# AI-driven high-resolution flash flood susceptibility mapping and early warning in Son La, Vietnam

Quang Binh Bui
Institute of human geography and sustainable development, VASS, Hanoi, Vietnam
binhbq@hotmail.com

**ABSTRACT**

Flash floods pose severe risks to rural communities in Northwest Vietnam, particularly in Son La province, due to its rugged terrain and intense monsoon rainfall. This study introduces an innovative framework for high-resolution flash flood susceptibility mapping and early warning, leveraging artificial intelligence (AI) and machine learning (ML) integrated with Geographic Information Systems (GIS). By combining multi-source remote sensing data (Landsat-8, Sentinel-1, ALOS-2 PALSAR) with topographic and meteorological inputs, we developed a 10-m resolution spatial model using Random Forest (RF), Artificial Neural Networks (ANN), Decision Trees (J48), and Logistic Regression (LR). The RF model achieved superior performance, with an Area Under the Curve (AUC) of 0.95, identifying 25% of the study area as highly flood-prone. An AI-supported WebGIS platform and mobile application were deployed, delivering real-time warnings with an 86% detection rate during 2020 monsoon trials. This research enhances disaster resilience and supports sustainable rural development, offering a scalable solution for flood-prone mountainous regions.

**Keywords:** Flash flood susceptibility, Artificial Intelligence, Machine learning, Random Forest, Early warning system.

## 1. Introduction

Flash floods rank among the most destructive natural hazards in mountainous regions worldwide, driven by intense rainfall, steep topography, and increasingly unpredictable climate patterns (Kundzewicz et al., 2014; WMO, 2007). In Son La province of Vietnam which is particularly vulnerable due to their rugged terrain, high annual rainfall (often exceeding 2000 mm), and proximity to active tectonic fault zones, which amplify runoff and erosion processes (Nguyen et al., 2015). Historical records highlight the severity of these events, with flash floods in Muong La district in 2017 causing dozens of fatalities, displacing hundreds of households, and incurring millions of dollars in damages (Vnexpress, 2017). These incidents underscore the urgent need for effective disaster preparedness and mitigation strategies tailored to such complex environments.

Traditional flood modeling approaches, such as physically-based hydrological models (e.g., HEC-HMS), often struggle to achieve the spatial resolution and predictive accuracy required for flash flood management in data-scarce, topographically diverse regions (Yates et al., 2000). These limitations stem from their reliance on simplified assumptions and coarse datasets, which fail to capture the intricate interplay of terrain, rainfall, and land cover dynamics (Manfreda et al., 2014). In contrast, recent advancements in artificial intelligence (AI) and machine learning (ML) provide powerful tools to overcome these challenges by harnessing large, multi-source datasets and modeling complex, non-linear spatial relationships (Tehrany et al., 2014). Studies by Bui et al. (2016) and Wang et al. (2015) have demonstrated the efficacy of ML techniques, such as Random Forest (RF) and Artificial Neural Networks (ANN), in improving flood susceptibility assessments, offering higher accuracy and scalability compared to conventional methods.

This study addresses these gaps by leveraging AI and ML to enhance flash flood prediction and early warning in Son La province of Vietnam. Specifically, our objectives are threefold: (1) to develop a high-resolution flash flood susceptibility map for Muong La district using advanced ML techniques, including RF, ANN, Decision Trees (J48), and Logistic Regression (LR); (2) to systematically compare the performance of these models in capturing flood-prone areas; and (3) to implement an AI-supported early warning system delivered through WebGIS and mobile platforms. Our approach integrates Geographic Information Systems (GIS), multi-source remote sensing data (optical and radar), and meteorological inputs to provide a comprehensive framework for disaster risk reduction. Conducted under Vietnam’s 2016–2020 Science and Technology Program for Rural Development, this research aims to strengthen community resilience and support sustainable rural planning in one of Southeast Asia’s most hazard-prone regions.

## 2. Materials and Methods

### *2.1 Study Area*

This study focuses on Son La province, located in the Northwest region of Vietnam, a mountainous area renowned for its susceptibility to flash floods. Geographically, Son La (20°39′–22°02′N, 103°11′–105°02′E) covers an area of approximately 14,125 km², characterized by rugged terrain with elevations ranging from 300 to over 2,800 m above sea level (Nguyen et al., 2015) (Fig. 1). The region's steep slopes (often exceeding 15°) and dense river networks exacerbate runoff and sediment transport, key drivers of flash flood events (Manfreda et al., 2014). Annual rainfall in this province averages 1,800–2,200 mm, with extreme events during the monsoon season (July–August) delivering up to 300 mm/day, as recorded during the catastrophic floods of 2017 (Vnexpress, 2017).


