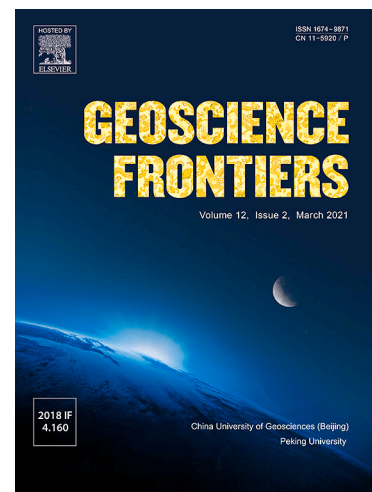


Research Paper

Flooding and its relationship with land cover change, population growth, and road density

Mahfuzur Rahman, Chen Ningsheng, Golam Iftekhar Mahmud, Md Monirul Islam, Hamid Reza Pourghasemi, Hilal Ahmad, Jules Maurice Habumugisha, Rana Muhammad Ali Washakh, Mehtab Alam, Enlong Liu, Zheng Han, Huayong Ni, Tian Shufeng, Ashraf Dewan



PII: S1674-9871(21)00088-8
DOI: <https://doi.org/10.1016/j.gsf.2021.101224>
Reference: GSF 101224

To appear in: *Geoscience Frontiers*

Received Date: 26 January 2021
Revised Date: 6 April 2021
Accepted Date: 30 April 2021

Please cite this article as: M. Rahman, C. Ningsheng, G. Iftekhar Mahmud, M. Monirul Islam, H. Reza Pourghasemi, H. Ahmad, J. Maurice Habumugisha, R. Muhammad Ali Washakh, M. Alam, E. Liu, Z. Han, H. Ni, T. Shufeng, A. Dewan, Flooding and its relationship with land cover change, population growth, and road density, *Geoscience Frontiers* (2021), doi: <https://doi.org/10.1016/j.gsf.2021.101224>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Research Paper

Flooding and its relationship with land cover change, population growth, and road density

Mahfuzur Rahman^{a,b,c,l}, Chen Ningsheng^{a,b,d,l}, Golam Iftekhar Mahmud^e, Md Monirul Islam^c, Hamid Reza Pourghasemi^{f,*}, Hilal Ahmad^{a,b}, Jules Maurice Habumugisha^{a,b}, Rana Muhammad Ali Washakh^{g,h}, Mehtab Alam^{a,b}, Enlong Liuⁱ, Zheng Han^j, Huayong Ni^k, Tian Shufeng^{a,b}, Ashraf Dewan^l

^a *Key Laboratory for Mountain Hazards and Earth Surface Process, Institute of Mountain Hazards and Environment (IMHE), Chinese Academy of Sciences (CAS), Chengdu 610041, China*

^b *University of Chinese Academy of Sciences (UCAS), Beijing 100049, China*

^c *Department of Civil Engineering, International University of Business Agriculture and Technology (IUBAT), Dhaka 1230, Bangladesh*

^d *Academy of Plateau Science and Sustainability, Xining 810016, China*

^e *Development Research Initiative, Dhaka 1216, Bangladesh*

^f *Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz 71454, Iran*

^g *School of Architecture, Neijiang Normal University, Neijiang 641100, China*

^h *Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China*

ⁱ *State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource and Hydropower, Sichuan University, Chengdu 610065, China*

^j *School of Civil Engineering, Central South University, Changsha 410075, China*

^k *Chengdu Institute of Geology and Mineral Resources, China Geological Survey, China*

¹ *School of Earth and Planetary Sciences, Curtin University, Kent St, Bentley, WA 6102, Australia*

*Corresponding author. E-mail address: hamidreza.pourghasemi@yahoo.com

¹These authors contributed equally to this work and should be considered as co-first authors.

Abstract

Bangladesh experiences frequent hydro-climatic disasters such as flooding. These disasters are believed to be associated with land use changes and climate variability. However, identifying the factors that lead to flooding is challenging. This study mapped flood susceptibility in the northeast region of Bangladesh using Bayesian regularization back propagation (BRBP) neural network, classification and regression trees (CART), a statistical model (STM) using the evidence belief function (EBF), and their ensemble models (EMs) for three time periods (2000, 2014, and 2017). The accuracy of machine learning algorithms (MLAs), STM, and EMs were assessed by considering the area under the curve—receiver operating characteristic (AUC-ROC). Evaluation of the accuracy levels of the aforementioned algorithms revealed that EM4 (BRBP-CART-EBF) outperformed (AUC > 90%) standalone and other ensemble models for the three time periods analyzed. Furthermore, this study investigated the relationships among land cover change (LCC), population growth (PG), road density (RD), and relative change of flooding (RCF) areas for the period between 2000 and 2017. The results showed that areas with very high susceptibility to flooding increased by 19.72% between 2000 and 2017, while the PG rate increased by 51.68% over the same period. The Pearson correlation coefficient for RCF and RD was calculated to be 0.496. These findings highlight the significant association between floods and causative factors. The study findings could be valuable to policymakers and resource managers as they can lead to improvements in flood management and reduction in flood damage and risks.

Keywords: Hydro-climatic disasters; Machine learning algorithms; Statistical model; Ensemble model; Relative change in flooding areas

Handling Editor: M. Santosh

Abbreviations

MLAs, machine learning algorithms; BRBP, Bayesian regularization back propagation; CART, classification & regression trees; STM, statistical model; EBF, evidence belief function; EMs, ensemble models; AUC-ROC, area under the curve-receiver operating characteristic; LCC, land cover change; PG, population growth; RD, road density; RCF, relative changes in flooding areas; GEE, Google Earth Engine; SLCDI, single land cover dynamic index; BLCDI, bilateral land cover dynamic index; LC, land cover; LCC, land cover change; IBLCDI, integrated bilateral land cover dynamic index; LGED, Local Government Engineering Department; SAR, synthetic aperture radar; NDVI, normalized difference vegetation index; IGR, information gain ratio; MCT, multicollinearity testing; TOL, tolerance; VIF, variance inflation factor; IL, input layer; HL, hidden layer; OL, output layer; SD, standard deviation; GWR, geographically weighted regression; MLR, multiple linear regression; MSE, mean square error; SRC, success rate curve; and PRC, prediction rate curve

1. Introduction

Owing to the immense destruction of properties, infrastructure and loss of lives, floods are one of the most destructive natural hazards. The estimated global economic damage associated with floods since the end of the 20th century was 386 billion USD (Wang et al., 2011). Of the natural disasters that occurred between 1995 and 2015, 47% were attributable to floods that directly affected 2.3 billion people, the majority of which lived in Southeast Asia (Wahlstrom and Guha-Sapir, 2015). Intensification of the hydrological cycle following climate warming is highly likely

to increase water-related hazards such as floods. Consequently, the exposure of people and properties to the risk of floods is a growing concern (Alfieri et al., 2017).

Frequent flooding occurs in Bangladesh as a result of severe overflow from three major rivers: the Ganges, Brahmaputra, and Meghna rivers (GBM), and heavy rainfall from upstream and inland locations (Chowdhury and Hassan, 2017). The flood events of 1988, 1998, 2004, 2007, 2014 and 2017 led to many fatalities and caused massive economic destruction (Dewan, 2015). In 2020, floods have occurred in a quarter of the country, affected more than four million people, and claimed 100 lives (NASA, 2020). In addition to the loss of lives and substantial damage to property, frequent flooding and outbreaks of vector-borne diseases, such as typhoid, significantly affect the economy.

Anthropogenic climate change is expected to increase the occurrence of flooding in Bangladesh (Viglione et al., 2016). However, previous studies have not comprehensively examined the influence of anthropogenic factors on flooding in Bangladesh. A few studies have reported that land cover change (LCC) is directly associated with an increase in flooding areas (Dewan and Yamaguchi, 2009; Adnan et al., 2020; Roy et al., 2020). Rapid population growth has resulted in significant expansion of urban lands and rapid depletion of agricultural land, as well as floodplains, waterbodies and wetlands, which have substantial impacts on flood dynamics. Additionally, urban and infrastructural growth are also associated with changes in hydrological and ecological systems, loss of drainage systems, and therefore, increased susceptibility of areas to floods. Furthermore, infrastructural development such as dense road construction influences inland flooding (Yin et al., 2016). Road density (RD) is closely related to land cover with respect to runoff and drainage in a floodplain (Kalantari et al., 2014; Pal et al., 2020; Bazai et al., 2021a).

Even though flooding cannot be fully controlled, its damage to people and property can be mitigated by identifying flood susceptible areas and influencing factors. In Bangladesh, many attempts have been made to map flood vulnerability and susceptible areas across the country; however, basin-scale risk assessment is still underdeveloped. The accurate identification of flood-prone areas in specific catchments may help susceptible communities and decision-makers to mitigate the damage caused by floods in the future, thereby alleviating flood risk (Arora et al., 2020). Different approaches have been applied to map flood-prone areas, including small-scale hydrodynamic modeling (Tingsanchali and Karim, 2005; Masood and Takeuchi, 2012; Hossain, 2015), and geospatial techniques (Islam and Sado, 2000). Although these studies are useful, they were carried out for relatively large areas, and are thus incapable of identifying flood dynamics at the basin scale.

Flooding is a dynamic process resulting from the complex interaction of watershed management and hydro-meteorological, hydrogeological, and geomorphological conditions. It is a complex and non-linear process that may never be effectively controlled through simple linear processes (Khosravi et al., 2020). In addition, past studies on flood mapping in Bangladesh have used numerical models that require input parameters, such as hydro meteorological data, river cross-sections, and discharges of upstream and inland catchments. Owing to the inadequate number of meteorological stations in Bangladesh, the mapping of flooding caused by local rainfall is difficult at a basin scale. Therefore, an alternative to numerical models is essential to fill the aforementioned knowledge gap. Moreover, floods are a regular phenomenon in the study area, northeast Bangladesh, especially during the monsoon season (May to September; Kamal et al., 2018). The lack of facilities and regular floods inflict enormous damage to the study area (Al Baky et al., 2019). According to the literature, no comprehensive studies can be found in this region. Precisely,

there is no study that relates flooding with anthropogenic factors, i.e., population growth, urban and infrastructural growth, etc. We took this shortcoming into account to explore how anthropogenic factors are associated with flooding in the northeast region (Surma River basin) of Bangladesh.

