7 Application of Hybrid Machine Learning Algorithms for Flood Susceptibility Modeling

Swapan Talukdar, Sujit Kumar Roy, Showmitra Kumar Sarkar, Susanta Mahato, Swades Pal, Atiqur Rahman, Bushra Praveen, and Tanmoy Das

Abstract

This study combines machine learning (ML) algorithms with statistical models to generate new hybrid models for flood susceptibility mapping (FSM) in the Teesta River basin of Bangladesh (LR). Two-hybrid ML algorithms, such as ANN-LR and RF-LR models for FSM, have been created by combining two ML techniques, such as artificial neural network (ANN) and random forest (RF), with one statistical approach, such as logistic regression. The FSMs were then validated using parametric and non-parametric receiver operating characteristic curves (ROC), such as empirical and binormal ROC. We evaluated the impact of the parameters on FSM using a Random forest-based sensitivity analysis. The extremely high (1023–1120 km^2) and high flood vulnerability zones were estimated using three methods and two hybrid models (521–674 km^2). Based on the ROC's area under the curve, the ANN-LR model (ROCe-AUC: 0.883; ROCb-AUC: 0.936) outperformed other models (AUC). According to the validation results, two hybrid models outperformed three machine learning and statistical models. The findings of this research will aid FSMs in building long-term flood control strategies by increasing their efficiency.

Keywords

Flood susceptibility · Machine learning · Statistical technique · ROC curve · Artificial neural network · Sensitivity analysis · Remote sensing

S. Talukdar · A. Rahman · T. Das
Faculty of Natural Science, Department of Geography, Jamia Millia Islamia, New Delhi 110025, India
e-mail: arahman2@jmi.ac.in

T. Das
e-mail: tanmoy2008953@st.jmi.ac.in

S. K. Roy
Institute of Water and Flood Management (IWFM), Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

S. K. Sarkar
Department of Urban and Regional Planning, Khulna University of Engineering and Technology, Khulna, Bangladesh

S. Mahato (✉)
Special Centre for Disaster Research, Jawaharlal Nehru University, New Delhi 110067, India
e-mail: susantamahato@jnu.ac.in

S. Pal
Department of Geography, University of Gour Banga, Malda, India

B. Praveen
School of Humanities and Social Science, Indian Institute of Technology Indore, Madhya Pradesh, India

B. Pradhan et al. (eds.), *Spatial Modelling of Flood Risk and Flood Hazards*, GIScience and Geo-environmental Modelling, https://doi.org/10.1007/978-3-030-94544-2_7

7.1 Introduction

Natural catastrophes have emerged as one of society's most critical global issues (Kabir and Hossain 2021). Flooding is one of the most prevalent, well-known, and everyday occurrences among all-natural catastrophes (Ahmed et al. 2021). Floods create a devastating situation and often occur during the monsoon, especially in the Indian subcontinent (Zhang et al. 2021), which receives nearly 80% of the annual rainfall in the monsoon. Flooding occurs in the surrounding land area because of heavy rainfall and massive river runoff (Yousefi et al. 2018). A flood becomes a catastrophe when it causes significant damage to people and their livelihood and habitat (De Silva and Kawasaki 2018). Floods hit Bangladesh every year in the Ganges–Brahmaputra-Meghna (GBM) basin (Uhe et al. 2019). Riverine, rainfall-induced, flash, tidal, and cyclonic/storm surgical floods have all been recorded (Uhe et al. 2019; Islam et al. 2021a). Being the GBM's outflow, Bangladesh has to deal with a lot of stream-flow during the rainy season (which begins in June and lasts till the end of October). Floods are more common and conspicuous in flatlands (which account for 79.1% of the geographical area) and hence get the most significant study and planning focus. Approximately 20–25% of the area is affected by flooding on an annual basis. Sixty percent of the region had been impacted by floods in 1987, 1988, and 1998 (Lin et al. 2019). The socioeconomic repercussions of these regular floods are enormous. Between 2009 and 2014, floods of various magnitudes affected 57.01% of Bangladeshi houses on average. In economic terms, the cost of damage was estimated to be 0.85 billion USD (De Silva and Kawasaki 2018; Rahman et al. 2020).

Despite the enormous financial and development costs, flood losses were unavoidable because of a lack of comprehensive flood assessment tools for improved preparation. The flood risk may be defined and assessed in a variety of ways. The mathematical modeling of flood susceptibility has been considered the most complicated process. It is possible to expect the geographical distribution of floods that have happened or are expected to occur in a particular location quantitatively and qualitatively. As a consequence, flood susceptibility mapping may help policymakers and other stakeholders establish disaster preparedness plans. While flood dangers cannot be eliminated, flood damages may be circumvented or considerably declined if flood inundated areas are correctly predicted (Shafizadeh-Moghadam et al. 2018).

As a consequence, determining flood susceptibility is crucial for disaster assistance. Researchers have presented a variety of flood risk assessment models. The bulk of recent models in geographic information systems (GIS) integrate hydrological, multi-criteria decision analysis, hydrodynamic, statistical models (SM), and ML algorithms (Jamali et al. 2020). GIS and remote sensing are one of the most significant databases and tools that have been extensively employed in hazard analysis (Pourghasemi and Rossi 2019).

The Frequency Ratio (FR) (Aditian et al. 2018), the Analytical Hierarchical Process (AHP) (Mukherjee and Singh 2020), the Analytical Network Process (ANP) (Abedi Gheshlaghi et al. 2020), Support Vector Machines (SVM) (Fan et al. 2020), and Random Forest (RF) (Probst et al. 2019) have all been used to study flood (Adnan et al. 2020). Every modeling approach has benefits and drawbacks. The accuracy, structure, and data of each model affect its performance. Consequently, a wide range of ensemble techniques to geohazard susceptibility and potentiality mapping has gained popularity quickly (Elmahdy et al. 2020).

Experimental hybrid techniques for FS modeling research have been researched in recent years since there is a necessity to evaluate existing prediction methods and procedures to get a better scientific basis and, as a result, more accurate conclusions (Eyoh et al. 2018). ANN-fuzzy logic, rough set-SVM, and other hybrid approaches created by combining SM with ML algorithms have been successfully employed in FS modeling (Chen et al. 2018; Jahangir et al. 2019). However, experts have emphasized that no model is ideal for modeling and that results

might differ from area to region and data. Thus, a new model should be designed and evaluated to get reliable results. As a result, in order to achieve the research gaps as mentioned earlier, the primary objectives are:

1. Construct a hybrid ensemble machine learning based FSM by merging LR with ML algorithms.
2. Conduct sensitivity modeling employing RF.
3. Verify the FS models using ROC curves.