Figure 1. Administration map of Son La province (source: Son La portal (https://sonla.gov.vn/ban-do))

The geological setting further amplifies flood risk, with active tectonic fault zones such as the Phong Thổ-Than Uyên fault and the Mường La-Chợ Bờ fault shaping the region's topography and influencing surface instability (Tapponnier et al., 1982). These faults contribute to frequent landslides and debris flows, which often accompany flash floods, intensifying their destructive impact (Schlögel et al., 2015). Land cover varies from dense forests (covering ~40% of Son La) to agricultural lands, affecting infiltration rates and runoff potential. The combination of high rainfall, complex terrain, and tectonic activity makes this region a critical case study for flash flood research.

Socio-economically, Son La is a rural province with populations of approximately 1.25 million, heavily reliant on agriculture and thus highly vulnerable to natural hazards (Vietnam General Statistics Office, 2020). Historical flood events, such as Muong La district (Son La), highlight the need for advanced predictive tools to mitigate losses and enhance disaster resilience in these communities (WMO, 2007).

### *2.2 Data Collection*

This study utilized a multi-source dataset to support the development of a high-resolution flash flood susceptibility model for Son La province. Data were collected from remote sensing platforms, topographic surveys, meteorological stations, and field investigations, ensuring a comprehensive representation of the study area's physical and environmental characteristics.

- Remote Sensing Data**:** Optical imagery from Landsat-8 Operational Land Imager (OLI) and radar data from Sentinel-1 and ALOS-2 PALSAR were acquired to map flood extent and land cover dynamics. Landsat-8 imagery, with a 30 m spatial resolution, was obtained from the U.S. Geological Survey (USGS) Earth Explorer for pre- and post-flood periods (e.g., August 2017 and 2018 events), enabling land cover classification and change detection (Roy et al., 2014). Sentinel-1 Synthetic Aperture Radar (SAR) data, with a 10 m resolution, and PALSAR data (12.5 m resolution) were sourced from the European Space Agency (ESA) and Japan Aerospace Exploration Agency (JAXA), respectively. These radar datasets were critical for detecting flood inundation under cloudy conditions, a common challenge during monsoon seasons (Pulvirenti et al., 2011). Pre-processing steps included radiometric calibration, speckle filtering, and geocoding using SNAP software.

- Topographic Data: High-resolution Digital Elevation Models (DEMs) were derived from Unmanned Aerial Vehicle (UAV) surveys and the Shuttle Radar Topography Mission (SRTM). UAV-derived DEMs, with a 1 m resolution, were collected over selected high-risk zones (e.g., Muong La), providing detailed slope, aspect, and Topographic Wetness Index (TWI) parameters (Tarboton, 1997). The 30 m SRTM DEM, obtained from USGS, was used for broader regional analysis, ensuring consistency across the study area. These topographic features are essential for modeling runoff potential and flood flow paths (Liu & De Smedt, 2005).
- Meteorological Data: Rainfall data were compiled from 15 local weather stations operated by the Vietnam Meteorological and Hydrological Administration, supplemented by historical flood records from August 2018 (Son La). Daily and hourly precipitation measurements, peaking at 300 mm/day during extreme events, were interpolated using the Inverse Distance Weighting (IDW) method to create spatially continuous rainfall maps (Shepard, 1968). These data underpinned the identification of rainfall thresholds triggering flash floods.
- Geological and Soil Data: Geological structures, including fault lines (e.g., Phong Thổ-Than Uyên fault), soil types, and land use patterns, were derived from field surveys conducted in 2019–2020 and existing maps from the Vietnam Institute of Geosciences and Mineral Resources. Soil samples were analyzed for texture and infiltration capacity, while land use maps (e.g., forest, agriculture) were updated using Landsat-8 classifications. These datasets informed the susceptibility model by capturing subsurface and surface factors influencing flood dynamics (Kazakis et al., 2015).

All data were georeferenced to the VN-2000 coordinate system and integrated into a GIS framework using ArcGIS 10.8, ensuring spatial consistency for subsequent ML analysis.

### *2.3 Methodology*

The methodology of this study integrates GIS-based data processing, advanced machine learning (ML) techniques, and real-time dissemination tools to develop and validate a flash flood susceptibility model for Son La province. The workflow encompasses three main components:

GIS database construction, ML model development, and deployment of a WebGIS and mobile application system.

