16(1):2529927. doi:10.1080/19475705.2025.2529927.
2. Godes CR, Lim OB, Rodrigazo SA, Kim Y, Kim Y, Lee S, et al. Advancing geohazard risk assessment—a digital tool for steep slope management. Prog Disaster Sci. 2025;26:100418. doi:10.1016/j.pdisas.2025.100418.
3. Zhang Z, Tyc J, Hensel M. An ecogeomorphological approach to land-use planning and natural hazard risk mitigation: a literature review. Land. 2025;14(9):1911. doi:10.3390/land14091911.
4. Li B, Liu K, Wang M, Zhu W, Jiang Z, Qiao N, et al. Spatiotemporal warning of rainfall-induced railway geohazards in China. Eng Geol. 2025;355(6):108234. doi:10.1016/j.enggeo.2025.108234.
5. Du W, Fu X, Sheng Q, Chen J, Zhou Y, Zheng S. Physics-based and data-driven long-term evolution of a landslide: from inversion to prediction. Eng Geol. 2025;355(1):108252. doi:10.1016/j.enggeo.2025.108252.
6. Li N, Hu X, Wang J, Huang J, Zheng H, Li W. A physics-based probabilistic method for assessment of landslide-induced wave run-up hazards. J Rock Mech Geotech Eng. 2025;94:S1674775525003609. doi:10.1016/j.jrmge.2025.06.018.
7. Kariminejad N, Amindin A, Sepehr A, Pourghasemi HR. Projecting the effect of climate change on multiple Geomorphological hazard using machine learning data driven approaches. Sci Rep. 2025;15(1):18333. doi:10.1038/s41598-025-03176-4.
8. Kisi O, Heddam S, Keshtegar B, Piri J, Adnan R. Predicting daily streamflow in a cold climate using a novel data mining technique: radial M5 model tree. Water. 2022;14(9):1449. doi:10.3390/w14091449.
9. Wang J, Sun P, Chen L, Yang J, Liu Z, Lian H. Recent advances of deep learning in geological hazard forecasting. Comput Model Eng Sci. 2023;137(2):1381–418. doi:10.32604/cmes.2023.023693.
10. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. doi:10.1023/A:1010933404324.
11. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13–17; San Francisco, CA, USA. p. 785–94. doi:10.1145/2939672.2939785.
12. Tosan M, Nourani V, Kisi O, Dastourani M. Evolution of ensemble machine learning approaches in water resources management: a review. Earth Sci Inform. 2025;18(2):416. doi:10.1007/s12145-025-01911-z.
13. He K, Chen X, Yu X, Dong C, Zhao D. Evaluation and prediction of compound geohazards in highly urbanized regions across China's Greater Bay Area. J Clean Prod. 2024;449:141641. doi:10.1016/j.jclepro.2024.141641.
14. Wang Z, Yu Y, Chen Y, Liu Z, He H, Guo Z, et al. Geohazard sensitivity evaluation in Xinning, Hunan, China, using random forest, artificial neural network, and logistic regression algorithms. Nat Hazards Rev. 2025;26(2):05024017. doi:10.1061/nhrefo.nheng-2138.

