https://doi.org/10.29227/IM-2025-02-76 *Received: 15 Jun 2025, Accepted: 20 Aug 2025, Published: 20 Oct 2025*

Advanced Machine Learning Models for Flood Susceptibility Mapping: A Case Study in Thai Nguyen Province (Old), Vietnam

CHIEU Dinh Vu[1], HANG Thi Ha[1], DUNG Ngoc Luong[1], QUYNH Duy Bui[1],
CHINH Dieu Thi Luu[2], HANH Hong Tran[3], TRAN Xuan Truong[3,*]

[1] Hanoi University of Civil Engineering, Hanoi, Vietnam
[2] Hanoi University of Civil Engineering, Hanoi, Vietnam
[3] Hanoi University of Mining and Geology, Hanoi, Vietnam
[*] *Corresponding author: tranxuantruong@humg.edu.vn*

Abstract: *Floods are considered one of the most dangerous natural disasters in the world. Due to its wide-ranging impact, floods cause significant damage to people, property, livelihoods, habitats, infrastructure, and economic development. To date, flood susceptibility maps are still considered an effective tool in flood damage management and prevention. Thus, this study proposed the use of three advanced machine learning models, including CatBoost, LightGBM, and NGBoost, to generate flood susceptibility maps based on the historical flood locations and 11 influencing factors. The ROC-AUC results were utilized to compare and evaluate the forecasting accuracy of these ML models. The LightGBM model demonstrated superior forecast performance and was selected to build the flood susceptibility map. This map indicates that the low-lying districts and concentrated river systems in Thai Nguyen province, such as Thai Nguyen City, Song Cong, Phu Binh, Dong Hy, and Pho Yen districts, fall in high and very high flood susceptibility zones. Mountainous districts situated on the edge of Thai Nguyen province, such as Vo Nhai, Phu Luong, and Dinh Hoa, fall in medium and very low flood susceptibility zones. The obtained map provides a visual view of future flood-prone areas, assisting local authorities in land-use planning and implementing effective priority investment strategies.*

Keywords: *flood susceptibility, Gradient Boosting model, Machine Learning, flood risk management, Thai Nguyen province, Vietnam*

1. Introduction

Flooding is a natural phenomenon in which river water overflows its banks and inundates adjacent lands, resulting in extensive damage to many aspects, such as economic development, human lives, transportation, ecological ecosystem, infrastructure, and cultural heritage sites (Shafizadeh-Moghadam et al., 2018). In recent decades, the frequency of floods has increased due to environmental degradation, population growth, deforestation, and urbanization (Bradshaw et al., 2007). It is estimated that the frequency of global floods has increased by more than 40% in the past two decades. Furthermore, it is also estimated that about 1.3 billion people will be living in flood-prone areas by 2050 (Gudiyangada Nachappa et al., 2020). In general, climate change, along with socio-economic development and urbanization, will directly impact and increase the formation of floods in the future (Chen et al., 2021). Therefore, identifying future flood-prone areas on flood susceptibility maps is an extremely important task for flood risk prevention and mitigation strategies (Uddin & Matin, 2021). These flood susceptibility maps can provide spatial information about flood-prone areas, helping to optimize resource allocation, refine emergency response plans, and enhance evacuation strategies (Mangaiyarkkarasi, 2024).

Today, thanks to advances in information technology, Machine Learning (ML) models are considered the most advanced and effective approach to disaster prediction and prevention. They focus on analyzing the ability of big data to extract disaster patterns from multi-source data and the innovative approaches to improve prediction accuracy (Mangaiyarkkarasi, 2024). Boosting is an ensemble learning method in which multiple weak learners are combined to form a strong learner to reduce training errors. Boosting algorithms differ in how they generate and aggregate weak learning models during sequential training (Ferreira & Figueiredo, 2012). In particular, the Gradient Boosting algorithm is a combination of gradient descent and boosting. This algorithm also trains models sequentially and learns on the residual errors of the previous model instead of adjusting the data weights, thereby improving prediction performance and reducing overfitting (Zhang & Haghani, 2015). Popular variations of gradient boosting algorithms include Extreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), Light Gradient Boosting Machine (LightGBM), and Natural Gradient Boosting (NGBoost) (Sahin, 2022). The biggest advantage of Gradient Boosting algorithms is their strong predictive performance. They can be deployed efficiently on datasets that require learning complex patterns and interactions (Zhang & Haghani, 2015). According to this, some of them have been widely employed in flood susceptibility assessment in recent years (Aydin & Iban, 2023; Lin et al., 2023; Xu et al., 2023). If Aydin and Iban (2023) compared the predictive performance of several variants of the Gradient Boosting model with the Random Forest (RF) algorithm in flood susceptibility mapping for Adana province on the Mediterranean coast of Türkiye using fourteen flood conditioning parameters representing the topography, meteorology, vegetation, lithology and human life in the area, Xu et al. (2023) and Lin et al. (2023) proposed a hybrid modeling approach for rapid urban flood forecasting based on a combination of hydrological-hydraulic model and light-enhanced machine model. Most of these studies have confirmed the superior predictive performance of the LightGBM model in flood forecasting.