The present study will aid decision-makers and governments in effectively regulating flood management.

7.2 Materials and Methods

7.2.1 Study Area

The Teesta sub-catchment is in northern Bangladesh and includes the districts of Nilphamary, Lalmanirhat, Rangpur, Kurigram, and Gaibandha (Fig. 7.1). Its drainage basin comprises several minor rivers with heights varying from 5 to 110 m, making it Bangladesh's most significant geomorphic unit. During floods, the river's usual gradient is between 0.46 and 0.56 m per kilometer, reflecting a very level area (Rahman et al. 2011). The research area's hydrological characteristics are complicated, and the region has a dense river network. This basin has a subtropical monsoonal climate. Monsoon (June–September) and dry (October–November) are the two major seasons in the basin (October–December). This basin gets around 1900 mm of precipitation per year on average (Islam et al. 2021b), with the monsoon season accounting for 80% of total annual precipitation.

7.2.2 Materials

In the present study, a variety of datasets have been obtained and used for flood susceptibility modeling. The details of datasets used for generating FSM were tabulated in Table 7.1.

7.2.3 Flood Inventory

For the creation of FSM maps, several researchers have started with the locations of previously flooded areas. It was determined that FSM was based on historical flooding regions. Two hundred ten flood points have been collected from different sources and a comprehensive site assessment for the inventory map. There must be a need to acquire non-flood data analogous to the flood data utilized in FSM (Islam et al. 2021a). The choice was based on the field survey, which included an identical amount of non-flood locations (210 points). Using a random separation to all flood and non-flood datasets, 80 (336):20 (84) training and testing datasets have been generated (Fig. 7.1). Training data has been employed to calibrate the model, while testing data has been employed to measure the accuracy of the models (Mallick et al. 2021). In the same way, inventory maps for other locations have been generated.

7.2.4 Methods for Preparing FS Conditioning Factors

Flood-influencing variables must be included as independent variables when creating a flood susceptibility map (El-Haddad et al. 2021; Azareh et al. 2019). According to Dodangeh et al. (2020), contributing factors utilized in one research area may have no effect in another, hence parameter identification must be event based. Among the highly significant and often utilized variables are elevation, aspect, TWI, SPI, STI, LULC, TRI, distance to the river (dR), curvature, soil condition, slope, and rainfall. The contributing variables have been translated into 30 m spatial resolution using resampling technique.

The topography and its derivative variables serve a vital part in determining FSM (Falah et al. 2019) (Fig. 7.2). Topography has a direct impact on runoff speed (Abdel-Fattah et al. 2017), and high slopes increase the speed of runoff and reduce infiltration rates (Abdel-Fattah et al. 2017). Furthermore, the dR has a significant

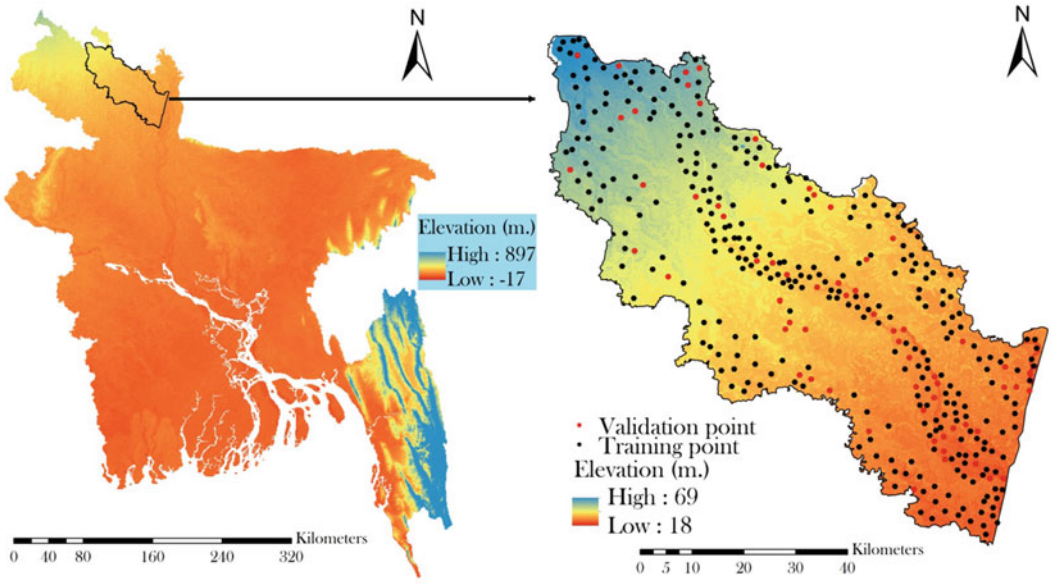

Fig. 7.1 The location of the study area

Table 7.1 Materials used for preparing FSM

Data types	Purpose	Resolution/scale	Source
ASTER DEM	For deriving different topographic and hydrologic parameters like slope, aspect, curvature, TPI, STI, TWI	Version 2, spatial resolution: 30 m	https://asterweb.jpl.nasa.gov/gdem.asp
Landsat 8 OLI	For generating LULC	Path/Row: 138/42, Spatial resolution: 30 m., date: 19/03/2019	https://earthexplorer.usgs.gov/
Rainfall	For generating rainfall map	–	Meteorological Department (BMC), Dhaka
Soil map	For preparing soil types map	1:50,000	Natural Resources Conservation Service of United States Department of Agriculture (USDA)

impact on flood magnitude. High drainage density causes a reduced capture rate and hence flow accumulation, which is one of the variables that substantially influences flood occurrence (Ogato et al. 2020; Bogale 2021).

Rainfall is recognized as the vital variable in determining FS levels (Wasko and Nathan 2019). Because heavy rain in a short period of time can create flash flooding (Wasko and Nathan 2019). In the ArcGIS 10.3 environment, acquired precipitation data from four Bangladeshi meteorological stations was utilized to create rainfall using the kriging interpolation procedure. Because the amount of data is so small, this approach is highly recommended (Islam et al. 2021b). The research area's annual rainfall, on the other hand, ranges from 361 to 550 mm (Fig. 7.3c).

