Jurnal Publikasi Teknik Informatika
Volume. 5 Nomor. 1 Januari 2026
e-ISSN: 2808-8972; p-ISSN: 2808-9367, Hal. 176-185
DOI: https://doi.org/10.55606/jupti.v5i1.6403
Tersedia: https://journalcenter.org/index.php/jupti

Geospatial Artificial Intelligence for Flood Disaster Mapping in Sumatra: A Review of Machine Learning Models, Data, and Computational Workflows

Aditya Dimas Dewanto
Informatics Study Program, Universitas Teknologi Yogyakarta, Indonesia

**Corresponding Author: adityadimas2121@gmail.com*

Abstract. *Flood disasters pose persistent socio-economic and environmental challenges, particularly in tropical regions such as Sumatra, Indonesia. Traditional hydrological and GIS-based approaches often struggle to capture complex interactions among terrain, rainfall, land use, and human activities. This review critically examines recent applications of Geospatial Artificial Intelligence (GeoAI) for flood disaster mapping, focusing on machine learning models, geospatial data sources, and computational workflows. Analysis of selected studies highlights that satellite imagery and digital elevation models remain dominant data inputs, while Random Forest, Support Vector Machines, Convolutional Neural Networks, and hybrid models are most frequently applied. Workflow patterns reveal recurring stages of data preprocessing, model training, and post-processing, yet gaps persist in model explainability, feature selection, and generalization across regions. The study underscores the importance of integrating multi-source data, standardizing workflows, and fostering interdisciplinary collaboration to enhance operational flood risk management. Findings provide a foundation for advancing GeoAI research and translating methodological innovations into practical flood preparedness and mitigation strategies.*

Keywords: *Flood Mapping; Geospatial Artificial Intelligence; Machine Learning; Sumatra; Workflow.*

1. INTRODUCTION

Flood disasters continue to impose significant socio-economic and environmental burdens in many tropical regions, especially where climate variability, land-use change, and rapid urbanization intersect. In Indonesia, recurrent flooding has disrupted livelihoods and critical infrastructure, underscoring persistent challenges for disaster risk management. Conventional flood mapping and risk assessment methods, often grounded in hydrological modeling and static GIS analyses, contribute important baseline insights but may fall short in capturing complex, non-linear interactions among terrain, rainfall, land cover, and anthropogenic influences. This shortfall is especially apparent in heterogeneous landscapes where data diversity and spatial dynamics play major roles in flood behavior and impacts (Bentivoglio et al., 2022).

In response to these challenges, Geospatial Artificial Intelligence (GeoAI) has garnered increasing attention as an analytical paradigm that marries machine learning (ML) with spatial data analysis. GeoAI moves beyond traditional spatial techniques by leveraging data-driven algorithms to discover patterns and relationships that are difficult to capture with rule-based models. Across recent studies, ML models such as Random Forest, Support Vector Machine, and deep learning have been used together with GIS and satellite data to produce flood susceptibility and hazard maps with improved predictive performance (Al-Ruzouq et al., 2024).

Naskah Masuk: 15 November 2025; Revisi: 18 Desember 2025; Diterima: 20 Januari 2026;
Terbit: 26 Januari 2026

These advancements reflect a broader shift in disaster science toward integrating AI tools with geospatial data for more adaptive and responsive analytical workflows.

Despite these methodological advances, the literature on GeoAI for flood disaster mapping reveals important gaps, particularly in how existing research structures analytical workflows from data acquisition to decision-ready outputs. Many studies focus narrowly on model accuracy or algorithm comparisons without detailing how data preprocessing, feature selection, and model integration interact to shape results. Moreover, existing review articles on related topics tend to address machine learning applications for specific tasks, such as flood depth estimation or predictive performance metrics, without explicitly synthesizing the broader interplay of models, data sources, and workflows that underpins end-to-end GeoAI practice (A Comprehensive Review of Machine Learning Approaches, 2025; Bentivoglio et al., 2022). This lack of synthesis limits the ability of researchers and practitioners to understand systematic patterns and methodological trade-offs in GeoAI implementations across different contexts.