15. Kisi O, Mirboluki A, Naganna SR, Malik A, Kuriqi A, Mehraein M. Comparative evaluation of deep learning and machine learning in modelling pan evaporation using limited inputs. Hydrol Sci J. 2022;67(9):1309–27. doi:10.1080/02626667.2022.2063724.
16. Wang L, Yang K, Wu C, Zhou Y, Liu J, Hu H. Novel approach to quantitative risk assessment of reservoir landslides using a hybrid CNN-LSTM model. Landslides. 2025;22(3):943–56. doi:10.1007/s10346-024-02398-3.
17. Sabzipour B, Arsenault R, Troin M, Martel JL, Brissette F, Brunet F, et al. Comparing a long short-term memory (LSTM) neural network with a physically-based hydrological model for streamflow forecasting over a Canadian catchment. J Hydrol. 2023;627(6):130380. doi:10.1016/j.jhydrol.2023.130380.
18. Ren X, Liu W, Yang W, Mao L, Li H. Landslide deformation uncertainty quantification using conformalized graph neural networks: a case study in Sichuan Province. China IEEE Access. 2025;13(3):86719–31. doi:10.1109/access.2025.3568273.
19. Yu B, Li H, Xing H, Ge W, Zhou L, Zhang J, et al. Geospatial SHAP interpretability for urban road collapse susceptibility assessment: a case study in Hangzhou. China Geomat Nat Hazards Risk. 2025;16(1):2491473. doi:10.1080/19475705.2025.2491473.
20. Yu B, Yan J, Li Y, Xing H. Risk assessment of multi-hazards in Hangzhou: a socioeconomic and risk mapping approach using the CatBoost-SHAP model. Int J Disaster Risk Sci. 2024;15(4):640–56. doi:10.1007/s13753-024-00578-2.
21. Nourani V, Tosan M, Huang JJ, Gebremichael M, Kantoush SA, Dastourani M. Advances in multi-source data fusion for precipitation estimation: remote sensing and machine learning perspectives. Earth Sci Rev. 2025;270(5):105253. doi:10.1016/j.earscirev.2025.105253.
22. Alqadhi S, Mallick J, Alkahtani M, Ahmad I, Alqahtani D, Hang HT. Developing a hybrid deep learning model with explainable artificial intelligence (XAI) for enhanced landslide susceptibility modeling and management. Nat Hazards. 2024;120(4):3719–47. doi:10.1007/s11069-023-06357-4.
23. Zarchi M, Nazari RA, Tee KF. Explainable multi-attribute machine learning via a hierarchical nature-inspired system toward predicting geological hazards. Model Earth Syst Environ. 2025;11(4):259. doi:10.1007/s40808-025-02421-z.
24. Dağdeviren U, Demir AS, Erden C, Kökçam AH, Kurnaz TF. Explainable AI using ensemble machine learning with integrated SHapley additive explanations (SHAP)-Borda approach for estimation of the safety factor against soil liquefaction. Environ Earth Sci. 2025;84(17):507. doi:10.1007/s12665-025-12466-z.
25. Razavi-Termeh SV, Sadeghi-Niaraki A, Ali Naqvi R, Choi SM. Investigating the efficacy of physics-based metaheuristic algorithms in combination with explainable ensemble machine-learning models for landslide susceptibility mapping. Stoch Environ Res Risk Assess. 2025;39(3):1109–41. doi:10.1007/s00477-025-02911-7.
26. Guang X, Xiang L, He Y, Zhang N, Zheng J, Qin Y, et al. Development and validation of a Geospatial eXplainable Artificial Intelligence (GeoXAI) framework for mapping mosquito density in metropolitans. Appl Geogr. 2025;181(3):103685. doi:10.1016/j.apgeog.2025.103685.
27. Jiang S, Sweet LB, Blougouras G, Brenning A, Li W, Reichstein M, et al. How interpretable machine learning can benefit process understanding in the geosciences. Earth's Future. 2024;12(7):e2024EF004540. doi:10.1029/2024EF004540.
28. Ke E, Zhao J, Zhao Y. Investigating the influence of nonlinear spatial heterogeneity in urban flooding factors using geographic explainable artificial intelligence. J Hydrol. 2025;648:132398. doi:10.1016/j.jhydrol.2024.132398.
29. Dong T, Wang R, Chen P, Sun C, Gan C, Liu J, et al. Submarine terrain generalization in nautical charts: a survey of traditional methods and graph neural network solutions. ISPRS Int J Geo Inf. 2025;14(7):257. doi:10.3390/ijgi14070257.
30. Huang J, Wen H, Hu J, Liu B, Zhou X, Liao M. Deciphering decision-making mechanisms for the susceptibility of different slope geohazards: a case study on a SMOTE-RF-SHAP hybrid model. J Rock Mech Geotech Eng. 2025;17(3):1612–30. doi:10.1016/j.jrmge.2024.03.008.
31. Ding Z, Zhang Y, Wang Z, Fang H, Chen Q. A metaverse-based visual analysis approach for 3D reservoir models. Open Geosci. 2025;17(1):20250873. doi:10.1515/geo-2025-0873.

32. Fazilova D, Magdiev K, Makhmudov M, Fazilov A. A multi-criteria GIS model for geohazard assessment in the Charvak reservoir area. Uzbekistan Egypt J Remote Sens Space Sci. 2025;28(3):587–96. doi:10.1016/j.ejrs.2025.09.003.

33. de Jesus Pugliese Viloria A, Folini A, Carrion D, Brovelli MA. Hazard susceptibility mapping with machine and deep learning: a literature review. Remote Sens. 2024;16(18):3374. doi:10.3390/rs16183374.

34. Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F. A review of statistically-based landslide susceptibility models. Earth Sci Rev. 2018;180(3):60–91. doi:10.1016/j.earscirev.2018.03.001.

35. Merghadi A, Yunus AP, Dou J, Whiteley J, ThaiPham B, Bui DT, et al. Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. Earth Sci Rev. 2020;207(11):103225. doi:10.1016/j.earscirev.2020.103225.

36. Dikshit A, Pradhan B, Alamri AM. Pathways and challenges of the application of artificial intelligence to geohazards modelling. Gondwana Res. 2021;100(6488):290–301. doi:10.1016/j.gr.2020.08.007.

37. Bentivoglio R, Isufi E, Jonkman SN, Taormina R. Deep learning methods for flood mapping: a review of existing applications and future research directions. Hydrol Earth Syst Sci. 2022;26(16):4345–78. doi:10.5194/hess-26-4345-2022.

38. Ghaffarian S, Taghikhah FR, Maier HR. Explainable artificial intelligence in disaster risk management: achievements and prospective futures. Int J Disaster Risk Reduct. 2023;98(3):104123. doi:10.1016/j.ijdrr.2023.104123.