Geospatial techniques are widely used in flood susceptibility modeling (FSM) (Dewan et al., 2007; Mentzafou et al., 2018; Thirumurugan and Krishnaveni, 2019). A number of studies have been conducted on flood prediction and FSM using MCDM (multiple-criteria decision making) techniques, statistical models, and machine learning algorithms (MLAs) integrated with geospatial tools including—AHP (analytical hierarchy process; Radwan et al., 2019; Malik et al., 2020a), SWARA (stepwise weight assessment ratio analysis; Tien Bui et al., 2018), VIKOR (Vise Kriterijuska Optimizacija I Komoromisno Resenje; Akay and Koçyiğit, 2020), TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution; An et al., 2020), SAW (Simple Additive Weighting; Jato-Espino et al., 2019), SE (Shannon entropy; Jothibasun and Anbazhagan, 2016), LR (logistic regression; Li et al., 2019), EBF (evidence belief function; Chowdhuri et al., 2020), WoE (weights-of-evidence; Paul et al., 2019), FR (frequency ratio; Ullah and Zhang, 2020), ANNs (artificial neural networks; Sarker et al., 2019), CART (classification & regression trees; Al-Abadi and Al-Najar, 2020), RF (random forest; Woznicki et al., 2019), DT (decision trees; Chen et al., 2020), and SVM (support vector machine; Band et al., 2020). Nevertheless, FSM has numerous limitations, including the need to determine a suitable methodological approach from a wide range of options, as each approach yields a distinct outcome. Many FSM methods exhibit certain limitations that can be addressed using the ensemble modeling approach (Band et al., 2020). Ensemble modeling is the process of integrating a standalone model into a combined model to improve the prediction accuracy (Bui et al., 2018; Hassan et al., 2021). Researchers have

previously introduced a wide range of ensemble MLAs to precisely carry out FSM (Choubin et al., 2019; Malik et al., 2020a; Shahabi et al., 2020a). Moreover, deep learning methods have witnessed significant progress in recent years, especially in the field of image analysis (Zhao and Du, 2016; Paoletti et al., 2018). Because of a deeper algorithmic structure, they may help to understand the performance of flood susceptibility modeling (Shahabi et al., 2020b; Wang et al., 2020; Yi et al., 2020). However, there is a lack of consensus regarding the selection of an appropriate approach for mapping flood susceptible areas, especially in data-poor countries such as Bangladesh. This study aimed at clarifying two research questions: (i) what are the influencing factors affecting flood hazards, and (ii) what is the contribution of anthropogenic influences to flood hazards, particularly in areas that are frequently inundated. To resolve these questions, the objectives are to: (i) evaluate the efficacy of MLAs integrated with geospatial techniques in predicting flood susceptibility; and (ii) identify the relationships between LCC, population growth (PG), RD, and flooding. Nonetheless, to the best of our knowledge, the BRBP algorithm and its ensemble with CART and EBF have not been employed in studying environmental hazards such as FSM. Thus, the use of these techniques can be regarded as effective tools for predicting flood susceptibility. Moreover, the results could provide the efficiency of the machine learning approach as opposed to traditional techniques in determining the spatial extent of floods. This could be a novel contribution to the scientific community. Although many studies have examined the effects of land use changes on flooding, the relationship between multiple anthropogenic variables (e.g., LCC, infrastructural growth, PG) and flooding remains unclear. This study aims to contribute to this area. The study is, therefore, novel, exciting, and is of general interest to readers/resource managers, dealing with flood risk management. The proposed methodology may help researchers

to examine how anthropogenic activities are driving floods to be worse. Furthermore, policymakers would greatly benefit from the findings of this study to manage flood-related issues.

2. Materials and methods

2.1. Study area

The Surma River basin of northeast Bangladesh located at 25°10'N and 91°30'E (Fig. 1), which includes the entire district of Sunamganj, was considered as the study area. It comprises nine unions, 268 villages, and 49,557 households, and covers an area of 286.36 km². Males and females constitute 51.27% and 48.73% of the total population, respectively (BBS, 2011). The area is situated near the bank of the Surma River, which is one of the longest rivers in Bangladesh. This area is rich in resources such as natural gas, oil, industrial raw materials, crops, and fish. The economy of Sunamganj is predominantly agriculture. Geologically, the Surma is a foreland basin and predominantly comprises siltstones, mudstones, and alternating sandstones (mid to late Miocene). The arenaceous Bhuban formation and argillaceous Bokabil formation are the two lithological units of the Surma group (Reimann, 1993). Tectonically, this area is influenced by the movement of the Eurasian plate, Indian plate, and Burma microplate. Aman and Boro are the major crops cultivated in floodplains and seasonal wetlands, respectively. The climate of the area is relatively warm (23 °C during August–October) in the summer and cool (7 °C, January) during the winter (Hossain and Paul, 2019). In addition, the study area is divided into two major geomorphic units: lowlying and hilly areas (Ovi et al., 2015). Topographically, surface elevation varies from 0 to 44 m above sea level (a.s.l.) and it is relatively flatland (Fig. 1). The annual minimum and maximum temperatures are 8.6 and 32.7 °C, and the mean monthly relative humidity is 73.7% (BBS, 2011). Heavy rainfall occurs during the monsoon (BBS, 2011), and the mean annual rainfall is approximately 3080 mm (Sharif et al., 2017).

Detailed information on specific aspects of the workflow is outlined below (Fig. 2).

2.2. Land cover analysis

In this study, the CART algorithm was used in the Google Earth Engine (GEE) (<https://code.earthengine.google.com/>) to develop multi-temporal land cover maps of the study area between 2000 and 2017. CART is a non-parametric algorithm, capable of detecting input-parameter relationships and splitting them into nodes based on spectral similarity. Further information on the CART algorithm is presented in section 2.5.5. The GEE aids in the storage of satellite images and the performance of complex calculations in the cloud. Calibrated top-of-atmosphere (TOA) reflectance products from Landsat 7 (2000) and Landsat 8 (2014 and 2017) were used to classify the land cover (LC). Then, using the JavaScript application programming interface (API) in GEE, a visual analysis of images was performed and features of individual land cover categories were collected (with 1133 element samples—318 for forest, 305 for built-up, 350 for agricultural land, and 160 for waterbodies). Furthermore, selected reference samples were randomly split into two categories: training (70%) and testing (30%) (Ghorbanian et al., 2020). Training datasets were used to generate land cover maps, and the remaining 30% of the data was used to examine accuracy and classification performance.

Following Calderón-Loor et al. (2021), green, red, blue, near-infrared, shortwave infrared 1, and shortwave infrared 2 bands of Landsat ETM+ were used in this study. Obtaining a high classification accuracy can be a challenging task with the use of only spectral bands (Naboureh et al., 2020), four spectral indices, that is, soil-adjusted vegetation index (SAVI), normalized difference vegetation index (NDVI), normalized difference built-up index (NDBI), normalized difference water index (NDWI; see Supplementary Data, Table S1) and topographic data (i.e., elevation, aspect, and slope) were used to enhance spectral classification. Furthermore, kappa

coefficients and producer and user accuracy metrics were calculated to assess the overall performance of the classification of land cover maps.

The single land cover dynamic index (SLCDI) and bilateral land cover dynamic index (BLCDI) are often used to quantify changes in the intensity of land cover conversion. These indices can be used to identify internal variations and transitions of LC by calculating the loss and gain of each LC category over a time period. In addition, the variation of LC intensity was also determined for each study period using the integrated bilateral land cover dynamic index (IBLCDI). SLCDI, BLCDI and IBLCDI can be defined as follows (Ji et al., 2014; Damtea et al., 2020):

$$SLCDI = \frac{(LA_E - LA_S)}{LA_S} \times \frac{1}{T} \times 100\% \quad (1)$$

$$BLCDI = \frac{(\sum LA_{S,E} + \sum LA_{E,S})}{LA_S} \times \frac{1}{T} \times 100\% \quad (2)$$

$$IBLCDI = \frac{(\sum LA_{S,E} + \sum LA_{E,S})}{\sum LA_S} \times \frac{1}{T} \times 100\% \quad (3)$$

In Eqs. (1) to (3), LA_S and LA_E are areas of land cover type at the beginning and end of the study period, respectively; T is the monitoring period; $LA_{S,E}$ is the area converted from S th to E th type, denotes loss; and $LA_{E,S}$ is the area converted from E th to S th type and denotes gain.

2.3. Population growth

Population data for 2000, 2014, and 2017 were obtained from the LandScan dataset (<https://landscan.ornl.gov/landscan-datasets>). Further, to calculate the PG, we divided the population by grid size (30 m \times 30 m) using resampling tools. Finally, the PG in each pixel is calculated using the following equation:

$$PG = \frac{P_E - P_S}{P_S} \times \frac{1}{T} \times 100\% \quad (4)$$

where P_S and P_E are the population per pixel at the start and end times in the study period, respectively, and T is the monitoring period.

2.4. Road density

Road data of the study area were obtained from the Local Government Engineering Department (LGED) at a scale of 1:800,000. It is worth noting that national highways, regional highways, and district roads were considered. RD values were calculated using the kernel density function.

2.5. Flood susceptibility modeling

2.5.1 Preparation of flood inventory map

The mapping of the flood extent aids in determining of its relationship with influencing factors. In this study, three flood inventory maps for 2000, 2014, and 2017 were prepared using Landsat 7 (ETM+), Landsat 8, and Sentinel-1 satellite images, respectively, and 1500 (500 for each year) flood points were considered. Landsat 7 and 8 images were obtained from the Earth Explorer (<https://www.earthexplorer.usgs.gov>) and processed in ArcGIS (v. 10.5) to develop flood inventories for 2000 and 2014. The 2017 flood inventory was prepared using JavaScript API in the GEE platform (<https://code.earthengine.google.com>) considering Sentinel-1 SAR: C-band (Synthetic Aperture Radar). To delineate flooded areas, satellite images consisting of pre- and post-flood periods were considered for each flooded year. Water and non-water pixels were combined to obtain the maximum waterbodies. The waterbodies that appeared in both images (pre- and post-floods) were considered permanent waterbodies (rivers, wetlands, ponds, etc.), and non-water pixels were deemed non-flooded areas. Hence, the areas depicting waterbodies in both images but non-waterbodies in the pre-flood image were considered to be flooded areas (Islam and Sado, 2000). Topographic data, historical flood maps were collected from the Department of

Disaster Management, Bangladesh and Flood Forecasting and Warning Center (FFWC), and flood points were cross-checked and validated. In this study, the appearance of a flood was assigned a value of 1, while the disappearance of flood from a grid was assigned a value of 0 to prepare flood inventories. In terms of MLAs, observation datasets should be classified into two categories: training and testing (Ali et al., 2019). Hence, these flood inventories were randomly split into two dissimilar categories for flood points: training dataset (60%, 900 points, where 300 points were assigned to each inventory) and testing dataset (40%, 600 points, where 200 points were assigned to each inventory) (Cearns et al., 2019). Similarly, we randomly selected non-flood points for training (60%) and testing (40%) datasets. The training dataset was used to build the MLAs and isolate key influencing factors for FSM, while the testing dataset was used to validate the outcomes of the MLAs (Costache et al., 2020c; Khosravi et al., 2020).

2.5.2 Identifying factors that influence floods

To determine flood-susceptible locations, a number of factors influencing the susceptibility of an area were studied. As the influencing factors may vary (Tehrany et al., 2019; Chen et al., 2020), a range of factors, including elevation, slope, aspect, NDVI, mean monsoonal rainfall, plan curvature, drainage density, population density, land cover, proximity to rivers, proximity to roads, geology, and soil texture, were selected based on existing literature, availability of data, and physical and environmental characteristics of the study area. The reasons for selecting these factors are explained in Table 1.

The factors derived from DEM, that is, elevation, slope, aspect, curvature, drainage density, etc., have a different spatial resolutions (30 m). At the same time, geology, soil texture, and proximity to roads were digitized at a scale of 1:100,000 and 1:800,000. Owing to these different spatial resolutions, it was challenging to avoid uncertainties. All factor maps were resampled to a $30\text{ m} \times$

30 m grid, assuming that the same grid size would be useful. This is also a common practice in geospatial data (Yi et al., 2020). However, based on the model performance, a spatial resolution of 30 m was deemed acceptable. In addition, Merghadi et al. (2020) mentioned that very high spatial resolution factors such as DEM and its derivatives may not improve the precision of susceptibility modeling (e.g., flood).