To address these gaps, this review critically examines recent applications of Geospatial Artificial Intelligence for flood disaster mapping, with a specific focus on relevance to the island of Sumatra. Rather than centering solely on algorithmic innovation, the review analyzes how machine learning models, geospatial data sources, and computational workflows are combined and operationalized in existing research. By highlighting dominant methodological patterns, recurring challenges, and promising directions for future work, this study aims to provide a clearer foundation for advancing GeoAI research and for informing its translation into practical flood risk management practices.

2. LITERATURE REVIEW

In the evolving field of flood disaster analysis, the integration of artificial intelligence with geospatial data has gained momentum as a response to the limitations of conventional hydrological and GIS-based methods. Traditional flood risk tools often depend on physics-based models that require extensive input data and can be computationally intensive, making near-real-time applications challenging in many contexts (Liu et al., 2025). Contemporary research increasingly emphasizes data-driven approaches where machine learning (ML) and deep learning (DL) methods are leveraged to process complex spatial relationships and heterogeneous geospatial datasets for flood mapping and prediction.

Recent studies demonstrate that GeoAI methods, which blend machine learning with geographic information systems and remote sensing, can enhance the accuracy and efficiency of flood susceptibility and hazard mapping. For example, hybrid models that incorporate multi-source geospatial data such as satellite imagery and digital elevation models achieve higher performance in delineating flood-prone areas than traditional models alone (Azeem et al., 2023; Destefanis et al., 2025). The use of multi-source datasets combined with classifiers like Random Forest, Convolutional Neural Networks (CNN), and XGBoost has shown improved predictive capability in urban and riverine flood contexts, highlighting the potential of data-driven spatial models to capture subtle patterns linked to hydrometeorological processes.

Despite these advances, several recurring methodological issues are identified across recent literature. Many empirical studies focus on evaluating individual algorithms or model accuracy metrics without elaborating how various stages of the analytical workflow interact, particularly from raw data acquisition to post-processing and interpretation. Moreover, core challenges such as model explainability, generalization to new geographic contexts, and integration of explainable AI tools for stakeholder communication remain underexplored (Liu et al., 2025). These limitations imply a need for systematic synthesis not only of model types but also of computational frameworks and data preparation strategies that can support both research comparability and operational applicability.

The broader scope of AI-enhanced flood risk research also points to a growing diversity of approaches beyond conventional ML classifiers. Advances in Earth Observation, including the integration of high-resolution satellite and 3D datasets, have expanded the capacity for real-time flood detection and depth estimation, while explainable AI and hybrid physics, data models are emerging as promising avenues to address interpretability and uncertainty issues in geospatial flood analysis. Together, these developments illustrate that the field is moving toward richer, multi-dimensional analytical frameworks that consider not only predictive performance but also data scarcity, model transparency, and decision support utility.

However, most existing reviews remain either domain-specific (e.g., focusing only on flood depth estimation or a subset of ML models) or oriented toward global trends without detailed synthesis of how individual studies operationalize machine learning models with geospatial data workflows. This fragmentation underscores a need for integrative review frameworks that systematically categorize methodological patterns, data characteristics, and workflow designs across studies, especially those with relevance to regional contexts such as Sumatra’s complex flood environments. Such synthesis can clarify prevailing methodological trade-offs and guide future research toward more cohesive and transferable flood risk intelligence systems.

3. RESEARCH METHODOLOGY

This study adopts a structured review approach to synthesize recent research on the application of Geospatial Artificial Intelligence (GeoAI) for flood disaster mapping, with particular attention to studies relevant to the Indonesian and Southeast Asian context. Rather than aiming for exhaustive coverage, the review emphasizes analytical depth by examining how machine learning models, geospatial data sources, and computational workflows are combined and implemented across studies. This approach allows the review to move beyond algorithm comparison and focus on recurring methodological patterns and practical implications, in line with recommendations for integrative and scoping-oriented reviews in applied geospatial research (Arksey & O'Malley, 2005; Liu et al., 2025).

