ORIGINAL PAPER

Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms

Swapan Talukdar[1] · Bonosri Ghose[2] · Shahfahad[3] · Roquia Salam[2] · Susanta Mahato[1] · Quoc Bao Pham[4,5] · Nguyen Thi Thuy Linh[6,7] · Romulus Costache[8,9] · Mohammadtaghi Avand[10]

Abstract

The flooding in Bangladesh during monsoon season is very common and frequently happens. Consequently, people have been experiencing tremendous damage to properties, infrastructures, and human casualties. Usually, floods are one of the devastating disasters from nature, but for developing nations like Bangladesh, flooding becomes worse. Due to the dynamic and complex nature of the flooding, the prediction of flooding sites was usually very difficult for flood management. But the artificial intelligence and advanced remote sensing techniques together could predict and identify the possible sites, which are vulnerable to flooding. The present work aimed to predict and identify the flooding sites or flood susceptible zones in the Teesta River basin by employing state-of-the-art novel ensemble machine learning algorithms. We developed ensembles of bagging with REPtree, random forest (RF), M5P, and random tree (RT) algorithms for obtaining reliable and highly accurate results. Twelve factors, which are considered as the conditioning factors, and 413 current and former flooding points were identified for flooding susceptibility modelling. The Information Gain ratio statistical technique was utilized to determine the influence of the factors for flooding. We applied receiver operating characteristic curve (ROC) for validation of the flood susceptible models. The Freidman test, Wilcoxon signed-rank test, Kruskal–Wallis test and Kolmogorov–Smirnov test were applied together for the first time in flood susceptibility modelling to compare the models with each other. Results showed that more than $800~km^2$ area was predicted as the very high flood susceptibility zones by all algorithms. The ROC curve showed that all models achieved more than 0.85 area under the curve indicating highly accurate flood models. For flood susceptibility modelling, the bagging with M5P performed superior, followed by bagging with RF, bagging with REPtree and bagging with RT. The methodology and solution-oriented results presented in this paper will assist the regional as well as local authorities and the policy-makers for mitigating the risks related to floods and also help in developing appropriate measures to avoid potential damages.

✉ Quoc Bao Pham
phambaoquoc@tdtu.edu.vn

1 Department of Geography, University of Gour Banga, Malda, West Bengal, India

2 Department of Disaster Management, Begum Rokeya University, Rangpur 5400, Bangladesh

3 Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia, 110025, New Delhi, India

4 Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam

5 Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam

6 Institute of Research and Development, Duy Tan University, Danang 550000, Vietnam

7 Faculty of Environmental and Chemical Engineering, Duy Tan University, Danang 550000, Vietnam

8 Research Institute of the University of Bucharest, 90-92 Sos. Panduri, 5th District, Bucharest, Romania

9 National Institute of Hydrology and Water Management, Bucureti-Ploieti Road, 97E, 1st District, 013686 Bucharest, Romania

10 Department of Watershed Management Engineering, College of Natural Resources, Tarbiat Modares University, P. O. Box 14115-111, Tehran, Iran

Graphic abstract
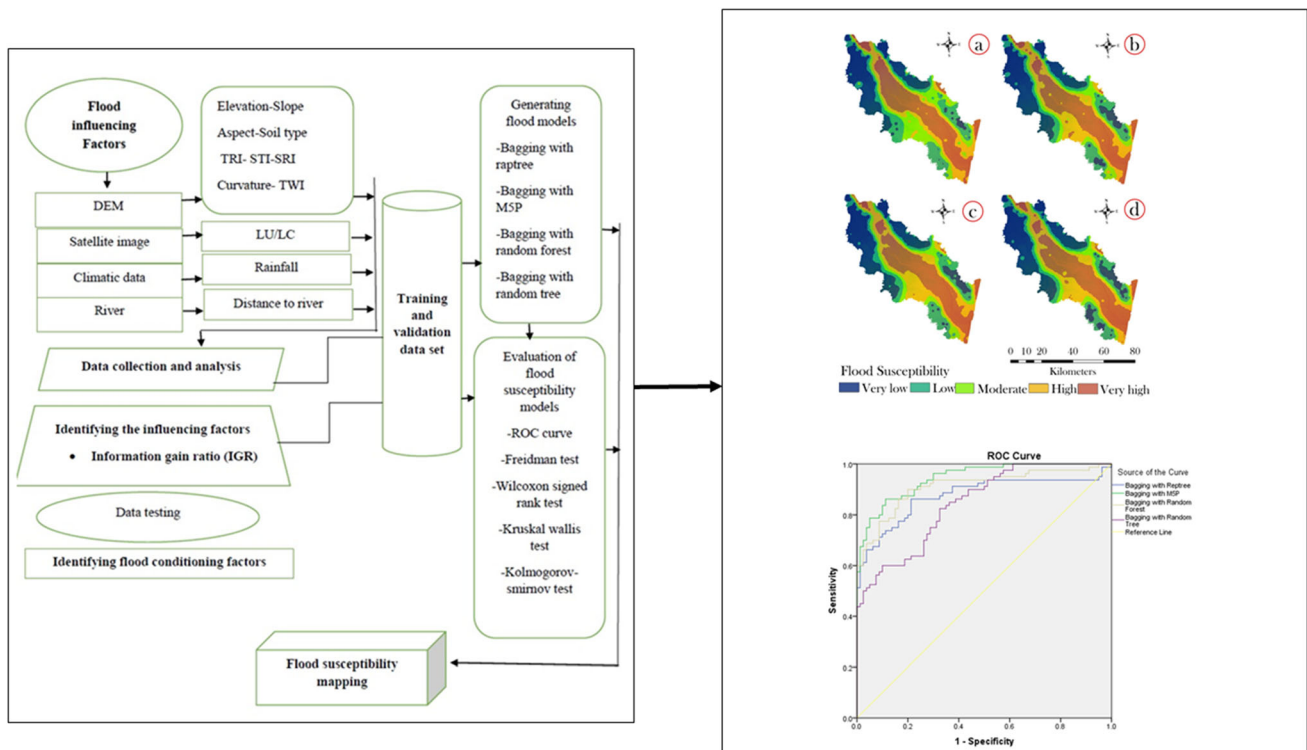
1 Introduction

During the past few decades, the frequencies of extreme weather events and related disasters have increased due to incessant changing climate and global warming (Hettiarachchi et al. 2018; Nam et al. 2015). Evidence suggests that the extreme weather events have occurred more frequently after the mid 20th century and have occurred in that region which does not have such in the history (Hoeppe 2016). Among all, the occurrences of flooding are maximum across the globe and causes higher damage than other natural disasters (Yang et al. 2018; Hirabayashi et al. 2013). It has been estimated that from 1995 to 2015, flood hazard affected more than 100 million people and the damages estimated at 75 billion USD in every year (Mohanty et al. 2020; Alfieri et al. 2017). In the countries like Bangladesh, the casualty due to flood is higher than any other natural calamities during recent past (Dewan 2015; Azad et al. 2013) as a large section of populations lives in the floodplains under the varying degree of vulnerability to

flooding, river erosion etc. (Tingsanchali and Karim 2005). Many parts of Bangladesh have experienced a number of devastating flood events during past decades which has caused huge loss of both property and lives (Ferdous et al. 2019). In northern Bangladesh, flash floods are periodic events that often happen in the downstream riparian areas, especially in the lower Teesta River basin. For instance, a very recent flash flood observed in the lower Teesta River basin, especially in Nilphamari, Lalmonirhat, and Kurigram districts in August-2017, which caused the submergence of considerable landmass, and five people were swept away. According to the network for information, response and preparedness activities on disaster (NIRAPAD), about 6.8 million people and over 560,000 hectares of croplands were badly damaged by the flash flood. FAO (2017) estimated that the destruction of the flash flood of August 2017 caused the damages of up to US$ 10 million (FAO 2017). Although natural along with anthropogenic activities are resulting in the flooding, the climate change has been identified as the principal cause behind the

manifestation of flooding in the world, which affects the pattern, intensity and magnitude of floods (Hettiarachchi et al. 2018; Zhao et al. 2018; Gill and Malamud 2017; de Kraker 2015; Taubenböck et al. 2011). Some natural factors such as elevation, soil texture, drainage density, distance, vegetation etc. act as the prompting factors of flood in different parts of the world (Azareh et al. 2019; Hosseini et al. 2020). The occurrences of flooding are natural, which cannot be stopped, however the damages done by floods can be mitigated by appropriate planning and management (Abebe et al. 2019). Hence, the prediction and delineation of flood prone areas is an important aspect of alleviation of flood hazards, which reduces the fatalities due to the flooding (Pyatkova et al. 2019; Sarhadi et al. 2012). Further, a study by Ward et al. (2013) pointed that the South and South-east Asia, especially India and Bangladesh has the highest share of population and GDP exposed to flood risk. Therefore, to cope with such looming conditions, the knowledge of vulnerability through the quantification and identification of spatiotemporal characteristics of flood prone areas are indispensable for the effective management and mitigation of flood hazards (Mohanty et al. 2020). Therefore, a flood susceptible model (FSM) is a useful tool which is required to identify regions at risk and to safeguard these high-risk regions and natural resources as well (Maaks et al. 2020).

The development of flood susceptibility maps is very challenging and difficult as several factors are involved. These factors are heterogeneous and very complex in nature (Ardıçlıoğlu and Kuriqi 2019; Kuriqi et al. 2020; Costache and Bui 2019). However, of late, regional data with very detailed information can be obtained from satellite images or remote sensing databases (Pourghasemi et al. 2020a; Nikolaos et al. 2019; Li et al. 2019; Talukdar and Pal 2017). Nowadays very high resolution data like synthetic aperture radar (SAR) and optical sensor images are available in some places, which can highly improve the flood susceptible maps (Bui et al. 2020a; Talha et al. 2019; Arora et al. 2019). However, these state-of-art techniques can handle spatial datasets and produce high resolution and prediction performances (Uthayakumar et al. 2020; Abba et al. 2020; Ma et al. 2020). Therefore, the integration of remote sensing databases and GIS technology have been widely applied to study the relationships between these factors and the occurrences of flood hazards (Choubin et al. 2019; Jahangir et al. 2019) and made the flood susceptible models less challenging and highly accurate. Consequently, researchers have used these technologies for predicting the natural hazards including flood susceptible models (Bui et al. 2020a, b; Wang et al. 2020; Pourghasemi et al. 2020a, b; Chen et al. 2020; Dodangeh et al. 2020).

Many scholars have developed and utilized several types of models and algorithms for preparing flood susceptible models (Sahana et al. 2020). Therefore, based on the previous literature, the models, have been used for preparing flood susceptibility maps, can be several types (Siahkamari et al. 2018; Chen et al. 2019a; Termeh et al. 2018; Hong et al. 2018a; Bui et al. 2019a; Mahmood et al. 2019), such as (1) expert knowledge based FSMs like analytical hierarchy process (AHP) (Costache et al. 2020a; Dano et al. 2020; Nachappa et al. 2020; Souissi et al. 2019), (2) bivariate and statistical based models, such as weights-of-evidence (Chen et al. 2019b; Paul et al. 2019), fuzzy logic (Wang et al. 2019; Sahana and Patel 2019), information value (Xu et al. 2013; Chen et al. 2014), frequency ratio (Chen et al. 2020a; Moghaddam et al. 2019; Khosravi et al. 2019a, Sahana et al. 2020), logistic regression (Tien Bui et al. 2019a; Shafapour Tehrany et al. 2019a, b; Pham et al. 2020a; Ali et al. 2020), analytical network process (Ali et al. 2020; Akay and Koçyiğit 2020), certainty factor (Costache et al. 2020a), neuro fuzzy logic (Termeh et al. 2018; Hong et al. 2018b) (3) machine learning algorithms (Shahabi et al. 2020; Dodangeh et al. 2020; Wang et al. 2020; Costache et al. 2020b; Costache and Bui 2020; Tang et al. 2020), and (4) hydrological models such as soil water assessment tool (SWAT) (Oeurng et al. 2011; Busico et al. 2020; Uniyal et al. 2020; Bhattacharya et al. 2020) and Hydraulic Engineering Centre-River Analysis System among others (Getahun and Gebre 2015; Joshi and Shahapure 2020; Huţanu et al. 2020). Recently, machine learning techniques have drawn more and more attention, which have been employed in FSMs by many researchers (Bui et al. 2020a; Chen et al. 2019a; Hong et al. 2018a; Wang et al. 2020). The most popular machine learning techniques are artificial neural networks (Falah et al. 2019; Moghaddam et al. 2019; Pham et al. 2020b; Bui et al. 2020b), random forest (Avand et al. 2019; Paul et al. 2019; Achour and Pourghasemi 2020; Chen et al. 2020b; Nhu et al. 2020; Vafakhah et al. 2020), support vector machines (Termeh et al. 2018; Khosravi et al. 2019a), and decision trees (Choubin et al. 2019; Moghaddam et al. 2019; Yariyan et al. 2020; Nhu et al. 2020; Chen et al. 2020c; Costache et al. 2020c), radial basis function (Choubin et al. 2019), which predict the areas at risk of flooding very accurately. However, FSM experiences many challenges, such as selecting the proper methods for modelling among vast numbers of methods, and each method produces different results (Costache et al. 2020a, b; Shafizadeh-Moghadam et al. 2018). Even, each of these models have some drawbacks for predicting the FSMs. Therefore, very recently, to overcome these limitations of several algorithms, researchers have been applied hybrid ensemble machine learning algorithms, which have shown better performance than the conventional and single models (Pham et al. 2016, 2017a; Wang et al. 2020; Shahabi et al. 2020; Nachappa et al. 2020; Costache et al. 2020a; Costache and Bui 2020). The widely applied and popular ensemble

machine learning algorithms, which showed very high accuracy to prepare flood susceptibility models, are random subspace (Pham et al. 2020b; Chen et al. 2019a), Reptree (Chen et al. 2019a; Ghasemain et al. 2020), bagging (Shahabi et al. 2020; Chen et al. 2019b; Yariyan et al. 2020), naive Bayes (Ali et al. 2020; Pham et al. 2020c; Tang et al. 2020), logistic tree (Chapi et al. 2019), ensemble of bootstrapping (Dodangeh et al. 2020), ensemble of boosted generalized linear model (Hosseini et al. 2020). The outstanding findings of the hybrid machine learning algorithms for several natural hazard's models inspire researchers to apply and develop the hybrid machine learning algorithms. However, no general agreement has been found on the selection of the best method for different types of natural hazards modelling such as landslide or flood susceptibility (Chen et al. 2019b). Researchers recommend developing and testing new models for flood susceptibility mapping and other kinds of natural hazards modelling (Chen et al. 2019a).

