https://doi.org/10.71573/fqxzdp48

Intelligent Flood Risk Management in Jeju Island: Grid-Based AI Prediction and Assessment

Hyeontae Moon [1], Kyung-Tak Kim [1] https://orcid.org/0000-0002-9596-2055 & Gilho Kim [1]*

[1] Department of Hydro Science and Engineering Research, Korea Institute of Civil Engineering and Building Technology, Goyang, 10223, South Korea
**Corresponding author email: ktkim1@kict.re.kr*

Abstract
Jeju Island, a volcanic island in Korea, exhibits distinct hydrogeomorphological features that pose challenges for disaster management. Traditional hydrological models and qualitative risk assessment methods developed for inland areas are insufficient for addressing Jeju Island's unique conditions. A localized approach is necessary to integrate hydrogeomorphological and urban drainage factors with hazard-related data. This study develops a grid-based AI flood risk assessment method tailored to Jeju Island's characteristics, considering urban drainage systems to improve prediction accuracy. The objectives are to: (1) establish a grid-type flood damage and flood risk influencing factor database that reflects Jeju Island's unique hydrogeomorphological and hazard characteristics, (2) create a binary classification deep learning model integrating hydrological, hazard, and urban drainage factors to assess flood risk, and (3) propose warning standards based on flood damage data and risk assessments, accounting for the specific disaster patterns of Jeju Island. This study seeks to overcome the limitations of existing disaster management approaches by leveraging AI and hydrological analysis technologies, addressing Jeju Island's increasing vulnerability to disasters. Through the integration of grid-based data and advanced ML models, this research aims to provide a comprehensive framework for predicting and managing flood risks, ultimately enhancing resilience and safety on the area.

Highlights
- This study develops a grid-based AI flood risk assessment model for Jeju Island, incorporating hydrogeomorphology and urban drainage factors.
- It leverages AI and hydrological analysis to assess flood risk and improve prediction accuracy.
- The research establishes flood damage databases and proposes warning standards to enhance disaster resilience.

Introduction
Flood risk assessment and disaster management have become critical challenges in regions with unique hydrogeographical and climatic characteristics, such as Jeju Island, Korea (Ali et al., 2017). Unlike inland areas, Jeju Island, a volcanic island, exhibits distinct hydrogeomorphological features that complicate the application of conventional hydrological models and qualitative risk assessment methods developed for inland regions (Kang & Yoo, 2020). Additionally, the vulnerability of Jeju Island is compounded by its geographical location, directly in the path of typhoons, and its rising population due to rapid urbanization and tourism growth (Lim et al., 2023). Furthermore, existing rainfall warning standards in Korea, which apply uniform thresholds nationwide, fail to reflect the specific disaster damage patterns observed on Jeju Island (Lim et al., 2023). This gap in localized disaster response capabilities underscores the urgent need for a grid-based flood risk assessment method that integrates

advanced technologies, including artificial intelligence (AI), to address Jeju Island's regional characteristics.

This study aims to develop a Hybrid AI Flood Risk Assessment and Prediction System that reflects Jeju Island's distinct hydrogeomorphological and environmental factors. By combining traditional hydrological analysis techniques with AI-based approaches, the study seeks to address the complex nature of flood risks on Jeju Island. Key objectives include:
1) Establishment of a Grid-Based Database: Building a flood damage and risk influencing factor database tailored to Jeju Island, including spatially distributed grid information such as hydrogeographical characteristics, damaged assets, and grid-specific runoff data.
2) Identification of Influencing Factors: Employing the recursive feature elimination method to analyze and identify the primary factors influencing flood risk.
3) Development of Binary Classification Models: Creating a deep learning-based method to classify flood risk levels using hydrogeomorphological factors and disaster data. The model was trained on flood trace maps, with each grid cell labeled as “flooded” (YES) or “not flooded” (NO) based on field-surveyed flood records.
4) Integration of MLOps Technology: Applying MLOps (Deep Learning Operations) to ensure the efficient and scalable implementation of the AI-driven flood risk assessment and prediction system.
The necessity of this research is further underscored by Jeju's increasing disaster vulnerability due to aging demographics, expanding urbanization, and shifting land use patterns, including the reduction of natural flood buffers like fields and forests. Through the proposed research, this study seeks to achieve two main outcomes: (1) the development of a grid-based AI flood risk prediction model that incorporates the spatial and temporal complexities of Jeju's flood events, and (2) the establishment of rainfall forecast and warning standards that reflect Jeju Island's unique disaster characteristics. By leveraging AI and hydrological analysis technologies, this study aims to provide a comprehensive, practical, and scalable flood risk assessment framework tailored to the specific needs of Jeju Island, contributing to enhanced disaster preparedness and resilience in this vulnerable region.