Literature Identification and Selection

Relevant literature was identified through systematic searches in major academic databases, including Scopus, Web of Science, and ScienceDirect. Search strings combined keywords related to flood disasters, GeoAI, machine learning, GIS, and remote sensing, using variations such as "flood mapping," "geospatial artificial intelligence," "machine learning," and "remote sensing." To ensure topical relevance, the review focused on peer-reviewed journal articles published within the last five years. Conference papers and technical reports were excluded unless they provided substantial methodological insights not available in journal publications. Titles and abstracts were first screened for relevance, followed by full-text review to confirm alignment with the study's analytical focus.

Analytical Framework and Review Process

To guide the synthesis process, the review employed an analytical framework that organizes each study around three interconnected components: geospatial data inputs, machine learning models, and computational workflows. This framework reflects the practical reality of GeoAI applications, where model performance is closely shaped by data characteristics and preprocessing strategies. Figure 1 illustrates the overall review framework, showing how literature selection feeds into data extraction, thematic categorization, and cross-study synthesis. By structuring the review in this way, the analysis highlights how methodological choices interact across stages, rather than treating each component in isolation.

The conceptual flow of the review, starting from the flood disaster context in Sumatra and the limitations of conventional flood mapping approaches. It then highlights the emergence of Geospatial Artificial Intelligence (Geo-AI) as an alternative paradigm, followed by the identification and analysis of machine learning-based flood mapping studies. The review synthesizes findings across three analytical dimensions, machine learning models, geospatial data sources, and computational workflows, to derive insights and implications for future flood risk mapping and management, as shown in Figure 1.

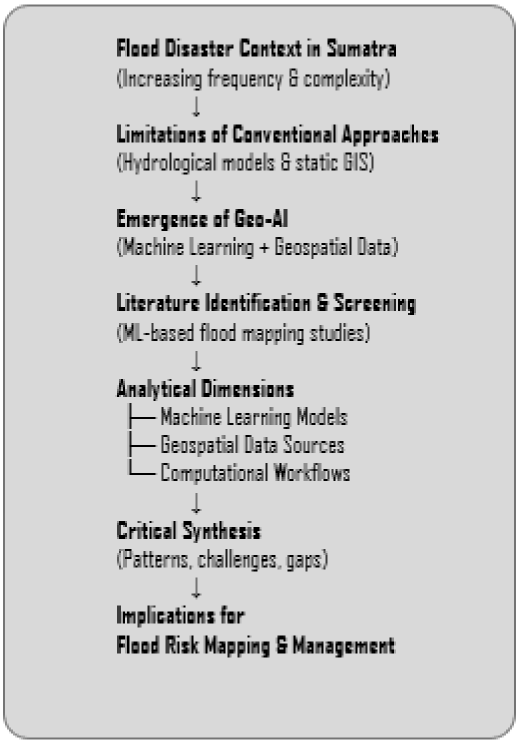

Figure 1. Research Framework of the Geo-AI Review on Flood Disaster Mapping.

Data Extraction and Classification

For each selected article, key methodological attributes were extracted, including study location, data sources, machine learning techniques, validation strategies, and reported limitations. These attributes were then organized into comparative categories to support cross-study analysis. Table 1 presents an overview of how reviewed studies were classified based on dominant data types, modeling approaches, and workflow emphasis. This tabular synthesis provides a concise overview of methodological diversity while supporting deeper qualitative interpretation in subsequent sections.

Table 1. Classification of Reviewed GeoAI Studies for Flood Disaster Mapping.