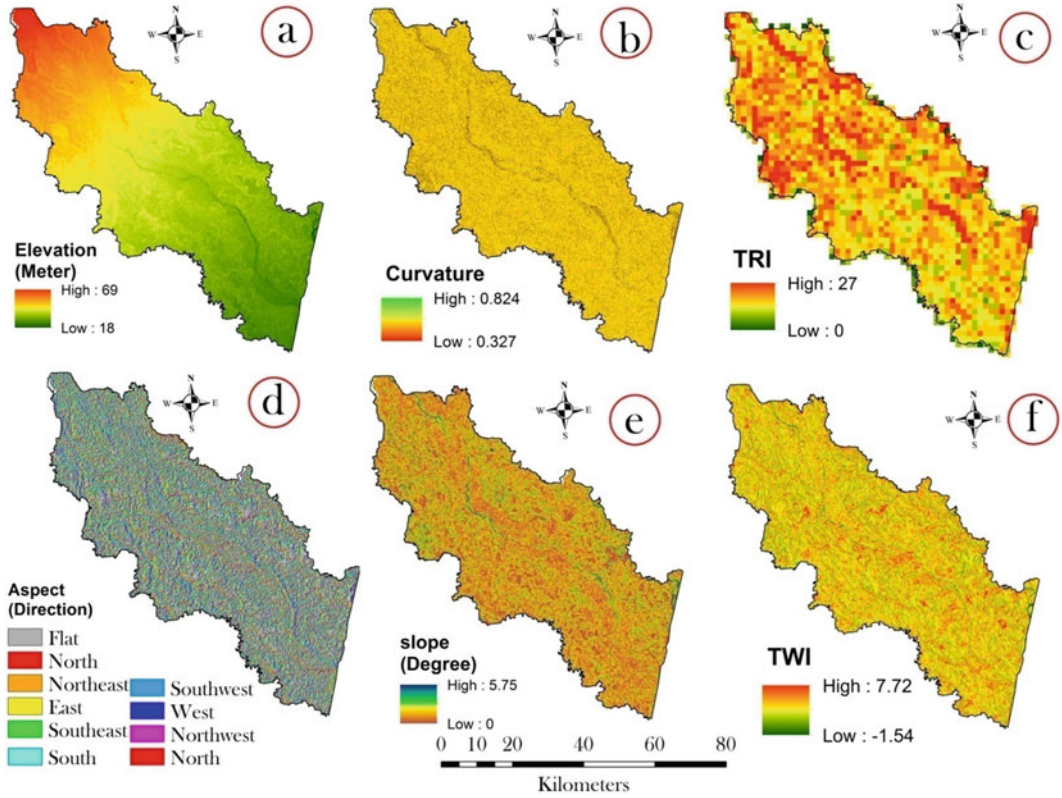

Fig. 7.2 Data layers for FSconditioning factors such as **a** elevation, **b** curvature, **c** TRI and **d** aspect, **e** slope, and **f** TWI

Soil property is the most critical factor for generating overflow (Yin et al. 2019). Flügel (1995) contends that soil quality regulates water penetration, affecting rainfall-runoff generation while indigenous climatic forms and erosion procedures also play a role. When the degree of penetration is high, susceptibility events are more likely to occur. According to USDA soil classification, the research area comprises 12 different types of soil (Fig. 7.3d).

LULC affects surface runoff, which in turn affects sediment flow, and hence has a substantial impact on the incidence of FS (Islam et al. 2021a). FS is relatively high in built-up areas because the LULC totally regulates surface flow generation and infiltration. Because different land use features preclude or help water for penetrating and generating surface water. The forest environment, on the other hand, increases water infiltration, resulting in lower FS (Islam et al. 2021b). The association between FS occurrences and plant density is adverse while comparing hydrological reactions at different temporal conditions. A LULC map has been created using the ANN model in ENVI software (version 5.3) (Talukdar et al. 2020) and classified into six classes (Fig. 7.3e).

7.2.5 Multicollinearity Analysis

The 12 conditioning variables were examined to see whether there was any association between them using the VIF and tolerance approach. If the variables are multicollinear, they are interrelated, and one of them could be predicted by other factors. Hence, it must be eliminated from the model. Among various models, such as Pearson's correlation coefficients, variance decomposition proportions, the VIF, and tolerance are the most frequent and extensively utilized. Multicollinearity is present if the VIF is more than ten or the tolerance is less than 0.1 (Talukdar et al. 2021).

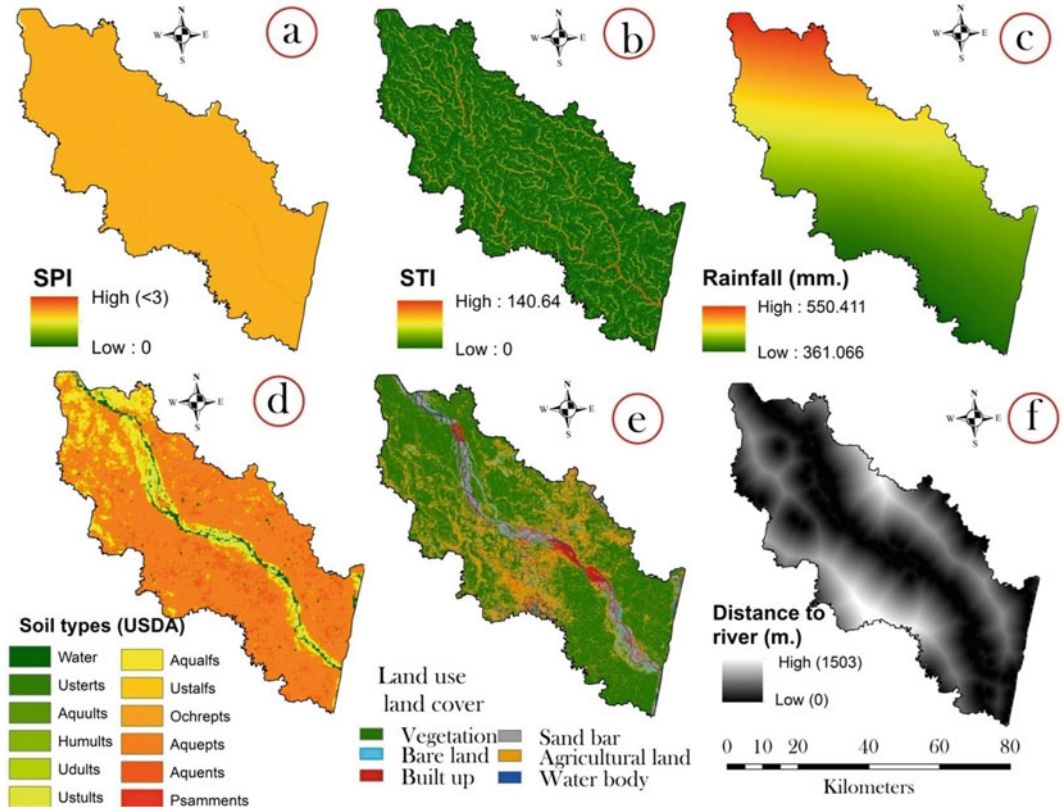

Fig. 7.3 Data layers for FS conditioning factors such as **a** SPI, **b** STI, **c** rainfall **d** soil types, **e** LULC, and **f** distance to river