The flood hazards and susceptibility mapping is not a new area of research in Bangladesh and researches have been already carried out (Hoque et al. 2019; Islam and Sado 2000). Further, the studies have been carried out in the upper Teesta River basin (Indian part of Teesta river) for flood susceptibility analysis (Roy et al. 2019; Mandal and Chakarbarty 2016), but no such comprehensive study for flood susceptibility analysis has been carried out in the Lower Teesta River basin of Bangladesh. To fill these gaps in research, this study designed to develop the new ensemble of bagging algorithms by integrating four other machine learning algorithms, which have not applied till date, for deriving the highly accurate prediction of the susceptibility to flood hazards in the Lower Teesta River basin of Bangladesh. The predicted models comparisons are highly recommendable work for exploring the performances of each model. For this, we used Kruskal–Wallis test and Kolmogorov–Smirnov test to compare the models with each other. Although we used two commonly used non-parametric tests, such as the Friedman test and Wilcoxon Signed-Rank test for model comparison. Based on this line of thinking, we set main objectives of this research were (1) to develop new ensembles of bagging algorithms for flood susceptibility analysis, (2) to delineate and prepare the flood susceptible zones in Lower Teesta River basin, and (3) to validate and compare the flood susceptibility models of the Lower Teesta River basin.

2 Study area

Lower Teesta River basin, which situated in the northern part of Bangladesh has been selected as the study area for this research which is located between 25° 30′ 02″ N and 26° 18′ 37″ N latitudes and 88° 52′ 58″ E and 89° 45′ 34″ E longitude (Fig. 1). The basin covers about 2284 km^2 area and includes five districts of Bangladesh namely Lalmanirhat, Nilphamary, Rangpur, Kurigram and Gaibandha. As per the 2011 Census, the total population of the area was about 10.42 million. The average elevation of the region varies between 05 and 100 meters and the slope is from north-west to south-east. The climate of the basin is subtropical monsoon type (*Koppen: CWA*) with highest temperature goes beyond 40° C during May, while lowest climate hardly goes below 15 °C during December. The mean annual rainfall of the region is more than 250 cm in which more than 80 per cent of the total rainfall occurs during monsoon season (June to September).

The flood-plain of the region is made up of the Teesta and several other small and medium sized rivers. This river deposits sediments each year during flooding period and makes its plain fertile and favorable for agriculture (Mandal and Chakarbarty 2016). The morphology of the lower Teesta basin is demarcated by the low depressions as well as the moribund river channel valley formed by long morphological alters in the basin pathways. Hence the basin is susceptible to flooding and flash flood is a common phenomenon and occurs each year during monsoon season. The sediments deposited are mostly recent making the surface a fertile alluvial plain which are composed of clay, silt and fine to medium sized alluvium (Saha et al. 2019).

3 Materials and methods

3.1 Materials and databases

As the study area has experienced frequent flooding each year; therefore, based on the field survey and local people perception, the historical flooding inventories were prepared. In the present study, we obtained several data types for flood susceptibility modelling. We obtained Landsat 8 operational land imager (OLI) for preparing land use land cover (LULC) maps (path/row: 138/42, spatial resolution: 30 m, date: 19/03/2019), which was downloaded from the United States Geological Survey (USGS) website. For deriving topographical factors and hydrological factors, we used ASTER GDEM (Version 2) (spatial resolution: 30 m). The rainfall data were collected from the Bangladesh Meteorological Department (BMC), Dhaka, Bangladesh. We used a soil taxonomy map, which was collected from the Natural Resources Conservation Service (NRCS)-United States Department of Agriculture (USDA). The drainage map was prepared by utilizing the topographic maps with a scale of 1:250,00 obtained from the Bangladesh water development board. A detailed procedure of methodology of this study is presented by Fig. 2.
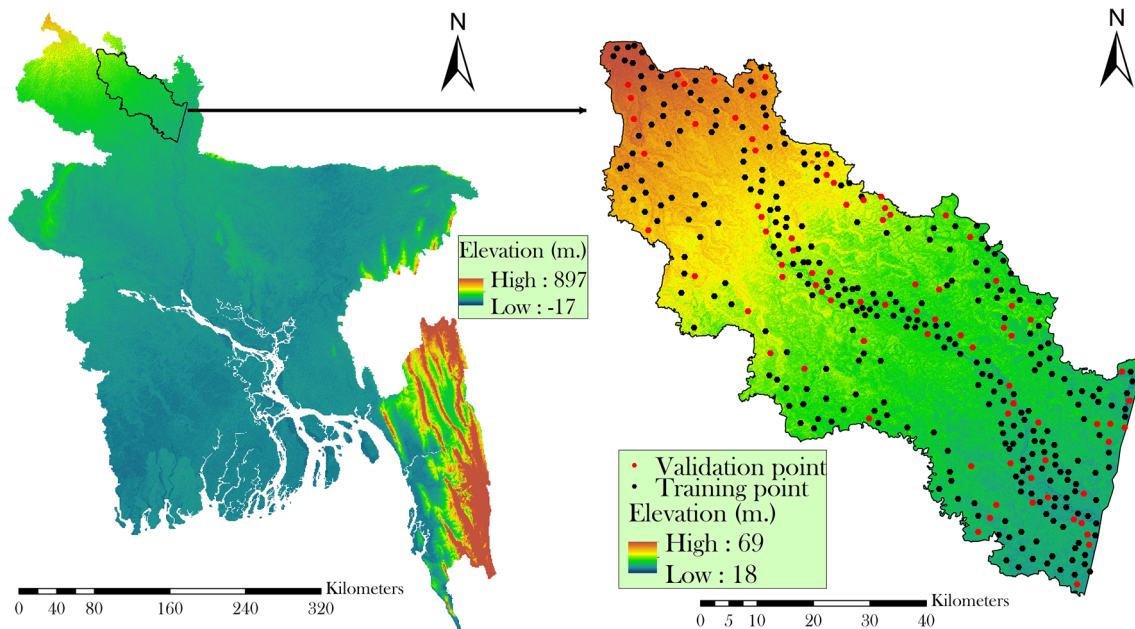
Fig. 1 The location of the study area having the training and validation flood points

3.2 Flood inventory

The primary step for preparing the flood susceptibility map is the creating of flood inventory maps of the study basin because the probable flood susceptible zones are predicted based on the mathematical relationship between the past flood events and its influencing factors (Bui et al. 2020a; Sarkar and Mondal 2020). However, for collecting the past common flood points in the present study area, we used the historical inundation maps, topographical map and survey on the perception of local people. We collected 207 flood points from the study area (Figs. 1, 3). Subsequently, the collected 207 flooded points were randomly partitioned into 80% (165 points) and 20% (42 points) groups to build and validate the flood susceptible models. However, flood susceptibility mapping is considered as the binary classification in which flood inventory has been classified into two classes, such as flood points and non-flood points. Therefore, in order to construct the training flood inventory, which is considered as the dependent factor for building the model, the binary values like 1 as flood points and 0 as non-flood points are required. Here, flood points have been considered as the exact points where frequent floods have been observed, while the non-flood points considered the points where floods were not recorded in the last few years. Similar to flood points, we had to obtain negative samples or non-flood points. To avoid bias, several researchers recommended choosing the similar number of non-flood points as the positive or flood samples (Tang et al. 2020). Therefore, we randomly collected 206 non-flood points based on the topographical map, historical

flood data, field survey and NDWI maps. Subsequently, we randomly classified the non-flood points into 80% (165 points) and 20% (41 points) groups. Thus, we prepared a dependent factor as training datasets which comprised 165 flood points as 1 and 165 non-flood points as 0.

Similarly, we also prepared testing datasets for evaluating the final models, which comprised 42 flood points as 1 and 41 non-flood points as 0. Both the training and validation datasets were shown in Fig. 1. We extracted data from twelve flood conditioning parameters (spatial datasets) based on the training datasets by using the 'extract values to point' tools in ArcGIS 10.5 software. Subsequently, we imported these datasets into WEKA (version 3.9.3) software, and the whole modelling was done over there.

3.3 Methods for generating the flood conditioning parameters

The flood susceptible model is usually very complex and comprehensive, as it requires several topographical and hydrological factors in geospatial format. We selected and prepared twelve flood conditioning parameters for the present work based on the previous literature on flood susceptibility modelling (Chen et al. 2019b; Sturzenegger et al. 2019; Bui et al. 2019b; Arabameri et al. 2019; Moghaddam et al. 2019; Paul et al. 2019; Janizadeh et al. 2019). The parameters were elevation, curvature, aspect, slope, topographic roughness index (TRI), topographic wetness index (TWI), stream power index (SPI), sediment transport index (STI), LULC, distance to the river, soil
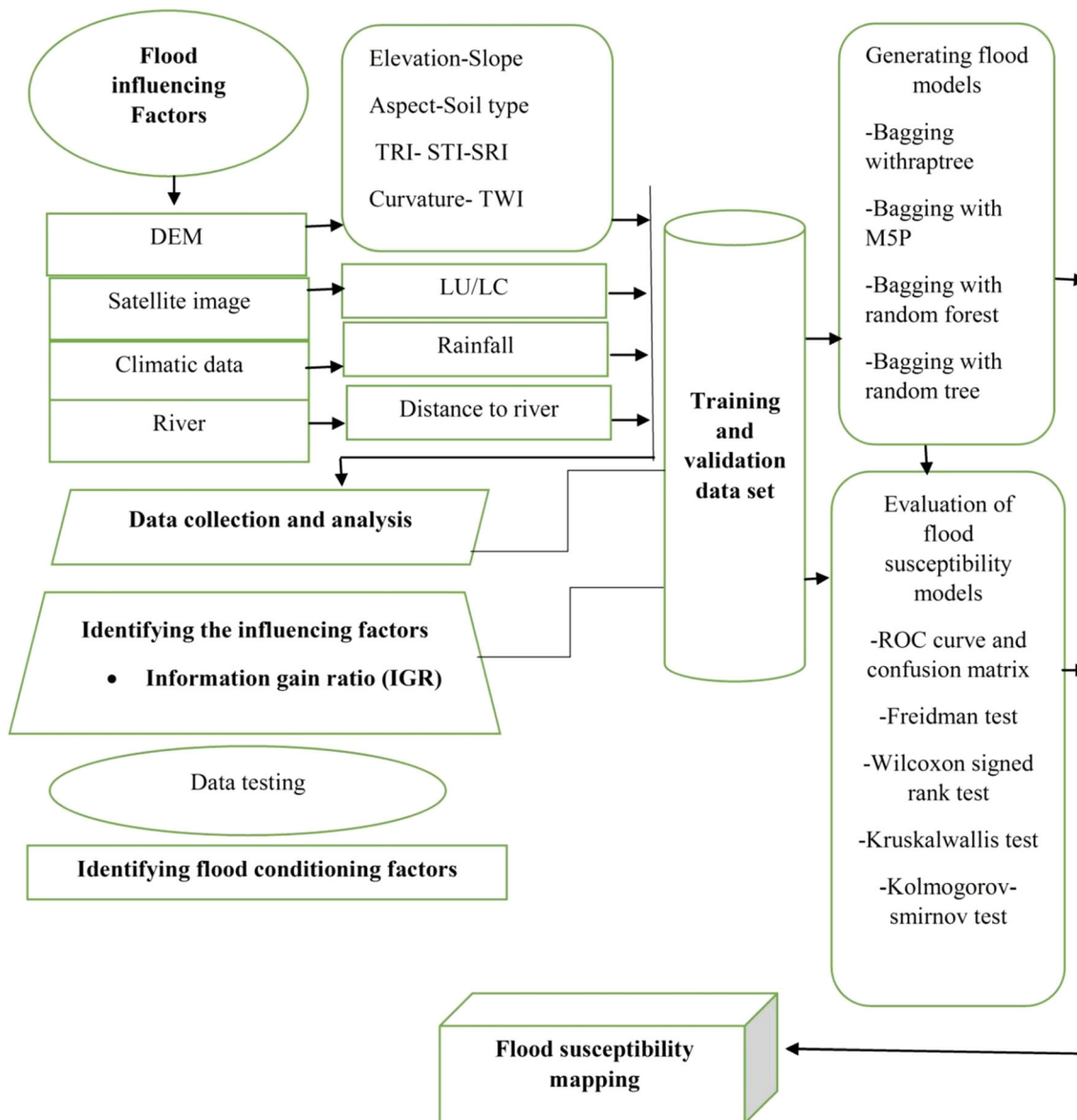
Fig. 2 Methodology flowchart of this study

type, and rainfall. As the collected parameters had different spatial resolution, we applied resampling technique to make them uniform (30 m spatial resolution). The details procedure for preparing the flood conditioning parameters were discussed as follows:

3.3.1 Elevation

The elevation has been identified as a major dominant factor for the modeling of flood (Choubin et al. 2019; Bui et al. 2016). The flood frequency and elevation are inversely related to each other, as one of them (flood) decreases with increase in another (elevation) and vice versa. The areas with low elevation are supposed to be more

susceptible to floods while the areas with higher elevation are supposed to be less susceptible to the floods (Khosravi et al. 2016a, b). The Teesta River basin is prone to the frequent flooding as it is located in a low elevation area with a flat topography.

