C207

Flood Risk Mapping Using Machine Learning: Case studies from Japan

〇Mohamed SABER・Sameh KANTOUSH・Tetsuya SUMI・Sohei KOBAYASHI・Ahmed EMARA. Emad MABROUK

Introduction

Japan faces considerable flood risks due to its climatic conditions, including heavy rainfall events and typhoons, which are exacerbated by climate change. In response, Japan has evolved its flood control planning and policies, shifting from traditional river channel-focused control to a more comprehensive, watershed-oriented management approach. This strategy, known as "River Basin Disaster Resilience and Sustainability by All," emphasizes integrated measures such as flood prevention, exposure reduction, and appropriate evacuation, involving all stakeholders within the river basin. Despite these efforts, challenges remain in managing flood risks in Japan. The increasing frequency and intensity of heavy rainfall events necessitate continuous adaptation and improvement of flood risk management strategies. This includes enhancing infrastructure resilience, improving early warning systems, and promoting community-based disaster preparedness to effectively mitigate the impacts of future flood events. Therefore, the objective of this study is to use machine learning approaches such as Random Forest (Saber et al. 2021, Saber et al 23) for predicting FFS in several basins in Japan, example of the Kyukitakami River (Fig. 1).

Approach and methods

In this paper, we used Random Forrest Model. Random Forest is a versatile and widely-used ensemble learning algorithm in machine learning, adept at handling both classification and regression tasks. It operates by constructing a multitude of decision trees during training and aggregating their outputs to enhance predictive accuracy and control overfitting. The algorithm introduces randomness by selecting random subsets of data and features for each tree, ensuring diversity among the trees and reducing variance in the model's predictions. This approach not only improves performance but also provides insights into feature importance, aiding in understanding the influence of different variables on the outcome. However, the complexity of Random Forests can lead to increased computational resources and longer training times, especially with large datasets. Despite these challenges, Random Forest remains a powerful tool due to its robustness, flexibility, and ability to handle missing data effectively.

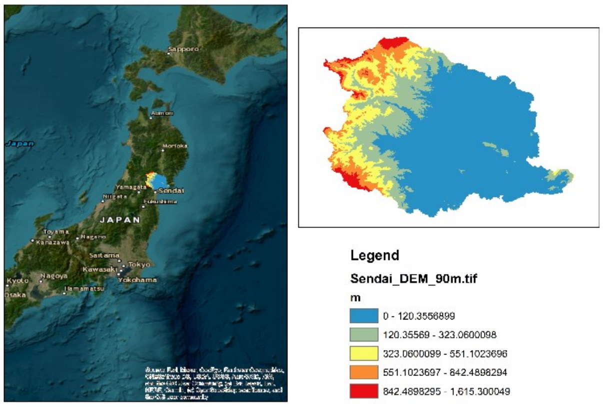

Fig. 1 (a) The location map of the Kyukitakami River basin.

Two main datasets were prepared for the flood susceptibility mapping using machine learning. First, the flood inventory map of the flood occurrences was prepared based on the post flood field survey. Additionally, the non-flooded points were randomly selected throughout the catchment using the geographic

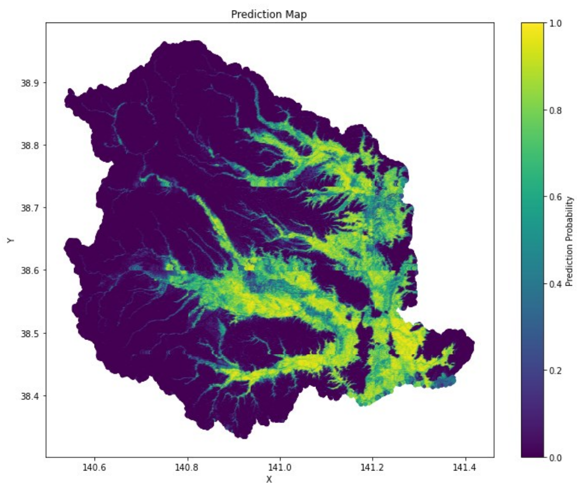

Fig. 2 Flood susceptibility map by the Random Forest.

information system (GIS) environment. Furthermore, a total of 10 FSS Factors were also prepared for modelling such as Elevation, Slope, Aspect, Plan Curvature, Hillshade, Horizontal flow distance, SPI, Geology, Rainfall, Land use/Land cover. In a later stage, the dataset was divided into two groups of training (70 %) and testing (30 %) through a random selection scheme. Spatial maps for each factor were produced using ArcGIS considering the consistency of spatial resolution. Afterwards, two methods of information gain ratio and multicollinearity test (VIF) were used to check the importance of the factors in FFS within the study area. Secondly, the implementation of the proposed machine learning approaches, the datasets have been divided into two main categories, training (70%) and validation (30%). The results of the model was assessed for the accuracy using different measures including the most famous one AUC.

Results and ongoing research

The flood susceptibility map presents a spatial prediction of flood-prone areas, with probabilities ranging from 0 (lowest susceptibility) to 1 (highest susceptibility). Areas with values close to 0.8–1.0 indicate regions with a very high likelihood of flooding, primarily located along river networks and low-lying floodplains. Zones with probabilities between 0.4–0.7 represent moderate susceptibility, often situated near tributaries or gently sloping areas. Regions with probabilities below 0.4 are considered low-risk, likely corresponding to elevated terrain or areas with effective drainage. This numerical representation aids in prioritizing high-probability zones for flood risk management, infrastructure planning, and disaster response, ensuring resources are focused where the potential for flooding is greatest. In the study area (Fig. 2), many of the downstream are showing high to medium flood susceptibility.

Conclusion

ML models are successfully used in predicting of flooding susceptibility in Japan. The outcomes of this study can be used as guidance for the planners and managers to mitigate the floods in the high prone flood susceptible regions. Our future plan is to develop flood risk map for the whole Japan

Acknowledgment

This research was funded by several projects, including DPRI International Collaborative Research (Project No. 2023LS-04) and Grants-in-Aid for Scientific Research, KAKENHI Project (C) Project ID (2 3 K 0 4 3 2 8).

References

1. Saber, M., Boulmaiz, T., Guermoui, M., Abdrado, K. I., Kantoush, S. A., Sumi, T., ... & Mabrouk, E. (2021). Examining LightGBM and CatBoost models for wadi flash flood susceptibility prediction. Geocarto International, 1-26.
2. Saber, M., Boulmaiz, T., Guermoui, M., Abdrabo, K. I., Kantoush, S. A., Sumi, T., ... & Mabrouk, E. (2023). Enhancing flood risk assessment through integration of ensemble learning approaches and physical-based hydrological modeling. Geomatics, Natural Hazards and Risk, 14(1), 2203798.