As climate change and extreme weather events increase, the frequency and intensity of natural disasters will continue to increase worldwide. Traditional methods of forecasting and preventing natural disasters are facing many challenges in terms of accuracy and speed of response (Qin et al., 2024). Additionally, since floods often occur after a long period of rain, heavy rain, or snowmelt, combined with adverse terrain conditions (Gudiyangada Nachappa et al., 2020), experiencing and developing forecast models in flood susceptibility mapping needs to be conducted continuously and regularly to improve and enhance forecast accuracy and identify suitable models (Kumar et al., 2023).

Thai Nguyen province (old) is one of the centers of the Viet Bac region, located in the regional planning of the capital Hanoi. Thai Nguyen has an important geographical location and is one of the important industrial, educational, and medical centers of Vietnam (Ha et al., 2024). In recent years, this province has frequently faced flooding due to heavy rains. In particular, the flood caused by Typhoon Yagi in September 2024 was considered the largest historical flood in the past 60 years, causing significant losses to human lives and property (ADINet, 2025). Meanwhile, there are no studies assessing flood susceptibility for this province using ML models, but only at the district level using a series of remote sensing images (Mai Sy et al., 2023).

Based on the reasons mentioned above, this study proposed the use of advanced machine learning models, specifically Categorical Boosting (CatBoost), Light Gradient Boosting Machine (LightGBM), and Natural Gradient Boosting (NGBoost), to create flood susceptibility maps for the study area. The ROC curve and AUC value were employed to compare and estimate the predictive performance of these ML models to select the most accurate model. The best flood susceptibility map was built based on the most accurate model. This flood susceptibility map can be an effective tool to support flood prevention, risk reduction, and local decision-making.

2. Study area

Thai Nguyen province (old) is situated in the midland and mountainous region of northern Vietnam. It serves as a gateway for socio-economic exchanges between the midland and mountainous areas and the Northern Delta. This province borders Bac Kan province to the North, Vinh Phuc and Tuyen Quang provinces to the West, Lang Son and Bac Giang provinces to the East, and Hanoi city to the South. Thai Nguyen province encompasses 9 administrative units (Thai Nguyen City, Song Cong, Pho Yen, Phu Binh, Dong Hy, Vo Nhai, Dinh Hoa, Dai Tu, and Phu Luong). In total, there are 178 communes, wards, and towns within the province. The province's natural area is 3,531.71 km^2, accounting for 1.07% of Vietnam's natural area, and its population in 2019 was 1,286,751 people. Thai Nguyen's climate is divided into two distinct rainy and dry seasons. The rainy season usually starts in May and lasts until early October, while the dry season starts from October to May of the following year. The average annual rainfall in Thai Nguyen is between 2000 and 2500 mm, with the highest rainfall occurring in August and the lowest in January. This province has a typical terrain of hills interspersed with low fields, mainly limestone mountains, and bowl-shaped hills. In addition, Thai Nguyen has two main rivers flowing through its territory, the Cau River and the Cong River, which significantly influence its hydrological regime (available at https://thainguyen.gov.vn/).

In recent years, Thai Nguyen province has experienced continuous flooding, while a trillion-VND flood prevention project on the Cau River has remained inactive. In early September 2024, Thai Nguyen City and surrounding districts faced a historic flood, the worst in over 20 years, which lasted for several days as the Cau River water level rose above alert level 3. Thousands of households and properties had to be evacuated urgently, resulting in estimated damages exceeding 970 billion VND (available at https://thainguyen.gov.vn/). According to a quick report from Thai Nguyen province, the most recent heavy rain on August 20-21, 2025, caused 373 houses to be flooded, 172.68 hectares of rice and crops to be flooded, 19 landslides at Km 17, National Highway 3, Route 257 to Cho Don Commune, and many inter-village and inter-hamlet roads (ADINet, 2025).

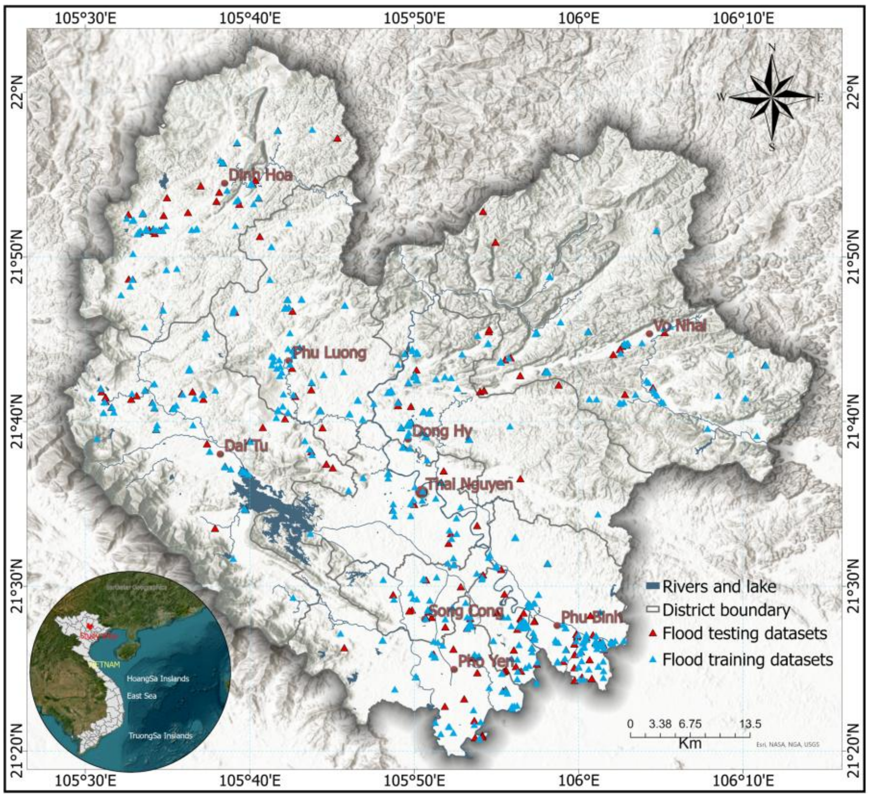

Fig. 1. Thai Nguyen province (old) in Vietnam.