3.3.2 Aspect

Maximum slope of the surface in a definite direction is known as aspect. Several studies considered aspect as significant parameters in the modeling of flood susceptibility (Bui et al. 2020; Bui et al. 2016; Chen et al. 2019). It determines the direction of flow of flood water and hence is an important parameter for flood study (Costache 2019a;

Fig. 3 Field photographs of the flooding situation in the Teesta river basin representing **a**, **c**, **d** flooded road, **b** damaged houses due to flooding, **e** flooded village, **f**, **h** destroyed road due to devastating flooding, **g** overflow on the culvert, and **i** camping on the national road by the people affected by flooding

Lei et al. 2020). The aspect has been prepared using the ASTER GDEM data.

3.3.3 Slope

Slope gradient is an important physiographic characteristic, which is directly related to the flooding as it contributes to the runoff velocity and vertical percolation of the water (Choubin et al. 2019; Rahmati et al. 2016). The chances of flooding increases with decline in the slope angle and decreases with an increase in slope angle(Costache 2019b). Therefore, the Teesta River basin is supposed to be more prone to flooding as it has flat topography with low elevation.

3.3.4 Curvature

The curvature is another determinant in flood susceptibility modeling which is prepared using ASTER GDEM in ArcGIS 10.2 domain. The convergent and divergent runoff regions were separated by the curvature. The activity of runoff is associated with the regions with negative value (Costache and Bui 2020) and these regions are highly susceptible to flooding.

3.3.5 Topographic roughness index (TRI)

Flooding also occurred by TRI which depends on the local topography of a basin. The occurrence of flooding highly depends on the TRI and the higher floods are always associated with lower TRI and the high TRI leads to either no or low floods (Tehrany et al. 2019a, b). In this research, stretch format was used for preparing TRI map with values ranging between 0 and 27 (Fig. 6a).

3.3.6 Topographic wetness index (TWI)

TWI is the indication of watersheds' wetness by spatial variation first proposed by Beven and Kirkby (1979). It is used to spatially represent the variation of wetness of a river basin (Meleset al. 2020). The TWI shows the quantity of water accumulated in a pixel size of a watershed area or basin and can be expressed as Eq. 1.

$$TWI = \frac{\ln(A_s)}{\tan \beta} \tag{1}$$

where, A_s represents the explicit catchment area ($m^2 \ m^{-1}$) and β represents the slope gradient (in degrees). The higher TWI values and the flood events have strong correlation with each other (Shit et al. 2020). In this research, the TWI value ranges between 0 and 7.72 (Fig. 6b).

3.3.7 Stream power index (SPI)

The SPI refers to the power of erosion (erosive power) of the flowing water and it has considerable impact on the fluvial systems (Tehrany et al. 2015). The sediment transportation capacity and erodibility of a river from its own bed is known as the SPI (Chen et al. 2020). The SPI was calculated using Eq. 2.

$$SPI = A_s \tan \beta \qquad (2)$$

where, A_s represents the specific catchment area and β represents the slope gradient.

3.3.8 Sediment transport index (STI)

Erosion as well as deposition processes in a specific basin are described by STI. It is used to reflect the erosive capacity of a surface/terrain and can be calculated using Eq. (3).

$$STI = \left(\frac{A_s}{22.13}\right)^{0.6} \left(\frac{\sin \beta}{0.0896}\right)^{1.3} \qquad (3)$$

where, β refers to the slope pixels while the A_s refers to the upstream area. The STI was calculated based on the geomorphologic as well as hydro-climatic attributes. The change in the channel's bed due to sediment deposition affects the water storing capability of the basin which leads to the increase in flood risk (Antoniazza et al. 2019). In the present research, the STI value ranges between 0 and 140.64 (Fig. 6d).

3.3.9 Land use land cover

The LULC affects the surface runoff and sediment transportation both directly and indirectly (Zhang et al. 2010). The flood events are more frequent in settlement areas than the forest and open areas because the built-up lands do not allow water to infiltrate and block surface runoff (Costache 2019c), while the forested and open surfaces does not put obstacles in the movement of water (Yin et al. 2017). In this research, the LULC mapping has been done using the Landsat 8 (OLI) dataset using ANN algorithm on ENVI software version 5.3. Six LULC classes have been identified in this study, i.e. built-up area, vegetation cover, bare land, agricultural land, sand bar and water body (Fig. 7a).

3.3.10 Distance to river

The areas next to the rivers are most exposed to the flooding and hence, the distance from the river is identified as an important flood conditioning factor. Chances of flooding increases with the decrease in distance from the river and decreases with increase in distance from the river

(Talukdar and Pal. 2019; Costache et al. 2020d; Binh et al. 2020). Topographic maps (scale 1:50,000) and Google Earth were used to prepare the distance to the river map.

3.3.11 Soil types

Flooding is also affected by soil as the properties soil determines the infiltration of water and surface runoff (Costache et al. 2019; Phillips et al. 2019). The infiltration rate and surface runoff are inversely related to the flooding. The soil map has been classified into 12 classes based on the USDA soil taxonomy classification using the USDA map as the base map.

3.3.12 Rainfall

Rainfall has been also identified as a major influencing factor of flooding as the intense rainfall for even a short time-period can cause flooding (Pham et al. 2019a, b; Ali et al. 2020; Costache et al. 2020e; Pourghasemi et al. 2020a, b). Data of rainfall was sourced from Bangladesh Meteorological Department and the spatial distribution of rainfall done by the well known interpolation technique kriging in ArcGIS version 10.3. The kriging method was employed because the rainfall data obtained was from only four meteorological stations and this technique has been suggested to plot less number of observations (Kourgialas and Karatzas 2011).

3.4 Methods for analyzing importance of flood conditioning parameters

Several spatial techniques as well as models have been proposed and applied for the mapping of flood susceptibility modeling and hazard zonation in order to delineate the flood prone areas. The preparation of flood hazard models involves the building of a set of parameters related to floods (Chen et al. 2019). The flood conditioning factors are used to enhance and increase the quality of the results. Total 12 factors have been used in this study as the flood conditioning factors; i.e. aspect, slope, curvature, stream power index (SPI), elevation, sediment transport index (STI), topographic roughness index (TRI), topographic wetness index (TWI), LULC, type of soil, distance to river, and rainfall. Further, to identify the parameters influencing the prediction of flood susceptibility modeling, the information gain ratio (IGR) has been used because of its ability to identify the significance of each factor influencing flood susceptibility modeling (Bui et al. 2020a, b; Al-Abadi 2018). The IGR values have been assigned based on the significance of the factor. The IGR was employed in the present study because of its efficacy and was calculated using Eq. (4).

$$\text{Gain ratio}(x, Z) = \frac{\text{Entropy }(Z) - \sum_1^n \sum_{i=1}^n \frac{|Z_i|}{|Z|} \text{Entropy }(Z_i)}{-\sum_{i=1}^n \frac{|Z_i|}{|Z|} \log \frac{|Z_i|}{|Z|}} \tag{4}$$

Further, identification of the importance of the factors responsible for flooding has been done by utilization of the Karl Pearson's correlation coefficient used by Xu and Li (2020) and the variance inflation factors (VIF) used by Javidan et al. (2020) techniques in this study. A VIF value more than 9 and the very low correlation coefficient shows the problem of multicollinearity in the factors employed. Therefore, it is recommended to exclude those conditioning factors with VIF more than 9 or very low coefficient of correlation in the modeling.

3.5 Methods for flood susceptibility mapping

3.5.1 Bagging

The bagging is a popular technique used for the construction of ensembles (Prasad et al. 2006). Bagging refers to an ensemble algorithm, which can constitute multiple models of different subsets of a training dataset. It combines the prophecy from all models. It is the application of bootstrapping and aggregating procedure to a high variance machine learning technique and was called Bagging by an American Statistician Breiman (1996).

For this study, a learning set C has been considered, which consists of n independent observations (Chen et al. 2018). Here, the independent observations are the flood conditioning factors, where $C = \{(X_i, Y_i), i = 1, 2, 3 \dots n\}$. For this, firstly, the set C_b ($b = 1, 2, 3 \dots n$) represents the bth bootstrap sample of training set C, acquired by illustration with substitution n components of the C. Later, to calculate the bootstrap estimator $g * (\cdot)$ by the plug in the code: $g * (\cdot) = h_n ((X_1, Y_1), \dots (X_n, Y_n)) (\cdot)$. Finally at last, replicate the above mentioned steps m times, in which the m could be either 50 or 100, based on the need, yielding $g^{*k}(\cdot)$ ($k = 1, 2, 3 \dots m$). Hence, the Bagging calculator will be as Eq. 5

$$gBag(\cdot) = \frac{\sum_{k=1}^m g^{*k}()}{m} \tag{5}$$

Further, the Bagging estimator can be illustrated as Eq. (6).

$$gBag(\cdot) = *[g * (\cdot)] \tag{6}$$

where the speculative quantity matches to $m = \infty$ and this infinite number m *directs* the precision of Monte Carlo estimation.

3.5.2 Ensembles of bagging

REPTree

The REPTree algorithm follows the idea of computing the information gain with entropy and minimizing the error occurring due to the variance (Witten and Frank 2005). Suggested by Quinlan (1987), this algorithm produces the regression tree by means of node statistics like information gain or the variance diminution calculated from the up-down phase, and trims it by using reduced-error cutting.

M5P

M5P is a tree based regression algorithm proposed by Quinlan (1992) which produced values at the trees' leaves for future prediction. The trees produced by this algorithm have some multivariate linear techniques. This algorithm can solve problems with high dimensionality equal to 100 characters. It is efficient and gives more accurate results by building comparatively smaller trees. This model works with continuous variables rather than discrete variables (Sihag et al. 2019).

Random forest

Random forest is a well admired ensemble learning algorithm proposed by Breiman (2001), which is a permutation of the decision trees for the classification as well as regression for making predictions. It is a combination of two subsets, i.e. bagging idea of Breimanand the random selection features of Ho. In a bagging ensemble, poor classifiers can give high accuracy by producing a number of strong classifiers with Random forest. A wide variety of samples were created along with generating various similar regression trees in the training phase by this ensemble. Then, based on the results of multiple classifiers, it classified the data. Lastly, it selects the classification, which has a majority vote over all trees in a forest.

Random tree

The random tree is also known as RTree, a regression model based on a decision tree algorithm. The trees are created by RTree considering randomly chosen attributes (K) at every node without pruning. Further, it gives an alternative to allocate the evaluation of the class probabilities on the basis of a hold-out set, i.e. Back-fitting.

3.6 Validation and comparisons of flood susceptibility models

3.6.1 Receiver operating curve

Receiver operating characteristics (ROC) curve is the graph of sensitivity along with 1-specificity, which produces an area under itself called AUC (Hajian-Tilaki

2013). AUC is very useful which helps to evaluate the performances of prediction accuracy and also for interpreting the results. ROC is also utilized for assessing the risk of any vulnerable situation or object. ROC curve is expressed by the following formula:

$$ROC = \frac{Sensitivity}{Specificity} \tag{7}$$

3.6.2 Confusion matrix

In the present study, for evaluating the accuracy assessment of the four flood susceptibility maps, we calculated confusion matrix apart from the ROC curve. We calculated sensitivity, specificity, Youden index J, predicted positive value, predicted negative value, and optimal criterion for validating the flood susceptible models (for details: Hong et al. 2020).

3.6.3 Friedman test

Friedman test is an ideal nonparametric test used for comparing several matching groups among them developed by Milton Friedman (Lindman 1974). It is a two way analysis. This test presumes that the sources of all the variables having equivalent continuous distribution and all variables are communally self-determining (Cieslak and Chawla 2009). The Friedman test is described by the following equation:

$$X^2 = \frac{12}{kn(k+1)} \sum_{j=1}^{k} r_j^2 - 3n(k+1) \tag{8}$$

where, X^2 is the probable p value, k is number of variables, n denotes number of examples under each variable and r denotes the rank.

3.6.4 Wilcoxon signed-rank test

Wilcoxon Signed-Rank test is the nonparametric technique for testing the variations of paired data based on ordinal scale (Suchmacher and Geller 2012). It is well known as the backup test of t test in which self-determining variables are binary based. This test is employed for detecting whether any variable is shifted by the influence of other variables. Four main steps should have been done for doing this test. First, calculate the variation of every pair of datasets. Secondly, rank the derived variation. Thirdly, assigning the respective sign (+ or − sign) of the ranked values. Finally, both sums (sums of positive sign and sums of negative sign) are computed.

3.6.5 Kruskal–Wallis test

Kruskal–Wallis test is the one way nonparametric test for evaluating the performance of several similar groups among them (Gibbons 1985). This test is recognized as a very useful test for performance evaluation. The statistical form of Kruskal–Wallis test is following below:

$$K = (N - 1) \frac{\sum_{i=1}^{g} n_i (\underline{r_i} - \underline{r})^2}{\sum_{i=1}^{g} \sum_{j=1}^{n=i} (r_{ij} - \underline{r})^2} \tag{9}$$

where, N means total samples, n_g denotes number of total samples in g group, r_{jg} denotes overall rank of j samples in group and r_g means rank of samples in g group and $\underline{r}$ denotes mean rank of samples among all samples.

3.6.6 Kolmogorov–Smirnov test

Kolmogorov–Smirnov test also known as KS test is a common nonparametric test which compares two observations on the basis of performances (Kolmogorov 1933; Smirnov 1939). It is a less sensitive model compared to other models, because it produces no assumptions on data distribution. Statistical equation of this test is noted below:

$$X^2 = \frac{4D^2 n_1 n_2}{n_1 + n_2} \tag{10}$$

where, X gives the values of p, n_1 and n_2 is the number of examples in two different observations.