7.2.6 Method for FS Modeling

7.2.6.1 Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) mimic the human nervous system and are able to acquire and simplify from instances to provide expressive solutions even while the input data includes errors or is inadequate (Jahangir et al. 2019; Khoirunisa et al. 2021). Many problems have been solved with ANNs (Kia et al. 2012). Reconstruction of the strange association between a set of input variables, such as rainfall and output variables, such as streamflow or groundwater level using ANN has been carried out (Fang et al. 2021) (Kia et al. 2012). This means that artificial neural networks can be used for flood predictions.

7.2.6.2 Random Forest (RF)

Random forest (Breiman 2001) could be summarized as a decision tree ensemble approach. It generates decision trees by randomly picking data from the training set. The decision trees were assessed independently during the training stage, with the best score being the average of the trees' outcomes. When generating decision trees, RF seeks to identify the most essential characteristics, hence feature selection is crucial. For RF, there are a few key factors to consider. The first is the forest's number of trees, which determines how many decision trees are formed during training. A large number of trees help in generalizing the model better, according to the general approach to setting this parameter. The number of features is another parameter that

refers to the depth of the decision trees. Accuracy can be improved by using a higher number of characteristics. It should be noted that bigger numbers take longer to compute, therefore time should be taken into account while parameterizing.

7.2.6.3 Logistic Regression

The LR method has been regarded as highly applied model in predicting natural hazards like flood, landslide, and drought susceptibility by many scholars throughout the world (Ali et al. 2020). In this work, the logistic regression model was chosen as one of the multivariate analytic methods to quantitatively measure flood susceptibility. The use of logistic regression to create a multivariate regression connection between a target variable and certain conditioning variables is effective for forecasting the presence or absence of the events like flood and landslide. By adding an appropriate link function to the conventional linear regression model, logistic regression allows the variables to be either continuous or discrete, or any mix of both, and they do not have to have normal distributions (Cao et al. 2020). The dependent variable in this study is a binary variable (0 or 1), with 1 indicating the existence of a flood and 0 indicating the absence of a flood.

7.2.7 Validation of the Models

The Receiver Operating Characteristic (ROC) curve is a common means of visually illustrating a marker's discriminating accuracy for differentiating between two populations. It has been utilized in radiology, psychology, epidemiology, factory inspection systems, and biomedical informatics, among other fields. The application of the ROC curve for natural hazards prediction has expanded recently, with researchers assessing the efficiency of models in discriminating between positive and negative functions of natural hazards (Tehrany et al. 2014). A depiction of the false positive rates versus the true positive rates for different diagnostic test cutoff settings is known as a ROC curve. The area under the ROC curve is the most often used metric for describing accuracy (AUC). The AUC may have a value of 0–1, with higher AUC values showing better accuracy (Yesilnacar and Topal 2005).

The nonparametric ROC techniques do not make assumptions about the distributions of diagnostic test results and do not produce a smooth ROC curve. On the other hand, parametric techniques presume that some function of the diagnostic test measures is normally distributed in both the positive and negative events, but with different means in each case. The parametric ROC approaches may create a smooth ROC curve. For the nonparametric ROC analysis in our work, we employed the empirical technique. We performed the parametric ROC analysis using the binormal approach.

7.3 Sensitivity Analysis

Mean decrease accuracy (MDA) and mean decrease Gini (MDG) coefficient are two significant measures for ranking variables and selecting variables in Random Forest. MDA analyzes the change in prediction accuracy when the values of a variable are randomly permuted compared to the original data, which determines the relevance of a variable. When a variable is used to produce a Random Forest split, MDG is the total of all Gini impurity reductions generated by that variable, normalized by the number of trees in the Random Forest (For details of RF, see method section).

7.4 Results and Discussion

7.4.1 Computation of the Multicollinearity Analysis and Importance of the Parameters

Findings of the multicollinearity test show the VIF and tolerance values of the variables are less than ten and greater than 0.2, respectively. Therefore, there is no requirement to concern

about multicollinearity between independent variables. Since all twelve flood-conditioning factors were included in this present research, FS maps included them all. The multicollinearity analysis findings for this research are as follows:

Elevation (VIF: 2.71, TOL: 0.635), slope (1.34, 0.805), curvature (1.08, 0.794), TWI (1.19, 0.653), SPI (1.56, 0.629), distance to river (1.3, 0.9), rainfall (2.76, 0.846), soil types (1.32, 0.804), and LULC (1.1, 0.803), aspect (1.08, 0.725), TRI (1.2, 0.821), STI (1.62, 0.87).

7.4.2 Description of the Parameters

The FS of an area may be affected by various factors (Bhattacharya et al. 2021). Influencing factors in this study were presented in Figs. 7.2 and 7.3. Low-lying locations, particularly those in the floodplain, retain higher soil moisture due to the continuous depressions, increasing the flooding probability. The elevation of the study area varied from 18 to 69 m above sea level (Fig. 7.2). In general, the capacity to recharge water is highest while the curvature has the feature of a concave surface. On the other hand, the water recharge capacity is less on the plain surface (Mishra et al. 2019). The DEM was used to create a curvature map with a range of 0.32–0.82 degrees of curvature (Fig. 7.2). The DEM has been employed to generate a curvature map with a range of 0.32–0.82 (See Fig. 7.2b.) A total of nine classifications were established: zero to twenty-five, twenty-five to sixty-five, seventy-five to ninety-five, and one hundred and fifty to three hundred and sixty-five (Fig. 7.2). The presence of a flat or mild slope is also advantageous since it helps to slow down water flow and enhance the sensitivity to recharge (Mahato and Pal 2019). The slopes that were employed in this study varied from 0 to 5.75°. TRI helps to determine how water flow was affected by competing for underlying surfaces (Fig. 7.2d) (Straatsma and Baptist 2008). Due to the rapid water flow created by the high hills around the Teesta River, it has lower TRI because of the rapid water flow, which suggests a greater likelihood of flooding (Mahato et al. 2021). According to the findings of this study, the highest TRI score was 27. (See Fig. 7.2 for an example.) A high TWI also guarantees that a person's susceptibility to infection is regenerated, which is beneficial. Between TWI levels and FS, there is a significant inverse relationship. Figure 7.3 depicts TWI value ranges from −1.54 to −7.72 linearly. Therefore, regions with higher SPI and STI values are more vulnerable to floods because of the greater water level represented by higher SPI and STI values (Bui et al. 2019). According to the findings of this investigation, the highest STI level measured was 140.64. Figure 7.3 depicts the distance between this location and the river, which was 1503 m in this case. It was possible to account for extra rainfall and water penetration by analyzing soil data (Johnson 2000). The 12 soil types were identified in this study area (Fig. 7.3).