39. Tosan M. Enhancing evapotranspiration estimation: a bibliometric and systematic review of hybrid neural networks in water resource management. Comput Model Eng Sci. 2025;142(2):1109–54. doi:10.32604/cmes.2025.058595.

40. Tugwell P, Tovey D. Prisma 2020. J Clin Epidemiol. 2021;134:A5–6. doi:10.1016/j.jclinepi.2021.04.008.

41. Tosan M, Dastourani M, Akbarpour A, Gharib MR. Global trend analysis of numerical simulation application in groundwater based on WoS database using VOSviewer and Biblioshiny between 1997 and 2023. Iran J Rainwater Catchment Syst. 2024;12(2):79–104.

42. Mardani M, Tosan M, Nasirian A, Dastourani M. A bibliometric analysis of research trends on the application of remote sensing in precipitation estimation with an emphasis on spatio-temporal analysis in Iran. Iran J Rainwater Catchment Syst. 2025;13(2):101–18.

43. Pranckutė R. Web of science (WoS) and Scopus: the titans of bibliographic information in today's academic world. Publications. 2021;9(1):12. doi:10.3390/publications9010012.

44. Akbarpour A, Dastourani M, Tosan M, Gharib MR. Performance analysis of finite element method in groundwater studies based on Web of Science using R Biblioshiny. J Aquifer Qanat. 2024;4(2):131–48.

45. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:1–10.

46. Andaryani S, Nourani V, Pradhan B, Ansarudi TJ, Ershadfath F, Haghighi AT. Spatiotemporal evaluation of future groundwater recharge in arid and semi-arid regions under climate change scenarios. Hydrol Sci J. 2022;67(6):979–95. doi:10.1080/02626667.2022.2050732.

47. Nourani V, Pradhan B, Ghaffari H, Sharifi SS. Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. Nat Hazards. 2014;71(1):523–47. doi:10.1007/s11069-013-0932-3.

48. Bilgilioğlu SS, Gezgin C, Iban MC, Bilgilioğlu H, Gündüz HI, Arslan Ş. Explainable sinkhole susceptibility mapping using machine-learning-based SHAP: quantifying and comparing the effects of contributing factors in Konya, Türkiye. Appl Sci. 2025;15(6):3139. doi:10.3390/app15063139.

49. Yu B, Xing H, Ge W, Yan J, Li YA. Explainable machine learning-based land subsidence susceptibility mapping: from feature importance to individual model contributions in ensembled system. Earth Sci Inform. 2025;18(2):407. doi:10.1007/s12145-025-01915-9.

50. Ibrahim B, Ahenkorah I, Ewusi A. Explainable risk assessment of rockbolts' failure in underground coal mines based on categorical gradient boosting and SHapley additive exPlanations (SHAP). Sustainability. 2022;14(19):11843. doi:10.3390/su141911843.

51. Li Y, Zhang C, Cui P, Hassan M, Duan Z, Bhattacharyya S, et al. Flash flood regionalization for the Hengduan Mountains Region, China, combining GNN and SHAP methods. Remote Sens. 2025;17(6):946. doi:10.3390/rs17060946.

52. Zheng D, Li Y, Yan C, Wu H, Yamashiki YA, Gao B, et al. Landslide susceptibility assessment using AutoML-SHAP method in the southern foothills of Changbai Mountain. China Landslides. 2025;22(6):1855–75. doi:10.1007/s10346-025-02462-6.

53. Nam K, Lee Y, Lee S, Kim S, Zhang S. Explainable artificial intelligence (XAI) for flood susceptibility assessment in Seoul: leveraging evolutionary and Bayesian AutoML optimization. Remote Sens. 2025;17(13):2244. doi:10.3390/rs17132244.

54. Alshayeb MJ, Hang HT, Ali A Shohan A, Ali Bindajam A. Novel optimized deep learning algorithms and explainable artificial intelligence for storm surge susceptibility modeling and management in a flood-prone island. Nat Hazards. 2024;120(6):5099–128. doi:10.1007/s11069-024-06414-6.

55. Yan X, Zhang D, Han Y, Li T, Zhong P, Ning Z, et al. Developing a hybrid model to enhance the robustness of interpretability for landslide susceptibility assessment. ISPRS Int J Geo Inf. 2025;14(7):277. doi:10.3390/ijgi14070277.

56. Purohit S, Pujari P. Enhancing flood segmentation in aerial imagery using a dual-decoder efficient U-Net with CBAM and post-hoc explainable AI technique. Eng Res Express. 2025;7(3):035296. doi:10.1088/2631-8695/ae0004.

57. Teke A, Kavzoglu T. Exploring the decision-making process of ensemble learning algorithms in landslide susceptibility mapping: insights from local and global explainable AI analyses. Adv Space Res. 2024;74(8):3765–85. doi:10.1016/j.asr.2024.06.082.

58. Sanderson J, Mao H, Abdullah MAM, Al-Nima RRO, Woo WL. Optimal fusion of multispectral optical and SAR images for flood inundation mapping through explainable deep learning. Information. 2023;14(12):660. doi:10.3390/info14120660.