4 Results and analysis

4.1 Importance of flood conditioning parameters

The determination of the influence of flood conditioning factors were evaluated by using the values of IGR for each parameter. It was calculated by using a tenfold cross validation technique. Figure 4 showed that the LULC (0.52), slope (0.495), DR (0.12) and elevation (0.11) were the most important flood conditioning factors with higher IGR values than TRI (0.03), SPI (0.03), STI (0.02), TWI (0.01), and curvature (0.01). Further, the IGR value of aspect factor was zero (0), hence it could be considered as the less influential parameter for flooding.

4.2 Characteristics of flood conditioning parameters

To explore the spatial relationship between natural hazard occurrences and influencing factors is significantly needed in modeling study (Pham et al. 2016). Flooding

Fig. 4 Determination of influence of flood conditioning parameters by using information gain ratio
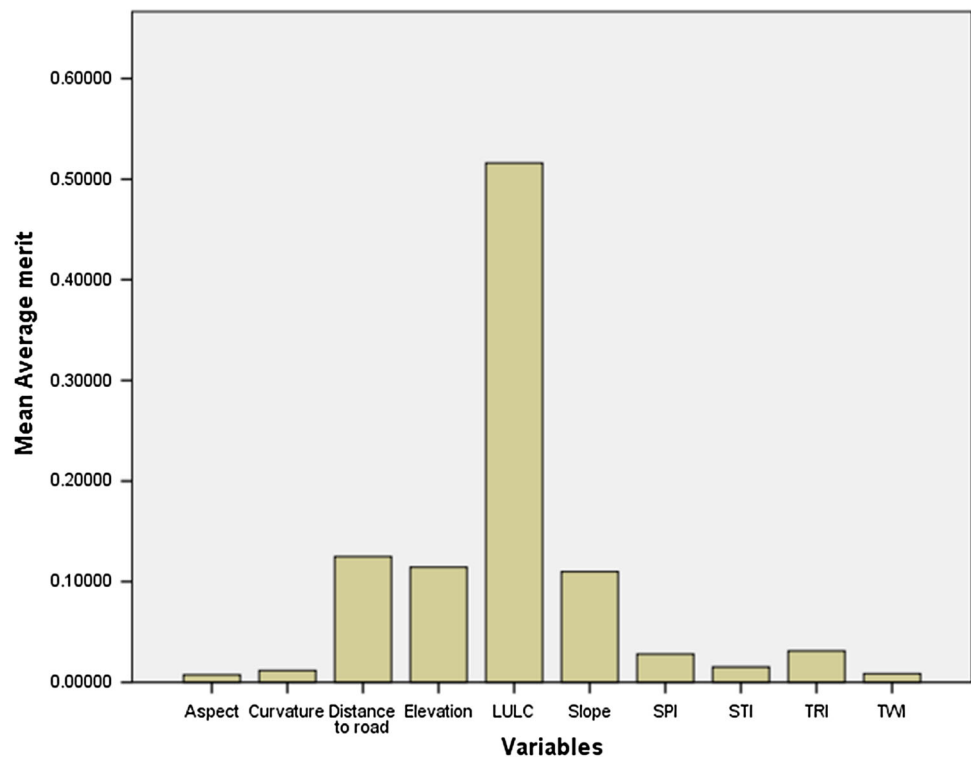
occurrences are influenced by several factors (Fernández and Lutz 2010; Pradhan 2010). In the present study, 12 influencing factors, such LULC, distance to road, elevation, slope, topographic wetness index, stream power index, sediment transport index, curvature, topographic roughness index, curvature and aspect were selected. The likelihood of flooding decreases with the increases of elevation. Elevations for the study area were ranged from 18 to 69 m, which remained in line with the flood occurrence (Fig. 5a). The ground surface, which is possessed by the curvature. The range of curvature value between 1.0 and 2.0 was considered as sensitive to flooding (Hudson and Kesel 2000). Curvature map, which was produced by using the DEM ranged from 0.32–0.82 (Fig. 5b). An aspect map was generated and classified into 9 categories: (0–22.5), (22.5–67.5), (67.5–112.5), (112.5–157.5), (157.5–202.5), (202.5–247.5), (247.5–292.5), (292.5–337.5), (337.5–360) (Fig. 5c). Due to the regional flood risk assessment, ground slope is a momentous element, which can increase runoff (Tehrany et al. 2015b). In this study, slope was ranged from 0 to 5.75 (Fig. 5d). TRI ascertained the confrontment pose on the water flow by the underlying surface (Straatsma and Baptist 2008). Teesta river located around the lowest TRI, which caused speedy water flow due to the hilly slopes around the river.

Consequently, the flooding has been happening in those regions, where the lowest TRI is observed. The highest value of TRI was 27 in this study (Fig. 6a). Flood plain is

strongly correlated with the high TWI values. In Fig. 6b, the range of TWI value was −1.54 to 7.72 (Fig. 6c). Flood occurrence is affected by the STI. The highest value of STI in this study was 140.64 (Fig. 6d).

In flood occurrences, LULC played a vital role, vegetated land turning into bare land resulting in the increase of runoff (García-Ruiz et al. 2008). In this study, LULC was classified into 6 categories, such as vegetation, bare land, built up, sand bar, agricultural land, and water body (Fig. 7a). For the flood discharge, river flow played a key role as a main track and caused flooding in those areas, which are near to the river (Opperman et al. 2009). Figure 7b showed that the highest distance from the river of this region was 1503 m. For accounting surplus precipitation and infiltration, soil data played a significant role (Johnson 2000). In this study, 12 soil types were found, such as water, usters, aqualts, humults, udults, ustults, aqualfs, ustalfs, ochrepts, aquepts, aquents, and psamments (Fig. 7c). For flood assessment, the amount of rainfall played a key role (Kay et al. 2006). The highest rainfall in this region was 550.411 mm (Fig. 7d).

4.3 Flood susceptibility mapping

Four novel ensemble machine learning algorithms, such as Bagging with Reptree, Bagging with M5P, Bagging with Random forest, and Bagging with Random tree methods

Fig. 5 Influencing factors of flood occurrence **a** elevation, **b** curvature, **c** aspect and **d** slope
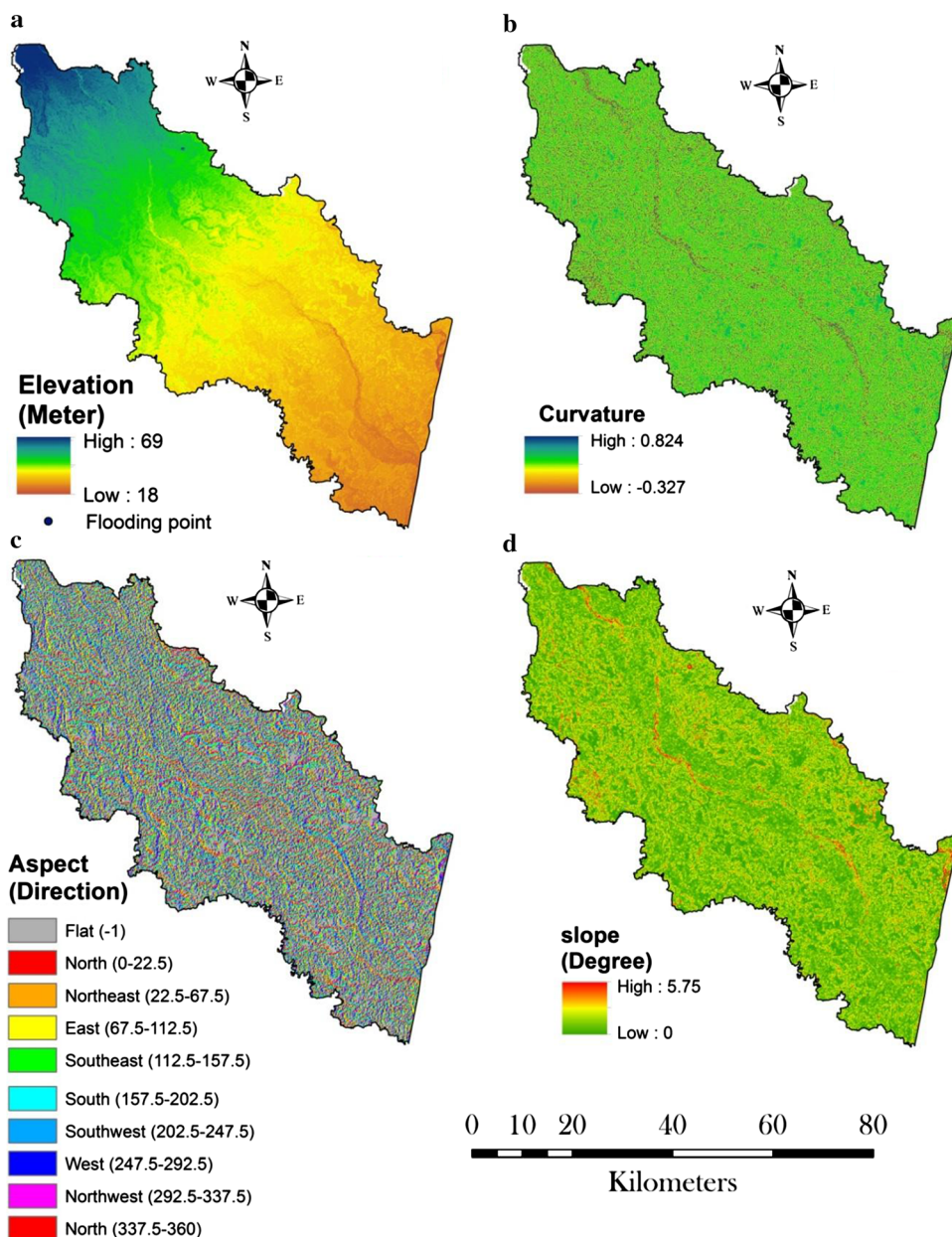
were developed and employed to predict the flood susceptibility areas in the Teesta flood region area.

We classified flood susceptibility zones into five classes, such as very low, low, moderate, high and very high (Fig. 8). 1071.71 km^2 area, largest area to the total area of the basin, was predicted as a very high susceptible zone and the moderate susceptibility zone was covered by the smallest area (395.49 km^2). These were predicted by Bagging with Random forest (Fig. 9c). Bagging with REPtree predicted 1045.72 km^2 and 521.65 km^2 area as very high and high flood susceptible zones (Fig. 9a). 1060.811 and 831.89 km^2 area as very high flood susceptible zone were predicted by Bagging with M5P and Bagging with random tree algorithms, while the very low

flood susceptibility zone was covered by 951.82 km^2 1038.31 km^2 respectively (Fig. 9b, d).

4.4 Evaluation and comparisons of flood susceptibility models

Four individual models (BgReptree, BgM5P, BgRf andBgRt) were used to implement and develop flood susceptibility maps in this study. The AUC and significant level of the ROC curve (Fig. 10) were used to assess the evaluation of these models. The values of 4 individual models were statistically significant (significant level, 0.00) in this study. Figure 10 showed that BgM5P model (AUC= 0.945) was the best performed model followed by BgRf

Fig. 6 Influencing factors of flood occurrence **a** TRI, **b** TWI, **c** SPI and **d** STI
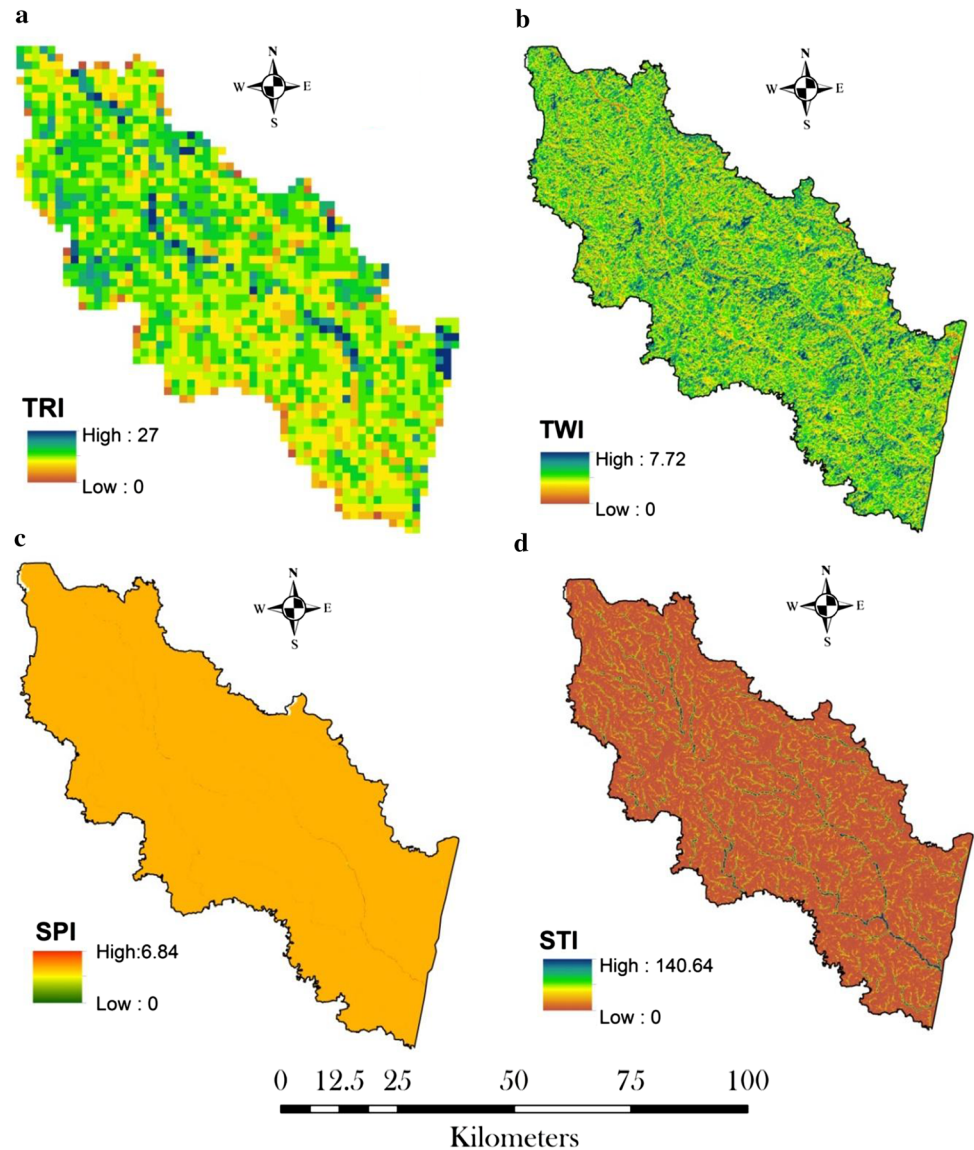
(AUC=0.912), BgReptree (AUC=0.876) and BgRt (AUC =0.844). Several measures of confusion matrix were calculated for validating the flood susceptible models (Table 1). Higher sensitivity (86.25), specificity (8.75), Youden index *J* (0.75), predicted positive (88.46%) and negative values (86.59%), and optimal criterion (>0.214) were calculated for the BgM5P based FSM model (Table 1). Based on the all values of different measures of confusion matrix, it could be stated that BgM5P algorithm selected as the representative for flood susceptible modelling in the present study area, followed by the BgRf, BgReptree, and BgRt.