7.4.3 FS Models and Their Validation

Figure 7.4 depicts the FS models built using hybrid machine learning techniques such as ANN, RF, LR, and combined ANN-LR and RF-LR. In the present study, we first applied standalone ML algorithms, and then we applied semi-machine learning algorithms, like logistic regression, for predicting flood susceptibility zones. Then, we developed hybrid models by integrating ML algorithms with the LR model. The LR model has been employed to define the weights for each parameter. The parameters are then allocated weights and turned into weighted parameters. We then used machine learning methods like ANN and RF to create hybrid models. In this way, ANN-LR and RF-LR hybrid models have been created. The weights derived using the LR model is as follows:

FSM = 18.30 + (Aspect * 0.0012) + (Curvature * 0.0325) + (Elevation * 0.0825) + (LULC * 0.3694) + (Rainfall * 0.156) + (Distance to river * 0.0884) + (Slope * 0.3029) + (Soil types * 0.0027) + (SPI * 0.0007) + (STI * 0.0032) + (TWI * 0.1975) + (TRI * 0.0761).

There are five categories in Jenkin's natural break method: very high, high, moderate, low, and very low flood susceptible zones. This

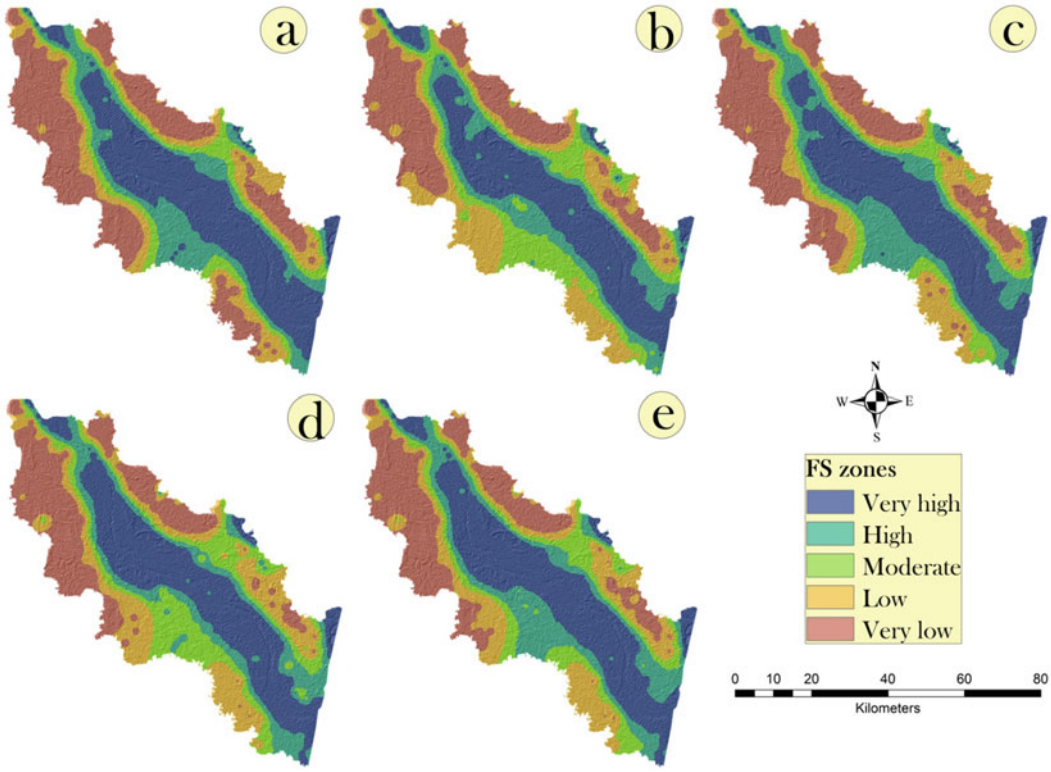

Fig. 7.4 FS models using **a** ANN, **b** RF, **c** LR, **d** ANN-LR, and **e** RF-LR

algorithm was used to classify the produced flood susceptible models (Fig. 4.4). Parallel to the drainage path of the watershed, the flood vulnerability zone runs northwest-southeast. Places with high susceptibility zones are concentrated in the south and southeast, whereas areas with low susceptibility zones are concentrated in the north and northwest of the country.

For ANN model, 2.26 and 36.69% of the entire basin area observed very "high" and "high" flood susceptibility, respectively (Table 7.2). However, all models anticipated 1023.99 km^2–1120.58 km^2 regions to be very susceptible to flooding. All models projected that 800.01 km^2–1103.01 km^2 regions were very low flood susceptible zone (Table 7.2). According to the models, most river catchment areas were classified as having high to very high FS zones. However, since the region's size varies, it is vital to choose the most accurate model.

The AUC of empirical and binormal ROC curves has been utilized to test the FS models based on the GPS locations that have been collected (Meten et al. 2015; Nahayo et al. 2019). The AUC under each ROC (empirical and binormal) for ANN, it is 0.874 and 0.912; for RF, it is 0.88 and 0.93; for LR, it is 0.861 and 0.873; for ANN-LR, it is 0.883 and 0.936; and for the RF-LR model, it is 0.89 and 0.92 (Fig. 7.5a–e). Results of ROC curves show that the ANN-LR model was the most effective, followed by the RF-LR model, the RF model, the ANN model, and the LR model. According to the binormal ROC curve, the best model was ANN-LR (AUCb: 0.936), followed by RF (AUCb: 0.93), RF-LR (AUCb: 0.92), ANN (AUCb: 0.912), and LR (AUCb: 0.912). Overall, all of the models performed well, with hybrid machine learning methods beating all of them in terms of overall performance.