2.5.3 Feature selection and multicollinearity analysis

Information gain ratio (IGR) is a technique that can be used for feature selection. It is one of the most common methods for evaluating significant factors affecting flooding, and improving the predictive efficiency and performance of the models (Al-Abadi, 2018). This study used Pearson's correlation and multicollinearity testing (MCT) of the input variables to determine the correlation between the variables of interest. A Pearson's correlation value of ≥ 0.70 , tolerance (TOL) of < 0.1 , and variance inflation factor (VIF) of > 5 are commonly considered signs of multicollinearity (Nguyen et al., 2020).

2.5.4 Bayesian regularization back propagation neural network

A BRBP neural network is a multi-layer neural network that typically consists of three layers: input layers (ILs), one or more hidden layers (HLs), and an output layer (OL). This MLN is computationally intensive but can result in a strong generalization of complex, limited, or noisy datasets. In this study, the "trainbr" algorithm with the command "nftool" in MATLAB (v. R2017a) was used to implement the BRBP neural network for FSM (Hateffard et al., 2019). The "trainbr" can train any network provided its weight, net input, and transfer functions are derivative. In addition, the BRBP neural network minimizes the linear combination of the square

errors and weights. Further details on the BRBP neural network have been reported by MacKay (1992) and Gouravaraju et al. (2020).

2.5.5 Classification and regression trees

CART is a non-parametric method proposed by Breiman et al. (1984). It can be used to establish a predictive relationship between input variables (categorical and continuous) (Choubin et al., 2018). A benefit of this is that large datasets can be minimized to limited datasets. Among numerous types of DT models, for example, ID₃, QUEST, C_{4.5}, CART, and CHAID, the CART algorithm performs well with nonlinear systemic processes with significant internal variations (Choubin et al., 2019). In the present study, the CART algorithm is implemented using “*RegTree*” command in MATLAB (v. R2015a; Hateffard et al., 2019). For details on CART, refer to studies conducted by Breiman et al. (1984) and Costache et al. (2020b).

2.5.6 Evidence belief function

EBF is a statistical model based on the Dempster-Shafer theory which identifies spatial integration based on a combination law (Dempster, 1968). The EBF model is an optimal approach for flood susceptibility mapping, but its use is limited (Tehrany and Kumar, 2018; Chowdhuri et al., 2020; Ramesh and Iqbal, 2020). The EBF statistical model can be quantified in two ways, data-driven and knowledge-driven models (Tehrany and Kumar, 2018). In this study, a data-driven EBF model was considered to yield the desired outcome. The data-driven EBF model has four components: (i) belief function (*Bel*), disbelief function (*DisBel*), (iii) uncertainty (*Unc*), and (iv) plausibility (*Pls*). The values of these four functions varied from zero to one (Supplementary Data, Fig. S3, Carranza and Hale, 2003). Tehrany and Kumar (2018) and Chowdhuri et al. (2020) discussed the statistical EBF data-driven model in detail. The combined EBFs for *Bel*, *DisBel*, *Unc*, and *Pls* maps of each factor influencing flooding were developed (Supplementary Data, Fig. S4).

2.5.7 Ensemble modeling

A few weaknesses have been observed in the aforementioned individual models, as they usually do not yield desirable outcomes; hence, many researchers recommend the use of ensemble models to overcome specific model limitations (Choubin et al., 2019; Chowdhuri et al., 2020). Ensemble modeling is a procedure that integrates the outputs of individual models into a single combined model. In this study, an ensemble model was developed using simple weighted averaging, based on the area under the curve—receiver operating characteristic (*AUC-ROC*) as expressed in the equation below:

$$EM = \frac{\sum_{i=1}^n (AUCROC_i^{Su} \times SM_i)}{\sum_{i=1}^n AUCROC_i^{Su}} \quad (5)$$

where *EM* indicates the ensemble model, and $AUCROC_i^{Su}$ represents the *AUCROC* value for the success rate of the standalone model (SM_i ; Fang et al., 2020).

2.5.8 Validation and performance of the models

In this study, the area under the curve—receiver operating characteristic (*AUC-ROC*) was plotted to validate flood susceptibility maps for both standalone and ensemble models. In the *AUC-ROC*, the “X” axis indicates the false positive (FP) rate or 1-specificity, while the “Y” axis indicates the true positive (TP) rate or sensitivity (Liuzzo et al., 2019). The *AUC-ROC* value varies from 0.00 to 1.00, where 0.50–0.60 denotes poor; 0.61–0.70 denotes average; 0.71–0.80 denotes good; 0.81–0.90 denotes very good; and 0.91–1.00 denotes excellent sensitivity (Yesilnacar and Topal, 2005). The *AUC-ROC* was separately calculated for training (success rate) and testing (prediction rate) datasets, and testing datasets were not used to develop the model.

Furthermore, the performance of predictive models (standalone and ensemble models) was graphically analyzed using a Taylor diagram (Yaseen et al., 2018), which is an effective technique

for illustrating the likelihood of the models to predict target variables (i.e., flood susceptible areas). In the Taylor diagram, statistical indices such as correlation coefficients and standard deviation (SD) were derived to compare model outputs and actual results.

2.5.9 Flood susceptibility mapping

Flood susceptibility maps of the study area were developed for three time periods (e.g., 2000, 2014, and 2017) considering the optimum model selected from the AUC-ROC and Taylor diagram. Finally, flood susceptibility was categorized into four classes: low, moderate, high, and very high susceptibility using the natural breaks classification method (Dent, 1993).

2.5.10 Relative changes in flooding areas

To analyze the correlations between flood susceptible areas with land cover, population, and RD, the relative changes in flooding (RCF) areas were calculated for the same period with a grid size of 30 m × 30 m. The RCF can be derived using the following equation (Costache et al., 2020a):

$$RCF = \frac{FS_E - FS_S}{FS_S} \times \frac{1}{T} \times 100\% \quad (6)$$

where FS_S indicates flood susceptibility in the initial year, FS_E denotes the ending year; and T is the monitoring period.

2.6. Geo-statistical analysis for spatial correlation

2.6.1 Moran's I index

The spatial correlation (Moran's I) was calculated from the residuals obtained from geographically weighted regression (GWR) model (Fotheringham et al., 2003). This method explains the level of correlation of the input data based on the feature distribution and the significance of attributes.

That is, it determines the similarity of features. Positive Moran's I denotes clustering, while a negative value represents dispersion (Moran, 1950).

2.6.2 Multiple linear regression analysis

Multiple linear regression (MLR) was performed to examine the relationship between different LCC scenarios and the relative changes in flood areas. Twelve scenarios (Sc.)—Sc. 1: agricultural land to built-up area, Sc. 2: agricultural land to forest, Sc. 3: agricultural land to waterbodies, Sc. 4: built-up area to agricultural land, Sc. 5: built-up area to forest, Sc. 6: built-up area to waterbodies, Sc. 7: forest to agricultural land, Sc. 8: forest to built-up area, Sc. 9: forest to waterbodies, Sc. 10: waterbodies to agricultural land, Sc. 11: waterbodies to built-up area, and Sc. 12: waterbodies to agricultural land—were selected as input variables to perform MLR analysis (Uyanık and Güler, 2013).

2.6.3 Rose diagram construction

The rose diagram was constructed based on the range value, obtained from the semi-variogram statistics. Additional details can be sourced from studies conducted by Mazzella and Mazzella (2013) and Obroślak and Dorozhynskyy (2017). The semi-variogram was plotted for different LCC scenarios with relative flood changes in angular directions (0° to 360°) and ranges were calculated to determine the degree of correlation between the LCC and RCF areas.

3. Results

3.1 Land cover classification and accuracy assessment

The classified land cover maps of the study area are shown in Fig. 3. The area (km^2) covered by each land-cover class for a given period is summarized in Supplementary Data (Table S2). Confusion matrix and kappa coefficients were calculated to determine the accuracy of satellite-

derived land-cover classification. The results revealed that the overall accuracy levels of the land cover for 2000, 2014, and 2017 were 89.57%, 97.75%, and 92.36%, respectively (Table 2). The kappa coefficients were 0.76, 0.95, and 0.83, respectively, indicating satisfactory performance (Table 2). Furthermore, the user and producer accuracies were calculated using the confusion matrix for the given years (Table 2).

3.2 Land cover change

The Markov transfer matrix was applied to detect land-cover transitions at different times (e.g., 2000–2014, 2014–2017, and 2000–2017) (Table 3). Diagonal values depict the proportion of each land cover class that remained unchanged within the year ranges. The transition results revealed that, for the period ranging from 2000 to 2014, 53.88% (150.36 km²) of land cover remained unchanged, while for the period ranging from 2014 to 2017 and 2000 to 2017, total land cover spanning 155.69 km² (55.78%) and 159.63 km² (57.20%), respectively, remained unchanged (Supplementary Data, Fig. S1). There was a strong indication of land cover transformation within the study area over a period of 17 years. Agricultural land is the predominant land cover class and occupies more than 50% of the total area for all three years studied, suggesting that agricultural land played a significant role in the socioeconomic growth of the study area (Fig. 3 and Supplementary Data, Table S2). Agricultural land areas generally decreased, particularly owing to conversion to forests for the periods ranging from 2000 to 2014 and 2000 to 2017, while an inverse relation was observed for the period 2014 to 2017 (Table 3 and Fig. 5). The built-up area was 7.32 km² (2.62%), 9.70 km² (3.47%), and 13.88 km² (4.97%) for 2000, 2014, and 2017, respectively, expanded by 2.38 km² (0.85%), 6.56 km² (2.35%), and 4.18 km² (1.50%) for 2000–2014, 2000–2017, and 2014–2017, respectively. This can be explained by the increase in population, industrial growth, expansion of road networks, and conversion of agricultural lands, forests, and waterbodies

into built-up areas. The reduction in waterbodies may be closely related to residential development. However, the water area increased from 2014 to 2017. The forest area expanded from 2000 to 2014 and from 2000 to 2017, while it decreased from 2014 to 2017 (Fig. 4).

Furthermore, the comparative LCC rate for forests from 2000 to 2017 was 163%, demonstrating a substantial increase in this category. Built-up areas changed from 2.62% in 2000 to 4.97% in 2017 at a comparative rate of 89.62% (Supplementary Data, Tables S2 and S3). By 2017, agricultural land reduced to 9.28% with an area reduction of approximately 25.94 km² and a rate of change of -14.58% (Supplementary Data, Tables S2 and S3). Subsequently, waterbodies reduced by approximately 11.53 km² (4.14%) from 2000 to 2017 with a comparative rate of change of -15.31% (Supplementary Data, Tables S2 and S3).

3.3 Dynamic index for land cover changes

Table 4 depicts the SLCDI for each time period (2000–2014, 2014–2017, and 2000–2017). Built-up and forest areas exhibited the highest SLCDI values over 17 years. From 2014 to 2017, SLCDI for forest areas decreased (10.43%), while it increased from 2000 to 2014 and 2000 to 2017 by 15.56% and 9.64%, respectively. The SLCDI values of waterbodies for long-term transition (2000–2014 and 2000–2017) exhibited a decreased by 1.49% and 0.90%, respectively, while for short-term transition (2014–2017), it increased to 4.54%. A similar pattern was observed for the agricultural land cover. Based on the BLCDI, LC transformation rates and the transitions of LC types during the short- and long-term transition were calculated and are summarized in Table 4. Throughout the short- and long-term transition periods, BLCDI for the built-up area was significantly higher; the newly developed built-up area was mostly transformed from agricultural land and waterbodies. This occurred as a result of the expansion of urban areas and construction of various types of infrastructure. More specifically, the increase in population resulted in a higher

demand for housing and industrial growth, and consequently, the built-up area increased at the expense of agricultural land and waterbodies. The results of IBLCDI showed that the overall long-term dynamic transition intensity was 13.32% (2000–2014) and 11.75% (2000–2017), while the short-term intensity was 99.91% (2014–2017) (Table 4). The integrated dynamic index for monitoring from 2014 to 2017 (short-term) was observed to be seven times higher than the index value calculated from long-term monitoring (2000–2017). Furthermore, the LC transition from the monitoring of 2014–2017 was faster than that of other monitoring years (2000–2014 and 2000–2017). The faster transition was attributable to higher bilateral dynamic changes (Table 4).

