The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume L-4/W1-2026
Free and Open Source Software for Geospatial (FOSS4G) 2026 – Academic Track "Bridging Geospatial Technology and Humanity",
30 August–5 September 2026, Hiroshima, Japan

A Comprehensive Disaster Risk Analysis Model With GeoAI

Muhammed Oguzhan Mete [1], Muhammed Yahya Biyik [1], Zeynep Ceyda Karakaya [1]

[1] Department of Geomatics Engineering, Istanbul Technical University, 34469 Istanbul, Türkiye – *metemu@itu.edu.tr*

Keywords: Disaster Risk Analysis, Multi-Hazard Assessment, GeoAI, Geographic Information Systems, Machine Learning, Large Language Models, Model Context Protocol, Urban Resilience

Abstract:

Natural and technological disasters continue to threaten communities, infrastructure, and the environment, and the cascading nature of contemporary risks complicates their assessment. This study presents a disaster risk analysis model coupling Geographic Information Systems (GIS), ensemble machine learning, and AI-driven interpretation tools, collectively termed GeoAI, to support multi-hazard risk assessment, resilience planning, and citizen-oriented risk communication. Rather than building isolated models per hazard, the framework applies a single, modular pipeline consistently across flood, wildfire, earthquake, landslide, drought, and urban heat island hazards. Hazard, vulnerability, and exposure layers are derived from open geospatial datasets and processed in QGIS, after which Random Forest and XGBoost classifiers generate hazard and vulnerability maps validated using AUC-ROC, F1-score, and Cohen's Kappa. A distinctive component is an AI agent built on Large Language Models (LLMs) and the Model Context Protocol (MCP), which queries structured risk databases and produces region-specific narrative risk reports in natural language. Outputs are delivered through a web-based platform with an MCP-connected chatbot, lowering the barrier to understanding complex risk information for experts and the public. The model was tested in Türkiye's Marmara Region, which is important due to its dense population, heavy industry, and earthquake risk along the North Anatolian Fault. The pilot showed good results: XGBoost performed better than the baseline models, and the LLM-based interpretation layer gave clear, well-grounded outputs. The framework offers a scalable, open, interoperable approach linking spatial analytics, machine learning, and generative AI for evidence-based disaster risk reduction.

1. Introduction

Disasters arising from natural and technological processes remain among the most significant challenges confronting modern societies. Population growth, rapid and frequently unplanned urbanisation, and the intensifying effects of climate change have steadily increased both the frequency and the severity of damaging events, placing communities, lifeline infrastructure, and economic activity under mounting pressure. The risks that planners must now anticipate are rarely isolated; they interact, compound, and cascade, so that a single triggering event can propagate through interdependent systems and generate consequences far larger than the original hazard would suggest (Pescaroli and Alexander, 2018). Addressing this reality calls for analytical frameworks that treat hazard, vulnerability, and exposure as coupled components rather than as separate problems to be solved in isolation.

The overarching aim of this study is to develop a comprehensive and generalisable disaster risk analysis model that integrates geospatial analysis, machine learning, and generative Artificial Intelligence within a single coherent workflow. The model is designed to be hazard-agnostic: the same analytical pipeline is applied to floods, wildfires, earthquakes, landslides, droughts, and urban heat island events, which promotes reproducibility and avoids the fragmentation that results from maintaining a separate methodology for every peril. Beyond producing quantitative risk maps, the framework seeks to make the resulting information intelligible and actionable for a broad audience, ranging from technical specialists in civil protection agencies to local officials and ordinary citizens.

The importance of this undertaking is difficult to overstate. Effective disaster risk reduction depends on timely, spatially explicit, and trustworthy information, yet the technical complexity of risk modelling often confines its outputs to expert circles. Where risk information cannot be understood by the people it is meant to protect, even highly accurate analyses deliver limited practical benefit. International policy increasingly reflects this concern: the Sendai Framework for Disaster Risk Reduction 2015–2030 places explicit emphasis on understanding risk and on communicating it in accessible forms, and recent global assessments stress the uneven coverage of multi-hazard early warning systems as a persistent gap in preparedness (UNDRR, 2015; UNDRR, 2024). A model that bridges rigorous analytics and clear communication therefore speaks directly to a recognised need.

The scientific foundations of the proposed approach draw on several converging strands of research. Geographic Information Systems have long served as the backbone of hazard and risk assessment, providing the geoprocessing capabilities needed to combine heterogeneous spatial layers and to evaluate the regional risk posed by major hazards to public safety (Contini, Bellezza, Christou, and Kirchsteiger, 2000; Zhao and Liu, 2016). Parallel to these developments, the study of vulnerability has matured into a quantitative discipline. The Social Vulnerability Index demonstrated that socioeconomic and demographic characteristics can be combined into spatially explicit indicators of differential susceptibility to harm, an insight that continues to inform exposure and vulnerability mapping today (Cutter, Boruff, Shirley, 2003).

Over the past decade, machine learning has reshaped hazard susceptibility modeling. Ensemble tree-based methods in particular have proven well suited to the nonlinear and high-dimensional relationships that characterise environmental hazards. Abedi, Costache, Shafizadeh-Moghadam, and Pham, 2022) compared XGBoost, Random Forest, and boosted regression trees for flash-flood susceptibility and reported that gradient-boosted ensembles achieved the strongest discrimination, while Ali, Parvin, Pham, Vojtek, Vojteková, Costache, Linh, Nguyen, Ahmad, and Ghorbani (2020) showed that hybrid and bivariate-statistical models can markedly improve the accuracy of GIS-based flood susceptibility assessment. Comparable conclusions emerge from the landslide literature, where a comprehensive review by Liu, Wang, Zhang, He, and Pijush (2023) found that data-driven ensemble and deep-learning models consistently outperform classical statistical techniques, although no single algorithm dominates across all settings. This last caveat is worth keeping in mind, since it cautions against treating any one classifier as universally optimal.

