European Water 87/88: 1-10, 2024.
© 2024 *E.W. Publications*

Fidelity of machine learning models in capturing flood inundation through geomorphic descriptors over Ganga sub-basin, India

V. Tripathi, M.P. Mohanty* and H. Singh

Department of Water Resources Development and Management, IIT Roorkee, Roorkee, Uttarakhand, India
* *e-mail: mohit.mohanty@wr.iitr.ac.in*

Abstract: Hydraulic Hydrodynamic Modeling (HHM) entails multidimensional data which results in huge computational costs and, therefore a constraint for low and middle-income nations. To overcome this, the present study blends Geomorphic Flood Descriptors (GFDs) and Machine Learning (ML) models to identify the fidelity of ML models in capturing the flood inundation in the floodplains of severely flood-prone Ganga sub-basin in India (geographical area ~ 93,854 km^2). GFDs have been widely recognized by the scientific community for assessing and mapping flood hazards across large geographical areas, being computationally penurious, and ease of availability make them ideal for large watersheds. An assortment of single and composite indices (GFDs) generated from a high-resolution CartoDEM (resolution~30m) was forced as input to different ML algorithms. To train ML models, the 100-yr flood hazard map developed by the Joint Research Centre (JRC) was considered. The ML models were optimized using a grid search approach. Four state-of-the-art ML models were utilized: Random Forest (RF), Adaptive Boosting (AdaBoost), Logistic Regression (LR), and k-nearest Neighbor (KNN). To enumerate the performance of ML models, the F1 score along with Cohen's kappa coefficient and AUC (ROC) values were used. The results obtained suggest the superiority of the RF model over others with a 'k' value of 0.70. The study presents an integrated top-down approach to flood mapping, which will pave the way for effective adaptation strategies for minimizing flood risks over large resource-constrained watersheds.

Key words: Geomorphic flood descriptors; machine learning; flood inundation; large watersheds.

1. INTRODUCTION

Flooding is the most frequent natural disaster that has caused significant economic losses and lives worldwide. The frequency and severity of these cataclysmic hazards have increased due to climate change associated with global warming (Davenport et al., 2021). United Nations Office for Disaster Risk Reduction reported that floods accounted for 43.4 % of all disasters recorded between 1998 and 2017 (UNDRR, 2020). Floods worldwide caused an economic loss of 20.4 billion USD (CRED, 2023). A devastating count of 1529 deaths and a staggering 10.2 million people were affected in India in 2023 due to flooding (CRED, 2023). Many studies have claimed that the combined effects of erratic rainfalls, intensive urbanization, and development in the flood zones have aggravated the flooding scenarios. While this growing trend is expected to continue and result in the extension of flood hazard areas in the future (Johnson et al., 2020), flood risk assessment is indispensable for ensuring appropriate resilience to vulnerable communities. The first step of flood risk assessment is the identification of flood-prone regions and the development of flood hazard maps within a watershed (Thakur et al., 2024). These geographical maps comprehensively inform the policymakers and other stakeholders on how to effectively use flood management techniques to minimize the number of casualties, economic losses, and social unrest. (Thakur and Mohanty, 2023).

Many past studies have utilized the advantage of Hydraulic-cum-Hydrodynamic modeling (HHM) to estimate flood inundation (Singh and Mohanty, 2023). They require multidimensional data like surface roughness, topography, and boundary conditions. Unfortunately, the paucity of gauging stations makes them less usable for low and middle-income nations. The computational requirements to simulate flood events using HHM are huge and the cost associated is significant, hence, a drawback for economically struggling nations (Namgyal et al., 2023). The limitation of HHM over large and data-stressed watersheds prompts to integrate basin's morphology. It denotes an intricate interplay of hydrological, geological, and anthropogenic disturbances over a period of time. The association of flooding and catchment attributes has been identified by previous studies (Manfreda et al., 2015; Samela et al., 2017). The factors unfolding aspects of flooding are known as Geomorphic Flood Descriptors (GFDs). GFDs are generally derived from the Digital Elevation Model (DEM) after a set of hydrological preprocessing procedures. These indicate the channel characteristics while determining the probability and severity of flooding in the basin using the topographical information. The fidelity of Geomorphic GFDs for flood inundation mapping has been recognized in recent times. Manfreda et al. (2014) conducted a study on the Tiber River in Central Italy to identify the flooded areas using DEM-based approaches. Samela et al. (2017) classified flooded and non-flooded pixels incorporating single and composite geomorphic indices using a threshold approach in the Ohio River basin. Various GFDs such as slope, geomorphic flood index (GFI), height above the nearest drainage (HAND), plan curvature, etc., have been exhaustively used in pin-pointing flooded or non-flooded locations (Samela et al., 2017; De Risi et al., 2018). Many past studies have coupled these with other factors like geology (Janizadeh et al., 2019) and land use land cover (Khosravi et al., 2018) in determining the information on floods. These studies used a linear binary classification approach using a threshold value, which is achieved by continuously iterating using a benchmark map. The non-linearity between GFDs and flood information i.e., flooded or non-flooded can not be estimated using a threshold value and it cannot be generalized for other large watersheds. To subdue this, the incorporation of Machine Learning (ML) models can be beneficial as they can handle the non-linearity in the data while identifying the hidden patterns or the relationships between the variables. The availability of the huge dataset has led to ML gains as they only need data without considering the physical processes involved (Tripathi and Mohanty, 2024). These models are forced with a sample of features from a training dataset to determine the label class in classification tasks. The catchment descriptors at a location are the features while their flooding status i.e. *flooded* or *non-flooded* is a label class in the binary classification problems. The incorporation of Machine Learning (ML) models into flood inundation mapping is widely regarded, however, their application to GFDs is still at a nascent stage (Debnath et al., 2023). The integration of catchment attributes into a machine learning model to quantify flood hotspots is still unknown and the choice of choosing an appropriate ML model is still not understood in large watersheds (Mishra et al., 2022). To enumerate this, the present framework utilizes the power of GFDs and integrates them into ML models to determine the fidelity of various ML models in flood inundation mapping for large resource-constrained watersheds. We applied this methodology over the severely flood-prone Ganga sub-basin (geographical area of 93,854 km^2). After complete hydrological processing over CartoDEM (resolution ~ 30 m), it has been utilized to derive the high-resolution spatial GFDs. Many ML model requires the pre-processing of datasets before directly incorporating them, hence to account for this the correlation matrix has been used to eliminate the correlated features. The present study uses the power of Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbor (KNN), and Adaptive Boosting (AdaBoost) to classify the pixels into flooded or non-flooded. A total of six performance metrics, namely precision, recall, Cohen's kappa coefficient, accuracy, F1 score, and ROC curve have been estimated to account for the best classification model. The central aim of this study is to develop an approach to identify flood hotspots in large watersheds where data scarcity persists and for low and middle-income nations that are struggling to create a flood atlas. To increase community adaptation and resilience to these hazards, our study also complies with the recommendations made by the National Disaster Management Authority (NDMA), the Sustainable Development Goals (SDG) 11.5, and the United Nations Office for Disaster Risk Reduction (UNDRR).