3.4 Feature selection and multi-collinearity test of flood influencing factors

The results of the IGR for three different periods among 13 influencing factors, elevation, slope, NDVI, aspect, mean monsoon rainfall, land cover, plan curvature, soil texture, drainage density, proximity to rivers, proximity to roads, and population density, are shown in Fig. 5. The IGR value was directly proportional to the level of importance. The IGR value for geology (2000, 2014, and 2017) was observed to be “null,” hence, geology was excluded from the FSM.

The results of the Pearson’s correlation indicated that the values did not exceed 0.70, which confirmed that the factors were free from multicollinearity (Fig. 6). In addition, the TOL and VIF were determined by conducting a multi-collinearity test, and the results revealed that all factors had values of $TOL > 0.1$ and $VIF < 5$, which are within the acceptable limits (Table 5).

3.5 Construction of flood susceptibility maps

In the BRBP neural network model, the training was stopped after 1000, 648, and 1000 iterations for the training datasets of 2017, 2014, and 2000, where the final structures of the neural network were confirmed as $ILs:12 \times HLs:15 \times OL:01$, $ILs:12 \times HLs:16 \times OL:01$, and $ILs:12 \times HLs:18 \times OL:01$.

The R^2 and mean square error (MSE) values were $7.02e-01$ and $1.27e-01$ for 2017, $8.30e-01$ and $8.18e-01$ for 2014, and $7.44e-01$ and $1.19e-01$ for 2000, respectively, for training datasets. The findings were consistent with our assumption regarding the precision of the flood susceptibility assessment. In the CART model, flood susceptibility assessment was performed using a tree-like structure. MSE and R^2 correspond to our expectations regarding the accuracy of the flood susceptibility analysis. The MSE was $3.27e-01$, $4.18e-01$, and $2.19e-01$ for 2017, 2014, and 2000, whereas, R^2 was $7.05e-01$, $8.01e-01$, and $7.23e-01$, respectively. In the EBF model, a flood susceptibility map was developed from the integrated maps of *Bel*, *DisBel*, *Unc*, and *Pls*. The MSE values were $4.13e-01$, $8.82e-01$, and $9.12e-01$ while the R^2 values were $4.53e-01$, $4.25e-01$, and $4.02e-01$ for the years 2017, 2014, and 2000, respectively.

Furthermore, the ensemble model (BRBP-CART-EBF) was used to generate the flood susceptibility map for the three time periods (2000, 2014, and 2017). The results showed that the MSE and R^2 values were $2.26e-01$ and $7.58e-01$ for 2017, $2.04e-01$ and $8.58e-01$ for 2014, and $1.79e-01$ and $7.68e-01$ for 2000, respectively, confirming that the ensemble model outperformed the standalone models. Finally, areas in maps generated from the standalone and ensemble models were categorized into four classes, low, moderate, high, and very high flood susceptibility using natural break classification (Fig. 7 and Supplementary Data, Fig. S5). The percentages of the specific flood susceptibility classes are summarized in Fig. 8.

3.6 Validation and model performance

The accuracies of the standalone and ensemble models were evaluated by plotting the (AUC-ROC). The model's training datasets (60%) were used to plot the success rate curve (SRC), but were unable to evaluate the model's predictive capability. Hence, the prediction rate curve (PRC) was used to determine the accuracy of flood prediction. The SRCs and PRCs for the 2017 flood

are shown in Fig. 9a and b, respectively. The AUC results for the SRCs and PRCs for 2014 and 2000 are summarized in Supplementary Data (Supplementary Data, Tables S5 and S6). The results revealed that the SRC values for the ensemble model (EM4) were 0.927, 0.960, and 0.934 for 2017, 2014, and 2000, which exhibited optimal accuracy compared to other standalone and ensemble models (Fig. 9a and Supplementary Data, Table S5). Likewise, the PRC values showed that EM4 had the highest predictive accuracy for the three years, with AUC values of 0.940, 0.959, and 0.943 for 2017, 2014, and 2000 (Fig. 9b and Supplementary Data, Table S6). The AUC values for both SRCs and PRCs were low for standalone and other ensemble models for 2017, 2014, and 2000 (Fig. 9a and b, Supplementary Data, Tables S5 and S6). Thus, EM4 can logically be inferred to have exhibited the highest accuracy (AUC > 90%) throughout the study period.

Furthermore, the aggregated accuracy of the models was checked using the Taylor diagram and the results showed that EM4 outperformed other standalone and ensemble models (Fig. 10).

3.7 Relationships between flooding with land cover change, population growth, and road density

3.7.1 Spatial pattern assessment

The results of the Moran's I index, derived from spatial correlation are presented in Table 6 which shows that the index values for three pairs (RCF-LC, RCF-PG, and RCF-RD) were positive and close to 0. These results revealed that the pairs had a clustered distribution (Fig. 11) and were significant for checking the spatial correlation. Furthermore, the derived critical values (z) exceeded 2.58, indicating a <1% probability that clustered patterns would have occurred randomly. Hence, relative changes in flood areas can be inferred to have been associated with LCC, increase in population, and growth of road networks.

3.7.2 Multiple linear regression

The MLR results as a part of the correlation analysis between the RCF areas and different scenarios of LCC are presented in Table 7. The standardized coefficients for scenarios 1, 3, and 7 showed a high degree of influence on the extent of floods, with coefficients of 0.707, 0.500, and 0.408, respectively. Meanwhile, scenarios 4, 5, 9, and 12 showed a negative influence on floods with corresponding coefficients of -0.010 , -0.011 , -0.012 , and -0.001 .

Table 8 summarizes rank-wise flood susceptible areas according to land cover categories from 2000–2017. For land cover scenarios of 2000, the low flood susceptible area was small (susceptible area of 28.78 km^2), the area ranked very high in flood susceptibility measured 59.32 km^2 , and the overall flood susceptible area was 286.36 km^2 . For 2014, the areas ranked low and very high in flood susceptibility measured 21.07 km^2 and 78.86 km^2 , respectively. For 2017, the area ranked low in flood susceptibility measured 21.72 km^2 and that ranked very high in flood susceptibility measured 71.02 km^2 . Based on Table 8, areas that were very highly susceptible to floods can be inferred to have increased over time as a result of changes in land cover from 2000 to 2017. In contrast, areas that are highly susceptible to flooding decreased by 9.94% in the short term (2014–2017) as a function of LCC.

3.7.3 Assessment of semi-variogram and rose diagrams

Semi-variogram plots were generated for four scenarios of LCC with RCF areas because these scenarios provided useful information regarding the relationship between flood and LCC (Supplementary Data, Table S7). These results were used to draw a rose diagram to ascertain the presence or absence of spatial correlations between the two variables. The rose diagrams revealed that LCC and its direction were closely related to changes in flooding areas for scenarios 1, 3, and 7; however, these variables were dissimilar in scenario 12 (Fig. 12). Based on the results, changes in land cover had a significant impact on flooding in the study area from 2000 to 2017.

3.7.4. Population growth and road density

An increased population within an area translates to increased anthropogenic activities, which enhances the likelihood of flood occurrence. A positive correlation between an increase in population and flooding was observed in this study. Specifically, people living downstream and adjacent to the river network were exposed to a relatively greater risk of flooding. Table 9 summarizes the PG and rank-wise flood susceptible areas in 2000–2017. For low flood susceptibility rank, PG values decreased by 19.71%, 1.47%, and 20.89% in 2000–2014, 2014–2017, and 2000–2017, respectively. Meanwhile, for moderate, high, and very high flood susceptibility ranks, PG values increased except for very highly susceptible areas during the period 2014–2017. These results suggest a significant correlation between flooding and the population in the study area.

Furthermore, the assessment of RD is an important factor in the occurrence of floods suggesting that the distribution of roads was directly related to flooding (Table 10). The Pearson correlation coefficient of the RCF areas and RD was found to be 0.496 (Table 10).

4. Discussion

Flooding is a natural phenomenon that cannot be avoided entirely, and noticeably affects livelihoods and socioeconomic development in affected areas. However, the lack of knowledge on the spatial variability of floods can result in poor understanding of its management. Therefore, flood susceptibility mapping is an important step toward flood risk management and decreasing flood damage. Several factors influence the occurrence of floods. The selection of influencing factors was based on IGR coefficients. Drainage density, rainfall, slope, NDVI, and LC were found to be the dominant factors affecting floods in the study area. However, geology has a minimal effect on floods. These findings are consistent with those of the studies conducted by Chowdhuri

et al. (2020) and Costache et al. (2020a, 2020b). According to our results, the standalone BRBP algorithm achieved the highest accuracy and outperformed other standalone models, whereas, the EBF bivariate statistical model yielded considerably lower performance ($AUC < 70\%$) in FSM for the three study periods (2000, 2014, and 2017). This can be explained by the fact that the EBF model cannot calculate the weight of the influencing factors and does not consider correlations among factors (Tehrany et al., 2019). More precisely, EBF requires a reasonable hypothesis, and is more dependent on influencing factors (Zhao and Chen, 2020). Meanwhile, MLAs use two approaches, that is, pre-pruning and post-pruning, to avoid over-fitting, which increases the precision (Khosravi et al., 2019). As both training and testing datasets performed well ($AUC > 90\%$ for both datasets in three years; 2000, 2014, and 2017) for the ensemble model (BRBP-CART-EBF) and outperformed standalone and other ensemble models, this model was used to further calculate flood susceptible areas. This can be justified by the fact that each model has limitations that can be minimized or eliminated in the ensemble model. The outcome of this study is in line with those of studies conducted by Choubin et al. (2019), Khosravi et al. (2019), Arora et al. (2020), and Talukdar et al. (2020), wherein ensemble models showed the best predictive ability for geographic phenomena such as floods. As the ensemble model reduces variance, bias, and noise in susceptibility modeling, modeling using this approach can yield precise flood susceptibility data (Al-Abadi, 2018; Islam et al., 2020; Malik and Pal, 2020; Malik et al., 2020b). Further, this study explored the relationship between LCC, PG, and increase in RD with relative changes in flood areas for the period between 2000 and 2017. Our results indicated that changes in land cover between 2000 and 2017 resulted in increased flood occurrence, especially in the downstream section of the study area. Flood incidents were more apparent in residential areas than in grasslands and barren land, because built-up areas hinder the infiltration of water and inhibit

surface runoff. Although open and grassland areas cannot hinder the flow of water (Yin et al., 2017; Costache, 2019; Talukdar et al., 2020), dynamic changes in land cover, especially an increase in built-up area and decrease in forest area, increased the susceptibility of the study area to flooding as impervious surfaces are relatively more likely to generate surface runoff (Zope et al., 2016; Hassan et al., 2017; Mojaddadi et al., 2017; Pal and Chowdhuri, 2019; Adnan et al., 2020).