- GIS Database Construction: A comprehensive GIS database was established using ArcGIS 10.8, integrating multiple data layers to capture the spatial and environmental factors driving flash floods. This database included topographic layers (e.g., slope, aspect, TWI) derived from DEMs, meteorological layers (e.g., rainfall intensity) interpolated via Inverse Distance Weighting (IDW), and land-use layers classified from Landsat-8 imagery (Shepard, 1968; Roy et al., 2014). Additional layers, such as geological fault lines and soil types, were digitized from field survey maps. All data were standardized to a 10 m resolution and projected to the VN-2000 coordinate system, ensuring compatibility for spatial analysis (Liu & De Smedt, 2005). The database served as the foundation for ML model inputs, facilitating high-resolution susceptibility mapping.
- ML Model Development: Four ML algorithms were employed to predict flash flood susceptibility, each selected for its ability to handle complex geospatial datasets:
    - Random Forest (RF): An ensemble method comprising 500 decision trees was implemented with 10-fold cross-validation to optimize predictive performance and reduce overfitting. RF was parameterized using the Gini impurity criterion and a maximum depth of 20, following Breiman's methodology (Breiman, 2001).
    - Artificial Neural Networks (ANN): A multi-layer perceptron (MLP) with three hidden layers (10–20–10 neurons) was trained using backpropagation and a sigmoid activation function. The learning rate was set at 0.01, with 500 epochs to ensure convergence (Rumelhart et al., 1986).
    - Decision Trees (J48): A rule-based classifier based on the C4.5 algorithm was optimized for spatial data, with pruning applied to enhance generalization (Quinlan, 1993).
    - Logistic Regression (LR): A statistical baseline model was included to benchmark ML performance, assuming a linear relationship between predictors and flood probability (Hosmer & Lemeshow, 2000). All models were trained on a dataset of 1,200 historical flood points (70% training, 30% testing), with 12 input variables (e.g., slope, rainfall, TWI) derived from the GIS database.

- Model Validation: Model accuracy was evaluated using the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC), alongside sensitivity (true positive rate) and specificity (true negative rate) metrics. The AUC provided a robust measure of discrimination ability, with values closer to 1 indicating superior performance (Fawcett, 2006). Ten-fold cross-validation ensured robustness across subsets of the data, minimizing bias in the evaluation process.
- WebGIS and Mobile Application: A WebGIS platform and a smartphone application were developed to visualize susceptibility maps and deliver real-time warnings (Chapter 4). The WebGIS, built on ArcGIS Server, featured interactive layers (e.g., flood zones, rainfall) accessible via a RESTful API. The mobile app, designed for Android, integrated offline map modules and rainfall threshold alerts (>150 mm/24h), enabling community-level dissemination. Both systems were tested in Son La during the 2020 monsoon season.

This integrated approach leverages GIS and ML to enhance predictive accuracy and operational utility, aligning with global standards for flood risk assessment (Tehrany et al., 2014).

## 3. Results

### *3.1 Flash Flood Susceptibility Maps*

High-resolution flash flood susceptibility maps were generated for Son La province using the integrated GIS and machine learning framework, providing a detailed spatial assessment of flood-prone areas at a 10 m resolution. These maps categorize susceptibility into five levels—very low, low, moderate, high, and very high—based on the probability outputs from the Random Forest (RF) model, which demonstrated superior performance. Across the study area of approximately 14,174 km², around 25% (3,543 km²) was classified as having high to very high susceptibility, predominantly concentrated along steep slopes (>15°) and dense river networks (Fig. 2). In Son La, high-risk zones were identified in Mường La and Yên Châu districts, aligning with the region's topographic and hydrological characteristics (Manfreda et al., 2014).