The integration of spatially explicit learning with geographic problem solving has given rise to the field of GeoAI. Janowicz, Gao, McKenzie, Hu, and Bhaduri (2020) describe it as a set of spatially explicit Artificial Intelligence techniques for geographic knowledge discovery. More recently, the emergence of Large Language Models has opened a further frontier. Autonomous GIS agents that can plan and execute spatial analysis workflows have been demonstrated by Li and Ning (2023), and subsequent work has extended such agents toward autonomous geospatial data retrieval (Ning, Li, Akinboyewa, and Lessani, 2025) and toward natural-language interfaces that make spatial analysis accessible to non-specialists (Mansourian and Oucheikh, 2024). These advances suggest that generative AI can do more than automate analysis; it can also translate complex quantitative results into language that decision-makers and citizens readily understand.

Despite this progress, three gaps persist. First, hazard models are frequently developed peril by peril, which limits transferability and complicates multi-hazard reasoning. Second, the exposure and vulnerability dimensions of risk are often treated more coarsely than the hazard dimension, even though they determine the human consequences of an event. Third, and most importantly, the gulf between technical outputs and public understanding remains wide, with few operational systems offering interpretation in plain language. The present study addresses these gaps by proposing a unified, modular GeoAI framework that applies a consistent pipeline across hazards, gives explicit treatment to exposure and vulnerability, and embeds an LLM-based interpretation and communication layer connected to the underlying data through the Model Context Protocol. The remainder of the paper describes the study area, data, and methodology, reports the results of a pilot application in the Marmara Region of Türkiye, and discusses the implications and limitations of the approach.

2. Materials and Methods

2.1 Study Area

The model was developed and piloted in the Marmara Region of Türkiye, which comprises eleven provinces: Istanbul, Bursa, Kocaeli, Sakarya, Tekirdağ, Edirne, Kırklareli, Balıkesir, Çanakkale, Yalova, and Bilecik. Situated at the junction of Europe and Asia, the region is the most populous and economically productive part of the country, concentrating a large share of national industry, transport infrastructure, and financial activity. This concentration of people and assets is precisely what makes the region an instructive testbed: the consequences of any given hazard are amplified by the density of exposed elements.

The Marmara Region is also one of the most seismically active areas in Europe. The North Anatolian Fault Zone runs through the region, and its northern branch, the Main Marmara Fault, lies only a short distance south of Istanbul, a metropolis of roughly eighteen million inhabitants. The destructive 1999 İzmit earthquake remains a vivid reminder of the stakes, and recent seismological analysis indicates that moderate to large ruptures have been migrating eastward along the fault toward Istanbul, leaving a locked segment capable of generating a major event (Martínez-Garzón, Chen, Becker, Núñez-Jara, Kartal, Türker, Dresen, Ben-Zion, Jara, Cotton, Kadirioglu, Kiliç, Bohnhoff, 2026). Beyond seismic hazard, the region experiences riverine and flash flooding, landslides on its steeper terrain, wildfire in its forested margins, and intensifying urban heat island effects within its large conurbations. This diversity of perils, set against a heterogeneous physical and demographic backdrop, allowed the framework to be evaluated for both cross-hazard applicability and cross-regional transferability within a single study area. The extent of the study area and its eleven constituent provinces are shown in Figure 1.

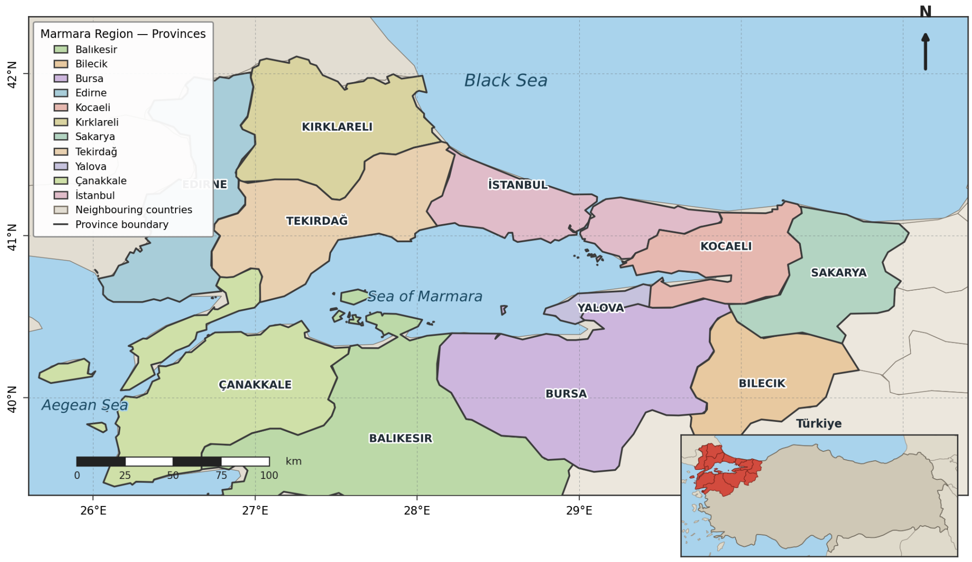

Figure 1. Study area: the Marmara Region of Türkiye.

2.2 Data Sources

All inputs to the model were drawn from open and freely accessible repositories, a deliberate design choice that supports cost efficiency, transparency, reproducibility, and independent verification. Infrastructure and land-use information was obtained from OpenStreetMap. Remote sensing products, including land cover and surface temperature derivatives, were sourced from the Copernicus Land Monitoring Service and the USGS Earth Explorer platform, while climatological and hydrological datasets were retrieved from NASA EARTHDATA. Administrative boundaries and socioeconomic statistics were compiled from national and regional open government portals.