Data

The data utilized in this study were structured in a grid format, adhering to the National Spatial Information Standard Grid Network with a resolution of 100 meters and employing the EPSG 5179 coordinate reference system. This structured grid-based approach ensures spatial consistency and supports the integration of multiple data sources. The primary dataset consists of flood trace maps derived from 23 flood events surveyed between 2006 and 2022 on Jeju Island. This relatively small number of labeled events reflects limitations in historical survey availability. The study incorporated 30 features expected to influence flood damage, tailored to reflect the unique hydrogeomorphological and environmental characteristics of Jeju Island.

Table 1. Selected flood damage influencing factors of Jeju Island.

ID	Definition	Data source	ID	Definition	Data source
F_1	Duration 3hr maximum rainfall	KMA	F_{16}	Total number of buildings	Public data portal
F_2	Duration 12hr maximum rainfall	KMA	F_{17}	Total population	Public data portal
F_3	Total 5-day antecedent rainfall	KMA	F_{18}	Public notice price	Public data portal
F_4	Maximum runoff	GRM result	F_{19}	Number of old buildings	Public data portal
F_5	Daily average groundwater level	Jeju provincial office	F_{20}	Dependent population	Public data portal
F_6	Maximum wind speed	KMA	F_{21}	Road density	Public data portal
F_7	Altitude	DEM	F_{22}	River density	Public data portal
F_8	Slope	DEM	F_{23}	Reservoir capacity	Jeju provincial office
F_9	Topographic wetness index	DEM	F_{24}	Distance to reservoir	Jeju provincial office
F_{10}	River Power Index	DEM	F_{25}	Distance to road	Public data portal

F_{11}	Curvature	DEM	F_{26}	Distance to river	Public data portal
F_{12}	Side curvature	DEM	F_{27}	Distance to natural disaster risk area	Jeju provincial office
F_{13}	Plane curvature	DEM	F_{28}	Distance to coastline	Public data portal
F_{14}	Land cover map	EGIS	F_{29}	Distance to spring water	Public data portal
F_{15}	Hydrological geological map	EGIS	F_{30}	Distance to volcanic cone	Public data portal