3. Methodology and data used

3.1 Methodology

The research flowchart in the study comprised the following steps: (1) Data preparation and collection; (2) Mapping flood inventory locations; (3) Developing advanced ML models using the sci-kit learn library in the Python programming language to build flood susceptibility maps; (4) Assessing the predictive accuracy of used models; (5) Selecting the best flood susceptibility map (Fig. 2).

3.2 Data used

3.2.1 Flood inventory data

A comprehensive flood inventory is essential in flood susceptibility modeling studies. It enables the understanding of vulnerable areas, historical trends, and the mitigation of flood impacts (Haltas et al., 2021). In this study, flood inventory positions were provided by the Thai Nguyen Provincial Steering Committee for Natural Disaster Prevention and Search and Rescue from 2016 to 2024. The flood inventory data included 833 locations (Fig. 1). These flood inventory data were randomly divided into two datasets: the 70% training dataset and the 30% validation dataset. According to this, 583 flood positions, representing 70% of the training dataset, were used to build the flood susceptibility maps, and 250 flood positions, comprising 30% of the validation dataset, were utilized to verify the predictive models.

3.2.2 Flood causative factors

Flood formation is influenced by a group of factors related to elevation, slope, hydrometeorology, drainage network, distance to river, land use, therefore, it is necessary to collect characteristic factors for flood susceptibility assessment (Shafizadeh-Moghadam et al., 2018). However, collecting all these elements for flood susceptibility mapping is not a simple task as it depends on their accessibility and availability. Based on the availability and ability to collect data, this study uses 11 flood causative factors, including altitude, slope, aspect, curvature, drainage density, terrain roughness, TWI, rainfall, distance to road, distance to river, and land use (Fig. 3).

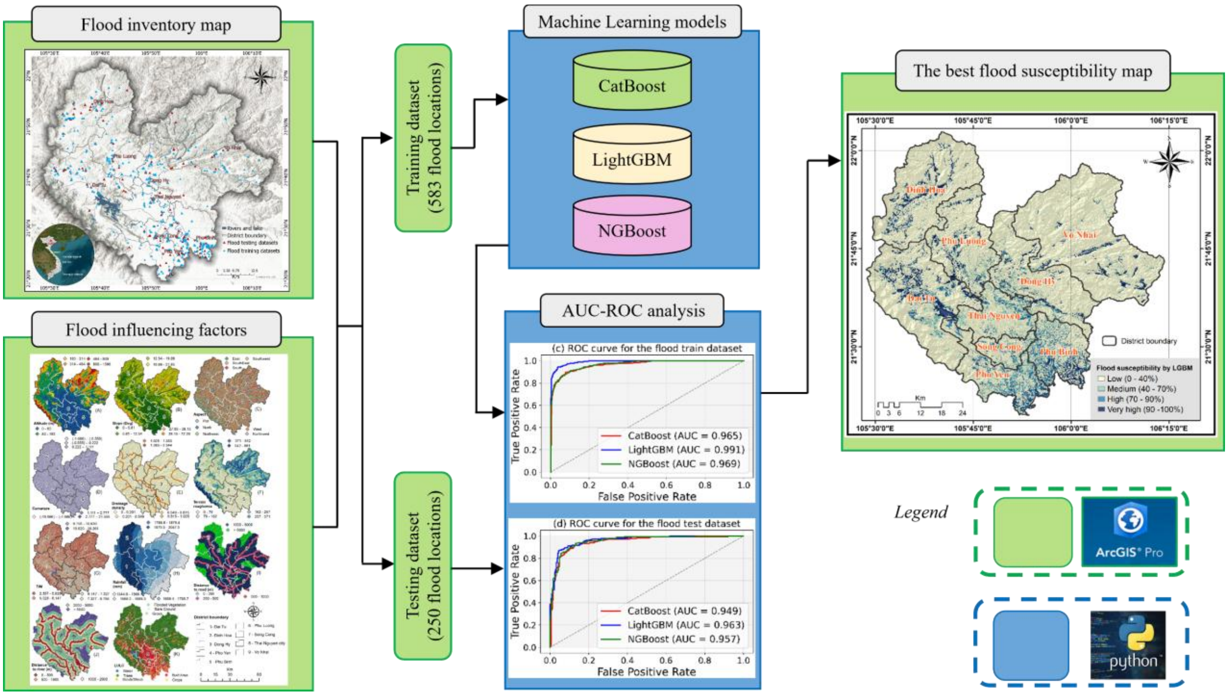

Fig 2. Methodology flowchart in this research.

