Flood Susceptibility Mapping of Chikkamagaluru District of India using Ensemble Technique

Sting Salvador Gonsalis[1], Shweta Vincent[2], Om Prakash Kumar[3], Vijay Mohan[4]
[1,2,4] Department of Mechatronics, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India
[3] Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India

Abstract.

Floods are an extensive and destructive natural occurrence that has an influence on agriculture, the environment, and the economy in addition to destroying property and human lives. Therefore, in order to minimize and limit damage, it is essential to identify sensitive zones using flood susceptibility mapping (FSM).This article shows the employment of ensemble machine learning model named as Light Gradient Boosting Machine (LightGBM) for categorization of flood-prone regions of the Chikkamagaluru district in India. Eight elements were utilized to develop the flood susceptibility model: slope, aspect, elevation, land use-land cover (LULC), normalized difference vegetation index (NDVI), normalized difference water index (NDWI), topographical wetness index (TWI), and drainage density (DD). In addition, flood inventory map (FIM) is utilized to determine the areas sensitive of flooding. The application of this ensemble machine learning model in mapping and estimating flood risk is investigated, and suggestions for additional research are made. The Area under the curve-Receiver operating characteristics (AUC-ROC) approach was used to evaluate the model's output, and LightGBM) was found to function well with an AUC score of 0.9666.

Keywords: Flood Suscpetibility Mapping (FSM), Flood Conditioning Factors (FCF), Flood Inventory Map (FIM), Ensemble Machine Learning, Light Gradient Boosting Machine (LightGBM), Area under the curve - Receiver operating characteristics (AUC-ROC)

1. INTRODUCTION

Effective disaster risk management depends mostly on flood susceptibility mapping, especially in areas like Chikkamagaluru, India, vulnerable to floods because of complicated topography and hydrological processes. Although total flood management is still unattainable, flood susceptibility maps (FSMs) provide priceless information for handling pre- and post-flood situations in face of both natural and human-induced obstacles. Coupled with the always rising global population, more frequent and severe downpour events brought on by climate change further increase flood hazards, maybe affecting millions of people.

In order to meet this urgent demand, this research investigates the creation of FSMs using a multi-criteria decision support system designed specifically for the South Indian area of Chikkamagaluru. By using ensemble machine learning, the research seeks to improve the precision of flood forecasts and identify regions that are vulnerable to floods. A strong flood risk map will be produced by using information gain theory to track down important conditioning factors causuating flooding.

Topographical, land use and hydrological indices are among the geoenvironmental elements that are analyzed as part of the study process. Historical flood data is included to train and verify the FSM model. Using the AUC-ROC approach and confusion matrix parameters, the algorithms' efficacy in producing FSMs will be thoroughly assessed.

1.1. THE PROBLEM OF FLOODS AND THE OBJECTIVE

Coastal position, urbanization, and changing rainfall patterns are complex interplaying elements that traditional flood susceptibility mapping in Chikkamagaluru District fails to represent due to inadequate data. Inaccurate flood maps are a result of these restrictions, which makes it harder to effectively plan for floods and manage risk.The recent floods in Chikkamagaluru have brought attention to the critical need for better flood forecast. In order to direct focused mitigation activities, current maps are not precise enough.Because floods are dynamic events affected by the strength and length of rainfall, traditional approaches have a hard time accounting for this fact and integrating varied geographical data (soil characteristics, land cover, drainage)[5]. Furthermore, they are ill-equipped to deal with the shifting ecosystems that are a direct result of global warming.

Our study aims to use ensemble machine learning to create a Flood Vlulnerability Map for Chikkamagaluru. The created map's efficacy would be judged by its performance of assessment generated by ensemble machine learning techniques. Various Flood Conditioning Factors (FCFs) have been explored for the production of the final FSMs.

1.2. Organization of this article

An overview of the literature on the different methods for creating FSMs and their associated accuracy is provided in the next section of this article. Our article's basic methodology is described in the third section, and the results and discussions are presented in the fourth. The article's fifth and final part concludes the article and outlines the future scope for study.