The increase in population density has caused rapid urbanization in Bangladesh. The changes in hydro-morphological cycles, reduction of natural drainage capacity and flooding in densely populated areas are mainly attributable to rapid urbanization. The findings of this study are consistent with those of previous studies conducted by Amarnath et al. (2015), Li et al. (2019), and Sarkar and Mondal (2020), which showed that floods had greater effects on densely populated areas than on sparsely populated areas. More specifically, an increase in population and urbanization increases the total surface runoff and streamflow of a river (Dewan and Yamaguchi, 2009). Increased flooding is significantly influenced by residential, manufacturing, farming, and other infrastructural demands, and is a significant issue for the economic viability and sustainable development of communities (Rawat et al., 2013; Bazai et al., 2021b). Additionally, roads generally act as artificial inhibitors of flooding in the area. Consequently, a higher RD can reduce the likelihood of flooding (Sarkar and Mondal, 2020). However, the current study found that flooding increased as RD increased. This dissimilarity with the previous study occurs because roads disrupt water storage by floodplains and affect the movement of natural sediments. Specifically, roads may affect the volume of infiltrated water. This finding is in agreement with those of studies conducted by Forman et al. (2003), Douven et al. (2009), and Kalantari et al. (2014), where RD was associated with flooding.

This study had three limitations. The first is the unavailability of relevant hydrological information such as flood depth, velocity, and discharge, precluding the development of a robust model. In addition, elevation data with 30 m spatial resolution appeared to be too coarse, especially in areas with flat topography. Unfortunately, a high-resolution elevation dataset is not available for monsoon regions such as Bangladesh. Finally, updated land use information for the study area is another constraint.

This study shows that flood hazards in Bangladesh are increasing as a result of increased floodplain development. The proposed algorithm can be a practical and novel approach to flood risk management in northeast Bangladesh and in similar settings. In addition, the findings could inform the development of adaption measures at a basin-scale and can guide the development of flood countermeasures by policymakers to minimize flood-related damage. Although the EM4 ensemble model was superior for flood susceptibility assessment in this case study, it can be tested in other environmental hazards such as landslides and debris flows.

As flooding is affected by growing anthropogenic activities (e.g., rapid changes in land cover) in combination with geographical factors (e.g., terrain characteristics), existing algorithms may be of little help. Hence, the proposed approach could provide a robust solution for quickly simulating flood maps in data-sparse regions. In addition, the primary focus of this work was to evaluate and compare the modeling performances of different approaches and to identify the best alternative. Although we applied this work to a specific area, the proposed method can be used in other areas with similar characteristics. Based on the available literature, we hope that this work will be helpful for future studies. Note that the accuracy of the outputs may be dependent on the environmental characteristics of the study area as well as input variables, wherein, hydrogeological and

topographical variations may exist. However, the method used in this study can be employed in other areas to understand its sensitivity.

5. Conclusions

This study used machine learning techniques and various factors to clarify flood-susceptible areas in the northeast region of Bangladesh. The results demonstrated that anthropogenic activities such as rapid LCC led to more frequent and severe flooding, and climate change may worsen the situation. Therefore, sophisticated techniques are required to determine areas that are more susceptible to flooding. Of the three standalone and four ensemble models, BRBP-CART-EBF (EM4) yielded the highest AUC values ($AUC > 90\%$ in the training and testing stages) for the three flood events evaluated (2000, 2014, and 2017). Thus, the flood maps produced from EM4 were considered accurate for delineating of areas that were susceptible to floods.

Furthermore, LCC that occurred in the downstream portion of the monitoring area was associated with changes in flooding areas between 2000 and 2017. In addition, flooding increases as RD and population increase. More specifically, a considerable decrease in waterbodies (+11.53% with $SLCDI = 0.858\%$) and agricultural land (+25.94% with $SLCDI = 0.90\%$) was observed, mostly due to rapid urbanization and population growth. This resulted in a decrease in the floodplain and disruption of floodwater by infrastructural development, resulting in greater flood susceptibility.

The resulting maps could inform strategies for reducing of flood vulnerability and aid in the development of adaptation measures. The major difference between the proposed approach and existing studies is that the ensemble algorithms, along with a standalone algorithm, are examined for the first time to predict floods. Hence, the novelty of this study contributes to new ensemble algorithms that can also be applied to other areas. In addition, the proposed ensemble algorithms are not only limited to floods; they can be utilized in different susceptibility mapping, for example,

landslides, debris flows, snow avalanches, etc. It is worth noting that the northeast region of Bangladesh has suffered severe damage caused by floods during monsoon months every year. However, no significant effort has been directed toward comprehending the anthropogenic influences, such as land cover change, population growth, and road infrastructure on floods. Therefore, this study aims to address this gap.

Likewise, the results of this study may serve as guidelines for sustainable land use and infrastructural planning for policymakers and city planners. Although rainfall in hilly areas of Bangladesh and beyond contributes to the extent and severity of flooding, inaccessibility to rainfall data at upstream points is an important limitation of this study. Future work could include this to make the work robust. In addition, the flood depth and velocity parameters within an integrated ensemble of hydrodynamic models and machine learning tools could be useful.

CRedit authorship contribution statement

Conceptualization: Rahman, M., Islam, M.M., and Dewan, A.; **Methodology:** Rahman, M.; **Software:** Rahman, M.; **Validation:** Rahman, M. and Pourghasemi, H.R.; **Formal analysis:** Rahman, M.; **Investigation:** Rahman, M., Ningsheng, C., Islam, M.M., Pourghasemi, H.R., and Dewan, A.; **Resources:** Rahman, M.; **Writing—original draft preparation:** Rahman, M.; **Writing—review and editing:** Ningsheng, C., Iftekhar, G.M., Islam, M.M., Pourghasemi, H.R., Ahmad, H., Habumugisha, J.M., Washakh, R.M.A., Alam, M., Enlong, L., Han, Z., Huayong, N., Dewan, A.; **Visualization:** Ningsheng, C., Islam, M.M., Pourghasemi, H.R., and Dewan, A.; **Supervision:** Ningsheng, C., Islam, M.M., Pourghasemi, H.R., and Dewan, A.; **Project administration:** Ningsheng, C.; **Funding acquisition:** Ningsheng, C.

Acknowledgments

This research was based on a PhD thesis by the first author: Mahfuzur Rahman, which was supervised by Professor Chen Ningsheng. This work was supported by the National Natural Science Foundation of China (Grant No. 41861134008), the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) of China (Grant No. 2019QZKK0902), the National Key Research and Development Program of China (Project No. 2018YFC1505202), and the Key R&D Projects of Sichuan Science and Technology (Grant No. 18ZDYF0329). The authors acknowledge and appreciate the provision of rainfall data by the Bangladesh Water Development Board, without which this study would not have been possible. Authors would like to thank the ‘Handling Editor: Prof. M. Santosh’ for his editorial comments and three anonymous reviewers for their careful review of the original manuscript and their valuable suggestions which helped us to improve the manuscript.

References

- Adnan, M.S.G., Abdullah, A.Y.M., Dewan, A., Hall, J.W., 2020. The effects of changing land use and flood hazard on poverty in coastal Bangladesh. *Land Use Policy*, 99, 104868. <https://doi.org/10.1016/j.landusepol.2020.104868>.
- Akay, H., Koçyiğit, M.B., 2020. Flash flood potential prioritization of sub-basins in an ungauged basin in Turkey using traditional multi-criteria decision-making methods. *Soft Comput.*, 1-13. <https://doi.org/10.1007/s00500-020-04792-0>.
- 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(2), 171-182. <https://doi.org/10.1002/2016EF000485>.
- Al-Abadi, A.M., 2018. Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study. *Arabian J. Geosci.*, 11(9), 218. <https://doi.org/10.1007/s12517-018-3584-5>.

- Al-Abadi, A.M., Al-Najar, N.A., 2020. Comparative assessment of bivariate, multivariate and machine learning models for mapping flood proneness. *Nat. Hazards*, 100(2), 461-491. <https://doi.org/10.1007/s11069-019-03821-y>
- Al Baky, M.A., Islam, M., Paul, S., 2019. Flood hazard, vulnerability and risk assessment for different land use classes using a flow model. *Earth Syst. Environ.*, 4, 225-244. <https://doi.org/10.1007/s41748-019-00141-w>
- Ali, S.A., Khatun, R., Ahmad, A., Ahmad, S.N., 2019. Application of GIS-based analytic hierarchy process and frequency ratio model to flood vulnerable mapping and risk area estimation at Sundarban region, India. *Model. Earth Syst. Environ.*, 5(3), 1083-1102. <https://doi.org/10.1007/s40808-019-00593-z>.
- Amarnath, G., Umer, Y.M., Alahacoon, N., Inada, Y., 2015. Modelling the flood-risk extent using LISFLOOD-FP in a complex watershed: case study of Mundeni Aru River Basin, Sri Lanka. *Proceedings of the International Association of Hydrological Sciences*. 370, 131-138. <https://doi.org/10.5194/piahs-370-131-2015>.
- An, Y., Tan, X., Gu, B., Zhu, K., 2020. Flood risk assessment using the CV-TOPSIS method for the Belt and Road Initiative: an empirical study of Southeast Asia. *Ecosyst. Health Sustainability*, 6(1), 1765703. <https://doi.org/10.1080/20964129.2020.1765703>.
- Arora, A., Arabameri, A., Pandey, M., Siddiqui, M.A., Shukla, U., Bui, D.T., Mishra, V.N., Bhardwaj, A., 2020. Optimization of state-of-the-art fuzzy-metaheuristic ANFIS-based machine learning models for flood susceptibility prediction mapping in the Middle Ganga Plain, India. *Sci. Total Environ.*, 750, 141565. <https://doi.org/10.1016/j.scitotenv.2020.141565>.
- Band, S.S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakraborty, R., Melesse, A.M., Mosavi, A., 2020. Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. *Remote Sens.*, 12(21), 3568. <https://doi.org/10.3390/rs12213568>
- Bazai, N.A., Cui, P., Zhou, K.J., Abdul, S., Cui, K.F., Wang, H., Zhang, G.T., Liu, D.Z., 2021a. Application of the soil conservation service model in small and medium basins of the mountainous region of Heilongjiang, China. *Int. J. Environ. Sci. Technol.* <https://doi.org/10.1007/s13762-021-03136-1>
- Bazai, N.A., Cui, P., Carling, P.A., Wang, H., Hassan, J., Liu, D., Zhang, G., Jin, W., 2021b. Increasing glacial lake outburst flood hazard in response to surge glaciers in the Karakoram. *Earth Sci. Rev.*, 212, 103432. <https://doi.org/10.1016/j.earscirev.2020.103432>