59. Sanderson J, Tengtrairat N, Woo WL, Mao H, Al-Nima RR. XFIMNet: an Explainable deep learning architecture for versatile flood inundation mapping with synthetic aperture radar and multi-spectral optical images. Int J Remote Sens. 2023;44(24):7755–89. doi:10.1080/01431161.2023.2288945.

60. Peng L, Gao L, Hong F, Sun J. Evaluating pavement deterioration rates due to flooding events using explainable AI. Buildings. 2025;15(9):1452. doi:10.3390/buildings15091452.

61. Gholami H, Mohammadifar A, Golzari S, Song Y, Pradhan B. Interpretability of simple RNN and GRU deep learning models used to map land susceptibility to gully erosion. Sci Total Environ. 2023;904(1):166960. doi:10.1016/j.scitotenv.2023.166960.

62. Choubin B, Jaafari A, Henareh J, Karimi O, Sajedi Hosseini F. Explainable artificial intelligence (XAI) for interpreting predictive models and key variables in flood susceptibility. Res Eng. 2025;27(3):105976. doi:10.1016/j.rineng.2025.105976.

63. Yao W, Zhang X, Li C, Lv Y, Fu Y, Criss RE, et al. Synergistic evolution of hydrological and movement characteristics of Majiagou landslide and identification of key triggering factors through interpretable machine learning. Bull Eng Geol Environ. 2025;84(2):91. doi:10.1007/s10064-025-04116-2.

64. Lv L, Chen T, Liu G, Dou J, Plaza A. A comparative study of model interpretability considering the decision differentiation of landslide susceptibility models. IEEE Trans Geosci Remote Sens. 2024;63(3):4401218. doi:10.1109/TGRS.2024.3522165.

65. Gao S, Wang Y. Explainable deep learning powered building risk assessment model for proactive hurricane response. Risk Anal. 2023;43(6):1222–34. doi:10.1111/risa.13990.

66. Slater L, Coxon G, Brunner M, McMillan H, Yu L, Zheng Y, et al. Spatial sensitivity of river flooding to changes in climate and land cover through explainable AI. Earth's Future. 2024;12(5):e2023EF004035. doi:10.1029/2023ef004035.

67. Wang J, Sanderson J, Iqbal S, Woo WL. Accelerated and interpretable flood susceptibility mapping through explainable deep learning with hydrological prior knowledge. Remote Sens. 2025;17(9):1540. doi:10.3390/rs17091540.

68. Dahal A, Lombardo L. Towards physics-informed neural networks for landslide prediction. Eng Geol. 2025;344(2):107852. doi:10.1016/j.enggeo.2024.107852.

69. Vogiatzoglou K, Papadimitriou C, Bontozoglou V, Ampountolas K. Physics-informed neural networks for parameter learning of wildfire spreading. Comput Meth Appl Mech Eng. 2025;434(31):117545. doi:10.1016/j.cma.2024.117545.

70. Taghizadeh M, Zandsalimi Z, Nabian MA, Shafiee-Jood M, Alemazkoor N. Interpretable physics-informed graph neural networks for flood forecasting. Comput Aided Civ Infrastruct Eng. 2025;40(18):2629–49. doi:10.1111/mice.13484.

71. Das T, Talukdar S, Shahfahad, Naikoo MW, Ahmed IA, Rahman A, et al. Integration of fuzzy AHP and explainable AI for effective coastal risk management: a micro-scale risk analysis of tropical cyclones. Prog Disaster Sci. 2024;23(1):100357. doi:10.1016/j.pdisas.2024.100357.

72. Yi X, Zhao J, Feng W, Guo C, Zhao Y, Xue Z, et al. Data-driven insights into the characteristics and drivers of the June 16, 2024 clustered shallow landslides in Southeastern China. Landslides. 2025;22(9):3049–64. doi:10.1007/s10346-025-02586-9.

73. Merabet K, Di Nunno F, Granata F, Kim S, Adnan RM, Heddam S, et al. Predicting water quality variables using gradient boosting machine: global vs. local explainability using SHapley Additive Explanations (SHAP). Earth Sci Inform. 2025;18(3):298. doi:10.1007/s12145-025-01796-y.

74. Cilli R, Elia M, D'Este M, Giannico V, Amoroso N, Lombardi A, et al. Explainable artificial intelligence (XAI) detects wildfire occurrence in the Mediterranean countries of Southern Europe. Sci Rep. 2022;12(1):16349. doi:10.1038/s41598-022-20347-9.

75. Jena R, Shanableh A, Al-Ruzouq R, Pradhan B, Gibril MBA, Ali Khalil M, et al. Explainable artificial intelligence (XAI) model for earthquake spatial probability assessment in Arabian Peninsula. Remote Sens. 2023;15(9):2248. doi:10.3390/rs15092248.