2. Literature Review

Flooding is an utmost natural hazard causing severe loss of life and economic damage. Techniques for predicting floods are essential for reducing these effects. This study looks at the research on mapping flood risk using Geographic Information System (GIS) and multi-criteria analysis. Because GIS makes it possible to create and analyze spatial data on flood-conditioning factors, it is essential to flood susceptibility mapping. Flood susceptibility is evaluated by integrating these aspects using multi-criteria analysis methodologies. Choosing the right flood-conditioning elements is essential. Topography (slope, elevation)[3], hydrology (drainage density, distance to/from river)[10], soil features, and land use and/or cover are examples of common factors[18]. Research places a strong emphasis on choosing parameters based on the data's availability and the research location.

Flood susceptibility mapping is done using a variety of approaches. These models evaluate how individual flood-conditioning parameters and flood events are related. Frequency Ratio (FR) and Information Value (IV) models, which are used by [14], [3], constitute examples. By using methods such as the Analytical Hierarchy Process (AHP), it is possible to give weights to various elements and incorporate expert knowledge [2]. [9] and [16] studies illustrate this strategy. The capacity of techniques such as Multi Criteria Decision Analysis (MCDA) [6], Logistic Regression (LR) [7], and Artificial Neural Networks[19] to handle intricate interactions between components has led to their growing application in flood susceptibility mapping. The use of remote sensing data, namely from Synthetic Aperture Radar (SAR), may improve mapping of flood vulnerability by offering insights into the magnitude of flooding and the presence of water [1]. For increased transparency and interpretability, research is looking at new strategies including Artificial Neural Network - Synthetic Minority Oversampling Technique (ANN-SMOTE)[19] and Alternating decision tree(ADT) classifiers [4].

[2], [9] highlight the usage of AHP in the construction of FSM. [14] illustrates how FR and other statistical models are used. Table 1 illustrates how flood risk is mapped using machine learning and other techniques in [4], [6], [7], [15], and [19].Analytic network process (ANP) and spatial analysis are discussed in [5], [12]. State-of-the-art FSM production uses a variety of soft-computing approaches, which are summarized in Table 1. Additionally, the table displays the evaluation parameters that were used to create and validate the FSMs.

The study makes it evident that a number of writers have investigated the use of machine learning techniques to the development of FSMs. However, our paper describes the development of the FSM for Chikkamagaluru, an area in India that is prone to flooding.

3. Methodology

This part explains the methods used for the building of the FSM of Chikkamagaluru. As shown in Figure 1, the Flood Inventory Map and Flood Conditioning Factors are created through these steps.

To produce the FIM and FCFs of Chikkamagaluru district, data is gathered from a variety of field study sources, including Google Earth, ESA, and USGS Earth Explorer. This study makes use of Landsat 8 data with a spatial resolution of 30 meters. The following FCFs were made using the ArcGIS program and the Chikkamagaluru Digital Elevation Model (DEM): "Drainage density" (DD) is the total length of a watershed's channels, including both natural and constructed stormwater drainage infrastructure, ‘Slope’ is the ground surface's rise or fall, "Aspect" is the degree of orientation of the slope, with 0 representing north, 90 east, 180 south, and 270 west and it ranges from 0 to 360 clockwise., ‘Elevation’ is defined as the distance above sea level, ‘Land use-land cover’ (LULC) in which the physical surface of the planet is known as land cover, while "land use" describes how land is used or utilized for human endeavors, "Normalized difference vegetation index" (NDVI) measures the amount and condition of ground-surface vegetation, ‘Normalized difference water index’ (NDWI) is an indicator for identifying and tracking changes in surface water content, and ‘Topographical wetness index’ (TWI) is a model that uses the land's topography to determine how moist the soil is in a given location.