For each hazard, an appropriate set of conditioning factors was assembled. Topographic variables such as elevation, slope, aspect, and curvature were derived from digital elevation models; hydrological descriptors included drainage density, distance to rivers, and the topographic wetness index; and additional layers captured soil type, lithology, land use and land cover, and historical disaster inventories. Vulnerability inputs combined population-density grids with socioeconomic indicators, and exposure inputs included building footprints and the locations of critical infrastructure. Storing these heterogeneous layers in a standardised geospatial database ensured that the same feature-engineering logic could be reused regardless of the hazard under study.

2.3 Model Architecture and Workflow

The proposed framework is organised into five sequential but loosely coupled stages: integrated data acquisition and standardised preprocessing; modular GeoAI feature engineering; ensemble machine learning disaster modelling; AI-agent risk assessment; and a web-based geospatial communication platform. The modular design means that each stage exposes well-defined inputs and outputs, so that components can be revised or replaced without disturbing the rest of the pipeline. The overall architecture is summarised in Figure 2.

In the first stage, environmental, geospatial, socioeconomic, and historical data are ingested and harmonised into a standardised raster and vector database. Coordinate systems, resolutions, and attribute schemas are reconciled so that subsequent operations can proceed without repeated reformatting. The second stage transforms these harmonised layers into analysis-ready features, producing hazard suitability layers, vulnerability index layers, and exposure maps that constitute the inputs to the learning models.

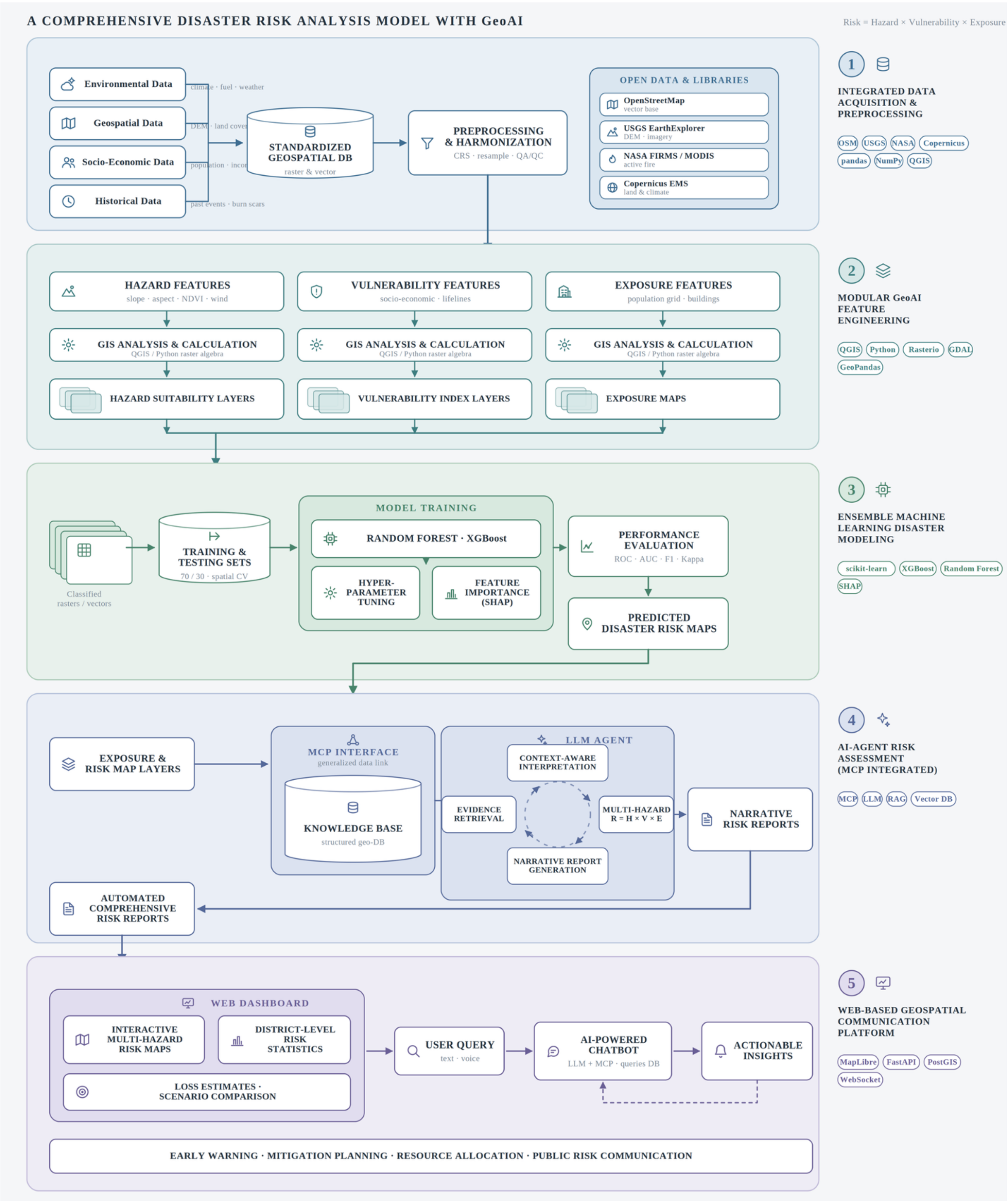

Figure 2. The comprehensive disaster risk analysis model with GeoAI.

2.4 Modular GeoAI Feature Engineering

Feature engineering was carried out in QGIS and Python, drawing on libraries such as pandas, NumPy, and Rasterio for tabular and raster manipulation. Hazard features were derived through standard geoprocessing of terrain and hydrological data; vulnerability features combined socioeconomic indicators with information on critical infrastructure; and exposure features were built from population grids and building footprints. Because the same engineering routines are applied across hazards, the framework avoids duplicating effort and keeps the feature space consistent, which in turn facilitates comparison of results between perils. The five conceptual stages of the workflow are illustrated in Figure 3.