The altitude, slope, aspect, curvature, TWI, and terrain roughness factors were calculated from a 30-meter ALOS DEM data. This ALOS DEM data was downloaded and acquired in June 2021 from https://www.eorc.jaxa.jp/ALOS/en/aw3d30/data/index.html. The remaining factors, including drainage density, distance to road, distance to river, and land use, were gathered from Vietnam's Ministry of Natural Resources and Environment and the Department of Agriculture and Rural Development of Thai Nguyen Province in 2017. The average annual rainfall was obtained from 11 hydrometeorological stations from 2011 to 2024 and 56 stations of Vrain and WeatherPlus from 2021 to 2024 in Thai Nguyen province. These flood influencing factors were normalized into raster layers at a 30-meter resolution to generate the thematic maps in the GIS environment for flood susceptibility modelling.

Tab. 1. Flood influencing factors and their categorization.

Factors	Categorization	Categorization technique
Altitude (m)	(1) [0 – 83], (2) [83 – 183], (3) [183 – 314], (4) [314 – 484], (5) [484 – 809], (6) [809 – 1590]	Natural Breaks
Slope angle (degree)	(1) [0- 5.81], (2) [5.81 – 12.54], (3) [12.54 – 19.88], (4) [19.88 – 27.85], (5) [27.85 – 38.15], (6) [38.15 – 72.39]	Natural Breaks
Aspect	(1) Plat, (2) North, (3) Northeast, (4) West, (5) Northwest (6) East, (7) Southeast, (8) South, (9) Southwest	Azimuth
Curvature	(1) [(-19.666) - (-1.666)], (2) [(-1.666) - (-0.555)], (3) [(-0.555) – 0.222], (4) [0.222 – 0.111], (5) [1.111 – 2.777], (6) [2.777 – 21.555]	Natural Breaks
Drainage density (km/km2)	(1) [0 – 0.201], (2) [0.201 – 0.549], (3) [0.549 – 0.815], (4) [0.815 – 1.025], (5) [1.025 – 1.383], (6) [1383 – 2.344]	Natural Breaks
Terrain roughness (m/km2)	(1) [0 - 79], (2) [79 - 162], (3) [162 - 257], (4) [257 - 371], (5) [371 - 542], (6) [542 - 981]	Natural Breaks
TWI	(1) [2.357 – 5.028], (2) [5.028 – 6.147], (3) [6.147 – 7.327], (4) [7.327 – 8.756], (5) [8.756 – 10.620], (6) [10.620 – 18.261]	Natural Breaks
Rainfall (mm)	(1) [1344.9 – 1568.1], (2) [1568.2 – 1689.3], (3) [1689.4 – 1785.7], (4) [1785.8 – 1879.4], (5) [1879.5 – 2047.5]	Geostatistical Kriging method
Distance to road (m)	(1) [0 - 200], (2) [200 - 500], (3) [500 - 1000], (4) [1000 - 5000], (5) [> 5000]	Natural Breaks
Distance to rivers (m)	(1) [0 - 200], (2) [200 - 500], (3) [500 - 2000], (4) [2000 - 5000], (5) [> 5000]	Natural Breaks
Land use	(1) Water, (2) Trees, (3) Scrub/Shrub, (4) Built areas, (5) Crops, (6) Flooded vegetation, (7) Bare ground, (8) Grass	Landcover categories