Wilcoxon signed rank tests were employed to evaluate the performance of the four models. The *p* value for the BgM5P-BgReptree, BgRf- BgReptree, BgRt- BgM5P, BgRt- BgRf placed at a 95% significant level (<0.05), whereas the values of Z exceeded the critical level (−1.96 and +1.96). The performance of the four models for flood susceptibility mapping was significantly different from each other. The Freidman test did not compare the differences between individual models and the Chi square value was 30.015 found by Friedman test (Table 2). The average ranking values of the Freidman tests for the four hybrid models (BgReptree, BgM5P, BgRf and BgRt) were 2.63, 2.12 2.38, 2.87, respectively. Therefore, the Wilcoxon Signed-Rank test was used for exploring the differences between individual models. The *z* value and *p* value of Wilcoxon signed rank test of Bg-Rt vs Bg-Reptree, Bg-Rf vs Bg-M5P were not exceed the critical level (−1.96 and + 1.96) and statistically significant (<0.05) (Table 3). This indicated that the performance of these two models were not significantly different.

Kruskal–Wallis test revealed that all the four ensemble models were significant at 0.01% level (Table 4) for flood

Fig. 7 Influencing factors of flood occurrence **a** land use land cover, **b** distance to river, **c** soil types and **d** rainfall
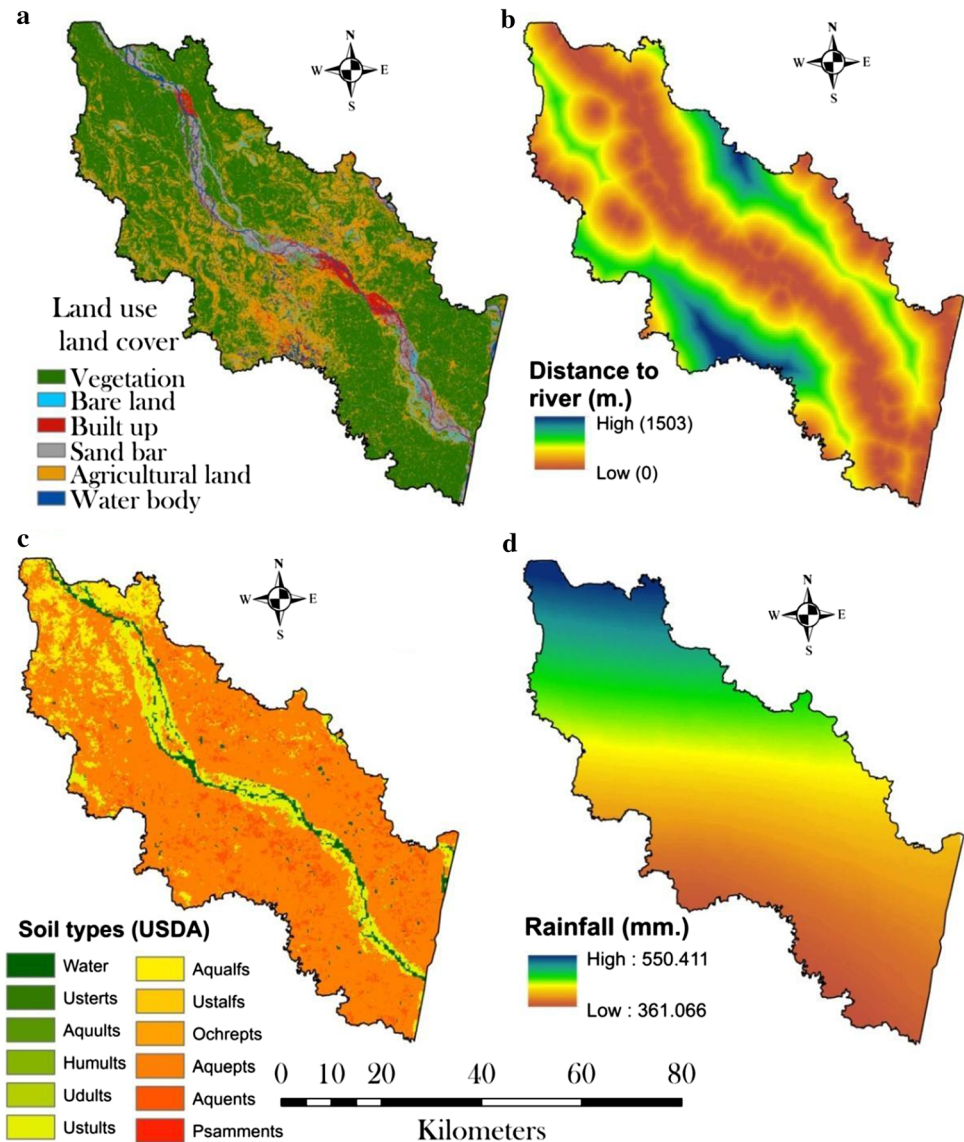
susceptibility modeling. Among four models, BgM5P performed best, because it achieved the lowest mean rank (44.90) in lower part and highest mean rank (116.10) in upper part and produced highest Chi Square value (94.462) compared to other three models. The order of the other models based on their mean rank and Chi Square values were Bg-Rf>Bg-Reptree>Bg-Rt. Following the mentioned three statistical tests, Kolmogorov–Smirnov test also explored that four Bagging with ensemble models provided significant ($p<0.01$) results of flood susceptibility modeling (Table 5). Most extreme differences of four models ranged from 0.500 to 0.750. z values produced by this test ranged from 3.162 to 4.743. BgM5P again outperformed in this test. Ranks of other models were Bg-Rf>Bg-Reptree> Bg-Rt.

5 Discussion

Extreme flood has been becoming a common miserable scene in the northern part of Bangladesh every year due to its geological structure and inappropriate law enforcement. It is a paramount need to take a prediction and mitigation approach in order to reduce property damage and loss of life. Flood modeling and flood susceptibility mapping are the essential approach to assess risk. Four Bagging ensemble models were used to make flood susceptibility maps in this study. In general, the flooding susceptible models cut off the exposed areas considering several flood conditioning parameters (Hong et al. 2018). Tehrany et al. (2015a) suggested that floods usually are affected by the particular area's morphological, geological, topographical and hydrological conditions. Therefore, choosing the appropriate flood-conditioning parameters is the most

Fig. 8 Flood susceptibility mapping using **a** bagging with Reptree, **b** bagging with M5P, **c** bagging with random forest, **d** bagging with random tree

Flood Susceptibility

Kilometers

▇ Very low ▇ Low ▇ Moderate ▇ High ▇ Very high
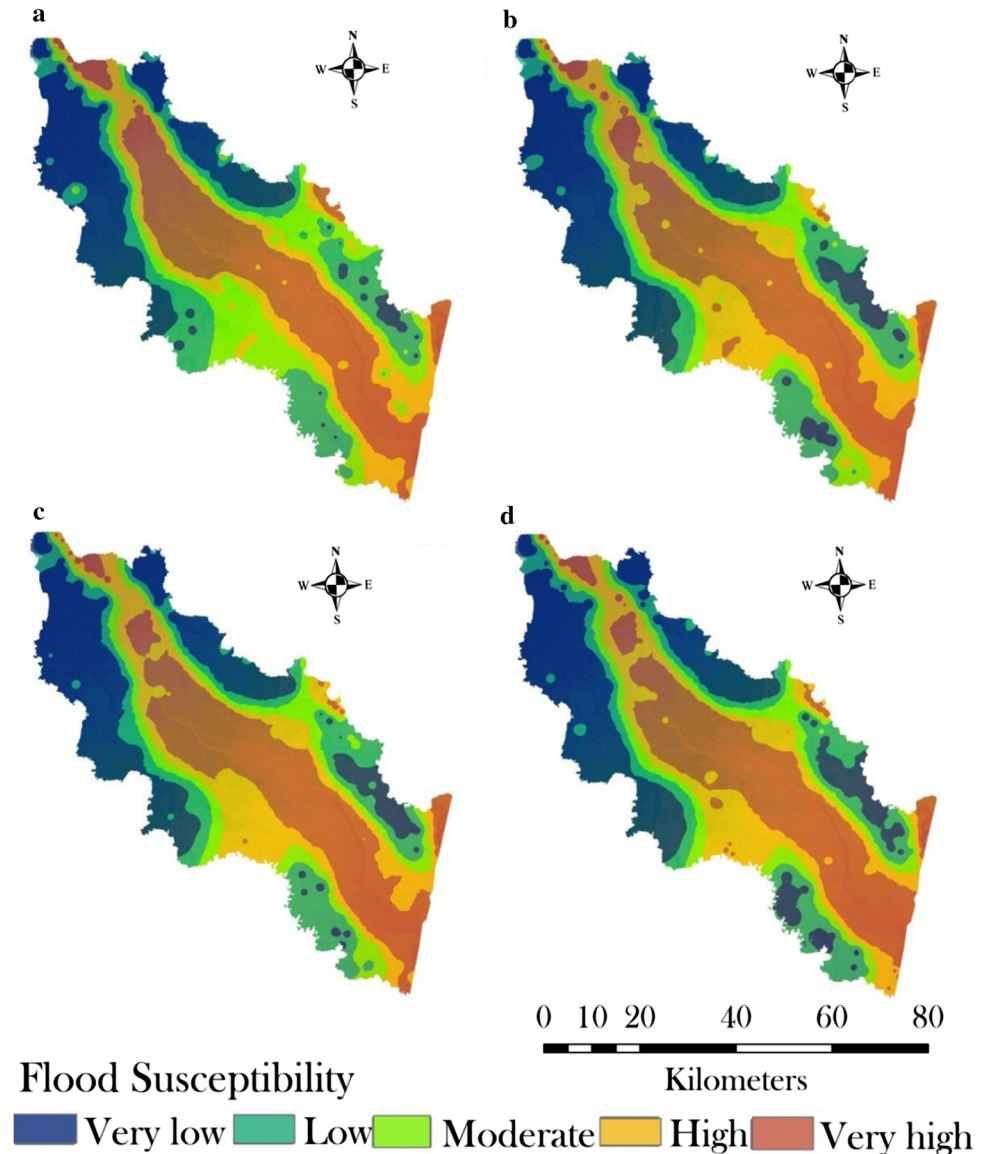
important and essential part in modeling of flood susceptibility. The IGR method, utilized in the present work, evaluated the influence of the selected parameters for flooding. Arora et al. (2019) pointed out that the importance of the flood conditioning factors varies from one, location to another. This is because the nature and cause of the flooding are not always similar at different locations (Rubinato et al. 2019). The result of IGR showed that the most effective factors were LULC, while, TRI, SPI, STI, TWI and curvature were evaluated as least effective conditioning factors and aspect had no effect in the flood susceptibility mapping in this study. This is because the study area lies in the lower part of Teesta River basin having a flat topography with low elevation, low slope angle and moderate drainage density (Mondal and Islam 2017). These findings are quite similar with the findings of

Khosravi et al. (2018), Khosravi et al. (2016a, b), Khosravi et al. (2019). They reported that the most important flood conditioning factor was altitude and less important factors were rainfall, SPI and curvature. Similar findings were found by Tehrany et al. (2015), Moghadam et al. (2018), Termeh et al. (2018), Chapi et al. (2017).