ASPECT	CATEGORIES
Geospatial Data	Satellite imagery, DEM, rainfall, land use
ML Models	RF, SVM, CNN, XGBoost, hybrid models
Workflow Focus	Data preprocessing, model training, validation
Application Goal	Flood susceptibility, hazard mapping, detection

Synthesis Strategy

The final stage of the review involved qualitative synthesis across the classified studies to identify dominant methodological trends, recurring challenges, and underexplored directions. Rather than ranking models based on accuracy metrics alone, the synthesis examined how studies balance data availability, computational complexity, and interpretability. Particular attention was given to issues frequently noted in the literature, such as data imbalance, limited model explainability, and challenges in transferring trained models across regions. This synthesis approach aligns with recent calls for more reflective and implementation-oriented GeoAI research that bridges methodological innovation with operational flood risk management needs (Bentivoglio et al., 2022; Liu et al., 2025).

4. RESULT AND DISCUSSION

Trends in Geospatial Data Usage

The review of selected studies indicates that satellite imagery and digital elevation models (DEM) remain the dominant geospatial inputs for flood disaster mapping. These datasets provide high spatial coverage and resolution, enabling the capture of terrain characteristics and hydrological patterns essential for model training. Rainfall data and land use/land cover information are increasingly incorporated to reflect dynamic environmental conditions, particularly in tropical and monsoonal regions such as Sumatra (Destefanis et al., 2025). While these datasets improve model comprehensiveness, uneven availability and inconsistent temporal resolution in local contexts can limit predictive performance. This finding emphasizes the need for flexible preprocessing workflows that can harmonize multi-source geospatial datasets, a trend evident in the majority of reviewed studies (Azeem et al., 2023).

Machine Learning Models and Their Application

Across the studies analyzed, Random Forest (RF) remains a widely applied model due to its balance between predictive power and interpretability. Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and hybrid architectures combining multiple algorithms are increasingly explored, reflecting growing sophistication in model design. However, model choice often depends less on systematic evaluation and more on data availability and prior familiarity, resulting in methodological heterogeneity. While some studies demonstrate improved flood susceptibility mapping using deep learning models with remote sensing inputs, others highlight that simpler models, when paired with careful feature selection, can achieve comparable results (Liu et al., 2025). This reinforces the idea that model effectiveness is inseparable from data quality and preprocessing workflows**,** echoing the rationale illustrated previously in Figure 1.

Workflow Patterns and Methodological Gaps

Analysis of the extracted workflows reveals three dominant stages: data preprocessing and feature extraction, model training and validation, and post-processing for spatial interpretation. To visualize trends across studies, Figure 2 summarizes the classification of reviewed studies by geospatial data types, machine learning models, and workflow focus.

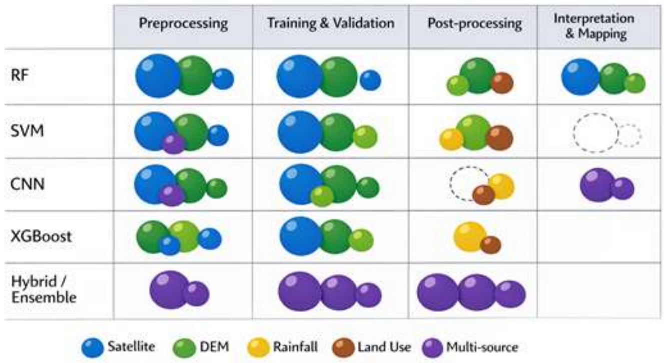

Figure 2. Classification of Reviewed GeoAI Studies for Flood Disaster Mapping.

Bubble size in Figure 2 corresponds to frequency of use across studies, while colors indicate data type, making it easy to identify which model–data–workflow combinations dominate and which remain underexplored. For instance, satellite imagery and DEM feeding Random Forest pipelines appear as large, dark-colored bubbles, whereas the integration of multi-source datasets with deep learning models in post-processing or interpretation stages is represented by smaller or lighter bubbles, highlighting underexplored areas.

Despite these recurring workflow stages, many studies lack detailed reporting on feature selection strategies, hyperparameter tuning, or model generalization across different spatial contexts. Explainability and uncertainty quantification are also underrepresented, even though these aspects are critical for operational decision-making in flood management (Bentivoglio et al., 2022). The combination of Table 1 and Figure 2 underscores that methodological gaps persist in both model design and workflow transparency, suggesting that future GeoAI research should prioritize standardization and reproducibility, particularly in complex regions like Sumatra.