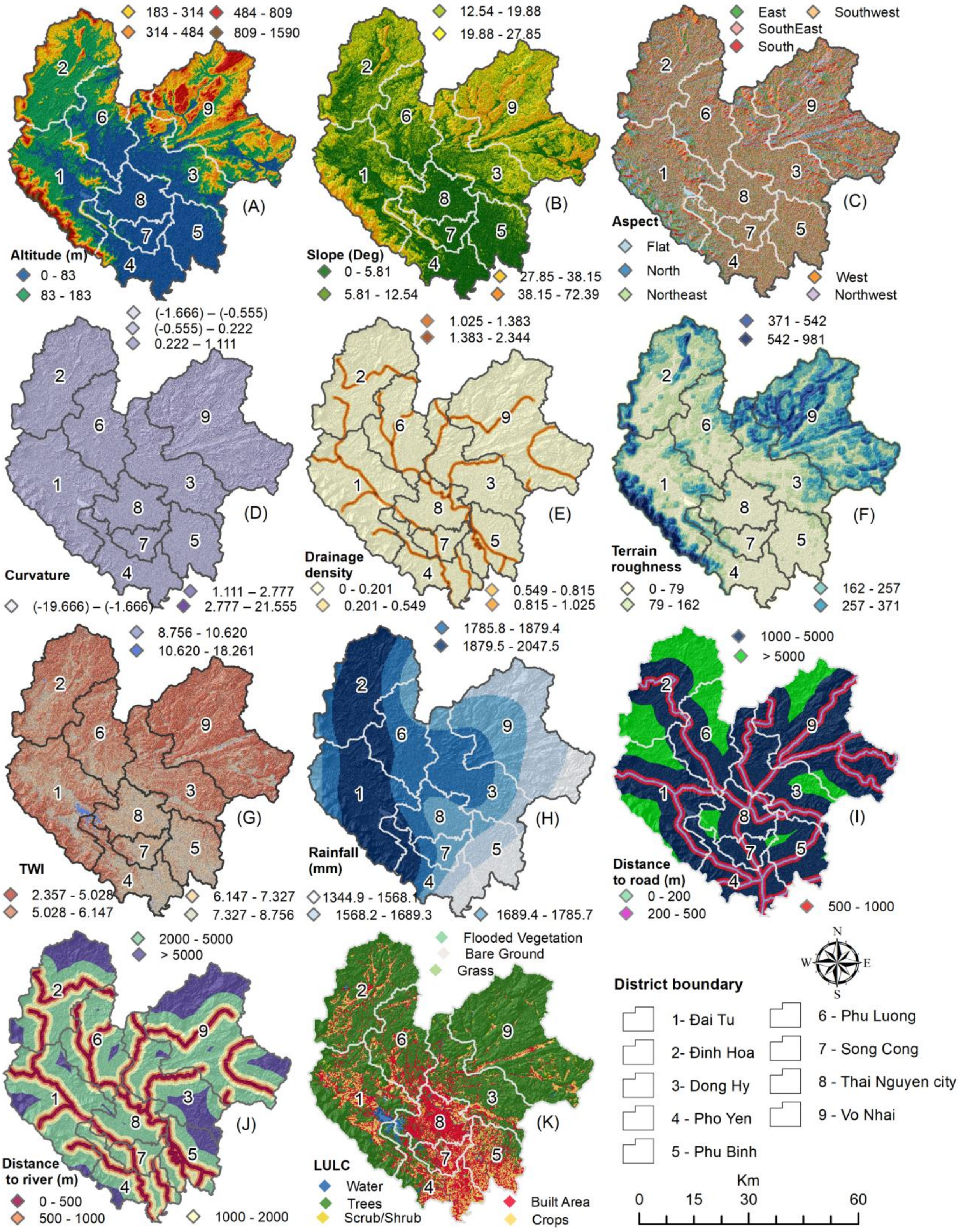

Fig 3. Flood influencing factors used in this study: (A) altitude, (B) slope, (C) aspect, (D) curvature, (E) drainage density, (F) terrain roughness, (G) TWI, (H) rainfall, (I) distance to road, (J) distance to river, and (K) land use

4. Machine learning models used

This study employed three advanced machine learning models, namely Categorical Boosting (CatBoost), Light Gradient Boosting Machine (LightGBM), and Natural Gradient Boosting (NGBoost), to build flood susceptibility maps for the study area.

4.1 Categorical Boosting (CatBoost)

Categorical Boosting (CatBoost) is developed based on the framework of the gradient boosting method, where decision trees are built sequentially to minimize errors and improve predictions (Sahin, 2022). The process works by building a decision tree and assessing how much error there is in predictions. After the first tree is built, the next tree will be created to correct the errors made by the previous one. This process continues iteratively with each new tree, focusing on improving the model's predictions by reducing previous errors. This process continues until a predefined number of iterations is met (van Hoof & Vanschoren, 2021). The obtained result is an ensemble of decision trees that work together to provide accurate predictions (Hancock & Khoshgoftaar, 2020). It can address various data formats, both categorical and numerical features. One of the crucial advantages of this algorithm is that it reduces overfitting by using the ordered boosting technique (Sahin, 2022).

4.2 Light Gradient Boosting Machine (LightGBM)

Light Gradient Boosting Machine (LightGBM) refers to an open-source high-performance framework developed by Microsoft. It can be used for ranking, classification, and regression issues (Sahin, 2022). This algorithm is an ensemble learning framework that utilizes the gradient boosting technique, which creates a strong learner by sequentially adding weak learners in a gradient descent manner (Alzamzami et al., 2020). Gradient-based One-Side Sampling (GOSS), histogram-based algorithms, and leaf-wise tree growth are unique strategies used for training and splitting the data in this algorithm (Sahin, 2022). Accordingly, various user-defined parameters can be tuned in LightGBM, like boosting type, max depth, learning rate, and number of leaves. In this way, this algorithm can speed up the training procedure and reduce computational complexity (Bentéjac et al., 2021).

4.3 Natural Gradient Boosting (NGBoost)

Natural Gradient Boosting (NGBoost) is an algorithm designed for generic probabilistic prediction using the gradient boosting technique (Duan et al., 2020). According to this, this algorithm uses the natural gradient technique to compensate for the probabilistic prediction limitations of current gradient boosting techniques (Kavzoglu & Teke, 2022). This algorithm can handle various tasks, like regression, classification, and survival prediction. The probabilistic forecasting characteristic of this algorithm enables obtaining probabilities about the whole possible results of any given search space, in place of determining one specific output as the forecast (Chen et al., 2022). The modular and scalable structure in the NGBoost can be expanded and comprises three main parameters, namely the scoring rule, the base learner, and the parametric probability distribution. Overall, the NGBoost algorithm outperforms existing probabilistic prediction methods in terms of flexibility, scalability, and usability (Kavzoglu & Teke, 2022).

4.4 Model verification and comparison

Validate the predictive performance of the used models to ensure their scientific significance and practical applicability (Luu et al., 2021). The area under the receiver operating characteristic (ROC) curve, known as AUC, is a single scalar value that measures the overall performance of a binary classifier (Pepe et al., 2006). The ROC curve is a two-dimensional measure of classification performance. AUC (Area Under ROC Curve) is the area under the ROC curve (Marzban, 2004). The AUC is a robust overall measure to evaluate the performance of score classifiers because its calculation relies on the complete ROC curve (Hand, 2009). The AUC value ranges from 0 to 1. The closer the AUC value is to 1, the better the overall performance of the predictive model (Ha et al., 2025). In this study, the ROC curve and AUC value are utilized to assess the overall performance of used ML models.