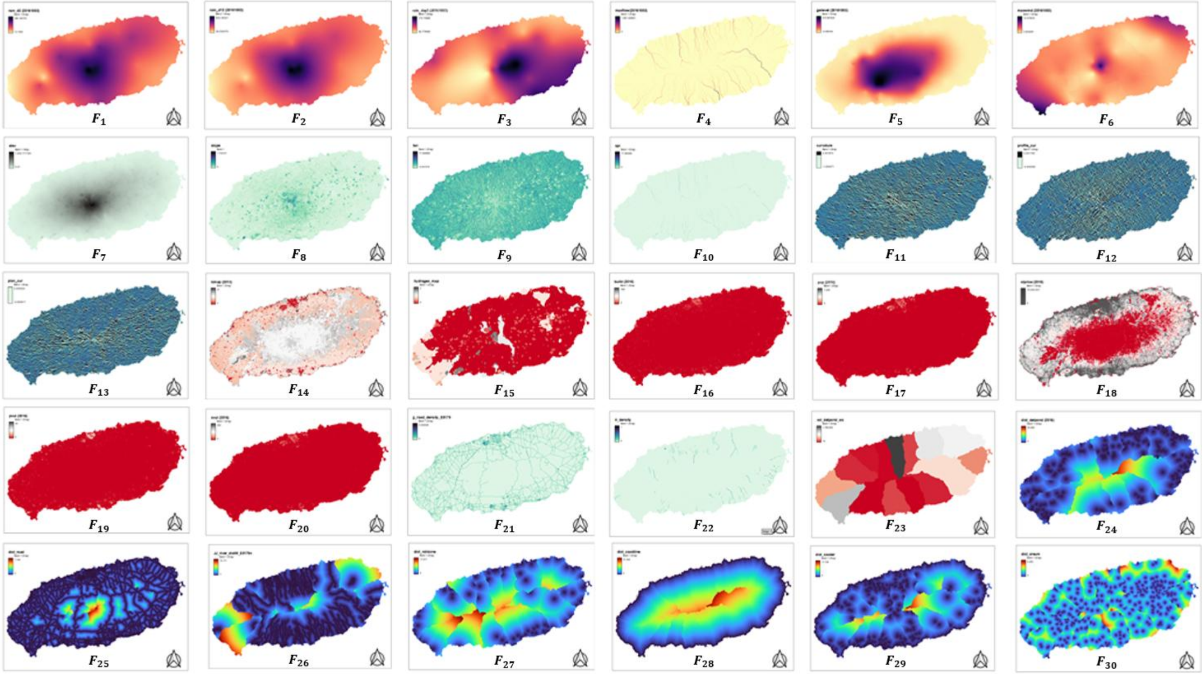


Figure 1. Flood risk influencing factors expected to affect flood damage in Jeju Island.

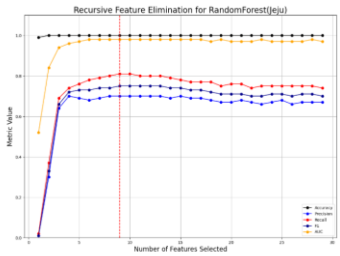
Deep learning modeling

The deep learning modeling framework for this study was designed to develop a robust flood risk prediction model that reflects the unique characteristics of Jeju Island (Gheisari et al., 2023). The process involved careful feature engineering (Verdonck et al., 2024), addressing class imbalance in the target data, and evaluating multiple tree-based deep learning models. Feature engineering began with the identification and elimination of highly correlated variables to reduce multicollinearity. Variables with a correlation coefficient greater than 0.9 were analyzed, and those with higher Variance Inflation Factors (VIF) were removed. After this process, 29 features were finalized for modeling. To address the imbalance in the target data, which is a common challenge in flood risk prediction, various thresholds between 0.1 and 0.5 were tested to balance sensitivity and specificity. Ultimately, a threshold of 0.2 was selected as the optimal value, ensuring improved representation of minority classes in the predictions. Following feature selection and preprocessing, five tree-based deep learning models were evaluated for their ability to handle nonlinear relationships and provide feature importance scores: Decision Tree (DT) (Tanveer et al., 2023), Random Forest (RF) (Verma, 2023), CatBoost (Khayatian et al., 2024), XGBoost (Abbasimehr et al., 2023), and LightGBM (Bian et al., 2023). These models were chosen for their demonstrated effectiveness in tabular datasets and their suitability for recursive feature selection. To further refine the model and enhance its performance, Recursive Feature Elimination (RFE) (Awad & Fraihat, 2023) was applied. The RFE process involved iteratively fitting the model with all available features, ranking them based on their importance, and removing the least important feature(s) in each iteration. This process was repeated until the desired number of features was reached, ensuring that only the most relevant variables contributing to flood risk prediction were retained.

Performance and flood risk impact factor assessment

The performance of the deep learning models was evaluated using metrics derived from the confusion matrix, including Accuracy, Precision, Recall, F1 Score, and AUC (Area Under the Curve). Among these, particular emphasis was placed on Recall and F1 Score. Recall was prioritized due to the critical importance of accurately predicting actual damaged areas in flood risk scenarios, where failure to identify high-risk zones can have severe consequences. The F1 Score was also considered as it provides a balance between Precision and Recall, offering a comprehensive measure of model performance. During the Recursive Feature Elimination (RFE) process, the optimal number of features and the corresponding deep learning model were determined based on the highest Recall and F1 Score values. Among the evaluated models, the Random Forest (RF) model demonstrated the best performance. It achieved optimal results using 9 features, delivering the highest F1 Score and Recall values across the grid-based dataset for Jeju Island. Consequently, the RF model was identified as the most reliable and effective for predicting flood risks.