2. STUDY AREA AND DATASET

The Ganga River Basin (GRB), which spans a geographic area of roughly 93,854 km^2, is the most flood-prone area. It is located between longitudes 79°–82° E and latitudes 25°–31° N (Figure 1). It encompasses 26.2% of India's total land area. According to The Himalayan Climate and Water Atlas (2015), there are roughly 179 and 152 wet days in the upper and lower basins, respectively, out of the approximately 1000 mm of precipitation that falls over GRB each year. The Himalayan Climate and Water Atlas (2015) states that the average lowest temperature in the basin varies from 21.5 °C in the summer to 6.4 °C in the winter. The average maximum temperature varies from 30.3 °C in the summer to 21.1 °C in the winter. The region's height varies from 5 to 7,184 meters. The GRB is retrieved from the HydroSHEDS (Hydrological Data and Maps Based on Shuttle Elevation Derivatives at Multiple Scales). Previous research has demonstrated that CartoDEM is more accurate and error-free than other globally accessible digital elevation models (DEMs) when it comes to hydrological modeling and mapping flood inundation (Mohanty et al., 2020). CartoDEM (horizontal resolution < 1 arc second, vertical accuracy ~ 8 m) (Muralikrishnan et al. 2013) has therefore been used to estimate GFDs. Nogherotto et al. 2022 reported superior performance of JRC over others. A 100-year flood inundation map (resolution ~ 30 arcsec/1 km) created by the Joint Research Centre (JRC) is used to train the machine learning models.