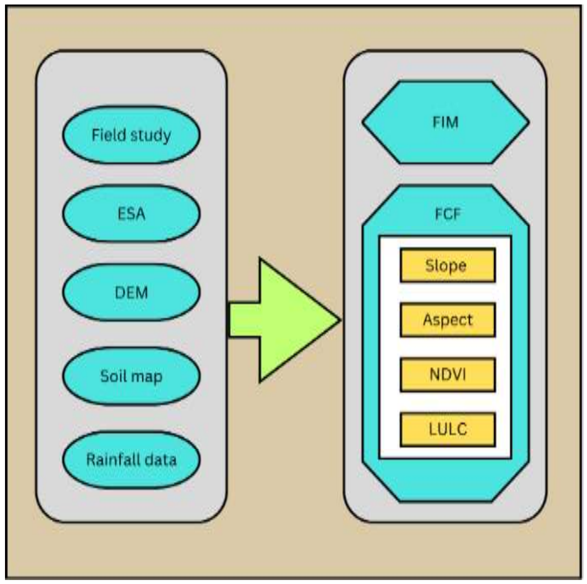

Figure 1: FIM and FCF creation

In order to acquire the data points, each of these thematic maps is compared to the others. The historical flood information is in the FIM. This is superimposed on top of the FCF maps, resulting in a thorough dataset for testing and training.

The use of these two maps, the FIM and FCF, in the final FSM generation is shown Figure 2.

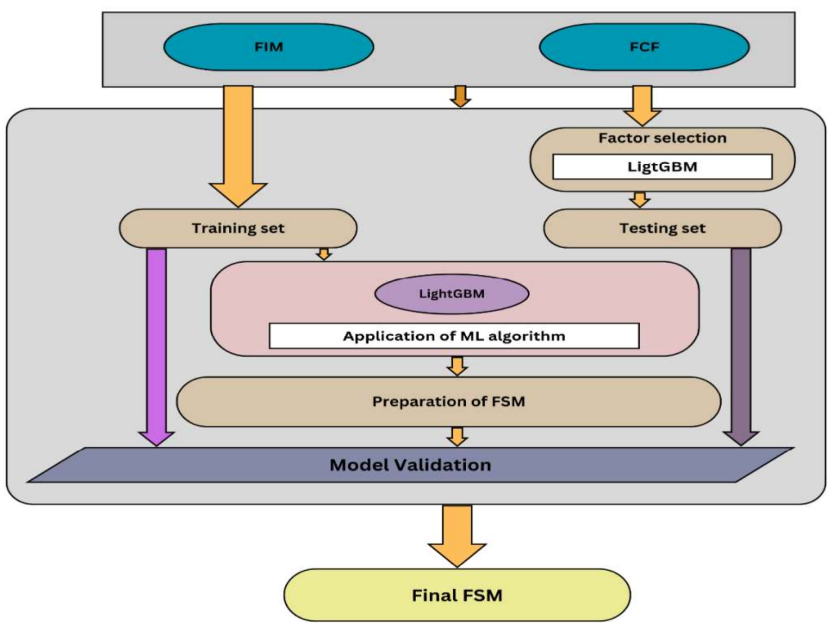

Fig. 2. Creation of FSM

The ensemble machine learning algorithm named LightGBM [20] receives the FIM and FCF as inputs, and outputs the FSM. Metaheuristics (which were not used in our research) may or may not be utilized to adjust the machine learning algorithm for weights. Following model creation, validation is performed and final AUC-ROC scores are produced.