Implications for Operational Flood Risk Management

The synthesis highlights that GeoAI has potential to significantly improve flood hazard mapping and risk assessment, particularly by integrating multi-source geospatial data with flexible machine learning pipelines. Nevertheless, bridging research outputs to operational use remains a challenge. Data scarcity, model transferability, and computational resource requirements constrain the direct adoption of advanced GeoAI models by local disaster management agencies. Addressing these challenges requires interdisciplinary collaboration, including geospatial scientists, AI specialists, and local practitioners, to design adaptive workflows that balance methodological rigor with feasibility. By systematically categorizing methodological patterns, data sources, and workflow designs, this review provides a foundation for translating GeoAI research into actionable insights for flood preparedness and mitigation (Liu et al., 2025; Destefanis et al., 2025).

5. CONCLUSION AND FUTURE DIRECTIONS

Conclusion

This review demonstrates that Geospatial Artificial Intelligence (GeoAI) has significant potential to enhance flood disaster mapping by integrating multi-source geospatial datasets with machine learning pipelines. Satellite imagery and DEM remain the backbone of most studies, while additional inputs such as rainfall and land use/land cover improve model comprehensiveness, particularly in dynamic regions like Sumatra (Destefanis et al., 2025). Random Forest, SVM, CNN, and hybrid models dominate the methodological landscape; however, their effectiveness is inseparable from data quality, preprocessing decisions, and workflow design, as highlighted in Figures 1 and 2.

The visual synthesis in Figure 2 illustrates both dominant trends and underexplored combinations, such as multi-source datasets coupled with deep learning models in post-processing or interpretation stages. Persistent gaps include limited reporting on feature selection strategies, hyperparameter tuning, model generalization, and explainability. Collectively, these findings suggest that the impact of GeoAI is determined not only by algorithmic sophistication but also by holistic integration of data, models, and computational workflows.

Future Directions

Future GeoAI research for flood risk management should focus on developing reproducible, end-to-end workflows that integrate data preprocessing, model training, and spatial interpretation, enhancing methodological transparency and comparability across studies. Incorporating explainable AI is also crucial to improve model interpretability and support actionable decision-making for local disaster management agencies. Models must be adapted to the unique hydrometeorological and socio-environmental conditions of regions like Sumatra to ensure transferability and operational relevance. Integrating multi-source datasets, including high-resolution satellite imagery, DEM, rainfall, land use, and crowdsourced data, can strengthen predictive performance while addressing data scarcity and heterogeneity.

Achieving these advances requires interdisciplinary collaboration among geospatial scientists, AI specialists, and local practitioners to develop adaptive workflows that balance technical rigor with practical feasibility. Collectively, these directions aim to translate methodological innovation into tangible contributions for flood preparedness and mitigation.

REFERENCE

Al Ruzouq, R., Shanableh, A., Yilmaz, A. G., Idris, A., Mukherjee, S., & Khalil, M. A. (2024). Integrating machine learning and geospatial data analysis for comprehensive flood hazard assessment. *Environmental Science and Pollution Research*, *31*, 15234–15252. https://doi.org/10.1007/s11356-024-34286-7

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Ashfaq, S., Rahman, A., Khan, M., & Ali, T. (2025). Flood susceptibility assessment and mapping using GIS-based multi-criteria decision analysis. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2025.125275

Atmaja, T., & Fukushi, K. (2022). Empowering geo-based AI algorithm to aid coastal flood risk analysis: A review and framework development. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, V-3*, 517–523. https://doi.org/10.5194/isprs-annals-V-3-2022-517-2022

Azeem, N., Naseer, R., & Khan, S. (2025). Integrating geospatial and machine learning approaches for flood risk assessment and disaster management: A case study of Sheikhupura District, Pakistan (2022–2024). *Frontiers in Computational Spatial Intelligence*, *3*(1), 11–20.