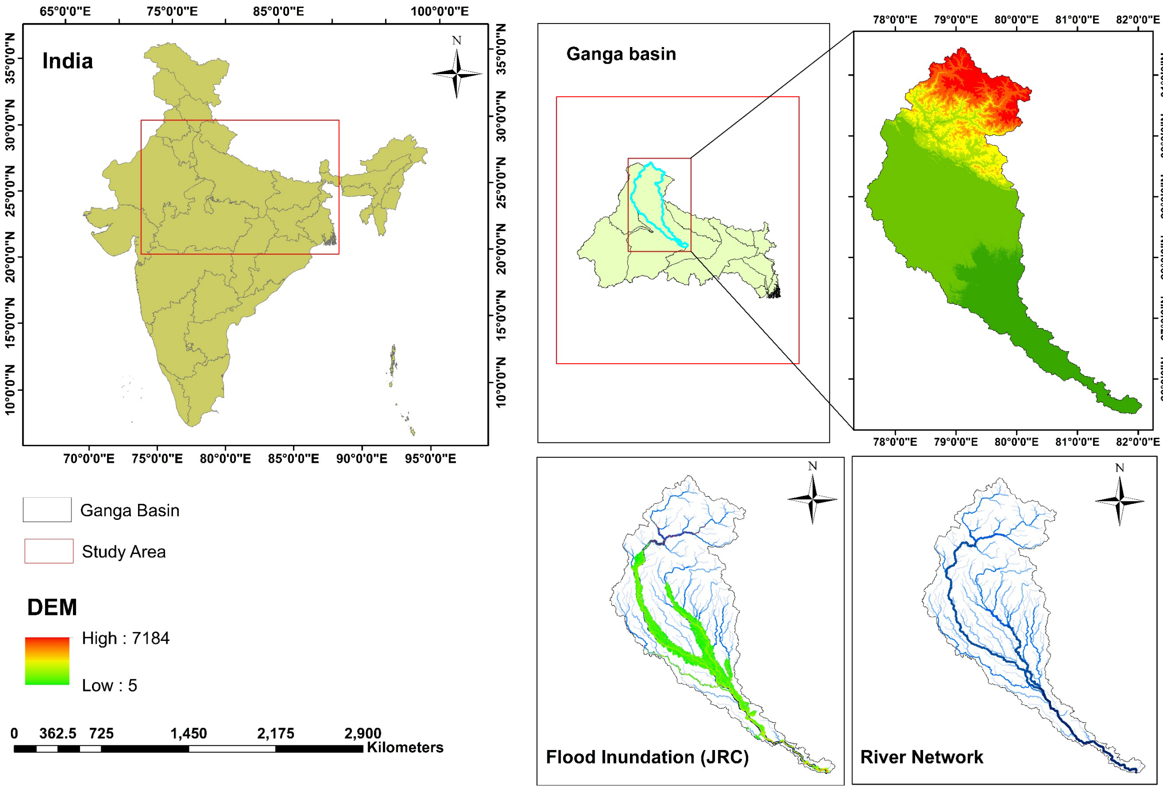

Figure 1. Description of the study area

3. METHODOLOGY

To address the above concerns, we develop ML models to quantify the flood hotspots using the channel characteristics as an input. We are also intrigued to account for the fidelity of ML models in flood susceptibility mapping. The comprehensive methodology is described below.

3.1 Selection of GFDs

The GFDs govern river flow dynamics and depict the flood patterns of a watershed, these represent the channel characteristics of flooding (Samela et al., 2017). GFDs are obtained from topographic data, their use in comprehending the dynamics of floodplains is fascinating (Samela et al., 2017). After an exhaustive literature review, a set of 15 GFDs was developed using CartoDEM (resolution ~1 arcsecond). A set of hydrological conditioning was applied to the DEM to remove the noise and errors in a GIS environment (ArcGIS *v*10.8 and QGIS *v*3.28.3). The values of GFDs were derived at a location along with their status of flooding, a dataset was constructed that fit into a binary categorization (Plataridis and Mallios, 2023). To make sure that all data inputs have the same size and dimensions, preprocessing data is necessary before using it in machine learning models. Therefore, the normalization of GFDs was performed using the following equation at a grid-scale:

$$GFD_{il}^{std} = \left\{ \frac{GFD_{il} - GFD_i^{min}}{GFD_i^{max} - GFD_i^{min}} \right\} \quad (1)$$

where GFD_{il}^{std} is normalised value and GFD_{il} is the actual value of i^{th} GFD at the l^{th} location, GFD_i^{min} and GFD_i^{max} are the minimum and maximum value of i^{th} GFD at all locations.

3.2 Feature selection

It is a method of choosing a subset of the features that have a significant impact on the target variable. Its main goal is to eliminate unnecessary predictors from the model. By avoiding overfitting and producing more comprehensible ML models, the integration of these techniques lowers computational costs and improves the performance, effectiveness, and resilience of the model. In the present framework, a Pearson's correlation matrix was calculated to examine the linear relationship between the independent and dependent variables.

3.3 Flood inundation using ML models

This study essentially considers flood hazard as a binary classification problem, i.e., the classification of a location as either "flooded" or "non-flooded" (Samela et al., 2017). The standard flood hazard map developed by the European Commission, Joint Research Centre (JRC) was considered the ground truth (Dottori et al., 2016). This dataset was segregated into "training" and "testing" in a ratio of 70:30. A grid search approach was adopted to estimate the best hyperparameter of these classification models, and hence testing results were calculated at these values to identify the best classifier. Python packages (scikit-learn, matplotlib, etc.) were used to develop these algorithms. The detailed description of ML models used in the study is:

- *Random Forest (RF):* it is an ensemble learning algorithm that takes advantage of multiple random decision trees in drawing a result. It was developed by Breiman (2001) to overcome the drawbacks of decision trees. It uses information gain theory in selecting a random feature which helps in identifying outliers and bootstrapping which reduces the variance in the model by introducing variability in the model. It has been widely used in classification and regression tasks because of its robustness against overfitting (Tripathi and Mohanty, 2024).
- *Adaptive Boosting (AdaBoost):* Freund and Schapire (1997) introduced Adaptive Boosting often referred to as AdaBoost. It is an ensemble learning algorithm that works on the principle of boosting while combining multiple weak or base learners (decision trees) into strong learners to classify the pixels into flooded or non-flooded. A base estimator is trained on samples selected randomly and weights are assigned. These weights are iteratively adjusted with higher weights being assigned to instances classified incorrectly. It decreases bias while increasing the variance which enhances its ability to handle huge and complex data and makes it one of the most powerful algorithms in binary classification problems (Aydin and Iban, 2023).
- *Logistic Regression (LR):* It is a type of statistical model often referred to as a logit model. It uses maximum likelihood estimation to determine the beta coefficients of the model. It also

uses regularization techniques to improve the numerical stability, making the model more robust. It also decreases the overfitting issue which is generally seen when the number of predictors is more. The logit transformation applied to odds is the ratio of the probability of success to failure. Because of its wide applicability and ease of use, it has been widely accepted in performing a classification task.

- *K-Nearest Neighbor (KNN):* It is one of the most basic classification algorithms in machine learning which uses a non-parametric classifier that makes predictions based on similarity of data points, i.e., it uses proximity to make classifications. Being non-parametric it does not assume the data distribution. It handles both numerical and categorical data making its applicability to any classification or regression problem. It uses Euclidean distance to find the K-nearest neighbors, the class is determined by voting of K neighbors. The hyperparameter to tune in this algorithm is the value of 'k'.