- BBS, 2011, Bangladesh Population and Housing Census-2011 (Zila Report: Sylhet), 2019. Bangladesh Bureau of Statistics.
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. CRC press.
- Bui, D.T., Ngo, P.T.T., Pham, T.D., Jaafari, A., Minh, N.Q., Hoa, P.V., 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. <https://doi.org/10.1016/j.catena.2019.04.009>.
- Bui, D.T., Panahi, M., Shahabi, H., Singh, V.P., Shirzadi, A., Chapi, K., Khosravi, K., Chen, W., Panahi, S., Li, S., 2018. Novel hybrid evolutionary algorithms for spatial prediction of floods. *Sci. Rep.*, 8(1), 15364. <https://doi.org/10.1038/s41598-018-33755-7>.
- Calderón-Loor, M., Hadjikakou, M., Bryan, B.A., 2021. High-resolution wall-to-wall land-cover mapping and land change assessment for Australia from 1985 to 2015. *Remote Sens. Environ.*, 252, 112148. <https://doi.org/10.1016/j.rse.2020.112148>
- Carranza, E.J.M., Hale, M., 2003. Evidential belief functions for data-driven geologically constrained mapping of gold potential, Baguio district, Philippines. *Ore Geol. Rev.*, 22(1-2), 117-132. [https://doi.org/10.1016/S0169-1368\(02\)00111-7](https://doi.org/10.1016/S0169-1368(02)00111-7).
- Cearns, M., Hahn, T., Baune, B.T., 2019. Recommendations and future directions for supervised machine learning in psychiatry. *Transl. Psychiatry*, 9(1), 1-12. <https://doi.org/10.1038/s41398-019-0607-2>.
- Chander, G., Markham, B.L., Helder, D.L., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.*, 113(5), 893-903. <https://doi.org/10.1016/j.rse.2009.01.007>.
- Chen, W., Hong, H., Li, S., Shahabi, H., Wang, Y., Wang, X., Ahmad, B.B., 2019. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J. Hydrol.*, 575, 864-873. <https://doi.org/10.1016/j.jhydrol.2019.05.089>.
- Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., Wang, X., Bian, H., Zhang, S., Pradhan, B., 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.
<https://doi.org/10.1016/j.scitotenv.2018.10.064>.
- 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. <https://doi.org/10.1007/s12665-018-7498-z>.
- Chowdhuri, I., Pal, S.C., Chakraborty, R., 2020. Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India. Adv. Space Res., 65(5), 1466-1489.
<https://doi.org/10.1016/j.asr.2019.12.003>.
- Chowdhury, E.H., Hassan, Q.K., 2017. Use of remote sensing data in comprehending an extremely unusual flooding event over southwest Bangladesh. Nat. Hazards, 88(3), 1805-1823. <https://doi.org/10.1007/s11069-017-2947-7>.
- Costache, R., 2019. 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.
<https://doi.org/10.1016/j.scitotenv.2018.12.397>.
- Costache, R., Bao Pham, Q., Corodescu-Roşca, E., Cimpianu, C., Hong, H., Thi Thuy Linh, N., Ming Fai, C., Najah Ahmed, A., Vojtek, M., Muhammed Pandhiani, S., 2020a. 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. <https://doi.org/10.3390/rs12091422>.
- Costache, R., Bui, D.T., 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. <https://doi.org/10.1016/j.scitotenv.2019.07.197>.
- Costache, R., Hong, H., Pham, Q.B., 2020b. 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. <https://doi.org/10.1016/j.scitotenv.2019.134514>.
- Costache, R., Pham, Q.B., Sharifi, E., Linh, N.T.T., Abba, S.I., Vojtek, M., Vojteková, J., Nhi, P. T.T., Khoi, D.N., 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.
<https://doi.org/10.3390/rs12010106>.

- Damtea, W.G., Kim, D., Im, S., 2020. Spatiotemporal analysis of land cover changes in the Chemoga Basin, Ethiopia, using Landsat and Google Earth Images. *Sustainability*, 12(9), 3607. <https://doi.org/10.3390/su12093607>.
- Dandapat, K., Panda, G.K., 2017. Flood vulnerability analysis and risk assessment using analytical hierarchy process. *Model. Earth Syst. and Environ.*, 3(4), 1627-1646. <https://doi.org/10.1007/s40808-017-0388-7>.
- Dempster, A. P., 1968. A generalization of Bayesian inference. *J. R. Stat Soc.: Series B.* 30(2), 205-232. <https://doi.org/10.1111/j.2517-6161.1968.tb00722.x>.
- Dent, B.D., 1993. *Cartography: Thematic Map Design, Guide*. WCB/McGraw-Hill.
- Dewan, A.M., Islam, M.M., Kumamoto, T., Nishigaki, M., 2007. Evaluating flood hazard for land-use planning in greater Dhaka of Bangladesh using remote sensing and GIS techniques. *Water Resour. Manage.*, 21(9), 1601. <https://doi.org/10.1007/s11269-006-9116-1>.
- Dewan, A.M., Yamaguchi, Y., 2009. Land use and land cover change in greater Dhaka, Bangladesh: using remote sensing to promote sustainable urbanization. *Appl. Geogr.*, 29(3), 390-401. <https://doi.org/10.1016/j.apgeog.2008.12.005>.
- Dewan, T.H., 2015. Societal impacts and vulnerability to floods in Bangladesh and Nepal. *Weather Clim. Extremes*, 7, 36-42. <https://doi.org/10.1016/j.wace.2014.11.001>.
- Douven, W., Goichot, M., Verheij, H., 2009. Best practice guidelines for the integrated planning and design of economically sound and environmentally friendly roads in the Mekong floodplains of Cambodia and Viet Nam, Synthesis Report of the 'Roads and Floods' Project (part of MRC-FMMP Component 2). MRC Technical Paper, 35.
- Fang, Z., Wang, Y., Peng, L., Hong, H., 2020. A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. *Int. J. Geog. Inf. Sci.*, 35(2), 321-347. <https://doi.org/10.1080/13658816.2020.1808897>
- Forman, R.T., Sperling, D., Bissonette, J.A., Clevenger, A.P., Cutshall, C.D., Dale, V.H., Fahrig, L., Heanue, K., France, R.L., Goldman, C.R., 2003. *Road Ecology: Science and Solutions*. Island Press.
- Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.
- Ghorbanian, A., Kakooei, M., Amani, M., Mahdavi, S., Mohammadzadeh, A., Hasanlou, M., 2020. Improved land cover map of Iran using Sentinel imagery within Google Earth Engine and a novel automatic workflow for

- land cover classification using migrated training samples. *ISPRS J. Photogramm. Remote Sens.*, 167, 276-288. <https://doi.org/10.1016/j.isprsjprs.2020.07.013>.
- Gouravaraju, S., Narayan, J., Sauer, R.A., Gautam, S.S., 2020. A Bayesian regularization-backpropagation neural network model for peeling computations. *arXiv preprint arXiv*. 2006.16409.
- Hassan, J., Kayastha, R. B., Shrestha, A., Bano, I., Ali, S.H., Magsi, H.Z., 2017. Predictions of future hydrological conditions and contribution of snow and ice melt in total discharge of Shigar River Basin in Central Karakoram, Pakistan. *Sci. Cold Arid Reg.*, 9(6), 0511-0524. [https:// DOI: 10.3724/SP.J.1226.2017.00511](https://doi.org/10.3724/SP.J.1226.2017.00511).
- Hassan, J., Chen, X., Muhammad, S., Bazai, N.A., 2021. Rock glacier inventory, permafrost probability distribution modeling and associated hazards in the Hunza River Basin, Western Karakoram, Pakistan. *Sci. Total Environ.*, 146833. <https://doi.org/10.1016/j.scitotenv.2021.146833>.
- Hateffard, F., Dolati, P., Heidari, A., Zolfaghari, A.A., 2019. Assessing the performance of decision tree and neural network models in mapping soil properties. *J. Mountain Sci.*, 16(8), 1833-1847. <https://doi.org/10.1007/s11629-019-5409-8>.
- Hossain, S., 2015. Local level flood forecasting system using mathematical model incorporating WRF model predicted rainfall. M.Sc. thesis, Bangladesh University of Engineering and Technology.
- Hossain, M.N., Paul, P., 2019. Impacts of climatic variability on agriculture and options for adaptation in the Surma River basin, Bangladesh. *Environ. Monit. Assess.*, 191(2), 111. <https://doi.org/10.1007/s10661-019-7256-z>.
- Islam, A.R.M.T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K.U., Pham, Q.B., Kuriqi, A., Linh, N.T.T., 2020. Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci. Front.*, <https://doi.org/10.1016/j.gsf.2020.09.006>.
- Islam, M.M., Sado, K., 2000. Development of flood hazard maps of Bangladesh using NOAA-AVHRR images with GIS. *Hydrol. Sci. J.*, 45(3), 337-355. <https://doi.org/10.1080/02626660009492334>.
- Jato-Espino, D., Lobo, A., Ascorbe-Salcedo, A., 2019. Urban flood risk mapping using an optimised additive weighting methodology based on open data. *J. Flood Risk Manage.*, 12, e12533. <https://doi.org/10.1111/jfr3.12533>.
- Ji, W., Wang, Y., Zhuang, D., Song, D., Shen, X., Wang, W., Li, G., 2014. Spatial and temporal distribution of expressway and its relationships to land cover and population: a case study of Beijing, China. *Transp. Res. Part D Transp Environ.*, 32, 86-96. <https://doi.org/10.1016/j.trd.2014.07.010>.

- Jothibasu, A., Anbazhagan, S., 2016. Flood susceptibility appraisal in Ponnaiyar River Basin, India using frequency ratio (FR) and Shannon's Entropy (SE) models. *Int. J. Adv. Rem. Sens. GIS*, 5(10), 1946-1962.
- Kalantari, Z., Nickman, A., Lyon, S.W., Olofsson, B., Folkeson, L., 2014. A method for mapping flood hazard along roads. *J. Environ. Manage.*, 133, 69-77. <https://doi.org/10.1016/j.jenvman.2013.11.032>.
- Kamal, A.M., Shamsudduha, M., Ahmed, B., Hassan, S.K., Islam, M.S., Kelman, I., Fordham, M., 2018. Resilience to flash floods in wetland communities of northeastern Bangladesh. *Int. J. Disaster Risk Reduct.*, 31, 478-488. <https://doi.org/10.1016/j.ijdrr.2018.06.011>.
- Khosravi, K., Panahi, M., Golkarian, A., Keesstra, S.D., Saco, P.M., Bui, D.T., Lee, S., 2020. Convolutional neural network approach for spatial prediction of flood hazard at national scale of Iran. *J. Hydrol.* 591, 125552. <https://doi.org/10.1016/j.jhydrol.2020.125552>.
- Khosravi, K., Shahabi, H., Pham, B.T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.B., Gróf, G., Ho, H.L., 2019. A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *J. Hydrol.*, 573, 311-323. <https://doi.org/10.1016/j.jhydrol.2019.03.073>.
- Li, X., Yan, D., Wang, K., Weng, B., Qin, T., Liu, S., 2019. Flood risk assessment of global watersheds based on multiple machine learning models. *Water*, 11(8), 1654. <https://doi.org/10.3390/w11081654>.
- Liuzzo, L., Sammartano, V., Freni, G., 2019. Comparison between different distributed methods for flood susceptibility mapping. *Water Resour. Manage.*, 33(9), 3155-3173. <https://doi.org/10.1007/s11269-019-02293-w>.
- MacKay, D.J., 1992. Bayesian interpolation. *Neural Comput.*, 4(3), 415-447. <https://doi.org/10.1162/neco.1992.4.3.415>.
- Malik, S., Pal, S.C., Chowdhuri, I., Chakraborty, R., Roy, P., Das, B., 2020a. Prediction of highly flood prone areas by GIS based heuristic and statistical model in a monsoon dominated region of Bengal Basin. *Remote Sens. Appl.: Soc. Environ.*, 19, 100343. <https://doi.org/10.1016/j.rsase.2020.100343>.
- Malik, S., Pal, S.C., 2020. Application of 2D numerical simulation for rating curve development and inundation area mapping: a case study of monsoon dominated Dwarkeswar River. *Int. J. River Basin Manage.*, 1-11. <https://doi.org/10.1080/15715124.2020.1738447>.