Li et al. (2012) reported that in the low elevated areas (areas having elevation lower than 300 m) the likelihood of flood occurrence was high which demonstrate that the study area of this study is very vulnerable to the susceptibility of flood occurrence as the elevation of this study area is low enough (69 m). Hong et al. (2017) also reported the similar kind of results in their study.. The curvature range found in this study was 0.82–0.32. Hudson and Kesel (2000) stated that the range of curvature value between 1.0 and 2.0 had the probability of flooding. Almost similar

Fig. 9 Area coverage of predicted flood susceptible models by **a** bagging with Reptree, **b** bagging with M5P, **c** bagging with random forest, **d** bagging with random tree
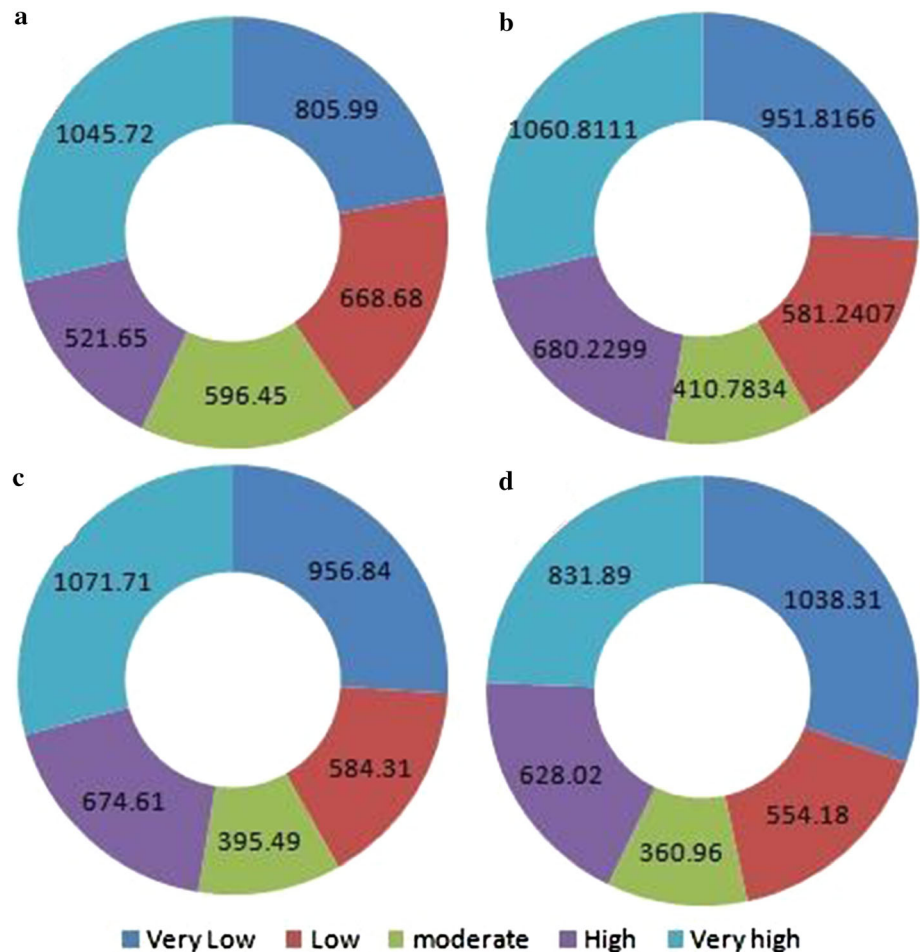
results were obtained by Cao et al. (2016), Chapi et al. (2017) and Khosravi et al. (2016a, b) in their studies. The aspect was ranged between 337.5 and 360 and the IGR value was 0.00. Therefore, present study excluded aspect for flood susceptibility modelling following Rahman et al. 2019. Khosravi et al. (2016a, b) also reported the similar results in their study. The slope angle found in this study ranges from 0 to 5.75° which determines the water velocity and Fernandez and Lutz (2010) described it as a vital parameter of causing flood. Rahmati and Pourghasemi (2017), Tehrany et al. (2014) reported that the probability of flood occurrence would be higher, if slope angle was lower. The findings of this study demonstrated that the study region had high likelihood of flood occurrence due to the lower slope angle.

The morphological factor of TRI is highly related with flooding (Werner et al. 2005). Findings showed that the study area had the highest value of TRI (27) which could cause flooding and this finding is similar to the findings of Tehrany and Kumar (2018). STI which is considering another flood occurrence factor defines the movement of the sediments in water bodies (Mojaddadi et al. 2017). The

highest STI value explored in this study was 14.64. Almost similar result of STI found in the work of Tehrany and Kumar (2018). SPI and TWI are two important hydrological factors responsible for the spatial variation of flooding. The TWI values ranged from 1.54 to 7.72 in this study. Topographical effects are quantified by the TWI (Lee et al. 2017). The LULC, distance to river, soil type and rainfall were the remarkable flood conditioning factors in this study. The findings of these parameters matched with the findings of Tehrany and Kumar (2018), Brath et al. (2006). Azareh et al. (2019) revealed that soil texture, land use, elevation and frequently occurring heavy rain storms were the most influential factors of flood in Iran which is analogous to this study. Hosseini et al. (2020) found that elevation (similar to this study); drainage density; vegetation and distance were the influential factors of flash flood in Iran. Five flood susceptible zones were predicted by BgReptree, BgM5P, BgRf and BgRt. High and very high flood susceptibility zones were covered by 20–29% areas of the total area of Teesta basin. Janizadeh et al. (2019) reported that 26.1% and 12.9% of area predicted as very high susceptibility according to QDA and ADT model

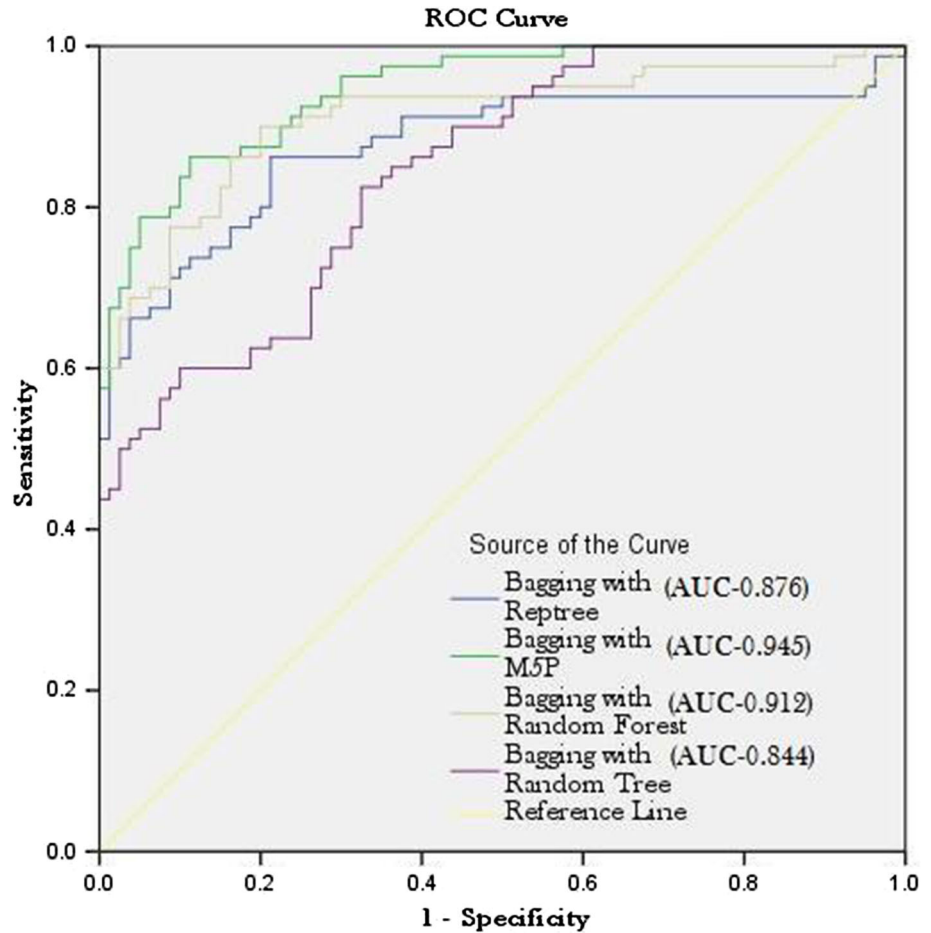**Fig. 10** Validation of flood susceptible models by using RoC curve

Table 1 Confusion matrix estimated for all flood susceptible models

Algorithms	Sensitivity	Specificity	ROC		Youden Index J	Positive predictive value	Negative predictive value	Optimal criterion
			Std. error	p value				
BgReptree	86.25	78.75	0.03	<0.0001	0.65	80.23	85.14	>0.112
BgM5P	86.25	88.75	0.016	<0.0001	0.75	88.46	86.59	>0.214
BgRf	90	80	0.024	<0.0001	0.7	81.82	88.89	>0.074
BgRt	82.5	67.5	0.029	<0.0001	0.5	71.74	79.41	>0.279

Table 2 Friedman test for all flood susceptible models

Models	Mean rank	Chi square	Significance
BgReptree	2.63	30.015	0.000
BgM5P	2.12		
BgRf	2.38		
BgRt	2.87		

respectively. According to fuzzy WofE-LR; WofE-RF and fuzzy WofE-SVM model, very high susceptibility area was 10.41%, 15.89%, and 17.65% in China, respectively (Hong et al. 2017). In their study, Choubin et al. (2018) reported that 80.6 km² and 10.1 km² area were, respectively, predicted as low and very high flood susceptibility classes revealed by using the MDA model. Pham et al. (2019) explored that very low, low, moderate, high, and very high susceptibility classes were covered by approximately 26%, 34%, 20%, 12%, and 8% area of the total land area, respectively, predicted by RSSFT model. Bui et al. (2019),

Table 3 Result of Wilcoxon signed-rank test for comparing the flood susceptible models

Pairwise comparison	Positive	Negative	z value	p value	Significance
BgM5P vsBgReptree	52	108	−5.050	0.000	Yes
BgRfvsBgReptree	78	82	−4.114	0.000	Yes
BgRtvsBgReptree	89	71	−.308	0.758	Yes
BgRfvs BgM5P	88	72	.000	1.000	No
BgRtvs BgM5P	105	55	−4.056	0.000	Yes
BgRtvsBgRf	105	55	−4.448	0.000	Yes

Table 4 Results of Kruskal–Wallis test for comparing the flood predictive models

Models	OBJECTID	N	Mean rank	Chi square value	df	Sig.
BgReptree	0	80	50.44	67.361	1	0.000 ($p<0.01$)
	1	80	110.56			
	Total	160				
BgM5P	0	80	44.9	94.462	1	0.000 ($p<0.01$)
	1	80	116.1			
	Total	160				
BgRf	0	80	47.53	81.045	1	0.000 ($p<0.01$)
	1	80	113.48			
	Total	160				
BgRt	0	80	52.98	56.469	1	0.000 ($p<0.01$)
	1	80	108.03			
	Total	160				

Kruskal–Wallis test

Grouping variable: OBJECTID

Table 5 Results of Kolmogorov–Smirnov test for all flood susceptible models

Models	Most extreme differences (absolute)	Kolmogorov–Smirnov z value	Sig. (two tailed)
BgReptree	0.650	4.111	0.000
BgM5P	0.750	4.743	0.000
BgRf	0.700	4.427	0.000
BgRt	0.500	3.162	0.000

Shahabi et al. (2020), Chen et al. (2019) reported that 15–24% area was predicted as high flood susceptible zone. Similar findings as this study were found by Tsakiri et al. (2018), Tehrany et al. (2019a, b), Ma et al. (2019), Costache et al. (2019). Therefore, it can be stated that the findings of the present study are highly correlated with the findings of previous literatures. These findings can be used as the basic foundation for flood management in the present study area.

To evaluate the model performance, several previous studies used the AUC values of the ROC curve (Bui et al. 2018; Choubin et al. 2018; Khosravi et al. 2019). The AUC values of the BgM5P and BgRf used in this study were 0.945 and 0.912, respectively. Bui et al. 2018 reported that the ANFIS-ICA (AUC=0.947) model performed better by comparing with the Bagging-LMT (AUC=0.940), BLR (AUC=0.936), LMT (AUC=0.934), ANFIS-FA (AUC=

0.917), LR (AUC=0.885) and RF (AUC=0.806) models. Choubin et al. (2018) used AUC for validation of the flooding susceptible models and considered the flood susceptible models as valid because of achieving the higher AUC values, such as ensemble model (AUC=0.91), followed by CART (AUC=0.83), SVM (AUC=0.88), MDA (AUC=0.89) models. Hosseini et al. (2020) used GLMBoost based Random forest and BayesGLM algorithms for flood modeling and revealed high performance accuracy of both the algorithms for modeling flood. Khosravi et al. (2019) found that the NBT had the highest predictive accuracy than the VIKOR (AUC=0.965), TOPSIS (AUC=0.968), SAW (AUC=0.97), NB (AUC=0.979) models. Hong et al. (2020) developed and applied the ensemble of bagging-LogitBoost alternating decision tree (LADT) and forest by penalizing attributes (FPA) for modelling the landslide susceptibility maps and reported

that bagging-LADT model achieved the very high accuracy for both training and testing datasets. Therefore, it could be stated that the ensemble of bagging could be improved significantly for any kinds of natural hazard prediction. Dodangeh et al. (2020) used BT-GAM, BT-MARS and BT-BRT ensemble algorithms for flood susceptibility prediction and BT-GAM (AUC=0.98) found as the outperformed model followed by BT-MARS (AUC=0.97) and BT-BRT (AUC=0.95). Therefore, we can state that the algorithms which were used in the present study had higher accuracy. For assessing the performance of the models, Wilcoxon signed-rank test, Friedman test, Kruskal–Wallis test and Kolmogorov–Smirnov test were also conducted in this study. The findings of these tests are identical with the work of Bui et al. (2018), Khosravi et al. (2018), Hong et al. (2017). Kruskal–Wallis test and Kolmogorov–Smirnov test found that all the four ensemble models performed significantly ($p<0.01$) for flood susceptibility mapping of Teesta River basin, Bangladesh.