❖ Sample : Stability of FS = Good



Importance order	DT	RF	CatBoost	XGBoost
1	maxwind	rain_d12	rain_day5	rain_d12
2	rain_d12	dist_detpond	maxwind	gwlevel
3	dist_detpond	rain_d3	gwlevel	maxwind
4	dist_ndrzone	maxwind	dist_detpond	dist_detpond
5	rain_d3	rain_day5	rain_d12	rain_day5

- Blue : 4 types of ML models included in the 1-5th most important features
- Yellow : 3 types of ML models included in the 1-5th most important features
- Green : 2 types of ML models included in the 1-5th most important features

Figure 2. Model performance and flood risk impact factors assessment

Conclusion

This study applied Recursive Feature Elimination (RFE) to analyze the factors influencing flood risk on Jeju Island, identifying key variables and optimizing the predictive capacity of the Random Forest (RF) model. The RF model was selected as the most effective model for flood risk assessment on Jeju Island, capable of achieving optimal performance with 9 key features when applied across the entire region. In terms of influencing factors, hazard-related variables and capacity-based measures, such as distances from critical points, were ranked highest. Traditional components of disaster vulnerability and risk assessment, including Exposure, Vulnerability, and Capacity, as defined in the Index-Based Approach (IBA) methodology, were ranked lower in importance. Regional meteorological characteristics further underscored the significance of location-specific factors. The proposed AI-based framework enables localized interpretation of flood risk and offers practical applications for disaster management, including early warning systems, zoning policies, and targeted infrastructure planning in Jeju Island.

Acknowledgement

This research was supported by the KICT Research Program (Project No. 20250221-001).

References

Abbasimehr H., Paki R., Bahrini A. (2023). A novel XGBoost-based featurization approach to forecast renewable energy consumption with deep learning models. Sustainable Computing: Informatics and Systems, 38, 100863

Ali K., Bajracharyar RM., Raut N. (2017). Advances and challenges in flash flood risk assessment: A review. J. Geogr. Nat. Disast, 7, 195

Awad M., Fraihat S. (2023). Recursive feature elimination with cross-validation with decision tree: Feature selection method for machine learning-based intrusion detection systems. Journal of Sensor and Actuator Networks, 12(5), 67

Bian L., Qin X., Zhang C., Guo P., Wu H. (2023). Application, interpretability and prediction of machine learning method combined with LSTM and LightGBM-a case study for runoff simulation in an arid area. Journal of Hydrology, 625, 130091

Gheisari M., Ebrahimzadeh F., Rahimi M., Moazzamigodarzi M., Liu Y., Dutta Pramanik PK., ... & Kosari S. (2023). Deep learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey. CAAI Transactions on Intelligence Technology, 8(3), 581-606

Kang M., Yoo C. (2020). Application of the SCS–CN Method to the Hancheon Basin on the Volcanic Jeju Island, Korea. Water, 12, 3350

Khayatian D., Maleki A., Nasiri H., Dorrigiv M. (2024). Histopathology image analysis for gastric cancer detection: a hybrid deep learning and catboost approach. Multimedia Tools and Applications, 1-27

Lim J., Kim Y., Park S., Ahn U. (2023). Testing heavy rainfall change in Jeju Island, Korea and its linkage to typhoon activity over East Asia during the Holocene. Palaeogeogr., Palaeoclimatol., Palaeoecol., 111715

Tanveer H., Adam MA., Khan MA., Ali MA., Shakoor A. (2023). Analyzing the Performance and Efficiency of Machine Learning Algorithms, such as Deep Learning, Decision Trees, or Support Vector Machines, on Various Datasets and Applications. The Asian Bulletin of Big Data Management, 3(2), 126-136

Verdonck T., Baesens B., Óskarsdóttir M., Broucke SV. (2024). Special issue on feature engineering editorial. Machine learning, 113(7), 3917-3928