76. Li H, Vulova S, Rocha AD, Kleinschmit B. Exploring the monthly contribution of drivers on European summer wildfires with explainable artificial intelligence (XAI). Ecol Indic. 2025;176(14):113605. doi:10.1016/j.ecolind.2025.113605.

77. Najafi H, He Z, Huang JJ. A climate-informed early warning framework for urban water pipe leakage: integrating environmental drivers with LSTM based risk prediction [Internet]. [cited 2025 Dec 1]. Available from: https://www.researchsquare.com/article/rs-7089419/v1.

78. Wen H, Yan F, Huang J, Li Y. Interpretable machine learning models and decision-making mechanisms for landslide hazard assessment under different rainfall conditions. Expert Syst Appl. 2025;270(4):126582. doi:10.1016/j.eswa.2025.126582.

79. Ma T, Wu L, Zhou J, Zhang H, Xiao H. An interpretable hybrid model for predicting step-like landslide displacement: a case study in the Three Gorges Reservoir. Nat Hazards. 2025;121(18):21441–58. doi:10.1007/s11069-025-07638-w.

80. Wang F, Zhou L, Zhao J, Liu Y, Chen J, Wen Z, et al. Selection of optimal factor combinations for typhoon-induced landslides susceptibility mapping using machine learning interpretability. Geomorphology. 2025;484(20):109855. doi:10.1016/j.geomorph.2025.109855.

81. Wang X, Xu Y, Xi X. Spatiotemporal risk assessment of H5 avian influenza in China: an interpretable machine learning approach to uncover multi-scale drivers. Animals. 2025;15(16):2447. doi:10.3390/ani15162447.

82. Zhang W, Zhao Y, Zhang F, Shi X, Zeng C, Maerker M. Understanding the mechanism of gully erosion in the alpine region through an interpretable machine learning approach. Sci Total Environ. 2024;949(5):174949. doi:10.1016/j.scitotenv.2024.174949.

83. Fang H, Shao Y, Xie C, Tian B, Shen C, Zhu Y, et al. A new approach to spatial landslide susceptibility prediction in karst mining areas based on explainable artificial intelligence. Sustainability. 2023;15(4):3094. doi:10.3390/su15043094.

84. Iban MC, Aksu O. SHAP-driven explainable artificial intelligence framework for wildfire susceptibility mapping using MODIS active fire pixels: an in-depth interpretation of contributing factors in Izmir, türkiye. Remote Sens. 2024;16(15):2842. doi:10.3390/rs16152842.