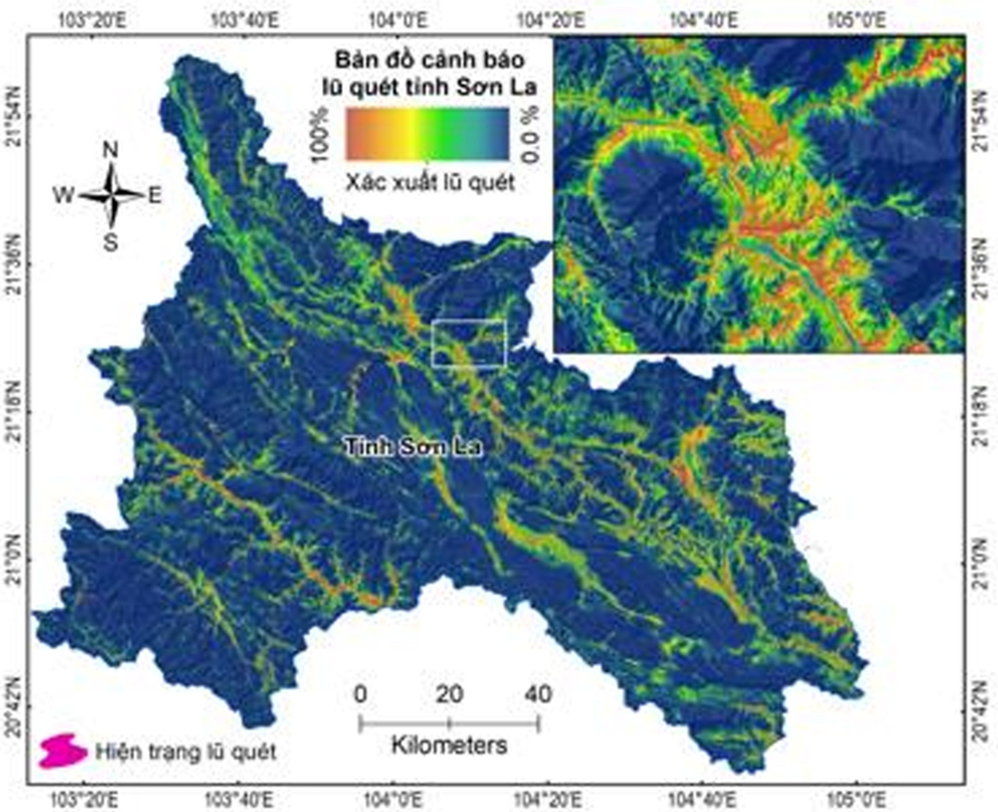
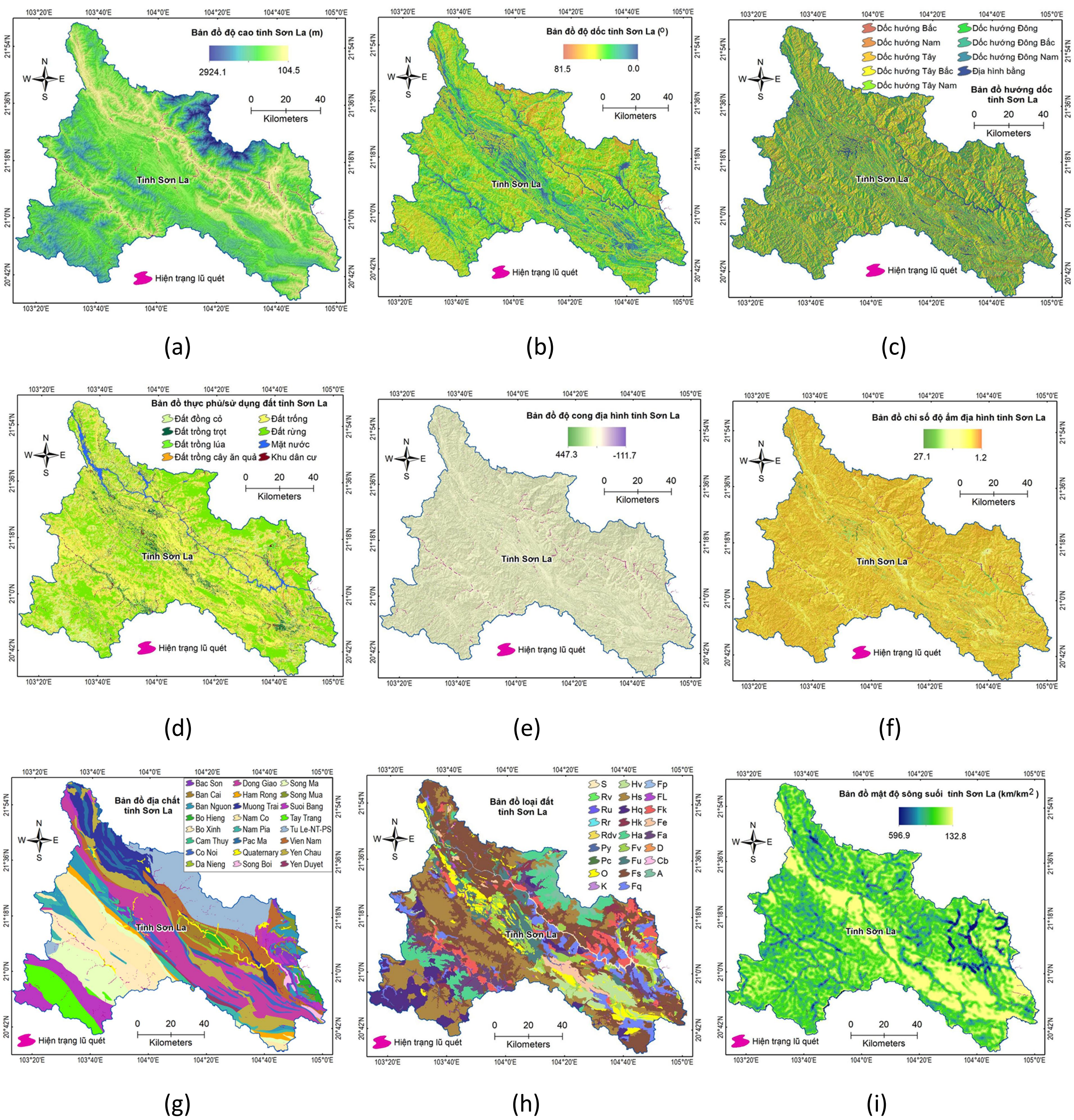