Table 7.2 Calculation of area for five FS zones

FS zones	Area (km^2)				
	ANN	RF	LR	ANN-LR	RF-LR
Very high	1102.21	1023.99	1071.70	1045.91	1120.58
High	584.87	546.11	674.62	521.65	627.35
Moderate	361.61	592.62	395.50	596.45	444.38
Low	507.79	722.03	584.31	668.68	642.29
Very low	1103.01	800.01	956.86	806.01	848.96

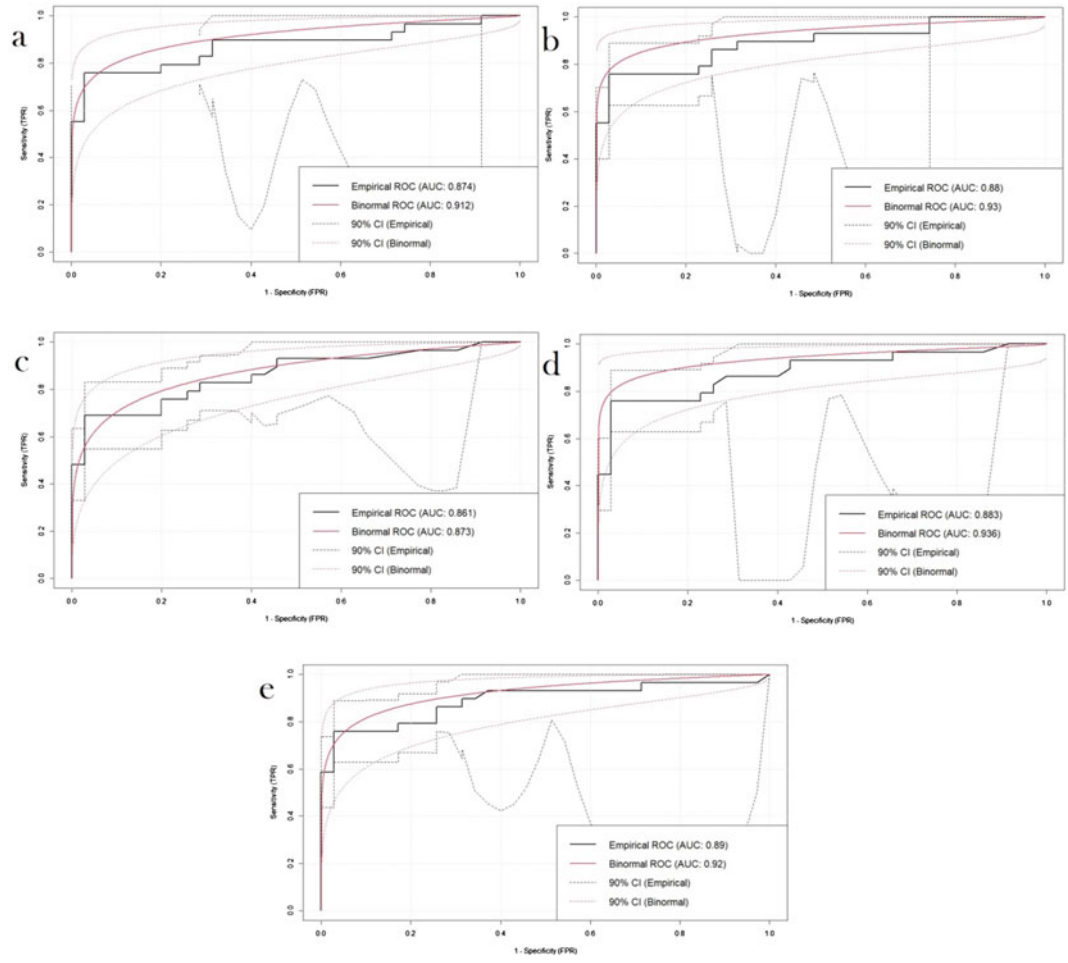

Fig. 7.5 Application of empirical and binormal ROC curve for validating the FS models based on **a** ANN, **b** RF, **c** LR, **d** ANN-LR, and **e** RF-LR

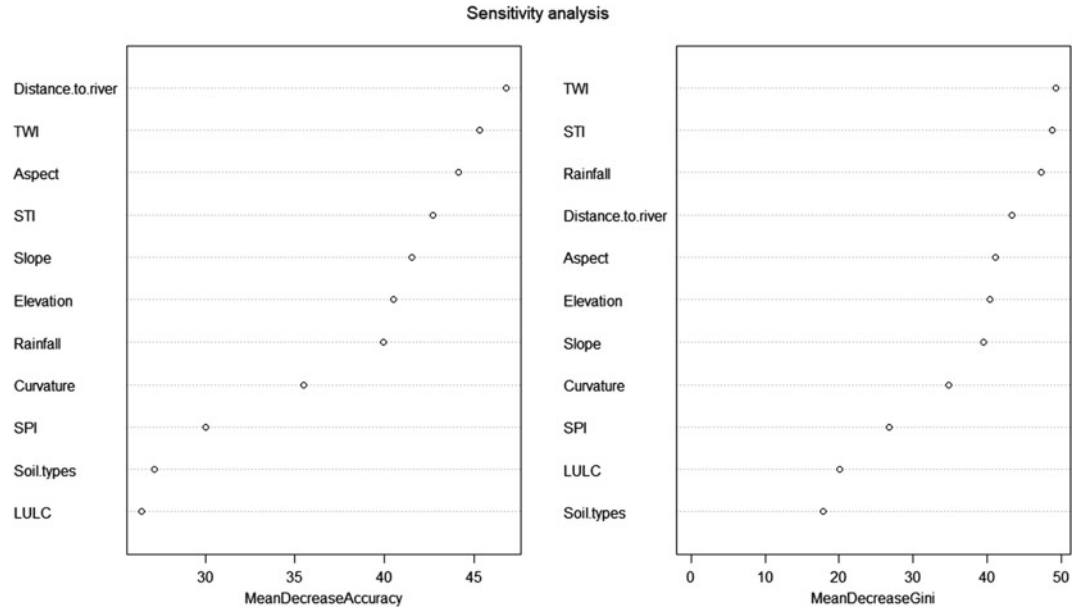

Fig. 7.6 Sensitivity analyses of susceptibility conditioning factors in terms of best FS models using **a** MDG, and **b** MDA

7.4.4 Sensitivity Analysis Using Machine Learning Algorithms

Advanced ensemble machine learning methods must be developed in order to map robust flood susceptibility zones. Dependent on the intricate mathematical association between historical floods and their conditioning factors, these algorithms can only depict the likely area around the future occurrence of a considerable and quantifiable amount of flooding. On the other hand, these models do not address the role that variables play in a given region during a flood event. Since several variables impact the possibility of flooding, the issue arises on how flood control strategies could be designed and implemented without this knowledge.