Ref No.	Algorithms / techniques used	Evaluation Parameters	Parameter values	Region	No. of FCFs used
[6]	Multi Criteria Decision Analysis (MCDA)	AUC-ROC	0.8724	Dodoma region, central Tanzania	7
[7]	Logistic Regression (LR	AUC-ROC	0.901	Mersin, Turkey	12
[13]	Shannon's Entropy (SE)	AUC-ROC	0.87	Bagmati basin, Bihar	12
[17]	Random Forest (RF)	AUC-ROC	0.956	The Bâsca Chiojdului river basin, Romania	10
[8]	Ensemble Artificial neural network (ANN)	AUC-ROC	0.936	Eastern Mediterranean basin, Türkiye	12
[1]	Random Forest (RF) CART (Classification and Regression Trees) SVM(Support vector Machines) XGBoost(Extreme Gradient Boosting)	AUC-ROC	0.807 0.780 0.756 0.727	The Metlili watershed, Morocco	12
[4]	Alternating decision tree (ADT) Functional tree (FT) Kernel logistic regression (KLR) Multilayer perceptron (MLP) Quadratic discriminant analysis (QDA)	AUC	0.972 0.959 0.956 0.95 0.95	Tafresh Watershed, Iran	8
[12]	Frequency ratio (FR) Weight of evidence (WofE) Random forest (RF) Multi-layer perceptron (MLP)	AUC	0.921 0.920 0.936 0.956	Gembong–Pekalen basin, Indonesia	11
[2]	K-Star (KS) algorithm K-Nearest Neighbor(KNN) KNN–AHP KS–AHP	AUC	0.887 0.852 0.896 0.878	Prahova river basin, Romania	10
[3]	Frequency ratio (FR) Weights of evidence (WoE) Certainty factor (CF) Shanon's entropy (SE) Information value (IV)	AUC	0.928 0.925 0.909 0.923 0.913	North-east depression, Bangladesh	11
[19]	Artificial Neural Network - Synthetic Minority Oversampling Technique (ANN-SMOTE) AHP Shannon Entropy	AUC-ROC	0.902 0.650 0.472	Ontario, Canada	6
[15]	AHP	AUC	0.812	Kathmandu metropolitan city	6

[16]	Galactic swarm optimization (GSO) Adagrad (Adaptive Gradient) Adaptive Moment Estimation (Adam)	AUC	0.887 0.900 0.882	Ha Giang, Vietnam	11
[18]	support vector regression (SVR) SVR-bat algorithm (SVR-BA) SVR-invasive weed optimization (SVR-IWO) SVR-firefly algorithm (SVR-FA)	AUROC	0.77 0.79 0.80 0.81	Ahwaz , Iran	11