5. Result and discussion

5.1 Model verification and comparison

In this study, three advanced ML models, including CatBoost, LightGBM, and NGBoost, were developed using the sci-kit learn library in the Python programming language to generate flood susceptibility maps. The ROC curve and AUC value were employed to assess the predictive accuracy of these ML models.

The results of ROC curve analysis on the training dataset indicated that the LightGBM model achieved the highest accuracy with an AUC of 0.991, followed by the NGBoost model with an AUC of 0.969, and the CatBoost model had the lowest AUC at 0.965. The results of ROC curve analysis on the testing dataset continued to show that the LightGBM model has the highest accuracy with an AUC of 0.963, followed by the NGBoost model with an AUC of 0.957, and the CatBoost model achieved the lowest AUC at 0.949 (Fig 4). These findings denote that the LightGBM model offers the best forecasting performance among the models used, and it was chosen to build the flood susceptibility map.

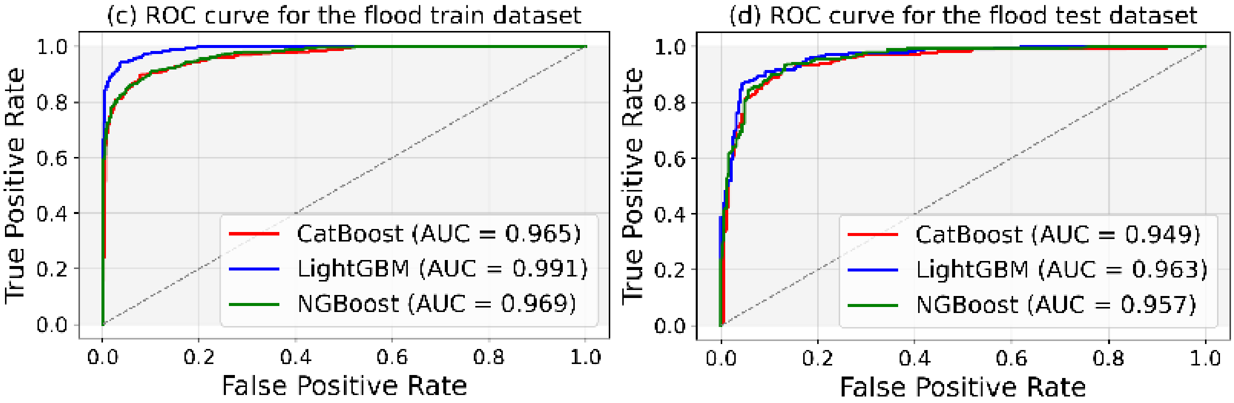

Fig 4. The ROC curve analysis on training dataset (left) and testing dataset (right) in flood susceptibility mapping.

5.2 Construction of the best flood susceptibility map

Based on the analysis of the ROC curve on both the training and testing datasets, the LightGBM model was selected to build the best flood susceptibility map for Thai Nguyen province (old), Vietnam. According to this, the most accurate flood susceptibility map for the Thai Nguyen province was created using the LightGBM model (Fig. 5). This map was classified into four classes, including low, medium, high, and very high, using the percentile technique in ArcGIS Pro software.