If the variables' influence on FS is uncertain, how management plans will be established and implemented. Floods may be less damaging if the variables of a particular area are identified that are associated with susceptibility zones. Consequently, it is essential to identify which factors have the most considerable impact. The significance of each triggering variable was determined using matrices from RF-based sensitivity models, such as MDG and MDA. Both the MDG and the MDA are significant metrics (Hollister et al. 2016). Results show that all the variables of the FSM model were taken into consideration, with dR, TWI, and aspect being the most important (Fig. 7.6). Among 12 variables, three of them were less influential than others in deciding the importance of the variables.

7.5 Conclusion

The present research focuses on creating hybrid models that use machine learning and statistical approaches to forecast flood susceptibility models. Three algorithms and two hybrid models were used to estimate the very high (1023–1120 km^2) and high flood susceptibility zones (521–674 km^2) flood susceptibility zones. The ANN-LR-based FSM (ROCe-AUC: 0.883; ROCb-AUC: 0.936) beat all FS models as per the AUC value. Results show that two hybrid models beat three ML and statistical models in terms of

performance. On the other hand, an RF-dependent sensitivity analysis was devised to assess the importance of the input factors for FSM. The distance to the river was the most dominating and sensitive variable for FSM, followed by slope, curvature, elevation, LULC, and SPI.

Hybrid models beat three ML and statistical models when it came to FS modeling. These findings lead the researchers to recommend adopting hybrid and ensemble algorithms for predicting natural hazards based on various factors in the future. Additionally, this study recommends that a few other hydrological, geological, and climatic variables be incorporated into the models for enhancing the robustness of the model findings. The Teesta River basin is acknowledged for flooding because of irregular rainfall, dams' construction, and other artificial issues. Therefore, these findings might be helpful in the creation of long-term flood control and farming techniques. Specifically, the paper states that land cover and vegetation conditions appeared as significant conditioning variables for FSM. Deforestation, conversely, is incontrovertible realities. Consequently, the maintenance of forest cover will benefit flood control. Further research is required for the scientific evaluation of floods at distinct susceptible zones to provide more accurate advice on what management techniques should be implemented from each possible zone.

References

Abdel-Fattah M, Saber M, Kantoush SA, Khalil MF, Sumi T, Sefelnasr AM (2017) A hydrological and geomorphometric approach to understanding the generation of wadi flash floods. Water 9(7):553

Abedi Gheshlaghi H, Feizizadeh B, Blaschke T (2020) GIS-based forest fire risk mapping using the analytical network process and fuzzy logic. J Environ Plan Manage 63(3):481–499

Aditian A, Kubota T, Shinohara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. Geomorphology 318:101–111

Adnan MSG, Abdullah AYM, Dewan A, Hall JW (2020) The effects of changing land use and flood hazard on poverty in coastal Bangladesh. Land Use Policy 99:104868

Ahmed N, Hoque MAA, Howlader N, Pradhan B (2021) Flood risk assessment: role of mitigation capacity in spatial flood risk mapping. Geocarto Int 1–23

Ali SA, Parvin F, Pham QB, Vojtek M, Vojteková J, Costache R, Linh NTT, Nguyen HQ, Ahmad A, Ghorbani MA (2020) GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naïve Bayes tree, bivariate statistics and logistic regression: a case of Topľa Basin, Slovakia. Ecol Indic 117:106620

Azareh A, Rafiei Sardooi E, Choubin B, Barkhori S, Shahdadi A, Adamowski J, Shamshirband S (2019) Incorporating multi-criteria decision-making and fuzzy-value functions for flood susceptibility assessment. Geocarto Int 1–21

Bogale A (2021) Morphometric analysis of a drainage basin using geographical information system in Gilgel Abay watershed, Lake Tana Basin, upper Blue Nile Basin Ethiopia. Appl Water Sci 11(7):1–7

Breiman, 2001. Breiman L (2001) Random forests. Machine Learn 45(1):5–32

Bui DT, Ngo PTT, Pham TD, Jaafari A, Minh NQ, Hoa PV, Samui P (2019) A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. Catena 179: pp 184–196

Cao Y, Jia H, Xiong J, Cheng W, Li K, Pang Q, Yong Z (2020) Flash flood susceptibility assessment based on geodetector, certainty factor, and logistic regression analyses in Fujian Province China. ISPRS Int J Geo-Inf 9(12):748

Chen W, Peng J, Hong H, Shahabi H, Pradhan B, Liu J, Zhu AX, Pei X, Duan Z (2018) Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. Sci Total Environ 626:1121–1135

De Silva MMGT, Kawasaki A (2018) Socioeconomic vulnerability to disaster risk: a case study of flood and drought impact in a rural Sri Lankan community. Ecol Econ 152:131–140

Dodangeh E, Choubin B, Eigdir AN, Nabipour N, Panahi M, Shamshirband S, Mosavi A (2020) Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. Sci Total Environ 705:135983

El-Haddad BA, Youssef AM, Pourghasemi HR, Pradhan B, El-Shater AH, El-Khashab MH (2021) Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin Egypt. Nat Hazards 105(1):83–114

Elmahdy SI, Mohamed MM, Ali TA, Abdalla JED, Abouleish M (2020) Land subsidence and sinkholes susceptibility mapping and analysis using random forest and frequency ratio models in Al Ain, UAE. Geocarto Int 1–17