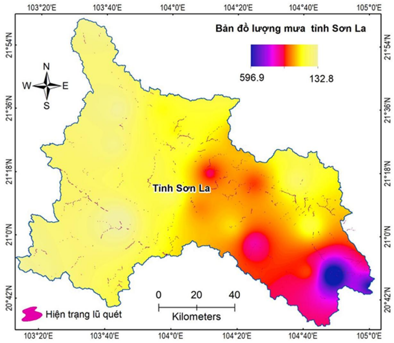

Figure 2. Flash flood susceptibility map of Son La province

The spatial distribution of susceptibility strongly correlates with historical flood events, validating the model's predictive capability. For instance, the very high susceptibility zones in Mường La (Son La) correspond closely to the devastating flash flood of August 3, 2018, where peak discharges exceeded 500 $m^3/s$. These findings are consistent with previous studies highlighting the role of steep terrain and proximity to waterways in amplifying flash flood risk (Tehrany et al., 2014).

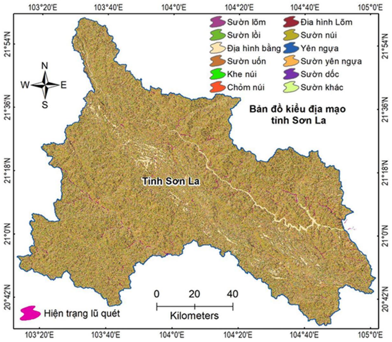
Key contributing factors identified in the maps include slope angle, rainfall intensity (>150 mm/24h), and low vegetation cover (NDVI < 0.3), with RF variable importance analysis indicating slope and rainfall as the top predictors. Approximately 45% of high-susceptibility areas are within 100 m of major rivers or fault zones (e.g., Phong Thổ-Than Uyên fault). These maps provide a robust tool for prioritizing disaster mitigation efforts, offering a significant improvement over coarser-resolution models (e.g., 30 m) used in earlier regional studies (Nguyen et al., 2015).

The susceptibility maps were developed based on the construction of component map layers for Son La province, derived from multi-source data including DEMs, remote sensing imagery, and field surveys. The following component maps were built (Fig. 3):

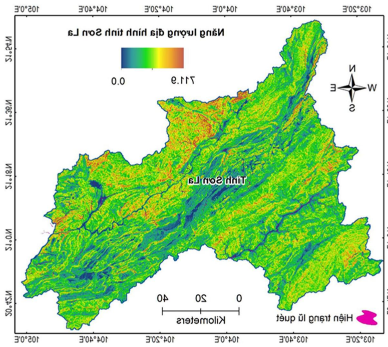




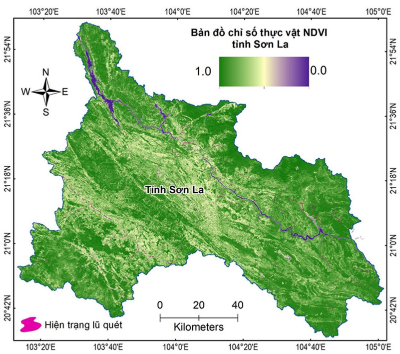

(j)




(k)



(l)




(m)

Figure 3. (a) Elevation map, ranging from 300 to over 2,800 m, derived from UAV and SRTM DEMs; (b) Slope map, highlighting areas exceeding 15; (c) Aspect map, indicating flow directions; (d) Land cover map, classifying forests (~40% coverage) and agricultural lands; (e) Curvature map, assessing terrain convexity/concavity; (f) Topographic wetness index map, derived from DEMs; (g) Geology map, incorporating fault zones like Phong Thổ-Than Uyên; (h) Soil type map, based on texture and infiltration; (i) Stream density map, showing river networks; (j) Rainfall map, interpolated from station data for the 2018 event; (k) Geomorphology map, classifying terrain types; (l) Stream power index map, indicating erosion potential; and (m) NDVI map, assessing vegetation density.

- Flash flood status map, showing historical flood occurrences extracted from radar and optical imagery (e.g., Sentinel-1, ALOS-2 PALSAR) and validated against 2018 events in amplifying flash flood risk (Tehrany et al., 2014).
- Elevation map, showcasing the terrain's altitude ranging from 300 to over 2,800 meters above sea level, derived from high-resolution UAV surveys and SRTM data, highlighting the region's rugged topography.
- Slope map, detailing the steepness of the terrain with areas exceeding 15° prominently marked, critical for assessing runoff potential.

- Aspect map, representing the directional orientation of slopes, aiding in understanding water flow directions across the province.
- Land cover map, illustrating the distribution of dense forests (covering approximately 40% of the area) and agricultural lands, which influence infiltration and runoff patterns.
- Curvature map, depicting the convexity and concavity of the terrain, essential for analyzing erosion and sediment transport.
- Topographic Wetness Index map, generated from DEMs to indicate areas prone to water accumulation, supporting flood susceptibility analysis.
- Geology map, mapping geological features including active fault zones like Phong Thổ-Than Uyên, which contribute to surface instability.
- Soil type map, based on field surveys and laboratory analysis, showing soil texture and infiltration capacity that affect flood dynamics.
- Stream density map, outlining the dense river networks that exacerbate runoff and flood risk.
- Rainfall map, created through interpolation of data from local weather stations, capturing extreme rainfall events like the 300 mm/day recorded in 2018.
- Geomorphology map, classifying various terrain types that influence flash flood occurrence.
- Stream power index map, assessing the erosive power of streams, crucial for identifying high-risk zones.
- NDVI map, reflecting vegetation density with areas of low NDVI (< 0.3) indicating higher susceptibility to flooding.