3.4 Evaluation of ML models

In this framework, six classification metrics like precision (or positive predictive value), recall (or true positive ratio), Cohen's kappa coefficient (κ), accuracy, F1 score, and area under the receiver operating characteristic curve (AUC) have been used to evaluate the performance of classifier ML models (Tripathi and Mohanty, 2024). The area under the receiver operating characteristics curve (AUC) is a threshold-independent metric that is widely used in flood susceptibility mapping (Lyu and Yin, 2023). It calculates the likelihood that an area will be accurately classified as flooded or not by a trained machine learning model. The ROC curves and AUC values are generated by varying, creating multiple confusion matrices, true positive rate (r_{tp}), and false positive rate (r_{fp}). The confusion matrices are generated using a threshold value (τ). The four values in the matrix are, as follows: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). TP is the quantity of known flooded pixels anticipated to fall into the flooded class, while FP represents non-flooded pixels classified to the flooded class. Similarly, TN is the quantity of known non-flooded pixels anticipated to fall into the non-flooded class, while FN represents flooded pixels classified to the non-flooded class.

Table 1. Performance metrics used for all classification models

Performance Statistic	Mathematical expression	Range	Description
Accuracy (p_o)	$\frac{(TP\ +\ TN)}{(TP\ +\ TN\ +\ FP\ +\ FN)}$	0 to 1	It is simply a ratio of correctly predicted data pixels to the total number of pixels.
Recall (r_{tp})	$\frac{(TP)}{(TP\ +\ FN)}$	0 to 1	It measures the proportion of actual positives that were correctly identified by the model.
Precision	$\frac{(TP)}{(TP+FP)}$	0 to 1	It depicts the proportion of positive predictions predicted by the model that are correct.
F1 score	$\frac{2(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}$	0 to 1	F1 score is the harmonic mean of precision and recall, hence widely accepted in classification analysis.
Kappa Coefficient (κ)	$\frac{p_o - p_e}{1 - p_e}$ $p_e = \frac{(TP+FN)(TP+FP)+(FP+TN)(FN+TN)}{(TP+TN+FN+FP)^2}$ $r_{fp} = \frac{FP}{(TN\ +\ FP)}$	-1 to 1	It measures the agreement between two evaluators.

4. RESULTS AND DISCUSSION

The present study evaluates the fidelity of Machine Learning models based on the values of different GFDs along with the identification of flooded or non-flooded pixels. The single indices considered are Height above the nearest drainage (H), tangential curvature (K_t), distance to the nearest stream (D), slope (S_l), flow accumulation (A_r), profile curvature (K_p), elevation (E), geomorphons (G), and plan curvature (K_h).