Eyoh I, John R, De Maere G, Kayacan E (2018) Hybrid learning for interval type-2 intuitionistic fuzzy logic

systems as applied to identification and prediction problems. IEEE Trans Fuzzy Syst 26(5):2672–2685
Falah F, Rahmati O, Rostami M, Ahmadisharaf E, Daliakopoulos IN, Pourghasemi HR (2019) Artificial neural networks for flood susceptibility mapping in data-scarce urban areas. In: Spatial modeling in GIS and R for earth and environmental sciences, pp 323–336. Elsevier
Fan J, Wu L, Ma X, Zhou H, Zhang F (2020) Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. Renew Energy 145:2034–2045
Fang Z, Wang Y, Peng L, Hong H (2021) Predicting flood susceptibility using LSTM neural networks. J Hydrol 594:125734
Flügel WA (1995) Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl Germany. Hydrol Process 9(3–4):423–436
Hollister JW, Milstead WB, Kreakie BJ (2016) Modeling lake trophic state: a random forest approach.Ecosphere 7(3): p e01321
Islam ARMT, Talukdar S, Mahato S, Kundu S, Eibek KU, Pham QB, Kuriqi A, Linh NTT (2021a) Flood susceptibility modelling using advanced ensemble machine learning models. Geosci Front 12 (3):101075
Islam M, Tamanna S, Amstel AV, Noman M, Ali M, Saadat S, Aparajita DM, Roy P, Tanha SR, Sarkar N, Ashiquzzaman M (2021b) Climate change impact and comprehensive disaster management approach in Bangladesh: a review. In: Bangladesh II: climate change impacts, mitigation and adaptation in developing countries, pp1–39
Jahangir MH, Reineh SMM, Abolghasemi M (2019) Spatial predication of flood zonation mapping in Kan River Basin, Iran, using artificial neural network algorithm. Weather Clim Extrem 25:100215
Jamali B, Bach PM, Deletic A (2020) Rainwater harvesting for urban flood management–an integrated modelling framework. Water Res 171:115372
Johnson RA, (2000) Habitat segregation based on soil texture and body size in the seed-harvester antsPogonomyrmex rugosus and P barbatus. Ecol Entomology 25(4): pp 403–412
Kabir MH, Hossain T (2021) Assessment on social vulnerability and response towards natural disaster in a disaster-prone coastal village: an example from Bangladesh. Int J Disaster Manag 4(1):39–60
Khoirunisa N, Ku CY, Liu CY (2021) A GIS-based artificial neural network model for flood susceptibility assessment. Int J Environ Res Public Health 18 (3):1072
Kia MB, Pirasteh S, Pradhan B, Mahmud AR, Sulaiman WNA, Moradi A (2012) An artificial neural network model for flood simulation using GIS: Johor River Basin Malaysia. Environ Earth Sci 67(1):251–264
Lin L, Di L, Tang J, Yu E, Zhang C, Rahman M, Shrestha R, Kang L (2019) Improvement and validation of NASA/MODIS NRT global flood mapping. Remote Sens 11(2):205
Mallick J, Talukdar S, Alsubih M, Ahmed M, Islam MT, Shahfahad AR, Thanh NV (2021) Proposing receiver operating characteristic-based sensitivity analysis with introducing swarm optimized ensemble learning algorithms for groundwater potentiality modelling in Asir region, Saudi Arabia. Geocarto Int 1–28
Mahato S, Pal S (2019) Groundwater potential mapping in a rural river basin by union (OR) and intersection (AND) of four multi-criteria decision-making models. Natural Resources Research 28(2): pp 523–545
Mahato S, Pal S, Talukdar S, Saha TK Mandal P (2021) Field based index of flood vulnerability (IFV): a new validation technique for flood susceptible models. Geoscience Frontiers 12(5): p 101175
Mukherjee I, Singh UK (2020) Delineation of groundwater potential zones in a drought-prone semi-arid region of east India using GIS and analytical hierarchical process techniques. Catena 194:104681
Ogato GS, Bantider A, Abebe K, Geneletti D (2020) Geographic information system (GIS)-Based multicriteria analysis of flooding hazard and risk in Ambo Town and its watershed, West shoa zone, oromia regional State, Ethiopia. J Hydrol RegNal Stud 27:100659
Pourghasemi HR, Rossi M (eds) (2019) Natural hazards GIS-based spatial modeling using data mining techniques. Springer
Probst P, Wright MN, Boulesteix AL (2019) Hyperparameters and tuning strategies for random forest. Wiley Interdiscip Rev Data Min Knowl Discov 9(3):e1301
Rahman M, Rani L, Hossain S (2020) Factors Affecting Workers Performance
Rahman MM, Arya DS, Goel NK, Dhamy AP (2011) Design flow and stage computations in the Teesta River, Bangladesh, using frequency analysis and MIKE 11 modeling. J Hydrol Eng 16(2):176–186
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. J Environ Manage 217:1–11
Talukdar S, Singha P, Mahato S, Pal S, Liou YA, Rahman A (2020) Land-use land-cover classification by machine learning classifiers for satellite observations—a review. Remote Sens 12(7):1135
Straatsma MW, Baptist MJ (2008) Floodplain roughness parameterization using airborne laser scanning and spectral remote sensing. Remote Sensing of Environ 112(3): pp 1062–1080
Talukdar S, Eibek KU, Akhter S, Ziaul S, Islam ARMT, Mallick J (2021) Modeling fragmentationprobability of land-use and land-cover using the bagging.random forest and random subspace in the Teesta River Basin, Bangladesh. Ecol Indicators 126: p 107612

Tehrany MS, Lee MJ, Pradhan B, Jebur MN, Lee S (2014) Flood susceptibility mapping using integrated bivariate and multivariate statistical models. Environ Earth Sci 72(10):4001–4015

Uhe PF, Mitchell DM, Bates PD, Sampson CC, Smith AM, Islam AS (2019) Enhanced flood risk with 1.5° C global warming in the Ganges–Brahmaputra–Meghna basin. Environ Res Lett 14(7):074031

Wasko C, Nathan R (2019) Influence of changes in rainfall and soil moisture on trends in flooding. J Hydrol 575:432–441

Yesilnacar E, Topal TAMER (2005) Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Eng Geol 79(3–4):251–266

Yin S, Bai J, Wang W, Zhang G, Jia J, Cui B, Liu X (2019) Effects of soil moisture on carbon mineralization in floodplain wetlands with different flooding frequencies. J Hydrol 574:1074–1084

Yousefi S, Mirzaee S, Keesstra S, Surian N, Pourghasemi HR, Zakizadeh HR, Tabibian S (2018) Effects of an extreme flood on river morphology (case study: Karoon River, Iran). Geomorphology 304:30–39

Zhang Z, Yao Q, Liu KB, Li L, Yin R, Wang G, Sun J (2021) Historical flooding regime along the Amur River and its links to East Asia summer monsoon circulation. Geomorphology 388:107782