6 Conclusion

In the present study, we developed and utilized four ensembles of bagging algorithms, such as bagging with REPtree, bagging with RF, bagging with M5P, and bagging with RT for the first time for modelling the flood susceptibility mapping in the Teesta River basin, Bangladesh (Northern). A total of 413 flooding points with twelve parameters, such as elevation, slope, curvature, aspect, SPI, TWI, STI, LULC, rainfall, distance to the river, TWI, and soil types, which affect the flooding, were selected for modelling. The importance of flood condition parameters were determined by employing the IGR technique. Based on the feature selection outcomes, aspect was not considered for flood susceptible modelling. The ROC curve was used to validate the flood susceptible models. The Friedman test, Wilcoxon signed-rank test, Kruskal–Wallis test and Kolmogorov–Smirnov test were employed to explore the differences of performance of the flood susceptible models with each other. The highest flexibility and predictive ability were obtained in case of the bagging with M5P algorithm, followed by bagging with RF, bagging with REPtree and bagging with RT. The application of the ROC Curve in the outcome validation phase depicted that bagging with the M5P algorithm had the highest efficiency in comparison with other models (AUC=0.945). However, the performance of all models for the mapping of flood susceptibility were excellent. The findings of the study stated that bagging with M5P and bagging with RF are one most capable tool for flood susceptible modelling. As an optimal model, a total area of 30% was identified as highly vulnerable to flooding. However, the major drawback is

that the application of these models did not consider the changes over time for some factors, including SPI and LULC, because these are dynamic. Based on the availability of temporal datasets of these factors, future research on the temporal scale will be performed. Furthermore, these models can be upgraded by performing the sensitivity analysis concerning various influential factors. Bagging with the M5P algorithm, in comparison with the other models, had advantages, including fewer candidate parameters, high optimization capability, and fast convergence for preparing flash flood susceptibility maps.

In recent times, the strategies for the management of flood are considered as the top priority, particularly in Bangladesh, where flash floods occur every year. However, other basins and regions have not yet been appraised for flood mitigation plans. Hence, the present study was taken place in the Teesta River basin using some advanced machine learning algorithms, which will provide valuable information concerning methods to be adopted for supporting the local authorities and other parties in developing efficient alleviation strategies of flash flood and land-use policy planning not only Bangladesh but also other basins of the world.

References

Abebe YA, Ghorbani A, Nikolic I, Vojinovic Z, Sanchez A (2019) Flood risk management in Sint Maarten: a coupled agent-based and flood modelling method. J Environ Manag 248:109317

Abba SI, Pham QB, Usman AG, Linh NTT, Aliyu DS, Nguyen Q, Bach QV (2020) Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant. J Water Process Eng 33:101081

Achour Y, Pourghasemi HR (2020) How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? Geosci Front 11(3):871–883

Akay H, Koçyiğit MB (2020) Flash flood potential prioritization of sub-basins in an ungauged basin in Turkey using traditional multi-criteria decision-making methods. Soft Comput 31:1–13

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 Ind 117:106620

Al-Abadi AM (2018) Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study. Arab J Geosci 11:218

Alfieri L, Bisselink B, Dottori F, Naumann G, de Roo A, Salamon P, Wyser K, Feyen L (2017) Global projections of river flood risk in a warmer world. Earth's Future 5:171–182

Antoniazza G, Bakker M, Lane SN (2019) Revisiting the morphological method in twodimensions to quantify bed-material transport in braided rivers. Earth Surf Proc Land 44:2251–2267

Arora A, Pandey M, Siddiqui MA, Hong H, Mishra VN (2019) Spatial flood susceptibility prediction in Middle Ganga Plain: comparison of frequency ratio and Shannon's entropy models. Geocarto Int. https://doi.org/10.1080/10106049.2019.1687594

Azad AK, Hossain KM, Nasreen M (2013) Flood-induced vulnerabilities and problems encountered by women in Northern Bangladesh. Int J Disaster Risk Sci 4(4):190–199

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 International, Milton Park, pp 1–21

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology / A Hydrol Sci Bull 24 (1): 43–69

Bhattacharya RK, Chatterjee ND, Das K (2020) Sub-basin prioritization for assessment of soil erosion susceptibility in Kangsabati, a plateau basin: a comparison between MCDM and SWAT models. Sci Total Environ 139474

Binh PT, Zhu X, Groeneveld RA, Ireland VC (2020) Risk communication. Policy. https://doi.org/10.1016/j.landusepol.2019.104436

Brath A, Montanari A, Moretti G (2006) Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty). J Hydrol 324(1–4):141–153

Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140

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

Bui DT, Pradhan B, Nampak H, Bui Q-T, Tran Q-A, Nguyen Q-P (2016) Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS. J Hydrol 540:317–330

Bui DT, Panahi M, Shahabi H, Singh VP, Shirzadi A, Chapi K, Khosravi K, Chen W, Panahi S, Li S, Ahmad BB (2018) Novel hybrid evolutionary algorithms for spatial prediction of floods. Sci Rep 8(1):1–14

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 179:184–196

Bui DT, Hoang N-T, Martínez-Álvarez F, ThiNgo P-T, Hoa PV, Pham TD, Samui P, Costache R (2020a) A novel deep learning neural network approach for predicting flash flood susceptibility: a case study at a high frequency tropical storm area. Sci Total Environ 701:134413

Bui QT, Nguyen QH, Nguyen XL, Pham VD, Nguyen HD, Pham VM (2020b) Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. J Hydrol 581:124379

Busico G, Colombani N, Fronzi D, Pellegrini M, Tazioli A, Mastrocicco M (2020) Evaluating SWAT model performance, considering different soils data input, to quantify actual and future runoff susceptibility in a highly urbanized basin. J Environ Manag 266:110625

Cao C, Xu P, Wang Y, Chen J, Zheng L, Niu C (2016) Flash flood hazard susceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas. Sustainability 8(9):948

Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui DT, Pham BT, Khosravi K (2017) A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environ Model Softw 95:229–245

Chen W, Shahabi H, Zhang S, Khosravi K et al (2018) Landslide susceptibility modeling based on GIS and novel bagging-based kernel logistic regression. Appl Sci 8(12):2540

Chen W, Hong H, Li S, Shahabi H, Wang Y, Wang X, Ahmad BB (2019) Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. J Hydrol 575:864–873

Chen W, Hong H, Li S, Shahabi H, Wang Y, Wang X, Ahmad BB (2019a) Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with random subspace and random subspace ensembles. J Hydrol 575:864–873

Chen W, Panahi M, Tsangaratos P, Shahabi H, Ilia I, Panahi S et al (2019b) Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. Catena 172:212–231

Chen W, Li Y, Xue W, Shahabi H, Li S, Hong H, Wang X, Bian H, Zhang S, Pradhan B, Ahmad BB (2020) Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. Sci Total Environ 701:134979. https://doi.org/10.1016/j.scitotenv.2019.134979

Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A (2019) An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Sci Total Environ 651:2087–2096

Choubin B, Zehtabian G, Azareh A, Rafiei-Sardooi E, Sajedi-Hosseini F, Kişi Ö (2018) Precipitation forecasting using classification and regression trees (CART) model: a comparative study of different approaches. Environ Earth Sci 77(8):314

Cieslak DA, Chawla NV (2009) A framework for monitoring classifiers' performance: when and why failure occurs? Knowl Inf Syst 18(1):83–108

Costache R (2019a) Flood susceptibility assessment by using bivariate statistics and machine learning models - a useful tool for flood risk management. Water Res Manag 33(9):3239–3256

Costache R (2019b) Flash-flood Potential Index mapping using weights of evidence, decision Trees models and their novel hybrid integration. Stochast Environ Res Risk Assess 33 (7):1375–1402

Costache R (2019c) Flash-Flood Potential assessment in the upper and middle sector of Prahova river catchment (Romania). A comparative approach between four hybrid models. Sci Total Environ 659:1115–1134

Costache R, Bui DT (2019) Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: a case study at the Putna river catchment of Romania. Sci Total Environ 691:1098–1118

Costache R, Bui DT (2020) Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. Sci Total Environ 712:136492

Costache R, Hong H, Wang Y (2019) Identification of torrential valleys using GIS and a novel hybrid integration of artificial intelligence, machine learning and bivariate statistics. Catena 183:104179

Costache R, Hong H, Pham QB (2020a) Comparative assessment of the flash-flood potential within small mountain catchments using bivariate statistics and their novel hybrid integration with machine learning models. Sci Total Environ 711:134514

Costache R, Pham QB, Avand M, Linh NTT, Vojtek M, Vojteková J, Lee S, Khoi DN, Nhi PTT, Dung TD (2020b) Novel hybrid models between bivariate statistics, artificial neural networks and

boosting algorithms for flood susceptibility assessment. J Environ Manag 265:110485

Costache R, Pham QB, Sharifi E, Linh NTT, Abba SI, Vojtek M, Vojteková J, Nhi PTT, Khoi DN (2020c) Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and gis techniques. Remote Sens 12(1):106

Costache R, Popa MC, Bui DT, Diaconu DC, Ciubotaru N, Minea G, Pham QB (2020d) Spatial predicting of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning. J Hydrol 585:124808

Costache R, Pham QB, Corodescu-Roşca E, Cîmpianu C, Hong H, Linh NTT, Fai CM, Ahmed AN, Vojtek M, Pandhiani SM, Minea G, Ciobotaru N, Popa MC, Diaconu DC, Pham BT (2020e) Using GIS, remote sensing, and machine learning to highlight the correlation between the land-use/land-cover changes and flash-flood potential. Remote Sens 12(9):1422

de Kraker AMJ (2015) Flooding in river mouths: human caused or natural events? Five centuries of flooding events in the SW Netherlands, 1500–2000. Hydrol Earth Syst Sci 19:2673–2684

Dewan TH (2015) Societal impacts and vulnerability to floods in Bangladesh and Nepal. Weather Clim Extreme 7:36–42

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

FAO (Food and Agriculture Organization of the United Nations) (2017) The state of world fisheries and aquaculture. Available from: http://www.fao.org/fishery/en. Accessed 23 Jan 2017

Ferdous MR, Wesselink A, Brandimarte L, Di Baldassarre G, Rahman MM (2019) The levee effect along the Jamuna River in Bangladesh. Water Int 44(5):496–519

Fernández DS, Lutz MA (2010) Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. Eng Geol 111(1–4):90–98

García-Ruiz JM, Regüés D, Alvera B, Lana-Renault N, Serrano-Muela P, Nadal-Romero E, Navas A, Latron J, Martí-Bono C, Arnáez J (2008) Flood generation and sediment transport in experimental catchments affected by land use changes in the central Pyrenees. J Hydrol 356(1–2):245–260

Gibbons JD (1985) Nonparametric statistical inference, 2nd edn. M. Dekker, New York City

Gill JC, Malamud BD (2017) Anthropogenic processes, natural hazards, and interactions in a multi-hazard framework. Earth Sci Rev 166:246–269

Hajian-Tilaki K (2013) Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Casp J Int Med 4 (2):627

Hettiarachchi S, Wasko C, Sharma A (2018) Increase in flood risk resulting from climate change in a developed urban watershed—the role of storm temporal patterns. Hydrol Earth Syst Sci 22:2041–2056

Hirabayashi Y, Mahendran R, Koirala S, Konoshima L, Yamazaki D, Watanabe S, Kim H, Kanae S (2013) Global flood risk under climate change. Nat Clim Change 3:816–821

Hoeppe P (2016) Trends in weather related disasters—consequences for insurers and society. Weather Clim Extremes 11:70–79

Hong H, Liu J, Zhu AX (2020) Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble. Sci Total Environ 718:137231

Hong H, Ilia I, Tsangaratos P, Chen W, Xu C (2017) A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area, China. Geomorphology 290:1–16

Hong H, Panahi M, Shirzadi A, Ma T, Liu J, Zhu AX, Chen W, Kougias I, Kazakis N (2018a) Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. Sci Total Environ 621:1124–1141

Hong H, Tsangaratos P, Ilia I, Liu J, Zhu A-X, Chen W (2018b) Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. Sci Total Environ 625:575–588

Hoque MA, Tasfia S, Ahmed N, Pradhan B (2019) Assessing spatial flood vulnerability at Kalapara Upazila in Bangladesh using an analytic hierarchy process. Sensors 19:1302

Hosseini FS, Choubin B, Mosavi A, Nabipour N, Shamshirband S, Darabi H, Haghighi AT (2020) Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: application of the simulated annealing feature selection method. Sci Total Environ 711:135161

Hudson PF, Kesel RH (2000) Channel migration and meander-bend curvature in the lower Mississippi River prior to major human modification. Geology 28(6):531–534

Huţanu E, Mihu-Pintilie A, Urzica A, Paveluc LE, Stoleriu CC, Grozavu A (2020) Using 1D HEC-RAS modeling and LiDAR data to improve flood hazard maps' accuracy: a case study from Jijia floodplain (NE Romania). Water 12 (6): 1624

Islam MM, Sado K (2000) Flood hazard assessment in Bangladesh using NOAA AVHRR data with geographical information system. Hydrol Process 14:605–620

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 Extremes 25:100215

Janizadeh S, Avand M, Jaafari A, Phong TV, Bayat M, Ahmadisharaf E, Prakash I, Pham BT, Lee S (2019) Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh Watershed, Iran. Sustainability 11(19):5426

Javidan N, Kavian A, Pourghasemi HR, Conoscenti C, Jafarian Z (2020) Data mining technique (maximum entropy model) for mapping gully erosion susceptibility in the gorganrood watershed, Iran. In: Gully erosion studies from India and surrounding regions. Springer, Cham, pp 427–448

Johnson LE (2000) Assessment of flash flood warning procedures. J Geophys Res Atmos 105(D2):2299–2313

Joshi MM, Shahapure SS (2020) Flood susceptibility mapping for part of Bhima River basin using two-dimensional HEC-RAS model. In: Techno-societal 2018. Springer, Cham, pp 595–605

Kay AL, Jones RG, Reynard NS (2006) RCM rainfall for UK flood frequency estimation. II. Climate change results. J Hydrol 318 (1–4):163–172

Khosravi K, Nohani E, Maroufinia E, Pourghasemi HR (2016a) A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Nat Hazards 83(2):947–987

Khosravi K, Pham BT, Chapi K, Shirzadi A, Shahabi H, Revhaug I, Prakash I, Bui DT (2018) A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. Sci Total Environ 627:744–755

Khosravi K, Pourghasemi HR, Chapi K, Bahri M (2016b) Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models. Environ Monit Assess 188(12):656

Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, Dou J, Ly HB, Gróf G, Ho HL, Hong H (2019) A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. J Hydrol 573:311–323