- Malik, S., Pal, S.C., Sattar, A., Singh, S.K., Das, B., Chakraborty, R., Mohammad, P., 2020b. Trend of extreme rainfall events using suitable Global Circulation Model to combat the water logging condition in Kolkata Metropolitan Area. *Urban Clim.*, 32, 100599. <https://doi.org/10.1016/j.uclim.2020.100599>.
- Masood, M., Takeuchi, K., 2012. Assessment of flood hazard, vulnerability and risk of mid-eastern Dhaka using DEM and 1D hydrodynamic model. *Nat. hazards*, 61(2), 757-770. <https://doi.org/10.1007/s11069-011-0060-x>.
- Mazzella, A., Mazzella, A., 2013. The importance of the model choice for experimental semivariogram modeling and its consequence in evaluation process. *J. Eng.*, 2013. <https://doi.org/10.1155/2013/960105>.
- Mentzafou, A., Markogianni, V., Dimitriou, E., 2018. The use of geospatial technologies in flood hazard mapping and assessment: case study from River Evros. In: Niedzielski T., Migala K. (Eds.), *Geoinformatics and Atmospheric Science. Pageoph Topical Volumes*. Birkhäuser, Cham, 221-242. https://doi.org/10.1007/978-3-319-66092-9_12.
- Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R., Abderrahmane, B., 2020. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth Sci. Rev.*, 103225. <https://doi.org/10.1016/j.earscirev.2020.103225>.
- Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., Ghazali, A.H. b., 2017. Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomatics Nat. Hazards Risk*, 8(2), 1080-1102. <https://doi.org/10.1080/19475705.2017.1294113>.
- Moran, P.A., 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37(½), 17-23. <https://doi.org/10.2307/2332142>.
- Mukherjee, F., Singh, D., 2020. Detecting flood prone areas in Harris County: a GIS based analysis. *GeoJournal*, 85, 647-663. <https://doi.org/10.1007/s10708-019-09984-2>.
- Naboureh, A., Li, A., Bian, J., Lei, G., Amani, M., 2020. A hybrid data balancing method for classification of imbalanced training data within Google Earth Engine: case studies from Mountainous Regions. *Remote Sens.*, 12(20), 3301. <https://doi.org/10.3390/rs12203301>.
- NASA, E., 2020. Intense Flooding in Bangladesh, 2020.
- Nguyen, V.N., Yariyan, P., Amiri, M., Dang Tran, A., Pham, T.D., Do, M.P., Thi Ngo, P.T., Nhu, V.H., Quoc Long, N., Tien Bui, D., 2020. A new modeling approach for spatial prediction of flash flood with biogeography

- optimized CHAID tree ensemble and remote sensing data. *Remote Sens.*, 12(9), 1373. <https://doi.org/10.3390/rs12091373>.
- Obroślak, R., Dorozhynskyy, O., 2017. Selection of a semivariogram model in the study of spatial distribution of soil moisture. *J. Water Land Dev.*, 35(1), 161-166.
- Ovi, M.M., Hossain, M.S., Mahfuzul, M., 2015. Geomorphic signatures of active tectonics from Sylhet city and adjoining areas Surma basin, Bangladesh. *Nepal Geol. Soc.*, 108.
- Pal, S.C., Das, B., Malik, S., Shit, M., Chakraborty, R., 2020. Flood frequency analysis and its management in selected part of Bardhaman district, West Bengal. In: Alam, A., Haldar, A., Satpati, L., Sarkar, R. (Eds.), *Habitat, Ecology and Ekistics Case Studies of Human-Environment Interactions in India*. Springer, p. 225-246.
- Pal, S.C., Chowdhuri, I., 2019. GIS-based spatial prediction of landslide susceptibility using frequency ratio model of Lachung River basin, North Sikkim, India. *SN Applied Sciences*, 1(5), 416. <https://doi.org/10.1007/s42452-019-0422-7>.
- Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A., 2018. A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS J. Photogramm. Remote Sens.*, 145, 120-147. <https://doi.org/10.1016/j.isprsjprs.2017.11.021>.
- Paul, G.C., Saha, S., Hembram, T.K., 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, 120-146. <https://doi.org/10.1007/s41976-019-00018-6>.
- Radwan, F., Alazba, A., Mossad, A., 2019. Flood risk assessment and mapping using AHP in arid and semiarid regions. *Acta Geophys.*, 67(1), 215-229. <https://doi.org/10.1007/s11600-018-0233-z>.
- Ramesh, V., Iqbal, S.S., 2020. Urban flood susceptibility zonation mapping using evidential belief function, frequency ratio and fuzzy gamma operator models in GIS: a case study of Greater Mumbai, Maharashtra, India. *Geocarto Int.*, 1-26. <https://doi.org/10.1080/10106049.2020.1730448>.
- Rawat, J., Biswas, V., Kumar, M., 2013. Changes in land use/cover using geospatial techniques: A case study of Ramnagar town area, district Nainital, Uttarakhand, India. *Egypt. J. Remote Sens. Space Sci.*, 16(1), 111-117. <https://doi.org/10.1016/j.ejrs.2013.04.002>.
- Reimann, K., 1993. *Geology of Bangladesh*. Gebruder Borntraeger, Berlin-Stuttgart.

- Roy, P., Pal, S.C., Chakraborty, R., Chowdhuri, I., Malik, S., Das, B., 2020. Threats of climate and land use change on future flood susceptibility. *J. Cleaner Prod.*, 272, 122757. <https://doi.org/10.1016/j.jclepro.2020.122757>.
- Sajedi-Hosseini, F., Choubin, B., Solaimani, K., Cerdà, A., Kavian, A., 2018. Spatial prediction of soil erosion susceptibility using a fuzzy analytical network process: Application of the fuzzy decision making trial and evaluation laboratory approach. *Land Degrad Dev.*, 29(9), 3092-3103. <https://doi.org/10.1002/ldr.3058>.
- Sarkar, D., Mondal, P., 2020. Flood vulnerability mapping using frequency ratio (FR) model: a case study on Kulik river basin, Indo-Bangladesh Barind region. *Appl. Water Sci.*, 10(1), 17. <https://doi.org/10.1007/s13201-019-1102-x>.
- Sarker, C., Mejias, L., Maire, F., Woodley, A., 2019. Flood mapping with convolutional neural networks using spatio-contextual pixel information. *Remote Sens.*, 11(19), 2331. <https://doi.org/10.3390/rs11192331>.
- Shahabi, H., Shirzadi, A., Ghaderi, K., Omidvar, E., Al-Ansari, N., Clague, J.J., Geertsema, M., Khosravi, K., Amini, A., Bahrami, S., 2020a. 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. <https://doi.org/10.3390/rs12020266>.
- Shahabi, H., Shirzadi, A., Ronoud, S., Asadi, S., Pham, B.T., Mansouripour, F., Geertsema, M., Clague, J.J., Bui, D.T., 2020b. Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. *Geosci. Front.*, 12(3), 101100. <https://doi.org/10.1016/j.gsf.2020.10.007>.
- Sharif, S.M.K., Sultana, S., Haq, F., Abdullah, A.S.M., 2017. Understanding the permutation of integrated agriculture approach for bringing resilience at Haor Basins in Bangladesh. *Int. J. Agri. Forest. Fish.*, 5(5), 67-74.
- Talukdar, S., Ghose, B., Salam, R., Mahato, S., Pham, Q.B., Linh, N.T.T., Costache, R., Avand, M., 2020. Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. *Stochastic Environ. Res. Risk Assess.*, 1-24. <https://doi.org/10.1007/s00477-020-01862-5>.
- Tehrany, M.S., 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. <https://doi.org/10.1007/s12665-018-7667-0>.

- Tehrany, M.S., Kumar, L., Shabani, F., 2019. A novel GIS-based ensemble technique for flood susceptibility mapping using evidential belief function and support vector machine: Brisbane, Australia. *PeerJ*, 7, e7653. <https://doi.org/10.7717/peerj.7653>.
- Thirumurugan, P., Krishnaveni, M., 2019. Flood hazard mapping using geospatial techniques and satellite images—a case study of coastal district of Tamil Nadu. *Environ. Monit. Assess.*, 191(3), 193. <https://doi.org/10.1007/s10661-019-7327-1>.
- Tien Bui, D., Khosravi, K., Li, S., Shahabi, H., Panahi, M., Singh, V.P., Chapi, K., Shirzadi, A., Panahi, S., Chen, W., 2018. New hybrids of ANFIS with several optimization algorithms for flood susceptibility modeling. *Water*, 10(9), 1210. <https://doi.org/10.3390/w10091210>.
- Tingsanchali, T., Karim, M.F., 2005. Flood hazard and risk analysis in the southwest region of Bangladesh. *Hydrol. Processes*, 19(10), 2055-2069. <https://doi.org/10.1002/hyp.5666>.
- Ullah, K., Zhang, J., 2020. GIS-based flood hazard mapping using relative frequency ratio method: a case study of Panjkora River Basin, eastern Hindu Kush, Pakistan. *PLoS One*, 15(3), e0229153. <https://doi.org/10.1371/journal.pone.0229153>.
- Uyanik, G.K., Güler, N., 2013. A study on multiple linear regression analysis. *Proc. Soc. Behavioral Sci.*, 106, 234-240. <https://doi.org/10.1016/j.sbspro.2013.12.027>.
- Viglione, A., Merz, B., Viet Dung, N., Parajka, J., Nester, T., Blöschl, G., 2016. Attribution of regional flood changes based on scaling fingerprints. *Water Resour. Res.*, 52(7), 5322-5340. <https://doi.org/10.1002/2016WR019036>.
- Wahlstrom, M., Guha-Sapir, D., 2015. The human cost of weather-related disasters 1995–2015. Geneva, Switzerland: UNISDR.
- Wang, Y., Li, Z., Tang, Z., Zeng, G., 2011. A GIS-based spatial multi-criteria approach for flood risk assessment in the Dongting Lake Region, Hunan, Central China. *Water Resour. Manage.*, 25(13), 3465-3484. <https://doi.org/10.1007/s11269-011-9866-2>.
- Wang, Y., Fang, Z., Hong, H., Peng, L., 2020. Flood susceptibility mapping using convolutional neural network frameworks. *J. Hydrol.*, 582, 124482. <https://doi.org/10.1016/j.jhydrol.2019.124482>.

- Woznicki, S.A., Baynes, J., Panlasigui, S., Mehaffey, M., Neale, A., 2019. Development of a spatially complete floodplain map of the conterminous United States using random forest. *Sci. Total Environ.*, 647, 942-953. <https://doi.org/10.1016/j.scitotenv.2018.07.353>.
- Yaseen, Z.M., Deo, R.C., Hilal, A., Abd, A.M., Bueno, L.C., Salcedo-Sanz, S., Nehdi, M.L., 2018. Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. *Adv. Eng. Software*, 115, 112-125. <https://doi.org/10.1016/j.advengsoft.2017.09.004>.
- Yesilnacar, E., Topal, T., 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. <https://doi.org/10.1016/j.enggeo.2005.02.002>.
- Yi, Y., Zhang, Z., Zhang, W., Jia, H., Zhang, J., 2020. Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region. *Catena*, 195, 104851. <https://doi.org/10.1016/j.catena.2020.104851>.
- Yin, J., Yu, D., Yin, Z., Liu, M., He, Q., 2016. Evaluating the impact and risk of pluvial flash flood on intra-urban road network: A case study in the city center of Shanghai, China. *J. Hydrol.*, 537, 138-145. <https://doi.org/10.1016/j.jhydrol.2016.03.037>.
- Yin, J., He, F., Xiong, Y.J., Qiu, G.Y., 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(1), 183. <https://doi.org/10.5194/hess-21-183-2017>
- Zhao, W., Du, S., 2016. Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS J. Photogramm. Remote Sens.*, 113, 155-165. <https://doi.org/10.1016/j.isprsjprs.2016.01.004>.
- Zhao, X., and Chen, W., 2020. Optimization of computational intelligence models for landslide susceptibility evaluation. *Remote Sens.*, 12(14), 2180. <https://doi.org/10.3390/rs12142180>.
- Zope, P., Eldho, T., Jothiprakash, V., 2016. Impacts of land use–land cover change and urbanization on flooding: a case study of Oshiwara River Basin in Mumbai, India. *Catena*, 145, 142-154. <https://doi.org/10.1016/j.catena.2016.06.009>.