Table 1. Literature Review

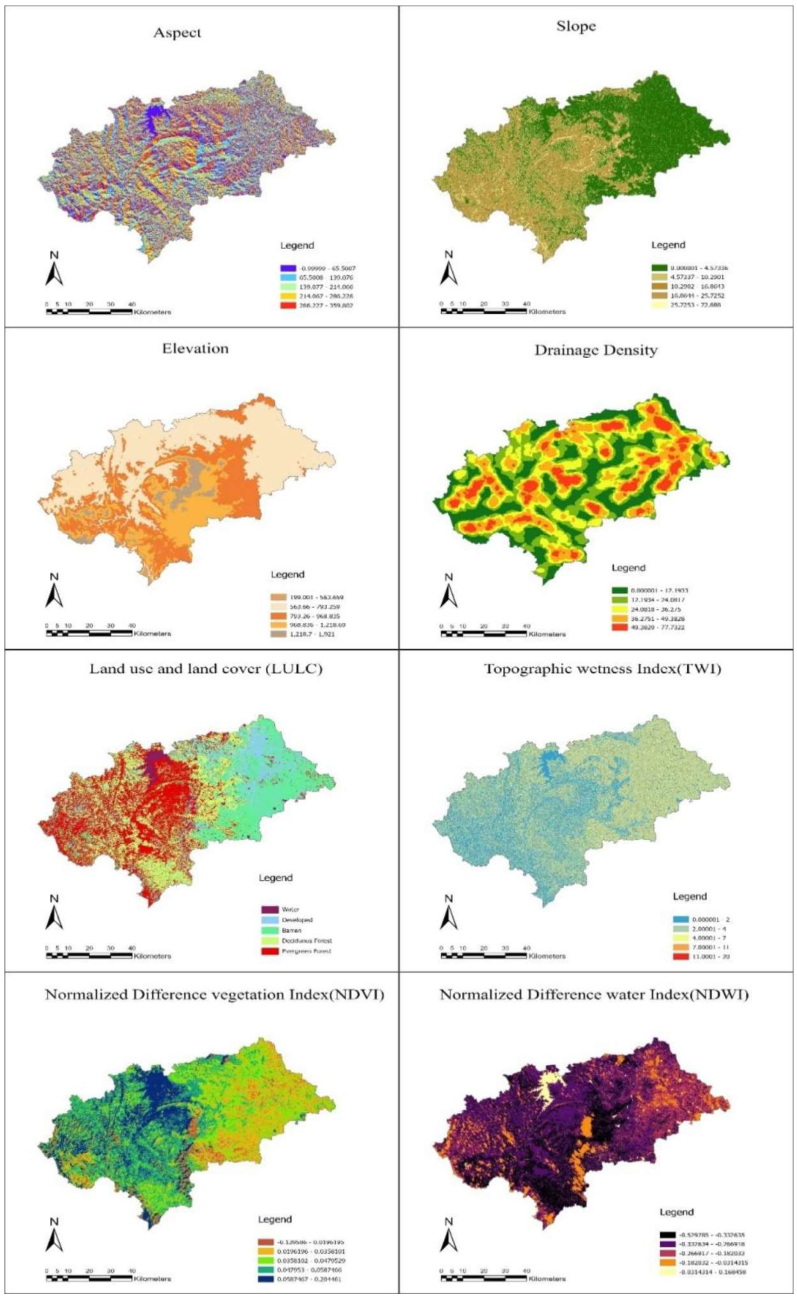

Fig. 3. FCFs generated for Chikkamagaluru

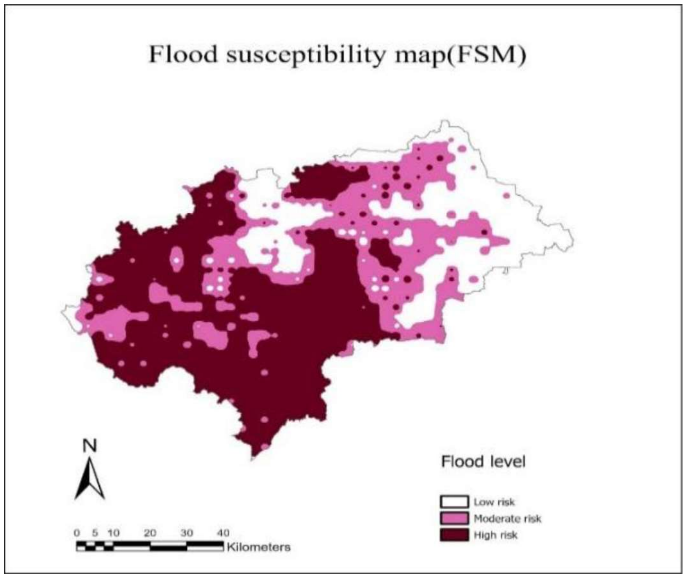

Fig. 4. Final FSM generated for Chikkamagaluru using LightGBM

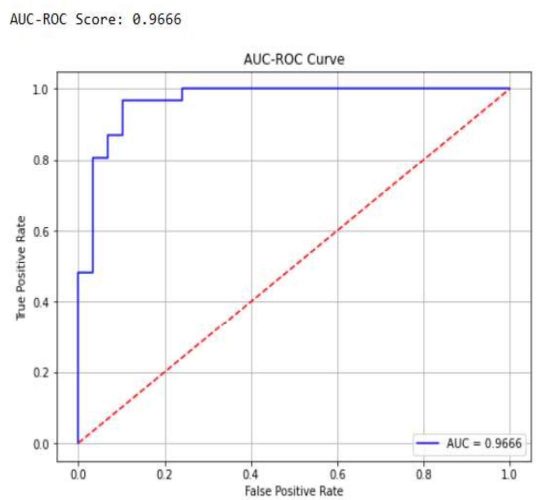

Fig. 5. AUC-ROC analysis of LightGBM model

4. RESULTS AND DISCUSSION

Figure 3 displays the several FCFs that have been produced using the previously discussed methodology. Figure 4 displays the final FSM produced by the LightGBM technique. Flood risk levels are divided into three categories based on the impact of floods: low, moderate, and high. It uses an algorithm based on histograms to find the best leaf split. Data must be categorized into histograms (bins) in order to identify the splitting point. This approach works with bins instead of individual data points, which drastically reduces the algorithm's temporal complexity. Figure 5 displays the AUC-ROC curves for the ensemble machine learning technique. The results of the AUC-ROC curve demonstrate how effectively the LightGBM algorithm works with the provided spatial data. Therefore, the LightGBM-based statistical assessment was shown to be accurate and successful in flood forecast in the current research.