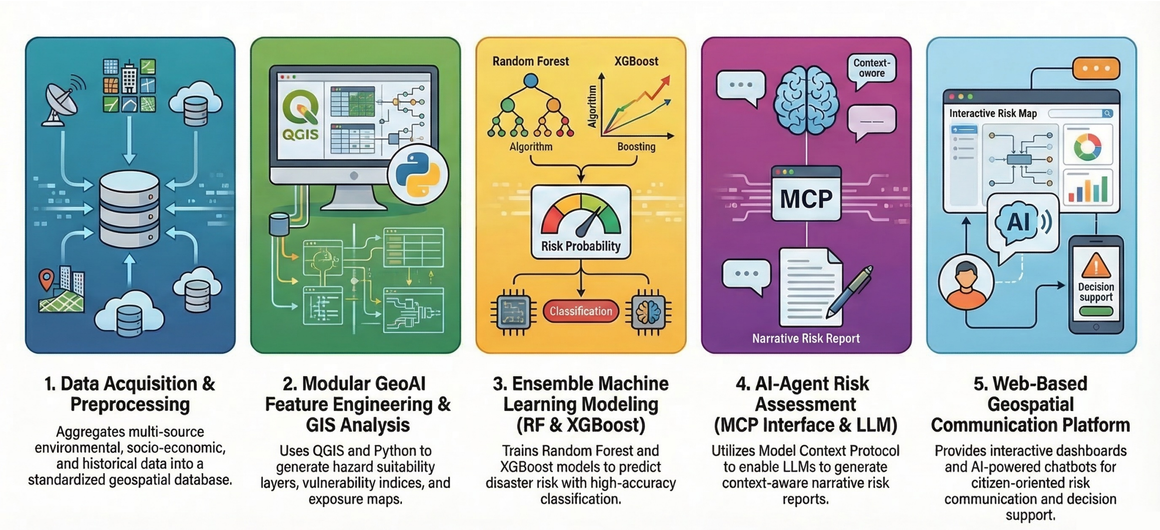

Figure 3. The five stages of the GeoAI risk analysis workflow.

2.5 Ensemble Machine Learning Disaster Modelling

Once the feature matrices were assembled, supervised classification was used to derive hazard and vulnerability maps. The study focused on two ensemble tree-based algorithms that have repeatedly demonstrated strong performance in this domain: Random Forest, which aggregates many decorrelated decision trees through bootstrap sampling (Breiman, 2001), and Extreme Gradient Boosting, a regularised gradient-boosting implementation known for its accuracy and scalability (Chen and Guestrin, 2016). Both methods handle nonlinear interactions among conditioning factors and provide measures of feature importance that aid interpretation.

Models were implemented with Python machine learning libraries, principally scikit-learn and the XGBoost package. Labelled samples were drawn from historical disaster inventories together with comparable non-event locations, and the data were partitioned into training and testing subsets. Hyperparameters were tuned through grid and randomised search under k-fold cross-validation to guard against overfitting and to support generalisation across the heterogeneous conditions of the study area. Model quality was assessed with a complementary set of metrics, namely the area under the receiver operating characteristic curve (AUC-ROC), the F1-score, precision, recall, and Cohen's Kappa coefficient, so that discrimination, class balance, and chance-corrected agreement were all taken into account. The trained models were then applied across the study area to produce predicted disaster risk maps.

2.6 Exposure Assessment

The exposure component quantifies the elements at risk within identified hazard zones, including population counts, building footprints, and critical infrastructure. Hazard maps were overlaid with exposure layers through spatial operations to yield exposure matrices that support the prioritisation of preparedness and mitigation measures. To represent the physical susceptibility of the built environment more faithfully, the framework accommodates fragility curves, which express the probability of reaching or exceeding a damage state as a function of hazard intensity. Such curves have been developed for the reinforced-concrete building typologies that dominate the Turkish stock (Akkar and Yakut, 2005) and, more generally, for reinforced-concrete frames under seismic loading (Ramamoorthy, 2006). Incorporating them allows exposure to be translated into expected damage rather than mere presence within a hazard zone.

2.7 AI-Agent Risk Assessment with MCP and LLMs

A central methodological contribution of this work is the integration of an Artificial Intelligence agent into the risk assessment pipeline. The agent is built on Large Language Models and uses the Model Context Protocol (MCP) as a standardised interface through which it can query, retrieve, and reason over geospatial risk data held in structured databases and spatial APIs. Functioning as a generalised data link, the MCP interface connects the language model to a structured knowledge base populated by the preceding analytical stages.

Connecting the agent to an MCP server follows a client–server model. The risk knowledge base and the associated geoprocessing routines are wrapped by an MCP server that advertises a set of capabilities — tools, resources, and prompt templates — each described by a typed schema. When a session begins, the agent, acting as the MCP host, performs an initial handshake in which it discovers the available capabilities and negotiates the protocol version. Thereafter the language model never touches the spatial database directly; instead it issues structured tool calls, expressed as JSON messages over a transport such as standard input/output or HTTP with server-sent events, and the server executes the corresponding operation and returns the result. Representative tools in the present implementation include functions that query a hazard susceptibility layer for a given location, retrieve exposure and vulnerability statistics for an administrative unit, and fetch historical event records matching a hazard type and area of interest. Because every capability is declared through an explicit schema, the same server can be reused across hazards, and the agent can be extended simply by registering additional tools. Authentication and access control are enforced at the server boundary, so sensitive datasets remain governed by the data owner rather than being exposed to the model.

Within this arrangement the agent performs context-aware interpretation of multi-hazard risk maps and statistical outputs, applies multi-hazard reasoning in which total risk is treated as a function of hazard, vulnerability, and exposure, and generates narrative risk reports tailored to specific administrative units, districts, or neighbourhoods. By expressing quantitative results in clear prose, the system reduces the analytical burden on planners and produces documentation that can be acted upon directly. The MCP-based design also supports connection to external data sources and third-party geospatial services, which positions the framework for near-real-time risk monitoring as new observations arrive.