85. Zhou R, Hu X, Xi C, He K, Deng L, Luo G. Unveiling dominant factors for gully distribution in wildfire-affected areas using explainable AI: a case study of Xiangjiao catchment, Southwest China. J Mt Sci. 2025;22(8):2765–92. doi:10.1007/s11629-025-9645-9.
86. Kowshal A, Das A, Lindenschmidt KE. Ice-jam flood predictions using an interpretable machine learning approach. Environ Model Softw. 2025;192(1):106534. doi:10.1016/j.envsoft.2025.106534.
87. Li M, Tian H. Insights from optimized non-landslide sampling and SHAP explainability for landslide susceptibility prediction. Appl Sci. 2025;15(3):1163. doi:10.3390/app15031163.
88. Majid SI, Kumar M, Bhadwal S. Application of explainable artificial intelligence to decode water-induced soil erosion in Lidder watershed of the Greater Himalayas. Stoch Environ Res Risk Assess. 2025;39(9):3903–36. doi:10.1007/s00477-025-03042-9.
89. Xu X, Qiang Y, Li L, Liang S, Chen T, Yang W, et al. A MaxEnt-TRIGRS hybrid model with dynamic safety factor mapping for enhanced debris flow susceptibility assessment in rainfall-triggered terrains. Sci Rep. 2025;15(1):26209. doi:10.1038/s41598-025-11284-4.
90. Singh R, Roy B. Explainable earthquake magnitude prediction with hybrid modeling and spatio-temporal data for scalability. Earth Sci Inform. 2025;18(2):355. doi:10.1007/s12145-025-01867-0.
91. Chu W, Zhang C, Li H, Zhang L, Shen D, Li R. SHAP-powered insights into spatiotemporal effects: unlocking explainable Bayesian-neural-network urban flood forecasting. Int J Appl Earth Obs Geoinf. 2024;131:103972. doi:10.1016/j.jag.2024.103972.
92. Alam MG, Tripathi V, Bhatt CM, Mohanty MP. A novel framework embedding Bayesian-optimized ensemble machine learning and explainable artificial intelligence (XAI) to improve flood prediction in complex watersheds. Environ Sustain Indic. 2025;27(2):100760. doi:10.1016/j.indic.2025.100760.
93. Liu C, Mostafavi A. FloodGenome: interpretable machine learning for decoding features shaping property flood risk predisposition in cities. Environ Res Infrastruct Sustain. 2025;5(1):015018. doi:10.1088/2634-4505/adb800.
94. Gebremichael M, Tien YC, Nourani V. Comprehensive evaluation of seasonal forecasts from NMME and statistical models over the Blue Nile Basin and the Grand Ethiopian Renaissance Dam (GERD) watershed. J Hydrol Reg Stud. 2025;57(11):102162. doi:10.1016/j.ejrh.2024.102162.
95. Thekdi S, Tatar U, Santos J, Chatterjee S. On the compatibility of established methods with emerging artificial intelligence and machine learning methods for disaster risk analysis. Risk Anal. 2025;45(4):863–77. doi:10.1111/risa.17640.
96. Pahde F, Wiegand T, Lapuschkin S, Samek W. Ensuring medical AI safety: interpretability-driven detection and mitigation of spurious model behavior and associated data. Mach Learn. 2025;114(9):206. doi:10.1007/s10994-025-06834-w.
97. Faddi Z, Da Mata K, Silva P, Nagaraju V, Ghosh S, Kul G, et al. Quantitative assessment of machine learning reliability and resilience. Risk Anal. 2025;45(4):790–807. doi:10.1111/risa.14666.
98. Wang T, Reiffsteck P, Chevalier C, Chen CW, Schmidt F. An interpretable model for bridge scour risk assessment using explainable artificial intelligence and engineers' expertise. Struct Infrastruct Eng. 2025;21(4):643–55. doi:10.1080/15732479.2023.2230564.
99. Jia Y, Opabola EA. Interpretable machine learning insights into wildfire damage drivers in California. USA Int J Disaster Risk Reduct. 2025;126:105610. doi:10.1016/j.ijdrr.2025.105610.
100. Nauta M, Trienes J, Pathak S, Nguyen E, Peters M, Schmitt Y, et al. From anecdotal evidence to quantitative evaluation methods: a systematic review on evaluating explainable AI. ACM Comput Surv. 2023;55(13s):1–42. doi:10.1145/3583558.
101. Kang S, Chen B, Yoo S, Lou JG. Explainable automated debugging via large language model-driven scientific debugging. Empir Softw Eng. 2024;30(2):45. doi:10.1007/s10664-024-10594-x.
102. Lin S, Liang Z, Guo H, Hu Q, Cao X, Zheng H. Application of machine learning in early warning system of geotechnical disaster: a systematic and comprehensive review. Artif Intell Rev. 2025;58(6):168. doi:10.1007/s10462-025-11175-0.
103. Khan D, Akram W, Ullah S. Enhancing landslide susceptibility predictions with XGBoost and SHAP: a data-driven explainable AI method. Geocarto Int. 2025;40(1):2514725. doi:10.1080/10106049.2025.2514725.