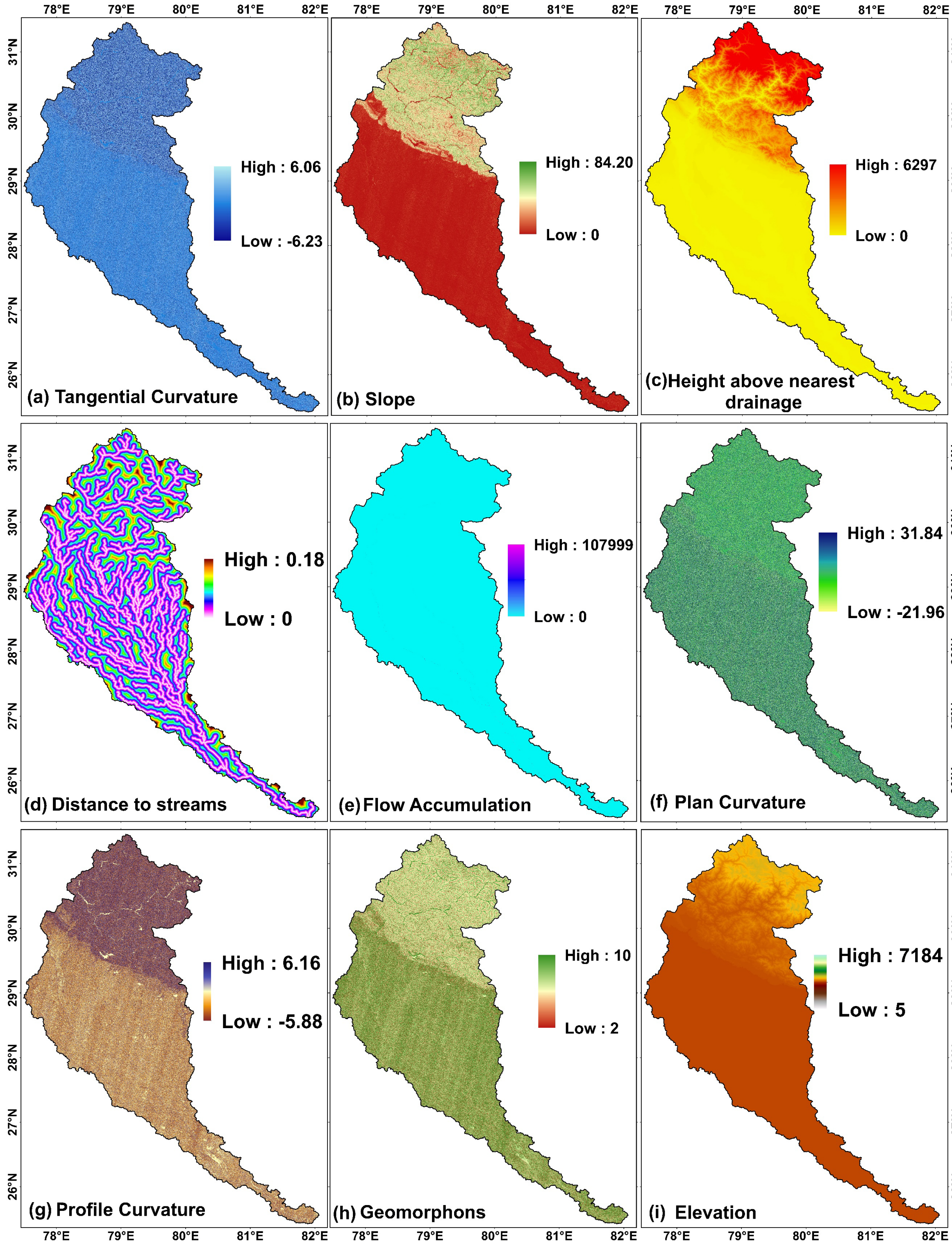

Figure 2. Description of single GFDs

Figure 2 shows the spatial maps of the single-feature GFDs. Downslope index (DI), multi-resolution valley bottom flatness index (MRVBF), geomorphic flood index (GFI), stream power index (SPI), convergence index (CI), and topographic wetness index (TI) are the composite feature GFDs considered for the current framework (Figure 3). K_t varied from -6.23 to 6.06 degrees from high mountains to low-lying areas respectively. GFI varied between −8.85 to 6.54 in the basin. TI consistently showcased its variability, spanning from −10.48 in the low -lying areas to 17.22 in major rivers and water bodies. 'H' one of the major factors in flood hazard studies varied from 0 in low-lying areas to 6297 in mountain ranges. SPI indicates the erosive power of flowing water, varied spatially from 0 in the flat terrains to as high as 6689 in the river streams. MRVBF captures the lowness and flatness of the terrain. The features capturing local flow behavior are Sl, Kh, Kp, Kt, and CI. However, the global behavior of the flow is represented by TWI, DI, H, and D.