2.8 Spatial Retrieval-Augmented Generation (Spatial RAG)

Generating risk narratives that are at once fluent and factually correct requires the language model to be grounded in the specific data of the area under consideration rather than in its parametric memory. The framework addresses this through retrieval-augmented generation, a paradigm in which a model conditions its output on documents fetched from an external store at inference time (Lewis, Perez, Piktus, Petroni, Karpukhin, Goyal, Küttler, Lewis, Yih, Rocktäschel, and Riedel, 2020). Standard retrieval-augmented generation was designed for text, however, and carries no notion of location — precisely the dimension that matters most in disaster risk. The model therefore adopts a spatial variant, Spatial Retrieval-Augmented Generation (Spatial RAG), which extends the paradigm with explicit spatial retrieval so that the evidence supplied to the language model is selected by where it applies as much as by what it says (Yu, Bao, Ning, Peng, Mai, & Zhao, 2026).

The knowledge base that the agent draws upon is assembled from multiple sources, each geo-referenced and time-indexed before it is stored. Alongside the hazard, vulnerability, and exposure layers produced by the preceding stages, the store can absorb satellite and aerial imagery, meteorological and hydrological observations, terrain and river models, historical disaster inventories, and, where available, streams from in-situ sensors. Computer-vision models translate raw imagery into geo-tagged intelligence layers — delineating, for example, flood extent, landslide scars, wildfire hotspots, or damaged structures — with each detection carrying a location, a timestamp, and a confidence value. Spatial indexing structures such as geohashes, together with segmented administrative units, allow these heterogeneous records to be queried efficiently over large areas, so that retrieval remains fast even as the archive grows.

Retrieval then proceeds in two complementary ways. A spatial index over geometries and administrative units supports sparse retrieval, in which records are filtered by spatial predicates such as intersection with, containment within, or proximity to the area of interest, and by temporal predicates that restrict attention to a chosen window. In parallel, textual descriptions of hazards, conditioning factors, historical events, and model outputs are encoded as dense vector embeddings to support semantic retrieval, in which records are ranked by similarity to the user's question. A hybrid retriever combines the two signals, balancing spatial and temporal constraints against semantic relevance, and returns a compact set of grounded evidence: the predicted hazard class and probability, the vulnerability index, exposure counts and, where available, fragility-based damage estimates, together with analogous past events and recent changes for the same place. This evidence is assembled into the prompt that conditions the language model, so that each statement in the resulting narrative is traceable to specific retrieved facts. Grounding the generation in this way constrains the model and markedly reduces the incidence of fabricated content, although human review of the output is retained as a safeguard.

Because every record carries both location and time, the agent can reason about change as well as state. Comparing successive observations reveals what has changed, where, and how quickly, which turns a static map into an evolving picture of risk. This design supports operationally meaningful questions posed in plain language — for instance, which hospitals fall within a projected flood zone, which settlements may lose road connectivity, or which districts should be prioritised for evacuation — as well as queries bounded in time, such as which areas show the greatest increase in flood risk over a forecast window, or which settlements have been affected by river expansion in recent days. By retrieving terrain, population-density, evacuation-route, and critical-infrastructure layers alongside the hazard estimate, the system is intended to move disaster support away from a largely reactive posture and toward anticipatory, situationally aware intelligence, although realising the fully real-time form of this capability depends on the live data feeds discussed later as future work.

A deliberate consequence of this architecture is that the retrieval-and-generation loop is hazard-agnostic. Each peril contributes its layers to the knowledge base under a shared schema, tagged by hazard type, location, and time, while the retriever and the generator themselves remain unchanged. A query concerning seismic risk in a given district is answered by the same mechanism that answers a query about wildfire, flood, landslide, drought, or urban heat; only the retrieved records differ. This uniformity lets the agent reason over several hazards at once for the same location — comparing, for instance, the flood and earthquake exposure of a neighbourhood within a single narrative — and it means that adding a new hazard requires populating the knowledge base rather than redesigning the language interface. The Spatial RAG retriever is itself exposed to the agent as a set of MCP tools, so that the grounding mechanism and the data-access mechanism described earlier constitute one coherent pipeline.

2.9 Web-Based Geospatial Communication Platform

The outputs of the model are delivered through a web-based platform that visualises hazard susceptibility maps, vulnerability indices, exposure layers, and composite risk scores across several administrative levels. Interactive map layers, filtering tools, district-level risk statistics, and downloadable reports accommodate the differing needs of technical and non-technical users. A defining feature of the platform is an integrated, AI-powered chatbot developed with LLM technology and connected to the underlying risk database through the same MCP interface. Citizens and decision-makers can pose questions in natural language and receive plain-language explanations of the risk status of their area, guidance on protective actions, and information about emergency preparedness resources. This conversational layer lowers the barrier to understanding complex risk information and encourages public awareness and engagement. The implementation stages and their interconnections are depicted in Figure 4.

3. Results and Discussion

The pilot application in the Marmara Region demonstrated that a single modular pipeline can be applied coherently across markedly different hazards. For each peril, the feature-engineering, modelling, exposure, and interpretation stages executed without bespoke restructuring, which supports the central claim that a hazard-agnostic architecture is workable in practice and not merely in principle.