In Chikkamagaluru, flooding is seen as a common threat that has severely damaged infrastructure, agricultural grounds, and human lives. However, because of a lack of data and other obstacles, a thorough evaluation of this danger has not yet been carried out in the research region. Therefore, LightGBM was used to map and evaluate this risk.The final FSM demonstrates that the areas most vulnerable to floods are those around the main waterbodies, or drainage networks. In order to adopt appropriate management strategies, planners and other government organizations may find the resulting FSM helpful in ascertaining flood-prone areas. The results will help with sustainable land management initiatives.

5. Conclusion and Future Scope

Flooding is an extensive disaster causing damage and loss of life. The ability of government officials to anticipate floods based on a variety of indicators is crucial for reducing the danger to property and human lives. One method for doing this is to use pre-conditioning factors, or FCFs, to generate FSMs for an area. This article described how an FSM for the Indian district of Chikkamagaluru was created using the ensemble machine learning approach. The FSM with the best AUC score of 0.9666 was produced using the LightGBM algorithm, it was found. A very high precision of the FSM would result in maximum safety with increased LightGBM performance, taken into consideration in detail since the maps' spatial resolution is 30m.

Metaheuristic methods will be used in the future to increase the accuracy of the maps by fine-tuning the categorization weights of the algorithms.

References

[1]. Hitouri, S., Mohajane, M., Lahsaini, M., Ali, S. A., Setargie, T. A., Tripathi, G., D'Antonio, P., Singh, S. K., & Varasano, A. (2024b). Flood susceptibility mapping using SAR data and machine learning algorithms in a small watershed in northwestern Morocco. *Remote Sensing*, *16*(5), 858. https://doi.org/10.3390/rs16050858

[2]. Costache, R., Pham, Q. B., Sharifi, E., Linh, N. T. T., Abba, S., Vojtek, M., Vojteková, J., Nhi, P. T. T., & Khoi, D. N. (2019). Flash-Flood susceptibility assessment using Multi-Criteria decision making and machine learning supported by remote sensing and GIS techniques. *Remote Sensing*, *12*(1), 106. https://doi.org/10.3390/rs12010106

[3]. Chowdhury, M. S. (2024). Flash Flood Susceptibility Mapping of North-east Depression of Bangladesh using Different GIS based Bivariate Statistical Models. *Watershed Ecology and the Environment*, *6*, 26–40. https://doi.org/10.1016/j.wsee.2023.12.002

[4]. Janizadeh, S., Avand, M., Jaafari, A., Van Phong, T., Bayat, M., Ahmadisharaf, E., Prakash, I., Pham, B. T., & Lee, S. (2019). Prediction Success of machine learning methods for flash flood susceptibility mapping in the Tafresh Watershed, Iran. *Sustainability*, *11*(19), 5426. https://doi.org/10.3390/su11195426

[5]. Dano, U., Balogun, A., Matori, A., Yusouf, K. W., Abubakar, I., Mohamed, M. S., Aina, Y., & Pradhan, B. (2019). Flood susceptibility mapping using GIS-Based Analytic Network Process: A case study of Perlis, Malaysia. *Water*, *11*(3), 615. https://doi.org/10.3390/w11030615

[6]. Msabi, M. M., & Makonyo, M. (2020). Flood susceptibility mapping using GIS and multi-criteria decision analysis: A case of Dodoma region, central Tanzania. *Remote Sensing Applications Society and Environment*, *21*, 100445. https://doi.org/10.1016/j.rsase.2020.100445

[7]. Özay, B., & Orhan, O. (2023). Flood susceptibility mapping by best–worst and logistic regression methods in Mersin, Turkey. *Environmental Science and Pollution Research*, *30*(15), 45151–45170. https://doi.org/10.1007/s11356-023-25423-9