104. Huang H, Wang Z, Liao Y, Gao W, Lai C, Wu X, et al. Improving the explainability of CNN-LSTM-based flood prediction with integrating SHAP technique. Ecol Inform. 2024;84:102904. doi:10.1016/j.ecoinf.2024.102904.
105. Wei X, Wang G, Grosser PF, Schmalz B. Using explainable artificial intelligence (XAI) methods to understand the nonlinear relationship between the Three Gorges Dam and downstream flood. J Hydrol Reg Stud. 2024;53(8):101776. doi:10.1016/j.ejrh.2024.101776.
106. Zhou X, Wen H, Li Z, Zhang H, Zhang W. An interpretable model for the susceptibility of rainfall-induced shallow landslides based on SHAP and XGBoost. Geocarto Int. 2022;37(26):13419–50. doi:10.1080/10106049.2022.2076928.
107. Al-Najjar HAH, Pradhan B, Beydoun G, Sarkar R, Park HJ, Alamri A. A novel method using explainable artificial intelligence (XAI)-based SHapley Additive Explanations for spatial landslide prediction using Time-Series SAR dataset. Gondwana Res. 2023;123(2):107–24. doi:10.1016/j.gr.2022.08.004.
108. Ye C, Wu H, Oguchi T, Tang Y, Pei X, Wu Y. Physically based and data-driven models for landslide susceptibility assessment: principles, applications, and challenges. Remote Sens. 2025;17(13):2280. doi:10.3390/rs17132280.
109. Zhang J, Ma X, Zhang J, Sun D, Zhou X, Mi C, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. J Environ Manage. 2023;332(2):117357. doi:10.1016/j.jenvman.2023.117357.
110. Wang Z, Kong Y, Xiao K, Cao C, Li Y, Wu Y, et al. An interpretable ensemble learning framework based on remote sensing for ecological-geological environment evaluation: the case of Laos. Remote Sens. 2025;17(18):3240. doi:10.3390/rs17183240.
111. Pradhan B, Dikshit A, Lee S, Kim H. An explainable AI (XAI) model for landslide susceptibility modeling. Appl Soft Comput. 2023;142(10):110324. doi:10.1016/j.asoc.2023.110324.
112. Song Y, Li Y, Zou Y, Wang R, Liang Y, Xu S, et al. Synergizing multiple machine learning techniques and remote sensing for advanced landslide susceptibility assessment: a case study in the Three Gorges Reservoir Area. Environ Earth Sci. 2024;83(8):227. doi:10.1007/s12665-024-11521-5.
113. Pradhan B, Lee S, Dikshit A, Kim H. Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model. Geosci Front. 2023;14(6):101625. doi:10.1016/j.gsf.2023.101625.
114. Yang H, Wang T, Fomin NI, Xiao S, Liu L. An integrated IKOA-CNN-BiGRU-Attention framework with SHAP explainability for high-precision debris flow hazard prediction in the Nujiang river basin. China PLoS One. 2025;20(6):e0326587. doi:10.1371/journal.pone.0326587.
115. Sun D, Chen D, Zhang J, Mi C, Gu Q, Wen H. Landslide susceptibility mapping based on interpretable machine learning from the perspective of geomorphological differentiation. Land. 2023;12(5):1018. doi:10.3390/land12051018.
116. Quevedo RP, Maciel DA, Uehara TDT, Vojtek M, Rennó CD, Pradhan B, et al. Consideration of spatial heterogeneity in landslide susceptibility mapping using geographical random forest model. Geocarto Int. 2022;37(25):8190–213. doi:10.1080/10106049.2021.1996637.
117. Wang N, Zhang H, Dahal A, Cheng W, Zhao M, Lombardo L. On the use of explainable AI for susceptibility modeling: examining the spatial pattern of SHAP values. Geosci Front. 2024;15(4):101800. doi:10.1016/j.gsf.2024.101800.
118. Jiang H, Ding M, Li L, Huang W. Global dynamic landslide susceptibility modeling based on ResNet18: revealing large-scale landslide hazard evolution trends in China. Appl Sci. 2025;15(4):2038. doi:10.3390/app15042038.
119. Razavi-Termeh SV, Sadeghi-Niaraki A, Ali F, Choi SM. Improving flood-prone areas mapping using geospatial artificial intelligence (GeoAI): a non-parametric algorithm enhanced by math-based metaheuristic algorithms. J Environ Manage. 2025;375(8):124238. doi:10.1016/j.jenvman.2025.124238.
120. Lai C, Liao Y, Yu H, Wang Z, Liao Y, Yang B, et al. Formation mechanism analysis and the prediction for compound flood arising from rainstorm and tide using explainable artificial intelligence. J Environ Manage. 2025;388(3):125858. doi:10.1016/j.jenvman.2025.125858.
121. Xing J, Sieber R. The challenges of integrating explainable artificial intelligence into GeoAI. Trans GIS. 2023;27(3):626–45. doi:10.1111/tgis.13045.
122. Mai G, Xie Y, Jia X, Lao N, Rao J, Zhu Q, et al. Towards the next generation of geospatial artificial intelligence. Int J Appl Earth Obs Geoinf. 2025;136(1):104368. doi:10.1016/j.jag.2025.104368.
123. Roussel C, Böhm K. Geospatial XAI: a review. ISPRS Int J Geo Inf. 2023;12(9):355. doi:10.3390/ijgi12090355.

124. Matin SS, Pradhan B. Earthquake-induced building-damage mapping using explainable AI (XAI). Sensors. 2021;21(13):4489. doi:10.3390/s21134489.

125. Graciosa JC, Corbi F, Capitanio FA. Uncovering deformation prior to analogue megathrust earthquakes with explainable artificial intelligence. Geophys Res Lett. 2025;52(12):e2024GL114428. doi:10.1029/2024GL114428.

126. Ghiurău D, Popescu DE. Distinguishing reality from AI: approaches for detecting synthetic content. Computers. 2025;14(1):1. doi:10.3390/computers14010001.

127. Mamalakis A, Barnes EA, Ebert-Uphoff I. Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. Artif Intell Earth Syst. 2022;1(4):e220012. doi:10.1175/aies-d-22-0012.1.

128. Emmert-Streib F, Yli-Harja O, Dehmer M. Explainable artificial intelligence and machine learning: a reality rooted perspective. Wires Data Min Knowl Discov. 2020;10(6):e1368. doi:10.1002/widm.1368.

129. Pradhan AMS, Ghimire P, Shrestha S, Lee JS, Lee JH, Park HJ. Extreme gradient boosting with Shapley Additive Explanations for landslide susceptibility at slope unit and hydrological response unit scales. Geosci Front. 2025;16(4):102081. doi:10.1016/j.gsf.2025.102081.

130. Liu X, Wang D, Guo M, Zhang X, Chen Z, Wan Z, et al. Determination of future gully erosion risk and its spatially quantitative interpretability of driving factors in regional scale using machine learning algorithms. Geoderma. 2025;459(8):117396. doi:10.1016/j.geoderma.2025.117396.

131. Chou YL, Moreira C, Bruza P, Ouyang C, Jorge J. Counterfactuals and causability in explainable artificial intelligence: theory, algorithms, and applications. Inf Fusion. 2022;81(10):59–83. doi:10.1016/j.inffus.2021.11.003.