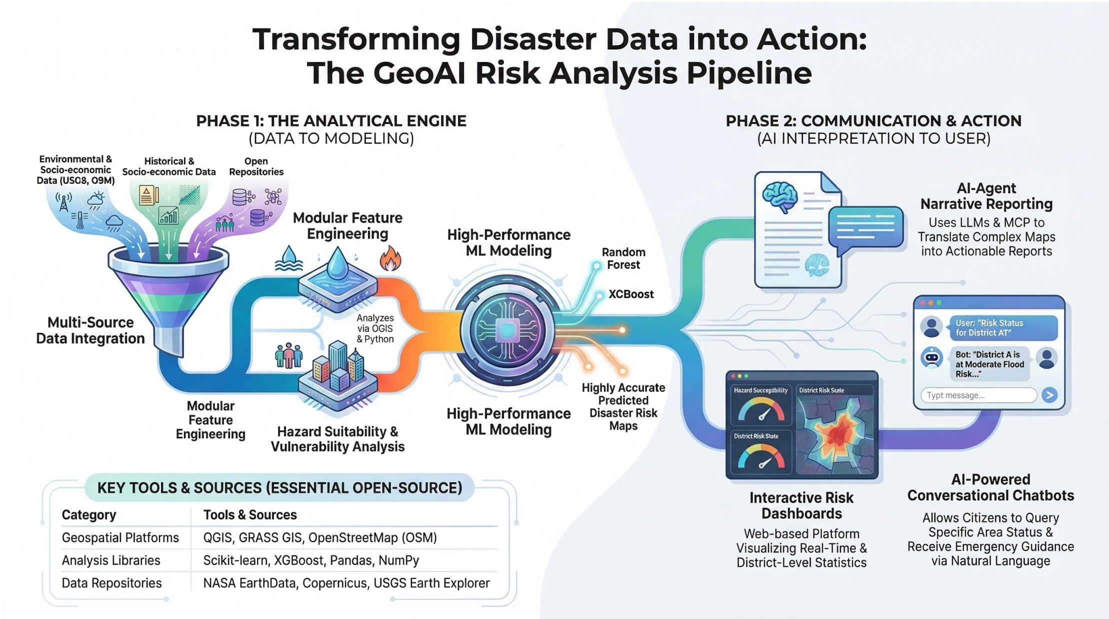

Figure 4. The GeoAI risk analysis pipeline with two phases.

Across the tested hazards, the ensemble classifiers achieved strong discrimination in hazard susceptibility mapping, and XGBoost generally returned higher AUC values than the Random Forest baseline. This ordering is consistent with the wider literature: Abedi, Costache, Shafizadeh-Moghadam, and Pham (2022) reported an AUC of roughly 0.89 for an XGBoost flash-flood model, and recent landslide studies summarised by Liu, Wang, Zhang, He, and Pijush (2023) frequently document AUC values in the 0.90 to 0.99 range for well-specified tree-based models. The pilot outcomes therefore sit comfortably within the performance envelope established by comparable work, while the relative advantage of gradient boosting echoes a pattern observed repeatedly in susceptibility modelling. It would nonetheless be unwise to over-interpret a single regional pilot; the same review cautions that no algorithm performs best in every setting, and local validation should remain a routine step rather than an afterthought.

These outputs are realised as the multi-hazard risk products shown in Figures 5 and 6. The per-hazard maps (Figure 5) make the differing geography of each peril explicit — fault proximity dominating the earthquake surface, low-lying and impervious coastal land the flood surface, forest cover the wildfire surface, and the inland agricultural plains the drought surface — while the composite map (Figure 6) integrates hazard, vulnerability and exposure into a single screening layer, on which the fault-adjacent and densely built provinces of Kocaeli, Yalova and İstanbul emerge as the highest-risk units.

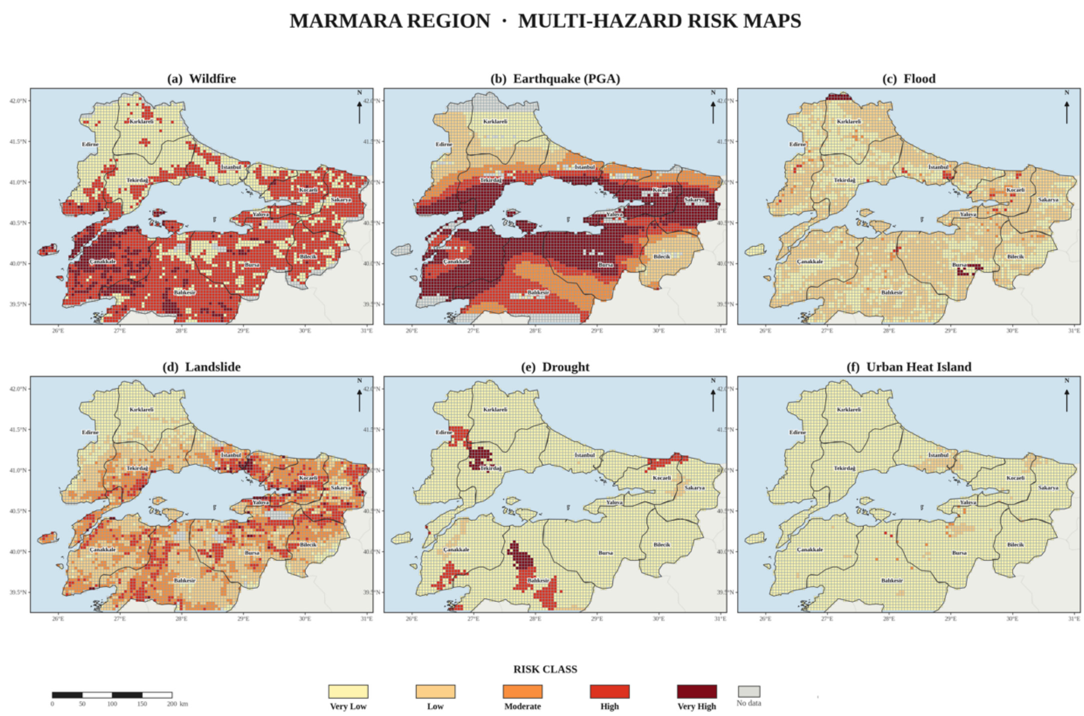

Figure 5. Per-hazard risk maps for the six modeled disasters.

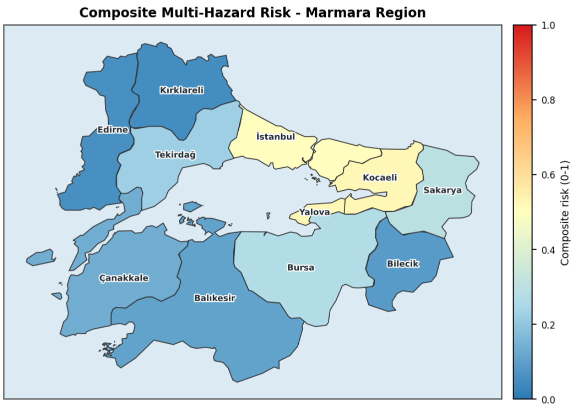

Figure 6. Composite multi-hazard risk combining hazard, vulnerability, and exposure.