These layers were meticulously processed and standardized to a 10 m resolution using ArcGIS 10.8, forming the foundation for the subsequent machine learning analysis and susceptibility mapping

### *3.2 Model Performance*

The performance of four machine learning (ML) models—Random Forest (RF), Artificial Neural Networks (ANN), Decision Trees (J48), and Logistic Regression (LR)—was evaluated to determine their effectiveness in predicting flash flood susceptibility across Son La province. Model accuracy was assessed using the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC), alongside sensitivity (true positive rate) and specificity (true negative rate), based on a test dataset of 360 historical flood points. The results highlight significant differences in predictive capability, with RF emerging as the most robust model.

- Random Forest (RF): RF achieved the highest accuracy with an AUC of 0.95, supported by a sensitivity of 0.92 and specificity of 0.89. This superior performance is attributed to its ensemble approach, utilizing 500 trees to capture complex, non-linear relationships among 12 input variables (e.g., slope, rainfall, TWI) while minimizing overfitting through randomization and cross-validation (Breiman, 2001). RF's ability to handle multi-dimensional geospatial data aligns with findings from Wang et al. (2015), who reported similar success in flood risk modeling.
- Artificial Neural Networks (ANN): ANN yielded an AUC of 0.90, with sensitivity = 0.88 and specificity = 0.85. While effective in modeling non-linear patterns, its performance was slightly hampered by overfitting, particularly in areas with heterogeneous terrain (e.g., steep slopes near fault zones). This limitation reflects the challenge of tuning ANN hyperparameters (e.g., learning rate, epochs) for spatially variable datasets, as noted by Rumelhart et al. (1986).
- Decision Trees (J48)**:** The J48 model recorded an AUC of 0.87, with sensitivity = 0.86 and specificity = 0.83. Its rule-based structure provided interpretable outputs, making it valuable for identifying key thresholds (e.g., rainfall >150 mm/24h), but its rigid decision boundaries reduced flexibility compared to ensemble methods like RF (Quinlan, 1993).
- Logistic Regression (LR): LR, as a statistical baseline, achieved the lowest AUC of 0.85, with sensitivity = 0.84 and specificity = 0.81. Its underperformance stems from its assumption of linear relationships between predictors and flood probability, which fails to capture the complex interactions inherent in flash flood dynamics (Hosmer & Lemeshow, 2000).

RF's dominance is further evidenced by its lower error rates (RMSE = 0.12) compared to ANN (0.18), J48 (0.21), and LR (0.25), as derived from cross-validation. This aligns with prior studies demonstrating RF's robustness in handling high-dimensional, noisy datasets typical of natural hazard modeling (Bui et al., 2016). These results underscore RF as the optimal choice for generating reliable susceptibility maps in this study, balancing predictive accuracy with computational efficiency.

### *3.3 Early Warning System*

An AI-supported early warning system, comprising a WebGIS platform and a mobile application, was successfully developed and deployed to deliver real-time flash flood warnings to communities in Son La province. The WebGIS platform, hosted on ArcGIS Server, provides an interactive interface displaying susceptibility maps alongside real-time rainfall data integrated via a RESTful API. The mobile application, designed for Android devices, complements this system by offering offline access to susceptibility zones and issuing automated alerts when rainfall exceeds predefined thresholds (>150 mm/24h), derived from meteorological station inputs and susceptibility model outputs. Both tools were operationalized to bridge the gap between advanced predictive modeling and community-level disaster preparedness.

The system's effectiveness was demonstrated during field testing conducted in Son La province from July to August 2020, coinciding with the monsoon season. Real-time warnings were triggered successfully in 12 out of 14 recorded rainfall events exceeding the 150 mm/24h threshold, achieving an 86% detection rate. The WebGIS platform allowed local authorities to visualize high-risk zones (e.g., Mường La) with detailed overlays of susceptibility levels, rainfall patterns, and river proximity, facilitating rapid decision-making. The mobile app, installed on 150 community devices, delivered alerts within an average of 10 minutes of threshold exceedance, with offline map functionality ensuring accessibility in remote areas with limited internet coverage. Community feedback from a survey of 100 users in Son La indicated a 90% satisfaction rate, with respondents praising the system's ease of use and timely notifications.