Kolmogorov AN (1933) On the empirical determination of a distribution function. In: (Italian) Giornaledell'InstitutoItalianodegliAttuari, vol 4, pp 83–91

Kourgialas NN, Karatzas GP (2011) Flood management and a GIS modelling method to assess flood-hazard areas—a case study. Hydrol Sci J J Sci Hydrol 56(2):212–225

Kuriqi A, Koçileri G, Ardiçlioğlu M (2020) Potential of Meyer-Peter and Müller approach for estimation of bed-load sediment transport underdifferent hydraulic regimes. Model Earth Syst Environ 6(1):129–137

Lee S, Kim JC, Jung HS, Lee MJ, Lee S (2017) Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. Geomat Nat Hazards Risk 8(2):1185–1203

Lei X, Chen W, Avand M, Janizadeh S, Kariminejad N, Shahabi H, Costache R, Shahabi H, Shirzadi A, Mosavi A (2020) GIS-based machine learning algorithms for gully erosion susceptibility mapping in a semi-arid region of Iran. Remote Sens 12(15):2478

Li LT, Xu ZX, Pang B, Liu L (2012) Flood risk zoning in China. ShuiliXuebao (J Hydraul Eng) 43(1):22–30

Li X, Cummings AR, Alruzuq A, Matyas CJ, Amanambu AC (2019) Combining water fraction and dem-based methods to create a coastal flood map: a case study of hurricane harvey. ISPRS Int J Geo-Information 8(5):231

Lindman HR (1974) Analysis of variance in complex experimental designs. WH Freeman & Co, New York

Maaks DLG, Starr NB, Brady MA, Cpnp-PC PR, Blosser CG, Gaylord NM et al. (2020). Burns' Pediatric Primary CareE-Book. Elsevier

Ma M, Liu C, Zhao G, Xie H, Jia P, Wang D, Wang H, Hong Y (2019) Flash flood risk analysis based on machine learning techniques in the Yunnan Province, China. Remote Sens 11(2):170

Ma J, Ding Y, Cheng JC, Jiang F, Tan Y, Gan VJ, Wan Z (2020) Identification of high impact factors of air quality on a national scale using big data and machine learning techniques. J Cleaner Prod 244:118955

Mandal SP, Chakarbarty A (2016) Flash flood risk assessment for upper Teesta River basin: using the hydrological modeling system (HEC-HMS) software. Model Earth Syst Environ 2:9

Mohanty MP, Vittal H, Yadav V, Ghosh S, Rao GS, Karmakar S (2020) A new bivariate risk classifier for flood management considering hazard and socio-economic dimensions. J Environ Manag 255:109733

Mojaddadi H, Pradhan B, Nampak H, Ahmad N, Ghazali AHB (2017) Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. Geomat Nat Hazards Risk 8(2):1080–1102

Mondal MSH, Islam MS (2017) Chronological trends in maximum and minimum water flows of the Teesta River, Bangladesh, and its implications. Jàmbá J Disaster Risk Stud 9(1):a373

Nam W-H, Hayes MJ, Svoboda MD, Tadesse T, Wilhite DA (2015) Drought hazard assessment in the context of climate change for South Korea. Agric Water Manag 160:106–117

Nhu V-H, Ngo P-TT, Pham TD, Dou J, Song X, Hoang N-D, Tran DA, Cao DP, Aydilek IB, Amiri M, Costache R, Hoa PV, Bui DT (2020) A new hybrid firefly–PSO optimized random subspace tree intelligence for torrential rainfall-induced flash flood susceptible mapping. Remote Sens 12(17):2688

Nikolaos S, Kleomenis K, Elias D, Panagiotis S, Panagiota L, Vagelis P, Christos C (2019) A robust remote sensing–spatial modeling–remote sensing (RMR) Approach for flood hazard assessment. In: Spatial modeling in GIS and R for earth and environmental sciences, Elsevier, pp 391–410

Opperman JJ, Galloway GE, Fargione J, Mount JF, Richter BD, Secchi S (2009) Sustainable floodplains through large-scale reconnection to rivers. Science 326(5959):1487–1488

Paul GC, Saha S, Hembram TK (2019) Application of the GIS-based probabilistic models for mapping the flood susceptibility in bansloi sub-basin of ganga-bhagirathi river and their comparison. Remote Sens Earth Syst Sci 2(2–3):120–146

Pham BT, Pradhan B, Bui DT, Prakash I, Dholakia MB (2016) A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). Environ Model Softw 84:240–250

Pham BT, Jaafari A, Prakash I, Bui DT (2019) A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. Bull Eng Geol Environ 78(4):2865–2886

Pham QB, Abba SI, Usman AG, Linh NTT, Gupta V, Malik A, Costache R, Vo ND, Tri DQ (2019) Potential of hybrid data-intelligence algorithms for multi-station modelling of rainfall. Water Res Manag 33(15):5067–5087

Phillips TH, Baker ME, Lautar K, Yesilonis I, Pavao-Zuckerman MA (2019) The capacity of urban forest patches to infiltrate stormwater is influenced by soil physical properties and soil moisture. J Environ Manag 246:11–18

Pourghasemi HR, Kariminejad N, Amiri M, Edalat M, Zarafshar M, Blaschke T, Cerda A (2020a) Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. Sci Rep 10(1):1–11

Pourghasemi HR, Razavi-Termeh SV, Kariminejad N, Hong H, Chen W (2020b) An assessment of metaheuristic approaches for flood assessment. J Hydrol 582:124536. https://doi.org/10.1016/j.jhydrol.2019.124536

Pradhan B (2010) Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. J Spatial Hydrol 9(2):1–18

Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystem 9:181–199

Pyatkova K, Chen AS, Butler D, Vojinović Z, Djordjević S (2019) Assessing the knock-on effects of flooding on road transportation. J Environ Manag 244:48–60

Quinlan J (1992) Learning with continuous classes. In: Adams A, Sterling L (eds) 'AI'92: proceedings of the 5th Australian joint conference on artificial intelligence, pp 343–348

Quinlan JR (1987) Generating production rules from decision trees. In: ijcai, vol 87, pp 304–307

Rahman M, Ningsheng C, Islam MM, Dewan A, Iqbal J, Washakh RMA, Shufeng T (2019) Flood susceptibility assessment in Bangladesh using machine learning and multi-criteria decision analysis. Earth Syst Environ 3:585–601

Rahmati O, Pourghasemi HR (2017) Identification of critical flood prone areas in data-scarce and ungauged regions: a comparison of three data mining models. Water Resour Manag 31(5):1473–1487

Rahmati O, Pourghasemi HR, Zeinivand H (2016) Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. Geocarto Int 31(1):42–70

Roy J, Saha S, Arabameri A, Blaschke T, Bui DT (2019) A novel ensemble approach for landslide susceptibility mapping (LSM) in Darjeeling and Kalimpong Districts, West Bengal, India. Remote Sens 11:2866

Rubinato M, Nicholas A, Peng Y, Zhang JM, Lashford C, Cai YP, Lin PZ, Tait S (2019) Urban and river flooding: comparison of flood risk management approaches in the UK and China and an assessment of future knowledge needs. Water Sci Eng 12(4):274–283

Saha S, Reza AHMS, Roy MK (2019) Hydrochemical evaluation of groundwater quality of the Tista floodplain, Rangpur, Bangladesh. Appl Water Sci 9:198

Sahana M, Rehman S, Sajjad H, Hong H (2020) Exploring effectiveness of frequency ratio and support vector machine models in storm surge flood susceptibility assessment: A study of Sundarban Biosphere Reserve, India. Catena, 189:104450

Sarhadi A, Soltani S, Modarres R (2012) Probabilistic flood inundation mapping of ungauged rivers: linking GIS techniques and frequency analysis. J Hydrol 458–459:68–86

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 Manag 217:1–11

Shahabi H, Shirzadi A, Ghaderi K, Omidvar E, Al-Ansari N, Clague JJ, Geertsema M, Khosravi K, Amini A, Bahrami S, Rahmati O (2020) Flood detection and susceptibility mapping using sentinel-1 remote sensing data and a machine learning approach: hybrid intelligence of bagging ensemble based on K-nearest neighbor classifier. Remote Sens 12(2):266

Shit PK, Pourghasemi HR, Bhunia GS (2020) Gully erosion susceptibility mapping based on bayesian weight of evidence. In: Shit P, Pourghasemi H, Bhunia G (eds) Gully erosion studies from India and surrounding regions. Advances in science, technology & innovation (IEREK interdisciplinary series for sustainable development). Springer, Cham

Siahkamari S, Haghizadeh A, Zeinivand H, Tahmasebipour N, Rahmati O (2018) Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. Geocarto Int 33(9):927–941

Sihag P, Karimi SM, Angelaki A (2019) Random forest, M5P and regression analysis to estimate the field unsaturated hydraulic conductivity. Appl Water Sci 9(5):129

Smirnov N (1939) On the estimation of the discrepancy between empirical curves of distribution for two independent samples (Russian). Bull Moscow Univ 2:3–16

Straatsma MW, Baptist MJ (2008) Floodplain roughness parameterization using airborne laser scanning and spectral remote sensing. Remote Sens Environ 112(3):1062–1080

Suchmacher M, Geller M (2012) Practical biostatistics: a friendly step-by-step approach for evidence-based medicine. Academic Press, Cambridge

Tang X, Li J, Liu M, Liu W, Hong H (2020) Flood susceptibility assessment based on a novel random Naïve Bayes method: a comparison between different factor discretization methods. CATENA 190:104536

Talha S, Maanan M, Atika H, Rhinane H (2019) Prediction of flash flood susceptibility using fuzzy analytical hierarchy process (Fahp)algorithms and Gis: a study case of guelmim region In Southwestern of Morocco. Int Arch Photogrammetry, Remote SensSpat Inf Sci 42(4/W19)

Talukdar S, Pal S (2017) Impact of dam on inundation regime of flood plain wetland of punarbhaba river basin of barind tract of Indo-Bangladesh. Int Soil Water Conserv Res 5(2):109–121

Talukdar S, Pal S (2019) Effects of damming on the hydrological regime of Punarbhaba river basin wetlands. Ecol Eng 135:61–74

Taubenböck H, Wurm M, Netzband M, Zwenzner H, Roth A, Rahman A, Dech S (2011) Flood risks in urbanized areas—multi-sensoral approaches using remotely sensed data for risk assessment. Nat Hazards Earth Syst Sci 11:431–444

Tehrany MS, Kumar L (2018) The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. Environ Earth Sci 77(13):490

Tehrany MS, Jones S, Shabani F (2019a) Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. CATENA 175:174–192

Tehrany MS, Kumar L, Jebur MN, Shabani F (2019b) Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. Geomat Nat Hazards Risk 10(1):79–101

Tehrany MS, Pradhan B, Jebur MN (2014) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J Hydrol 512:332–343

Tehrany MS, Pradhan B, Jebur MN (2015a) Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stoch Environ Res Risk Assess 29(4):1149–1165

Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015b) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. CATENA 125:91–101

Termeh SVR, Kornejady A, Pourghasem HR, Keesstra S (2018) Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. Sci Total Environ 615:438–451

Tingsanchali T, Karim MF (2005) Flood hazard and risk analysis in the southwest region of Bangladesh. Hydrol Process 19:2055–2069

Tsakiri K, Marsellos A, Kapetanakis S (2018) Artificial neural network and multiple linear regression for flood prediction in Mohawk River, New York. Water 10(9):1158

Uniyal B, Jha MK, Verma AK, Anebagilu PK (2020) Identification of critical areas and evaluation of best management practices using SWAT for sustainable watershed management. Sci Total Environ 140737

Uthayakumar J, Metawa N, Shankar K, Lakshmanaprabu SK (2020) Financial crisis prediction model using ant colony optimization. Int J Inf Manag 50:538–556

Wang Y, Hong H, Chen W, Li S, Panahi M, Khosravi K, Shirzadi A, Shahabi H, Panahi S, Costache R (2019) Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. J Environ Manag 247:712–729

Wang Y, Fang Z, Hong H, Peng L (2020) Flood susceptibility mapping using convolutional neural network frameworks. J Hydrol 582:124482

Ward PJ, Jongman B, Weiland FS, Bouwman A, van Beek R, Bierkens MF, Ligtvoet W, Winsemius HC (2013) Assessing flood risk at the global scale: model setup, results, and sensitivity. Environ Res Lett 8 (4): 044019

Werner MGF, Hunter NM, Bates PD (2005) Identifiability of distributed floodplain roughness values in flood extent estimation. J Hydrol 314(1–4):139–157

Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques with java implementations, 2nd edn. Morgan Kaufmann Publishers, Burlington

Xu M, Li C (2020) Influencing factors analysis of water footprint based on the extended STIRPAT model. In: Application of the water footprint: water stress analysis and allocation. Springer, Singapore, pp 105–126

Yariyan P, Janizadeh S, Phong TV, Nguyen HD, Costache R, Le HV, Pham BT, Pradhan B, Tiefenbacher JP (2020) Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping. Water Res Manag 34(9):3037–3053

Yang W, Xu K, Lian J, Bin L, Ma C (2018) Multiple flood vulnerability assessment approach based on fuzzy comprehensive evaluation method and coordinated development degree model. J Environ Manag 213:440–450

Yin J, He F, Xiong YJ, Qiu GY (2017) Effects of land use/land cover and climate changes on surface runoff in a semi-humid and semi-arid transition zone in northwest China. Hydrol Earth Syst Sci 21:183–196

Zhang X, Cao W, Guo Q, Wu S (2010) Effects of landuse change on surface runoff and sediment yield at different watershed scales on the Loess Plateau. Int J Sedim Res 25(3):283–293

Zhao G, Pang B, Xu Z, Yue J, Tu B (2018) Mapping flood susceptibility in mountainous areas on a national scale in China. Sci Total Environ 615:1133–1142

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.