132. Medina-Ortiz D, Khalifeh A, Anvari-Kazemabad H, Davari MD. Interpretable and explainable predictive machine learning models for data-driven protein engineering. Biotechnol Adv. 2025;79(3):108495. doi:10.1016/j.biotechadv.2024.108495.

133. Dey S, Das S, Roy SK. Landslide susceptibility assessment in eastern Himalayas, India: a comprehensive exploration of four novel hybrid ensemble data driven techniques integrating explainable artificial intelligence approach. Environ Earth Sci. 2024;83(22):641. doi:10.1007/s12665-024-11945-z.

134. Hu W, Yang Z, Yang J, Li Q, Deng J, Zhao S, et al. Scale effects in landslide susceptibility assessment: integrating slope unit division and SHAP-based interpretability in a typical river basin. Water. 2025;17(13):1877. doi:10.3390/w17131877.

135. Bowers C, Serafin KA, Baker JW. Uncovering drivers of atmospheric river flood damage using interpretable machine learning. Nat Hazards Rev. 2024;25(3):04024022. doi:10.1061/nhrefo.nheng-1995.

136. Nguyen Van L, Nguyen GV, Yeon M, Thi-Tuyet Do M, Lee G. Unveiling environmental drivers of soil erosion in South Korea through SHAP-informed machine learning. Land Use Policy. 2025;155:107592. doi:10.1016/j.landusepol.2025.107592.

137. Han J, Guzman JA, Chu ML. Prediction of gully erosion susceptibility through the lens of the SHapley Additive exPlanations (SHAP) method using a stacking ensemble model. J Environ Manage. 2025;383(9):125478. doi:10.1016/j.jenvman.2025.125478.

138. Yang X, Wang ZH, Wang C, Lai YC. Finding causal gateways of precipitation over the contiguous United States. Geophys Res Lett. 2023;50(4):e2022GL101942. doi:10.1029/2022GL101942.

139. Iglesias-Suarez F, Gentine P, Solino-Fernandez B, Beucler T, Pritchard M, Runge J, et al. Causally-informed deep learning to improve climate models and projections. J Geophys Res Atmos. 2024;129(4):e2023JD039202. doi:10.1029/2023JD039202.

140. Collini E, Ipsaro Palesi LA, Nesi P, Pantaleo G, Nocentini N, Rosi A. Predicting and understanding landslide events with explainable AI. IEEE Access. 2022;10:31175–89. doi:10.1109/access.2022.3158328.

141. Moeineddin A, Seguí C, Dueber S, Fuentes R. Physics-informed neural networks applied to catastrophic creeping landslides. Landslides. 2023;20(9):1853–63. doi:10.1007/s10346-023-02072-0.

142. Sakar C, Schwartz N, Moreno Z. Physics-informed neural networks trained with time-lapse geo-electrical tomograms to estimate water saturation, permeability and petrophysical relations at heterogeneous soils. Water Resour Res. 2024;60(8):e2024WR037672. doi:10.1029/2024WR037672.

143. El Montassir R, Pannekoucke O, Lapeyre C. HyPhAICC v1.0: a hybrid physics-AI approach for probability fields advection shown through an application to cloud cover nowcasting. Geosci Model Dev. 2024;17(17):6657–81. doi:10.5194/gmd-17-6657-2024.

144. Degen D, Ziegler M, Heidbach O, Henk A, Reiter K, Wellmann F. About the trustworthiness of physics-based machine learning-considerations for geomechanical applications. Solid Earth. 2025;16(6):477–502. doi:10.5194/se-16-477-2025.

145. Jiang S, Zheng Y, Solomatine D. Improving AI system awareness of geoscience knowledge: symbiotic integration of physical approaches and deep learning. Geophys Res Lett. 2020;47(13):e2020GL088229. doi:10.1029/2020GL088229.

146. Zhang Z, Wang D, Mei Y, Zhu J, Xiao X. Developing an explainable deep learning module based on the LSTM framework for flood prediction. Front Water. 2025;7:1562842. doi:10.3389/frwa.2025.1562842.

147. Kakhani N, Taghizadeh-Mehrjardi R, Omarzadeh D, Ryo M, Heiden U, Scholten T. Towards explainable AI: interpreting soil organic carbon prediction models using a learning-based explanation method. Eur J Soil Sci. 2025;76(2):e70071. doi:10.1111/ejss.70071.

148. Zerouali B, Almaliki AH, Santos CAGães. Flood susceptibility mapping in arid urban areas using SHAP-enhanced stacked ensemble learning: a case study of Jeddah. J Environ Manage. 2025;393(1):127128. doi:10.1016/j.jenvman.2025.127128.

149. Nikoo MR, Tosan M, Nazari R, Vahabzadeh M, Karimi M, Zarrinabadi E. Assessing the fidelity of multi-satellite precipitation estimates for drought monitoring in a mountain water tower to arid basin system. J Arid Environ. 2026;232(5):105519. doi:10.1016/j.jaridenv.2025.105519.