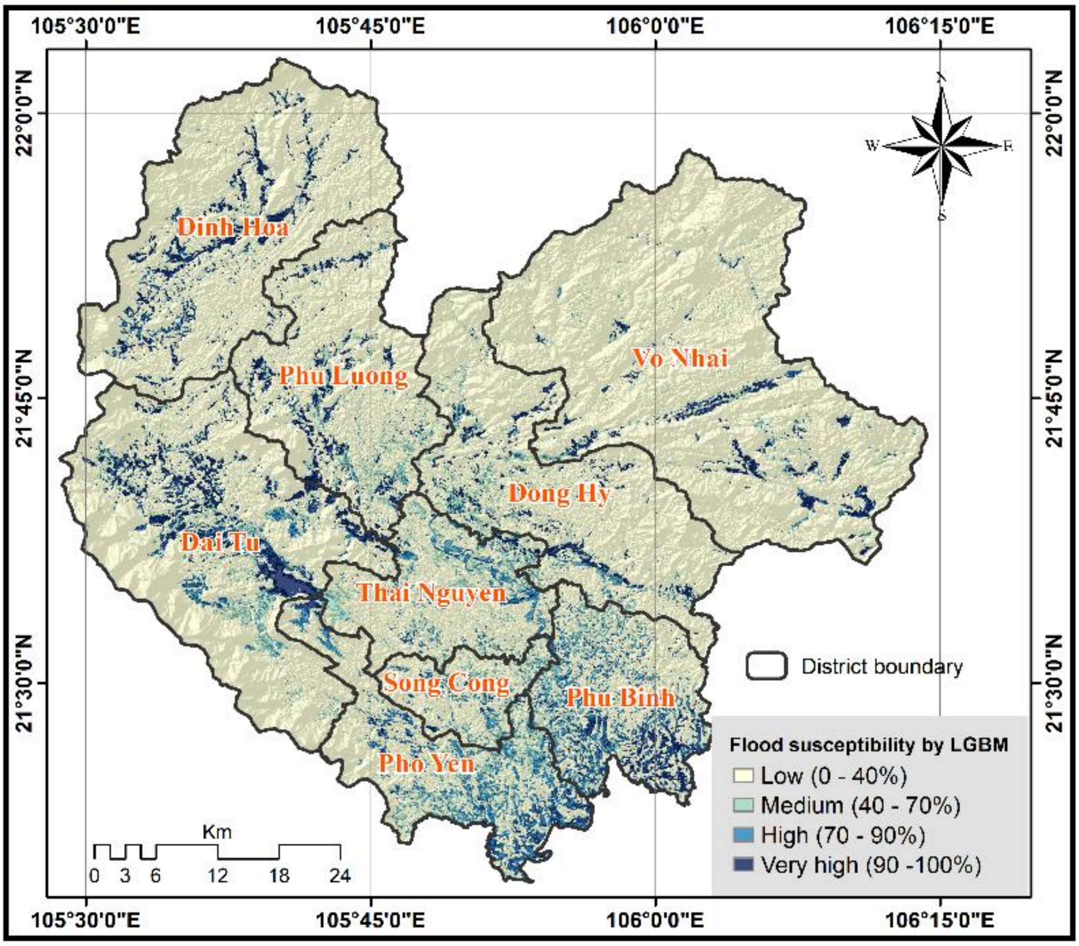

Fig 5. The best flood susceptibility map using the LightGBM model.

The resulting map indicated that high flood susceptibility zones are mainly concentrated in the districts of Phu Binh, Pho Yen, and Song Cong, as well as parts of southern Dong Hy and Thai Nguyen city. The very high flood susceptibility zones are primarily distributed along the main river and stream systems, including the Cau River, Cong River, and their tributaries, particularly in the districts of Pho Yen, Song Cong, Phu Binh, Dong Hy, and Thai Nguyen city. Generally, these areas have a high density of rivers and streams and are located at lower elevations compared to other districts in Thai Nguyen province. The flooding hazard has posed significant challenges for local authorities and communities, especially in these districts that have a high population density, rapid urbanization, and developed economic infrastructure. Meanwhile, most areas in the Dinh Hoa, Phu Luong, and Vo Nhai districts, as well as the high mountainous regions in the North and Northeast of Thai Nguyen province, fall in the flood susceptibility from medium to very low. These districts are located in mountainous terrain along the arcs of the Ngan Son and Gam Rivers in the northern part of Thai Nguyen province. Additionally, the steepness of the terrain contributes to a lower flood hazard in these areas (Ha et al., 2024).

6. Conclusion

In this study, three advanced ML models, namely CatBoost, LightGBM, and NGBoost, were employed by using the sci-kit learn library in the Python programming language to create flood susceptibility maps for Thai Nguyen province. The analysis of the ROC curve on both the training and testing datasets indicated that the LightGBM model had the highest predictive accuracy, so this model was chosen to create the flood susceptibility map. The flood susceptibility map demonstrates that the low-lying districts and concentrated river systems in Thai Nguyen province, such as Thai Nguyen City, Song Cong, Phu Binh, Dong Hy, and Pho Yen districts, are in high and very high flood susceptibility zones. Mountainous districts situated on the edge of Thai Nguyen province, such as Vo Nhai, Phu Luong, and Dinh Hoa, are in medium and very low flood susceptibility zones. The results of this study not only provide important spatial information to identify areas at risk of future flooding, but also serve as an effective disaster risk management tool, allowing for the planning of reasonable urban development plans, flood prevention, and response plans.

Acknowledgment

This research is funded by the Ministry of Education and Training under grant number B2024-MDA-06.

References

ADINet. (2025). Viet Nam, Flooding and Landslides in Son La, Thai Nguyen, Bac Ninh, and Dong Nai Retrieved 9 from https://adinet.ahacentre.org/report/viet-nam-flooding-and-landslides-in-son-la-thai-nguyen-bac-ninh-and-dong-nai-20250820

Alzamzami, F., Hoda, M., & Saddik, A. E. (2020). Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation. IEEE Access, 8, 101840-101858. https://doi.org/10.1109/ACCESS.2020.2997330

Aydin, H. E., & Iban, M. C. (2023). Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations. Natural Hazards, 116(3), 2957-2991. https://doi.org/10.1007/s11069-022-05793-y

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54(3), 1937-1967. https://doi.org/10.1007/S10462-020-09896-5

Bradshaw, C. J., Sodhi, N. S., PEH, K. S. H., & Brook, B. W. (2007). Global evidence that deforestation amplifies flood risk and severity in the developing world. Global change biology, 13(11), 2379-2395. https://doi.org/10.1111/j.1365-2486.2007.01446.x

Chen, S.-Z., Feng, D.-C., Wang, W.-J., & Taciroglu, E. (2022). Probabilistic machine-learning methods for performance prediction of structure and infrastructures through natural gradient boosting. Journal of Structural Engineering, 148(8), 04022096. https://doi.org/10.1061/(ASCE)T.1943-541X.0003401

Chen, X., Zhang, H., Chen, W., & Huang, G. (2021). Urbanization and climate change impacts on future flood risk in the Pearl River Delta under shared socioeconomic pathways. Science of The Total Environment, 762, 143144. https://doi.org/10.1016/j.scitotenv.2020.143144

Duan, T., Anand, A., Ding, D. Y., Thai, K. K., Basu, S., Ng, A., & Schuler, A. (2020). NGBoost: Natural Gradient Boosting for Probabilistic Prediction Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research. https://proceedings.mlr.press/v119/duan20a.html