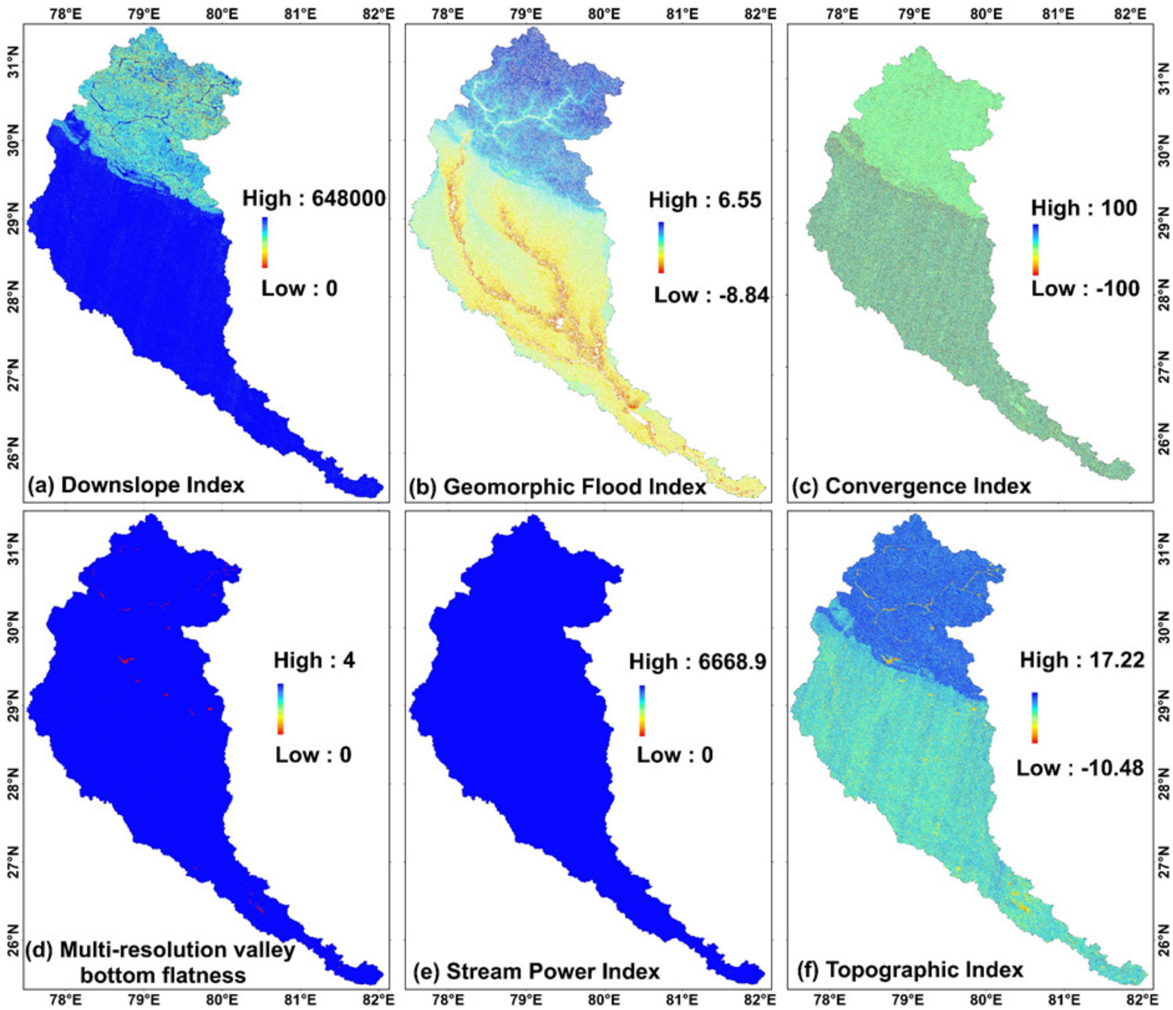

Figure 3. Description of composite GFDs

4.1 Selection of the best ML model

The ML model's performance in the testing dataset is used to determine its dependability. The training dataset's performance measures assess how well the machine learning model replicates the intended map. It assesses the model's capacity to learn from the unseen dataset using the testing dataset (Wang et al., 2021). The entire dataset was normalized using Equation (1). A Pearson's correlation matrix was created to evaluate a linear relationship between the variables. It was discovered that there was a strong positive association between "E" and "H," "DI" and "S_l," and "K_h" and "K_t," but a negative correlation between "CI" and "G" (Figure 4). The correlation coefficient indicates how two variables are dependent (Habibi et al., 2023). The highly correlated features were not considered as an input to ML Classifiers in determining their efficacy in deriving the flood inundation. The performance of the ML Classifiers is highlighted in Table 2. RF exhibited the highest accuracy 94% followed by AdaBoost (93%), LR (93%), and KNN (93%) in the testing phase. The high value of precision of AdaBoost indicates that the rate of false positives is low in the model, while the low rate of false negatives in the model is denoted by high recall values.

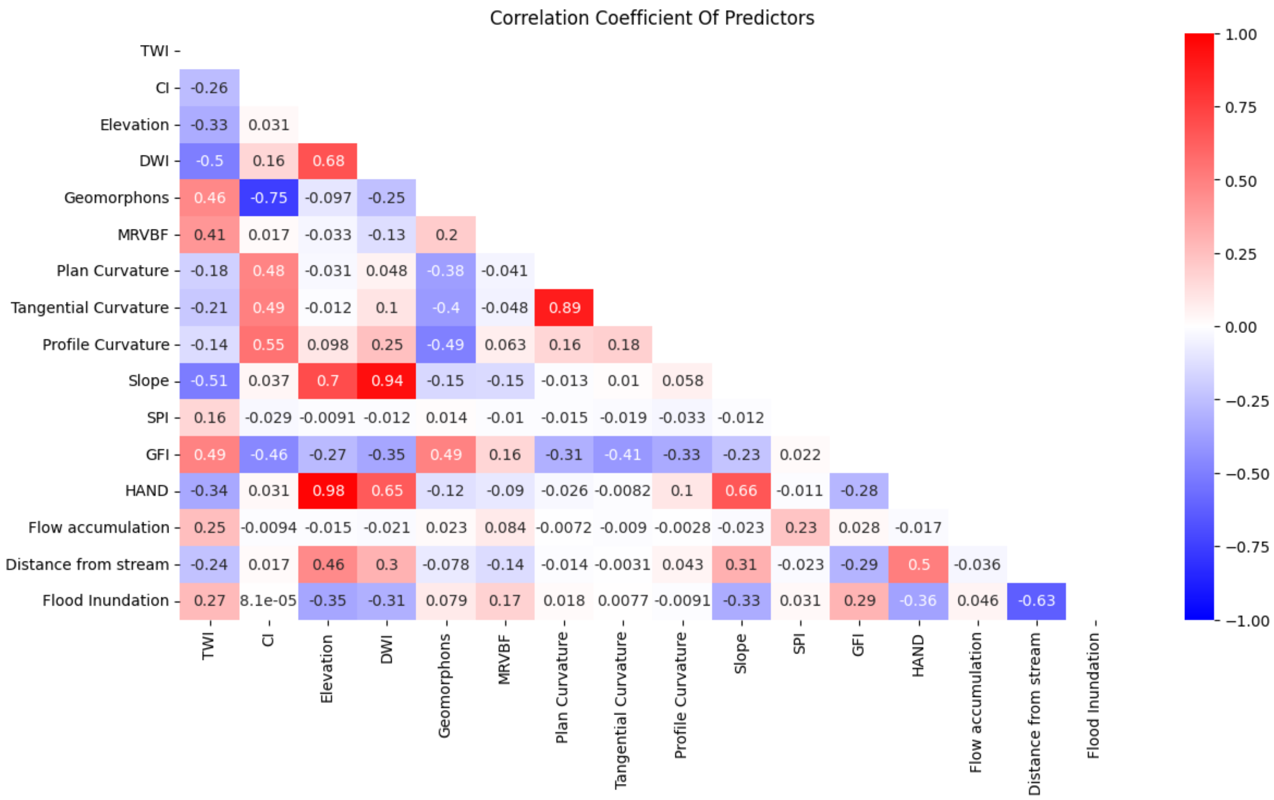

Figure 4. Correlation matrix of GFDs

In flood inundation mapping, false negatives should be minimized as they represent flooded locations as non-flooded. The F1 score is the harmonic mean of precision and recall, providing an overall index for the models. RF represented the highest value of F1 score (0.73) among the models with a promising value of 'k' as 0.70 in the testing phase, followed by AdaBoost.