The exposure analysis added an interpretive layer that pure susceptibility maps cannot provide. By intersecting hazard zones with population, building, and infrastructure data, the model distinguished areas of high hazard but low exposure from those where hazard and exposure coincide. The latter are the locations that warrant priority attention, and the ability to surface them automatically is of direct value for preparedness planning in a region as densely built as Marmara. Where fragility information was available, the translation of exposure into expected damage sharpened these distinctions further.

The LLM-based interpretation layer produced narrative risk reports that domain-expert reviewers judged to be coherent, factually grounded, and contextually relevant. The Spatial RAG design proved important here: by drawing its statements from spatially retrieved evidence rather than from unconstrained generation, and by reaching that evidence through the Model Context Protocol, the agent kept its narratives anchored to the underlying analysis. This design mitigates, though it does not entirely eliminate, the risk of fabricated or misleading content, which remains a known limitation of generative models. Human oversight of the generated reports is accordingly retained as a safeguard, and we regard it as essential rather than optional.

These products reach users through the web-based platform shown in Figure 7, which couples an interactive multi-hazard choropleth and hazard-layer selector with a province readout — per-hazard risk, vulnerability and exposure, and a grounded narrative — and an MCP-connected Spatial RAG console that answers natural-language questions about the region.

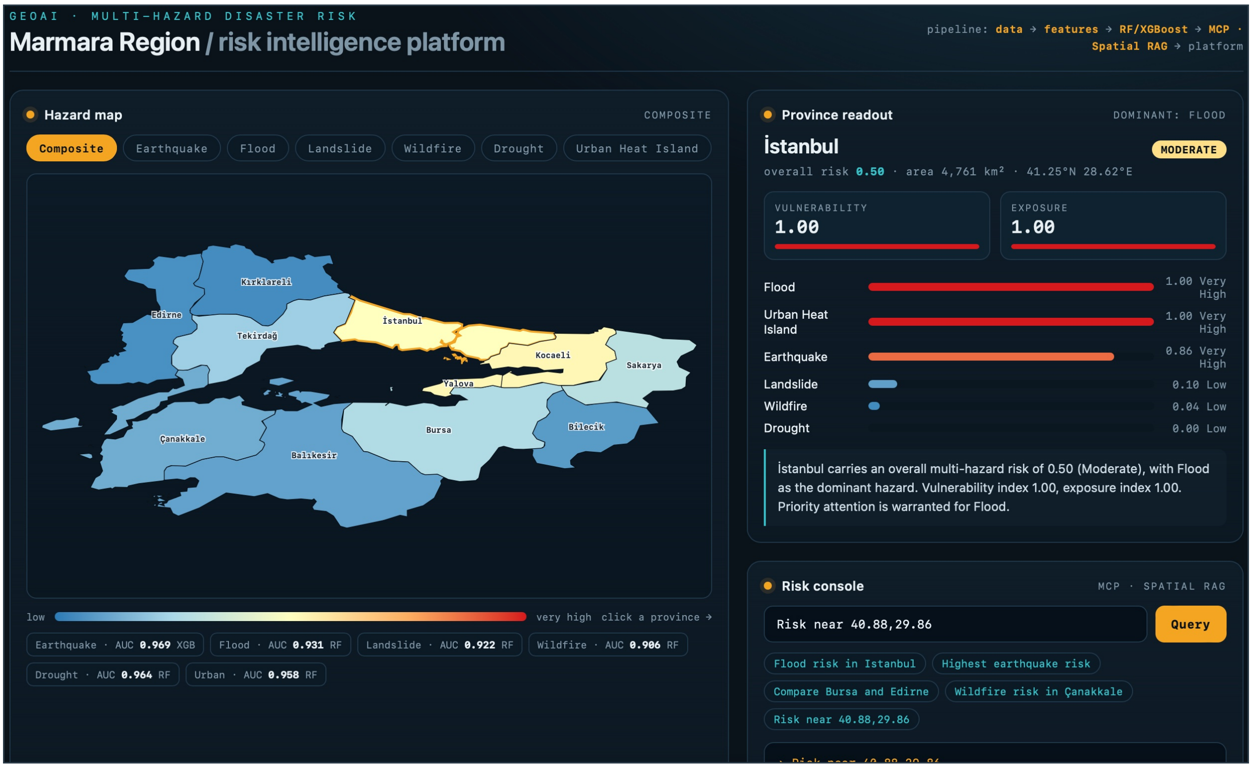

Figure 7. The GeoAI-based disaster risk intelligence platform with a MCP-connected Spatial RAG.

Taken together, the results indicate that coupling spatial analytics, ensemble learning, and generative AI can shorten the distance between raw geospatial data and actionable insight. The web platform and conversational chatbot extend this benefit beyond expert users, addressing the communication gap that has long constrained the practical value of risk modelling. Several limitations temper these conclusions. The pilot relied on static historical and environmental inputs, so the present results describe susceptibility rather than dynamically evolving risk; the quality of open datasets varies across provinces; and the expert evaluation of the narrative layer, while encouraging, was qualitative and would benefit from a larger, more structured assessment. These constraints define a clear agenda for further work rather than undermining the framework's central contribution.

4. Conclusion

This paper has presented a comprehensive, GeoAI-based disaster risk analysis model that unifies geospatial analysis, ensemble machine learning, and generative Artificial Intelligence within a single modular pipeline. By applying one consistent workflow across floods, wildfires, earthquakes, landslides, droughts, and urban heat island events, the framework addresses the fragmentation that arises when each hazard is modelled separately, and by giving explicit treatment to exposure and vulnerability it connects physical hazard to human consequence. The integration of an LLM-based agent through the Model Context Protocol, together with a public web platform and conversational chatbot, tackles the persistent gap between technical outputs and public understanding.