Ferreira, A. J., & Figueiredo, M. A. (2012). Boosting algorithms: A review of methods, theory, and applications. Ensemble machine learning, 35-85. https://doi.org/10.1007/978-1-4419-9326-7_2

Gudiyangada Nachappa, T., Tavakkoli Piralilou, S., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O., & Blaschke, T. (2020). Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. Journal of Hydrology, 590, 125275. https://doi.org/10.1016/j.jhydrol.2020.125275

Ha, H., Bui, Q. D., Nguyen, V.-P., Tran, X. T., Nguyen, D. Q., & Luu, C. (2025). Assessing housing exposure to landslide hazards using hybrid machine learning and spatial modeling. Modeling Earth Systems and Environment, 11(4), 251. https://doi.org/10.1007/s40808-025-02428-6

Ha, H., Bui, Q. D., Vu, D. C., Hoang, T., Pham, D.-H., Nguyen, V.-P., Do, M. H., Tran, X. T., & Luu, C. (2024). Landslide spatial prediction utilizing fuzzy unordered rules induction ensemble models: a case study in Thai Nguyen, Vietnam. Modeling Earth Systems and Environment, 10(2), 2201-2219. https://doi.org/10.1007/s40808-023-01888-y

Haltas, I., Yildirim, E., Oztas, F., & Demir, I. (2021). A comprehensive flood event specification and inventory: 1930–2020 Turkey case study. International Journal of Disaster Risk Reduction, 56, 102086. https://doi.org/10.1016/j.ijdrr.2021.102086

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. Journal of big data, 7(1), 94. https://doi.org/10.1186/s40537-020-00369-8

Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. Machine learning, 77(1), 103-123. https://doi.org/10.1007/s10994-009-5119-5

Kavzoglu, T., & Teke, A. (2022). Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). Arabian Journal for Science and Engineering, 47(6), 7367-7385. https://doi.org/10.1007/s13369-022-06560-8

Kumar, V., Azamathulla, H. M., Sharma, K. V., Mehta, D. J., & Maharaj, K. T. (2023). The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management. Sustainability, 15(13), 10543. https://doi.org/10.3390/su151310543

Lin, K., Sheng, S., Chen, H., Zhou, Y., Luo, Y., Xiong, L., Guo, S., & Xu, C.-Y. (2023). Exploring an intelligent adaptation method of hydrological model parameters for flood simulations based on the light gradient-boosting machine. Journal of Hydrology, 626, 130340. https://doi.org/10.1016/j.jhydrol.2023.130340

Luu, C., Bui, Q. D., Costache, R., Nguyen, L. T., Nguyen, T. T., Van Phong, T., Van Le, H., & Pham, B. T. (2021). Flood-prone area mapping using machine learning techniques: a case study of Quang Binh province, Vietnam. Natural Hazards, 108(3), 3229-3251. https://doi.org/10.1007/s11069-021-04821-7

Mai Sy, H., Luu, C., Bui, Q. D., Ha, H., & Nguyen, D. Q. (2023). Urban flood risk assessment using Sentinel-1 on the google earth engine: A case study in Thai Nguyen city, Vietnam. Remote Sensing Applications: Society and Environment, 31, 100987. https://doi.org/10.1016/j.rsase.2023.100987

Mangaiyarkkarasi, J. (2024). Machine Learning Algorithms for Natural Disaster Prediction and Management. In Utilizing AI and Machine Learning for Natural Disaster Management (pp. 24-38). IGI Global. https://doi.org/10.4018/979-8-3693-3362-4.ch002

Marzban, C. (2004). The ROC curve and the area under it as performance measures. Weather and Forecasting, 19(6), 1106-1114. https://doi.org/10.1175/825.1

Pepe, M. S., Cai, T., & Longton, G. (2006). Combining predictors for classification using the area under the receiver operating characteristic curve. Biometrics, 62(1), 221-229. https://doi.org/10.1111/j.1541-0420.2005.00420.x

Qin, F., Huang, C., & Lin, Z. (2024). Big data and artificial intelligence-driven natural disaster prediction and prevention: Technological advances and prospects. Geographical Research Bulletin, 3, 381-398. https://doi.org/10.50908/grb.3.0_381

Sahin, E. K. (2022). Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. Geocarto International, 37(9), 2441-2465. https://doi.org/10.3390/toxins15100608

Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K., & Shirzadi, A. (2018). Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. Journal of Environmental Management, 217, 1-11. https://doi.org/10.1016/j.jenvman.2018.03.089

Uddin, K., & Matin, M. A. (2021). Potential flood hazard zonation and flood shelter suitability mapping for disaster risk mitigation in Bangladesh using geospatial technology. Progress in Disaster Science, 11, 100185. https://doi.org/10.1016/j.pdisas.2021.100185

van Hoof, J., & Vanschoren, J. (2021). Hyperboost: Hyperparameter optimization by gradient boosting surrogate models. arXiv preprint arXiv:2101.02289. https://doi.org/ 10.48550/arXiv.2101.02289

Xu, K., Han, Z., Xu, H., & Bin, L. (2023). Rapid prediction model for urban floods based on a light gradient boosting machine approach and hydrological–hydraulic model. International Journal of Disaster Risk Science, 14(1), 79-97. https://doi.org/10.1007/s13753-023-00465-2

Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies, 58, 308-324. https://doi.org/10.1016/j.trc.2015.02.019