[8]. Özdemir, H., Koçyiğit, M. B., & Akay, D. (2023). Flood susceptibility mapping with ensemble machine learning: a case of Eastern Mediterranean basin, Türkiye. *Stochastic Environmental Research and Risk Assessment*, *37*(11), 4273–4290. https://doi.org/10.1007/s00477-023-02507-z

[9]. Shah, R. K., & Shah, R. K. (2023). GIS-based flood susceptibility analysis using multi-parametric approach of analytical hierarchy process in Majuli Island, Assam, India. *Sustainable Water Resources Management*, *9*(5). https://doi.org/10.1007/s40899-023-00924-0

[10]. Gharakhanlou, N. M., & Perez, L. (2022). Spatial Prediction of current and future flood susceptibility: Examining the implications of changing climates on flood susceptibility using machine learning models. *Entropy*, *24*(11), 1630. https://doi.org/10.3390/e24111630

[11]. Express News Service, & Express News Service. (2019, August 21). *Long route to recovery for flood-hit Chikkamagluru and its coffee estates*. The New Indian Express.

https://www.newindianexpress.com/galleries/nation/2019/Aug/21/long-route-to-recovery-for-flood-hit-chikkamagluru-and-its-coffee-estates-102485.html

[12]. Hidayah, E., Indarto, N., Lee, W., Halik, G., & Pradhan, B. (2022). Assessing coastal flood susceptibility in East Java, Indonesia: comparison of statistical bivariate and machine learning techniques. *Water*, *14*(23), 3869. https://doi.org/10.3390/w14233869

[13]. Islam, S., Tahir, M., & Parveen, S. (2021). GIS-based flood susceptibility mapping of the lower Bagmati basin in Bihar, using Shannon's entropy model. *Modeling Earth Systems and Environment*, *8*(3), 3005–3019. https://doi.org/10.1007/s40808-021-01283-5

[14]. Penki, R., Basina, S. S., & Tanniru, S. R. (2022). Application of geographical information system-based analytical hierarchy process modeling for flood susceptibility mapping of Krishna District in Andhra Pradesh. *Environmental Science and Pollution Research*, *30*(44), 99062–99075. https://doi.org/10.1007/s11356-022-22924-x

[15]. Chaulagain, D., Rimal, P. R., Ngando, S. N., Nsafon, B. E. K., Suh, D., & Huh, J. (2023). Flood susceptibility mapping of Kathmandu metropolitan city using GIS-based multi-criteria decision analysis. *Ecological Indicators*, *154*, 110653. https://doi.org/10.1016/j.ecolind.2023.110653

[16]. Chou, T., Hoang, T., Fang, Y., Nguyen, Q., Lai, T. A., Pham, V., Vu, V., & Bui, Q. (2020). Swarm-based optimizer for convolutional neural network: An application for flood susceptibility mapping. *Transactions in GIS*, *25*(2), 1009–1026. https://doi.org/10.1111/tgis.12715

[17]. Abedi, R., Costache, R., Shafizadeh-Moghadam, H., & Pham, Q. B. (2021). Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto International*, *37*(19), 5479–5496. https://doi.org/10.1080/10106049.2021.1920636

[18]. Mehravar, S., Razavi-Termeh, S. V., Moghimi, A., Ranjgar, B., Foroughnia, F., & Amani, M. (2023). Flood susceptibility mapping using multi-temporal SAR imagery and novel integration of nature-inspired algorithms into support vector regression. *Journal of Hydrology*, *617*, 129100. https://doi.org/10.1016/j.jhydrol.2023.129100

[19]. Khalid, R., & Khan, U. T. (2024). Flood susceptibility mapping using ANNs: a case study in model generalization and accuracy from Ontario, Canada. *Geocarto International*, *39*(1). https://doi.org/10.1080/10106049.2024.2316653

[20]. Zhang, J., Mucs, D., Norinder, U., & Svensson, F. (2019b). LightGBM: an effective and scalable algorithm for prediction of Chemical Toxicity–Application to the TOX21 and Mutagenicity data sets. *Journal of Chemical Information and Modeling*, *59*(10), 4150–4158. https://doi.org/10.1021/acs.jcim.9b00633