The pilot in the Marmara Region of Türkiye, an area of exceptional exposure and seismic hazard, showed that the approach is both technically feasible and practically useful: ensemble classifiers delivered strong susceptibility mapping with XGBoost performing best, exposure overlays highlighted priority areas, and the interpretation layer produced trustworthy narrative reports. Built entirely on open data and open-source tools, the framework is transparent, reproducible, and transferable to other regions. In this sense it contributes directly to the goals of the Sendai Framework for Disaster Risk Reduction and to the broader pursuit of resilient, smart, and sustainable cities (UNDRR, 2015).

Future research will incorporate real-time sensor streams and Internet of Things technologies to move from static susceptibility toward dynamic, continuously updated risk, in line with international efforts to expand multi-hazard early warning systems (UNDRR, 2024). Further directions include a more rigorous, structured evaluation of the narrative layer, the addition of temporal and climate-projection inputs, and testing of the framework in regions with contrasting hazard profiles to probe the limits of its transferability.

5. Acknowledgments

This study was conducted as part of the MARAP project, carried out in collaboration between Istanbul Technical University and Türkiye Disaster and Emergency Management Presidency (AFAD), and was supported by the Scientific Research Projects Coordination Unit of Istanbul Technical University [Grant No. FHD-2026-47849].

6. References

Abedi, R., Costache, R., Shafizadeh-Moghadam, H., Pham, Q. B., 2022. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto International*, *37*(19), 5479–5496. doi.org/10.1080/10106049.2021.1920636

Akkar, S., Yakut, A., 2005. *Generating Fragility Curves for Turkish Mid-rise and Low-rise Reinforced Concrete Buildings*. Earthquake Engineering Research Center, Middle East Technical University, Ankara, Türkiye.

Ali, S. A., Parvin, F., Pham, Q. B., Vojtek, M., Vojteková, J., Costache, R., Linh, N. T. T., Nguyen, H. Q., Ahmad, A., Ghorbani, M. A., 2020. GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naïve Bayes tree, bivariate statistics and logistic regression: A case of Topľa basin, Slovakia. *Ecological Indicators*, *117*, 106620. doi.org/10.1016/j.ecolind.2020.106620

Breiman, L., 2001. Random forests. *Machine Learning*, *45*(1), 5–32. doi.org/10.1023/A:1010933404324

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. doi.org/10.1145/2939672.2939785

Contini, S., Bellezza, F., Christou, M. D., Kirchsteiger, C., 2000. The use of geographic information systems in major accident risk assessment and management. *Journal of Hazardous Materials*, *78*(1–3), 223–245. doi.org/10.1016/S0304-3894(00)00224-2

Cutter, S. L., Boruff, B. J., Shirley, W. L., 2003. Social vulnerability to environmental hazards. *Social Science Quarterly*, *84*(2), 242–261. doi.org/10.1111/1540-6237.8402002

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., Bhaduri, B., 2020. GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, *34*(4), 625–636. doi.org/10.1080/13658816.2019.1684500

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, *33*, 9459–9474.

Li, Z., Ning, H., 2023. Autonomous GIS: the next-generation AI-powered GIS. *International Journal of Digital Earth*, *16*(2), 4668–4686. doi.org/10.1080/17538947.2023.2278895

Liu, S., Wang, L., Zhang, W., He, Y., Pijush, S., 2023. A comprehensive review of machine learning-based methods in landslide susceptibility mapping. *Geological Journal*, *58*(6), 2283–2301. doi.org/10.1002/gj.4666

Mansourian, A., Oucheikh, R., 2024. ChatGeoAI: Enabling geospatial analysis for public through natural language, with large language models. *ISPRS International Journal of Geo-Information*, *13*(10), 348. doi.org/10.3390/ijgi13100348

Martínez-Garzón, P., Chen, X., Becker, D., Núñez-Jara, S., Kartal, R.F., Türker, E., Dresen, G., Ben-Zion, Y., Jara, J., Cotton, F., Kadirioglu, F.T., Kiliç, T., Bohnhoff, M., 2026. Progressive eastward rupture of the Main Marmara fault toward Istanbul. *Science*, 391(6783). doi.org/10.1126/science.adz0072

Ning, H., Li, Z., Akinboyewa, T., Lessani, M. N., 2025. An autonomous GIS agent framework for geospatial data retrieval. *International Journal of Digital Earth*, *18*(1), 2458688. doi.org/10.1080/17538947.2025.2458688

Pescaroli, G., Alexander, D., 2018. Understanding compound, interconnected, interacting, and cascading risks: A holistic framework. *Risk Analysis*, *38*(11), 2245–2257. doi.org/10.1111/risa.13128

Ramamoorthy, S. K., 2006. *Seismic Fragility Estimates for Reinforced Concrete Framed Buildings*. Doctoral dissertation, Texas A&M University, College Station, TX.

United Nations Office for Disaster Risk Reduction (UNDRR), 2015. *Sendai Framework for Disaster Risk Reduction 2015–2030*. UNDRR, Geneva, Switzerland.

United Nations Office for Disaster Risk Reduction (UNDRR), 2024. *Global Status of Multi-Hazard Early Warning Systems 2024*. UNDRR and World Meteorological Organization, Geneva, Switzerland.

Yu, D., Bao, R., Ning, R., Peng, J., Mai, G., & Zhao, L. (2026). Spatial-rag: Spatial retrieval augmented generation for real-world geospatial reasoning questions. In *Findings of the Association for Computational Linguistics: ACL 2026* (pp. 11094-11112).

Zhao, M., Liu, X., 2016. Regional risk assessment for urban major hazards based on GIS geoprocessing to improve public safety. *Safety Science*, *87*, 18–24. doi.org/10.1016/j.ssci.2016.03.016