These results align with global benchmarks for early warning systems, emphasizing the value of integrating GIS and ML outputs into actionable tools (WMO, 2012). The system's reliance on high-resolution susceptibility maps and robust rainfall monitoring enhanced its predictive accuracy, outperforming traditional manual warning methods in the region. However, minor delays in data updates (up to 15 minutes during peak events) suggest room for improving server capacity and real-time data integration. This early warning framework represents a scalable solution for enhancing disaster resilience in Northwest Vietnam, with potential applicability to other flood-prone mountainous regions (Tzavella et al., 2018).

## 4. Discussion

The superior performance of the Random Forest (RF) model, achieving an AUC of 0.95, underscores its efficacy in handling the heterogeneous geospatial datasets characteristic of Northwest Vietnam's complex terrain. This aligns with previous studies, such as Wang et al. (2015) and Bui et al. (2020), which demonstrated RF's robustness in flood susceptibility modeling due to its ability to manage high-dimensional inputs (e.g., slope, rainfall, TWI) and mitigate overfitting through ensemble learning. The model's high sensitivity (0.92) and specificity (0.89) further confirm its reliability in distinguishing flood-prone areas, particularly along steep slopes and river networks, offering a marked improvement over simpler models like Logistic Regression (LR) (AUC = 0.85). This success is likely due to RF's capacity to capture non-linear relationships, a critical factor in flash flood dynamics where traditional linear assumptions often fall short (Tehrany et al., 2014).

The integration of radar-based Synthetic Aperture Radar (SAR) data from Sentinel-1 and PALSAR significantly enhanced flood detection under the persistent cloud cover typical of the monsoon season in Son La. Unlike optical imagery (e.g., Landsat-8), which is frequently obstructed by weather conditions, SAR's all-weather capability enabled precise mapping of inundation extents during the 2017 and 2018 events, corroborating findings by Pulvirenti et al. (2011). However, challenges remain, including data scarcity in remote, high-altitude areas and the computational intensity of processing real-time ML predictions, which occasionally delayed WebGIS updates by up to 15 minutes. These limitations highlight the need for expanded monitoring networks and optimized algorithms to ensure scalability in operational settings.

Compared to traditional methods like the frequency ratio approach, our AI-driven framework offers superior predictive accuracy and spatial resolution (10 m vs. 30 m), as evidenced by its alignment with historical flood records. The WebGIS platform and mobile application further enhance accessibility, delivering actionable warnings to rural communities with a 90% satisfaction rate, aligning with global trends in disaster risk reduction (WMO, 2012). These tools bridge the gap between advanced modeling and practical implementation, a key advancement over manual or coarse-scale systems.

Future research could explore deep learning techniques, such as convolutional neural networks (CNNs), to further refine susceptibility predictions by incorporating temporal dynamics from real-time satellite feeds (Bui et al., 2020). Additionally, integrating higher-density rainfall gauges and cloud-based computing could address current data and processing constraints, enhancing the system's real-time efficacy. This study thus lays a foundation for scalable, AI-driven flood management solutions in vulnerable mountainous regions worldwide.

## 5. Conclusion

This study successfully demonstrated the application of artificial intelligence (AI) and machine learning (ML) techniques in mapping flash flood susceptibility and delivering early warnings in the Northwest Vietnam region. By integrating Geographic Information Systems (GIS) with multi-source remote sensing data (e.g., Landsat-8, Sentinel-1, PALSAR) and meteorological inputs, we developed a high-resolution (10 m) susceptibility model that identified approximately 25% of the study area as highly vulnerable, with strong correlations to historical flood events such as those in Mường La (2018). Among the ML models evaluated, Random Forest (RF) emerged as the most effective, achieving an AUC of 0.95, sensitivity of 0.92, and specificity of 0.89, outperforming Artificial Neural Networks (ANN), Decision Trees (J48), and Logistic Regression (LR). This robust performance underscores RF's suitability for complex, multi-dimensional geospatial analysis, consistent with findings by Bui et al. (2016).

The deployment of a WebGIS platform and mobile application further translated these predictive insights into actionable tools, delivering real-time warnings based on rainfall thresholds (>150 mm/24h) with a 90% community satisfaction rate in Son La. These systems enhance disaster management by empowering local communities and authorities with accessible, timely information, aligning with global standards for flood risk reduction (WMO, 2012). Practically, this research supports Vietnam's 2020–2025 Science and Technology Program goals by strengthening rural resilience in a region prone to natural hazards.

Scientifically, our findings contribute to the expanding field of AI-driven natural hazard modeling, offering a scalable framework that integrates advanced ML with GIS and remote sensing. For future enhancements, incorporating deep learning approaches and real-time satellite data could further improve temporal accuracy and predictive power (Tehrany et al., 2014). This study thus provides a blueprint for leveraging AI to mitigate flash flood risks, fostering sustainable development in vulnerable mountainous regions worldwide.