Table 2. Performance metrics used for all classification models

Performance metrics	RF		AdaBoost		LR		KNN	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Precision	0.83	0.81	0.83	0.86	0.85	0.82	0.84	0.79
Recall	0.69	0.67	0.70	0.62	0.72	0.69	0.70	0.66
κ	0.73	0.70	0.72	0.68	0.71	0.67	0.72	0.68
F1 score	0.75	0.73	0.76	0.72	0.78	0.75	0.76	0.72
Accuracy	0.93	0.94	0.94	0.93	0.94	0.93	0.92	0.93

The excellent performance of RF is due to the bootstrap aggregation. The area under ROC represents the model performance and how exactly a trained model can classify between flooded and non-flooded locations. Random Forest performed better in the training phase (AUC = 0.98) as compared to other models, but in the testing phase, AdaBoost was found slightly better than the rest of the models (AUC=0.95), followed by Random Forest (AUC=0.94), Logistic Regression (AUC=0.89) and K-nearest neighbour (AUC=0.70).

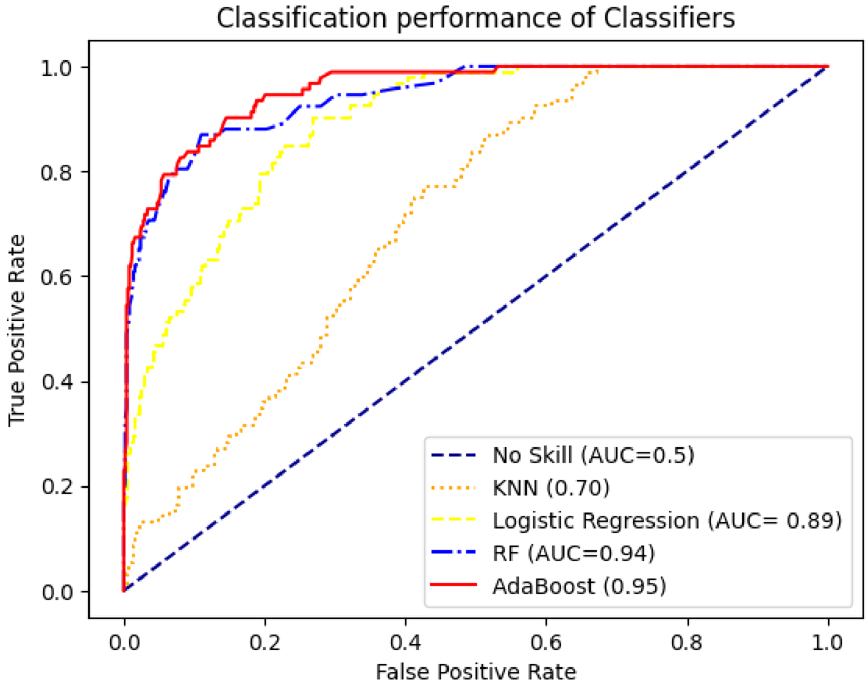

Figure 5. AUC(ROC) value of ML Classifiers

5. CONCLUSION

An essential first step in managing flood risk is having information on flood hazards. Flood inundation maps serve as a guide for legislators, government employees, aid agencies, and urban planners. They may put into practice more effective catastrophe risk reduction and management techniques by coordinating their efforts and making well-informed judgments based on this data. One of the biggest obstacles to accounting for flood threats across large watersheds is the need for computationally intricate hydrological-cum-hydrodynamic models that can only operate well in the presence of massive datasets. This obstacle drives the quest for equally inventive solutions, which are governed by the geomorphic flood descriptors (GFDs) unique to each river basin. Disaster management has gained new life thanks to the incorporation of machine learning (ML) models in computational and data-driven statistics. Experts recognize the promise of ML models because they simplify the process of making difficult judgments based on a variety of data sources that are relevant to disaster management. For the first time, the current study offers a thorough framework to combine data-driven methodologies from the ML model with GFDs to define flood hotspots across the vast, flood-prone Ganga sub-basin. According to our research, Random Forest outperforms other machine learning models, and this might be taken into consideration going forward when creating flood inundation maps. The study advises against using machine learning (ML) models excessively for mapping flood inundation, particularly in large regions where hydrodynamic modeling of floods is a difficult undertaking. As a result, there will be very little uncertainty created in the flood risk and hazard dimensions.

REFERENCES

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.

Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.
CRED (Centre for Research on the Epidemiology of Disasters) (2023) EM-DAT: the emergency events database. From https://www.emdat.be/. Accessed 10 June 2024
Davenport, F. V., Burke, M., & Diffenbaugh, N. S. (2021). Contribution of historical precipitation change to US flood damages. Proceedings of the National Academy of Sciences, 118(4), e2017524118.
De Risi, R., Jalayer, F., De Paola, F., & Lindley, S. (2018). Delineation of flooding risk hotspots based on digital elevation model, calculated and historical flooding extents: the case of Ouagadougou. Stochastic environmental research and risk assessment, 32, 1545-1559.
Debnath, J., Sahariah, D., Mazumdar, M., Lahon, D., Meraj, G., Hashimoto, S. et al. (2023). Evaluating Flood Susceptibility in the Brahmaputra River Basin: An Insight into Asia's Eastern Himalayan Floodplains Using Machine Learning and Multi-Criteria Decision-Making. Earth Systems and Environment, 7(4), 733-760.
Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F. A., & Feyen, L. (2016). Development and evaluation of a framework for global flood hazard mapping. Advances in water resources, 94, 87-102.
Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1), 119-139. https://doi.org/10.1006/jcss.1997.1504
Habibi, A., Delavar, M. R., Sadeghian, M. S., Nazari, B., & Pirasteh, S. (2023). A hybrid of ensemble machine learning models with RFE and Boruta wrapper-based algorithms for flash flood susceptibility assessment. International Journal of Applied Earth Observation and Geoinformation, 122, 103401.
Janizadeh, S., Avand, M., Jaafari, A., Phong, T. V., Bayat, M., Ahmadisharaf, E. et al. (2019). Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh watershed, Iran. Sustainability, 11(19), 5426.
Johnson, K. A., Wing, O. E., Bates, P. D., Fargione, J., Kroeger, T., Larson, W. D. et al. (2020). A benefit–cost analysis of floodplain land acquisition for US flood damage reduction. Nature Sustainability, 3(1), 56-62.
Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I. et al. (2018). A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. Science of the Total Environment, 627, 744-755.
Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrological Processes, 27(15), 2171-2186.
Lyu, H. M., & Yin, Z. Y. (2023). Flood susceptibility prediction using tree-based machine learning models in the GBA. Sustainable Cities and Society, 104744.
Manfreda, S., Nardi, F., Samela, C., Grimaldi, S., Taramasso, A. C., Roth, G., & Sole, A. (2014). Investigation on the use of geomorphic approaches for the delineation of flood prone areas. Journal of Hydrology, 517, 863-876.
Manfreda, S., Samela, C., Gioia, A., Consoli, G. G., Iacobellis, V., Giuzio, L. et al. (2015). Flood-prone areas assessment using linear binary classifiers based on flood maps obtained from 1D and 2D hydraulic models. Natural Hazards, 79, 735-754.
Mishra, B. P., Ghose, D. K., & Satapathy, D. P. (2022). Geospatial modeling using hybrid machine learning approach for flood susceptibility. Earth Science Informatics, 15(4), 2619-2636.
Mohanty, M. P., Vittal, H., Yadav, V., Ghosh, S., Rao, G. S., & Karmakar, S. (2020). A new bivariate risk classifier for flood management considering hazard and socio-economic dimensions. Journal of environmental management, 255, 109733.
Muralikrishnan, S., Pillai, A., Narender, B., Reddy, S., Venkataraman, V. R., & Dadhwal, V. K. (2013). Validation of Indian national DEM from Cartosat-1 data. Journal of the Indian Society of Remote Sensing, 41, 1-13.
Namgyal, T., Thakur, D. A., Rishi, D. S., & Mohanty, M. P. (2023). Are open-source hydrodynamic models efficient in quantifying flood risks over mountainous terrains? An exhaustive analysis over the Hindu-Kush-Himalayan region. Science of the Total Environment, 897, 165357.
Nogherotto, R., Fantini, A., Raffaele, F., Di Sante, F., Dottori, F., Coppola, E., & Giorgi, F. (2022). A combined hydrological and hydraulic modelling approach for the flood hazard mapping of the Po river basin. Journal of Flood Risk Management, 15(1), e12755.
Plataridis, K., & Mallios, Z. (2023). Flood susceptibility mapping using hybrid models optimized with Artificial Bee Colony. Journal of Hydrology, 129961. https://doi.org/10.1016/j.jhydrol.2023.129961.
Samela, C., Troy, T. J., & Manfreda, S. (2017). Geomorphic classifiers for flood-prone areas delineation for data-scarce environments. Advances in Water Resources, 102, 13-28.
Singh, H., & Mohanty, M. P. (2023). Can atmospheric reanalysis datasets reproduce flood inundation at regional scales? A systematic analysis with ERA5 over Mahanadi River Basin, India. Environmental Monitoring and Assessment, 195(10), 1143.
Thakur, D. A., & Mohanty, M. P. (2023). A synergistic approach towards understanding flood risks over coastal multi-hazard environments: Appraisal of bivariate flood risk mapping through flood hazard, and socio-economic-cum-physical vulnerability dimensions. Science of The Total Environment, 901, 166423.
Thakur, D. A., Mohanty, M. P., Mishra, A., & Karmakar, S. (2024). Quantifying flood risks during monsoon and post-monsoon seasons: An integrated framework for resource-constrained coastal regions. Journal of Hydrology, 630, 130683.
The Himalayan Climate and Water Atlas (2015) retrieved from The Himalayan Climate and Water Atlas | GRID-Arendal (grida.no) on 22 July 2023.
Tripathi, V., & Mohanty, M. P. (2024). Can geomorphic flood descriptors coupled with machine learning models enhance in quantifying flood risks over data-scarce catchments? Development of a hybrid framework for Ganga basin (India). Environmental Science and Pollution Research, 1-20.
UNDRR. (2020). The human cost of disasters: An overview of the last 20 years. United Nations Office for Disaster Risk Reduction (UNDRR).
Wang, Y., Fang, Z., Hong, H., Costache, R., & Tang, X. (2021). Flood susceptibility mapping by integrating frequency ratio and index of entropy with multilayer perceptron and classification and regression tree. Journal of Environmental Management, 289, 112449.