Badillo Rivera, E. (2025). Flood susceptibility mapping in El Niño phenomenon areas: Hybrid machine learning and SHAP analysis. *Frontiers in Environmental Science*. https://doi.org/10.3389/fenvs.2025.1672107

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: A review of existing applications and future research directions. *Hydrology and Earth System Sciences*, *26*, 4345–4378. https://doi.org/10.5194/hess-26-4345-2022

Comprehensive review of machine learning approaches for flood depth estimation. (2025). *International Journal of Disaster Risk Science, 16*, 433–445. https://doi.org/10.1007/s13753-025-00639-0

Derman, M. T., & Mete, M. O. (2025). Flood risk analysis with explainable geospatial artificial intelligence (GeoAI) techniques. *Systems*, *13*(11), 1007. https://doi.org/10.3390/systems13111007

Destefanis, T., Guliyeva, S., Boccardo, P., & Fissore, V. (2025). Advancing flood detection and mapping: A review of Earth observation services, 3D data integration, and AI-based techniques. *Remote Sensing, 17*(17), 2943. https://doi.org/10.3390/rs17172943

Feizbahr, M., Rahmati, O., Tiefenbacher, J. P., & Lee, S. (2025). Flood susceptibility mapping using machine learning and Sentinel-1 SAR data for enhanced early warning systems. *Remote Sensing, 17*(20), 3471. https://doi.org/10.3390/rs17203471

Flood susceptibility mapping: Integrating machine learning and GIS for enhanced risk assessment. (2024). *Applied Computing and Geosciences, 23*, 100183. https://doi.org/10.1016/j.acags.2024.100183

Islam, T., Zeleke, E. B., Afroz, M., & Melesse, A. M. (2025). A systematic review of urban flood susceptibility mapping: Remote sensing, machine learning, and other modeling approaches. *Remote Sensing, 17*(3), 524. https://doi.org/10.3390/rs17030524

Li, W., Lee, H., Wang, S., Hsu, C. Y., & Arundel, S. T. (2025). Assessment of a new GeoAI foundation model for flood inundation mapping. *arXiv*. https://arxiv.org/abs/2510.23364

Mustikaningrum, R., Nugroho, A., & Suryani, E. (2024). Flood susceptibility mapping using machine learning in Kening River sub-watershed of Bengawan Solo, Indonesia. *Urban & Environmental Technology, 8*(2), 199–210.

Putra, A. R., Nugraheni, D., & Saputra, M. A. (2025). Flood prediction using machine learning model integrated with GIS. *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika, 11*(1), 45–54.

Rahman, Z. U., Khan, A., Ali, S., & Ahmad, M. (2025). Flood susceptibility mapping using supervised machine learning models. *Environmental Technology & Innovation*. https://doi.org/10.1080/19475705.2025.2516728

Rezvani, S. M. H. S., Silva, M. J. F., & Almeida, N. M. d. (2025). Mapping geospatial AI flood risk in national road networks. *ISPRS International Journal of Geo-Information, 13*(9), 323. https://doi.org/10.3390/ijgi13090323

Santos, L. B. L., Pereira, J., & Costa, R. (2025). Machine learning-based hydrological models for flash floods: Recent advances and challenges. *Water*. https://doi.org/10.1007/s44268-025-00071-9

Venegas Quiñones, H. L., García Chevesich, P., Valdés Pineda, R., Ferré, T. P. A., Gupta, H., Groenendyk, D., Valdés, J. B., McCray, J. E., & Bakkensen, L. (2024). Creating sustainable flood maps using machine learning and free remote sensing data in unmapped areas. *Sustainability, 16*(20), 8918. https://doi.org/10.3390/su16208918

Zhang, J., Guo, W., Chang, S. W., Nguyen, D. D., & Ngo, H. H. (2025). Data-driven innovations in flood hazard assessment with machine learning. *Environmental Sustainability*. Article 2507001017.