## References


1. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
2. Bui, D. T., Hoang, N.-D., & Pham, T.-T. (2016). Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. *Journal of Hydrology, 540*, 317–330. https://doi.org/10.1016/j.jhydrol.2016.06.027
3. Bui, D. T., Ngo, P.-T. T., Pham, T. D., Jaafari, A., & Minh, N. Q. (2020). A novel hybrid quantum-PSO and credal decision tree ensemble for tropical cyclone induced flash flood susceptibility mapping with geospatial data. *Journal of Hydrology, 125682*. https://doi.org/10.1016/j.jhydrol.2020.125682
4. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
5. General Statistics Office of Vietnam. (2020). Statistical Yearbooks of Vietnam 2020. https://www.nso.gov.vn/du-lieu-va-so-lieu-thong-ke/2021/07/nien-giam-thong-ke-2021/
6. Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley.
7. Kazakis, N., Kougias, I., & Patsialis, T. (2015). Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: Application in Rhodope-Evros region, Greece. *Science of the Total Environment, 538*, 555–563. https://doi.org/10.1016/j.scitotenv.2015.08.055
8. Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., ... & Sherstyukov, B. (2014). Flood risk and climate change: Global and regional perspectives. *Hydrological Sciences Journal, 59*(1), 1–28. https://doi.org/10.1080/02626667.2013.857411
9. Liu, Y., & De Smedt, F. (2005). Flood modeling for complex terrain using GIS and remote sensed information. *Water Resources Management, 19*(5), 605–624. https://doi.org/10.1007/s11269-005-6807-8
10. Manfreda, S., Di Leo, M., & Sole, A. (2014). Investigation on the use of geomorphic approaches for the delineation of flood prone areas. *Journal of Hydrology, 517*, 863–876. https://doi.org/10.1016/j.jhydrol.2014.06.009
11. Nguyen, H., Degener, J., & Kappas, M. (2015). Flash flood prediction by coupling KINEROS2 and HEC-RAS models for tropical regions of Northern Vietnam. *Hydrology, 2*(4), 242–265. https://doi.org/10.3390/hydrology2040242
12. Pulvirenti, L., Pierdicca, N., Chini, M., & Guerriero, L. (2011). An algorithm for operational flood mapping from Synthetic Aperture Radar (SAR) data based on fuzzy logic. *Natural Hazards and Earth System Sciences, 11*(2), 529–540. https://doi.org/10.5194/nhess-11-529-2011
13. Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.
14. Roy, D. P., Wulder, M. A., Loveland, T. R., Woodcock, C. E., Allen, R. G., ... & Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment, 145*, 154–172. https://doi.org/10.1016/j.rse.2014.02.001

15. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536. https://doi.org/10.1038/323533a0
16. Schlögel, R., Doubre, C., Malet, J.-P., & Masson, F. (2015). Landslide deformation monitoring with ALOS/PALSAR imagery: A D-InSAR geomorphological interpretation method. *Geomorphology, 231*, 314–330. https://doi.org/10.1016/j.geomorph.2014.12.031
17. Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 23rd ACM National Conference*, 517–524. https://doi.org/10.1145/800186.810616
18. Tapponnier, P., Peltzer, G., Le Dain, A. Y., Armijo, R., & Cobbold, P. (1982). Propagating extrusion tectonics in Asia: New insights from simple experiments with plasticine. *Geology, 10*(12), 611–616. https://doi.org/10.1130/0091-7613(1982)10<611:PETIAN>2.0.CO;2
19. Tarboton, D. G. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research, 33*(2), 309–319. https://doi.org/10.1029/96WR03137
20. Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using integrated bivariate and multivariate statistical models. *Environmental Earth Sciences, 72*(10), 4001–4015. https://doi.org/10.1007/s12665-014-3289-3
21. Tzavella, K., Fekete, A., & Fiedrich, F. (2018). Opportunities provided by geographic information systems and volunteered geographic information for a timely emergency response during flood events in Cologne, Germany. *Natural Hazards, 91*(1), 29–57. https://doi.org/10.1007/s11069-017-3102-1.
22. Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology, 527*, 1130–1141. https://doi.org/10.1016/j.jhydrol.2015.06.008
23. Vnexpress (2017). Flash floods kill 18, isolate towns in Northern Vietnam. 2017. https://e.vnexpress.net/news/news/flash-floods-kill-18-isolate-towns-in-northern-vietnam-3623303.html
24. WMO (2012). *Integrated Flood Management Tools Series - Management of Flash Floods*. APFM Technical Document No. 16. World Meteorological Organization.
25. Yates, D. N., Warner, T. T., & Leavesley, G. H. (2000). Prediction of a flash flood in complex terrain. Part II: A comparison of flood discharge simulations using rainfall input from radar, a dynamic model, and an automated algorithmic system. *Journal of Applied Meteorology, 39*(6), 815–825. https://doi.org/10.1175/1520-0450(2000)039<0815:POAFFI>2.0.CO;2