Figures caption

Fig. 1 Location of the study area.

Fig. 2 Workflow of this study.

Fig. 3 Land cover maps of the study area: (a) 2000, (b) 2014, and (c) 2017.

Fig. 4 Increase and decrease in land cover categories from 2000 to 2017.

Fig. 5 Importance of the flood influencing factors using information gain ratio (IGR) coefficients: (a) 2000, (b) 2014, and (c) 2017.

Fig. 6 Pearson correlation: (a) 2000, (b) 2014, and (c) 2017. Abbreviations: Elev: elevation; Slp: slope; NDVI: normalized difference vegetation index; Aspt: aspect; Rain: mean monsoon rainfall; LC: land cover; Curv: plan curvature; Soil: soil type; DD: drainage density; Pri: proximity to rivers; Pro: proximity to roads; and Pop: population density.

Fig. 7 Flood susceptibility maps of the study area via ensemble modeling (BRBP-CART-EBF).

Fig. 8 Model-wise flood susceptibility percentage of individual classes between 2000 and 2017.

Fig. 9 Area under the curves for the 2017 flood: (a) success rate curves (SRCs), (b) prediction rate curves (PRCs). SM1: BRBP, SM2: CART, SM3: EBF, EM1: BRBP-CART, EM2: BRBP-EBF, EM3: CART-EBF, and EM4: BRBP-CART-EBF.

Fig. 10 Taylor diagram showing model performance. SM1: BRBP, SM2: CART, SM3: EBF, EM1: BRBP-CART, EM2: BRBP-EBF, EM3: CART-EBF, and EM4: BRBP-CART-EBF.

Fig. 11 Distribution of spatial correlation and significance levels.

Fig. 12 Rose diagram, showing land cover change and relative change of flooding area.

Tables caption

Table 1 Flood influencing factors.

Table 2 Accuracy of land cover maps.

Table 3 Land cover area transition matrix from 2000 to 2017.

Table 4 Dynamic index of LC changes (%) from 2000 to 2017 in the study area.

Table 5 Tolerance (TOL) and variance of inflation (VIF) for the flood influencing factor.

Table 6 Spatial correlation metrics.

Table 7 Relationships between relative changes in flood areas and different land cover change scenarios .

Table 8 Flood susceptible areas according to land cover, 2000–2017.

Table 9 Population growth (PG) between 2000 and 2017.

Table 10 Correlation between the relative change in flooding (RCF) areas and road density (RD), 2000–2017.

Supplementary Data

Supplementary Figures

Fig. S1 Land use transition

Fig. S2 Flood influencing factors

Fig. S3 Relationships of EBF

Fig S4 Schematic diagram of the Dempster's rule of combination

Fig. S5 Flood susceptibility maps for standalone models

Supplementary Tables

Table S1 Computed spectrometric indices and their formulae

Table S2 Land cover statistics during 2000, 2014 and 2017

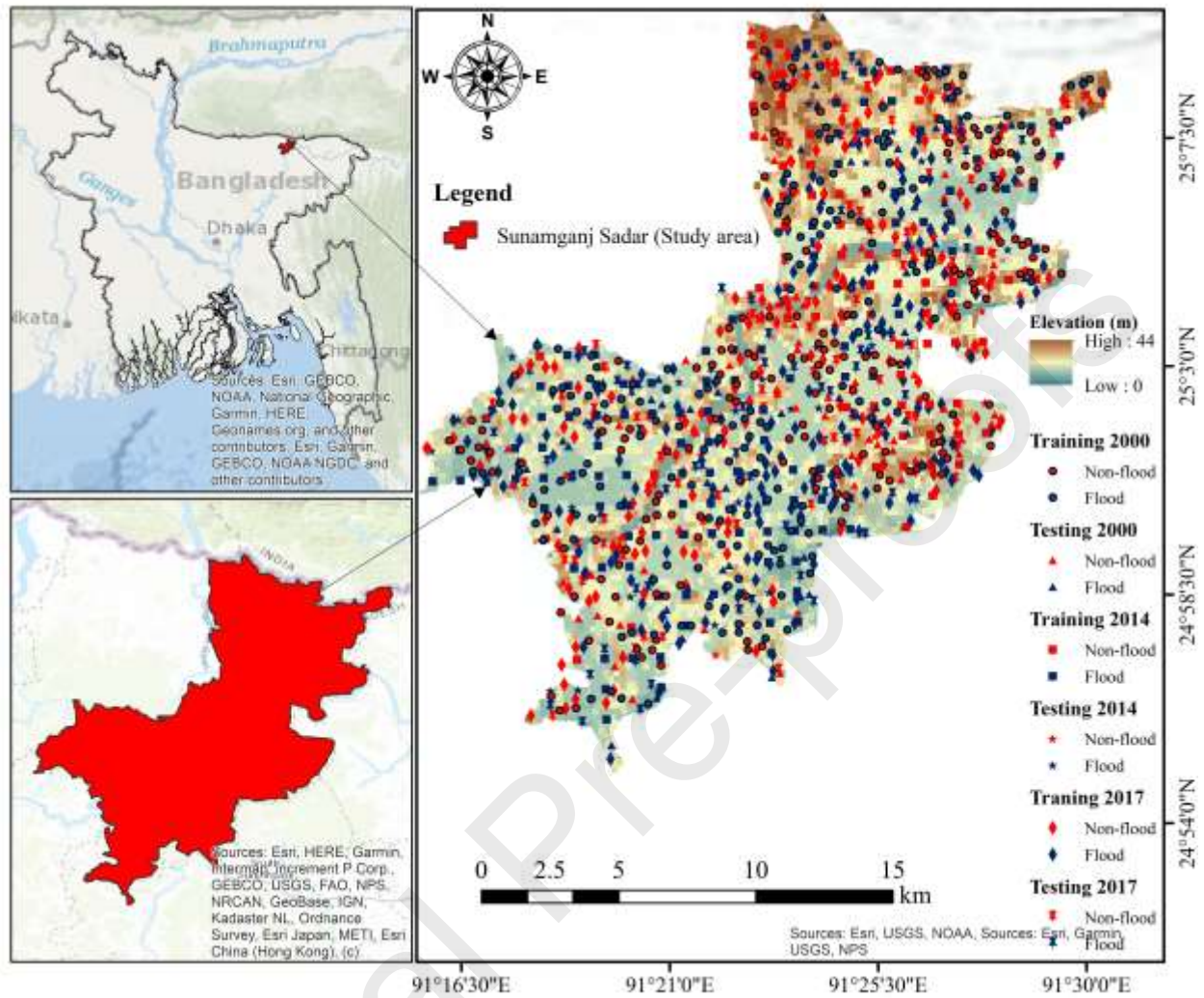
Table S3 Rate of land cover change, 2000-2017

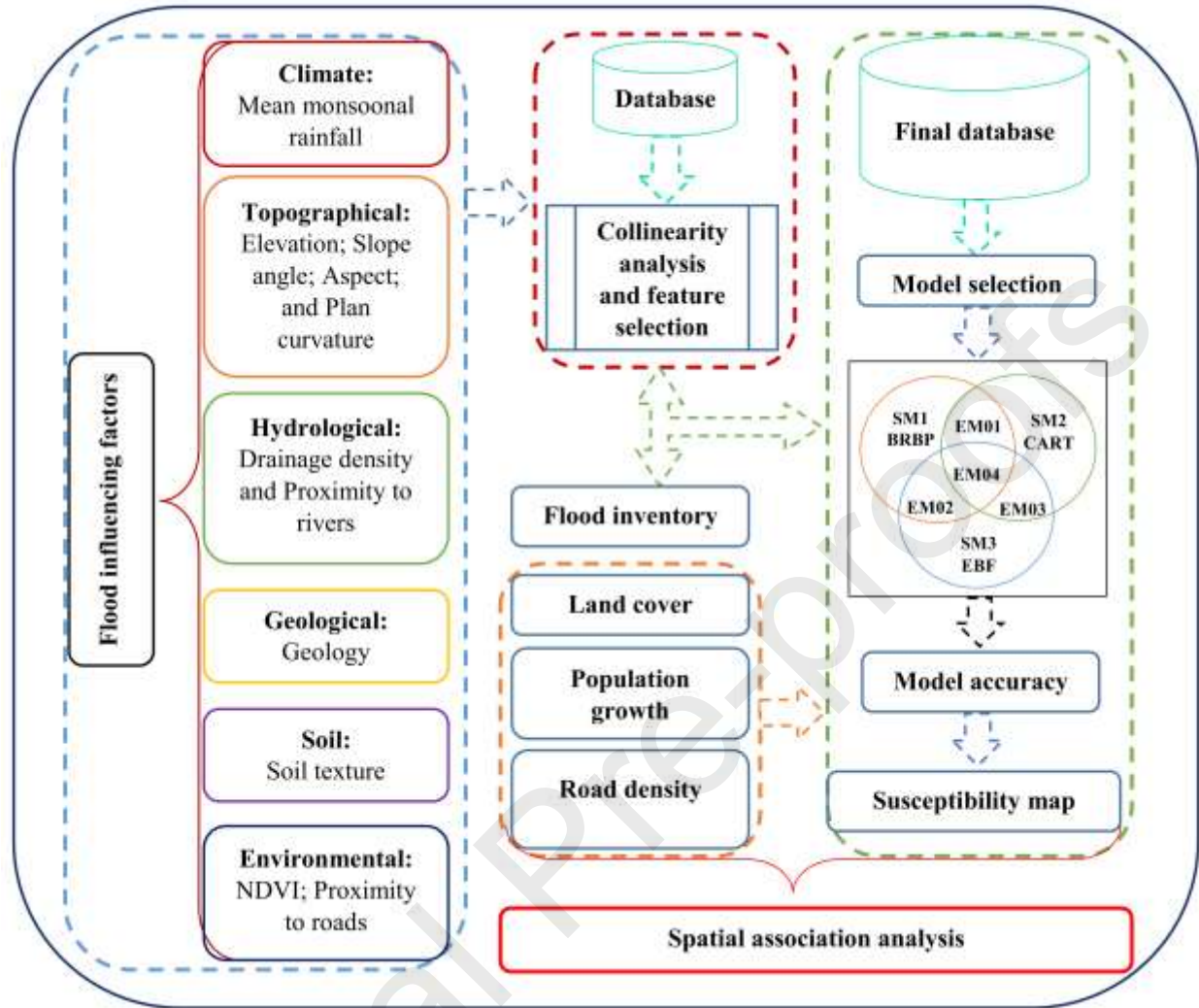
Table S4 Statistics for evidential belief function models

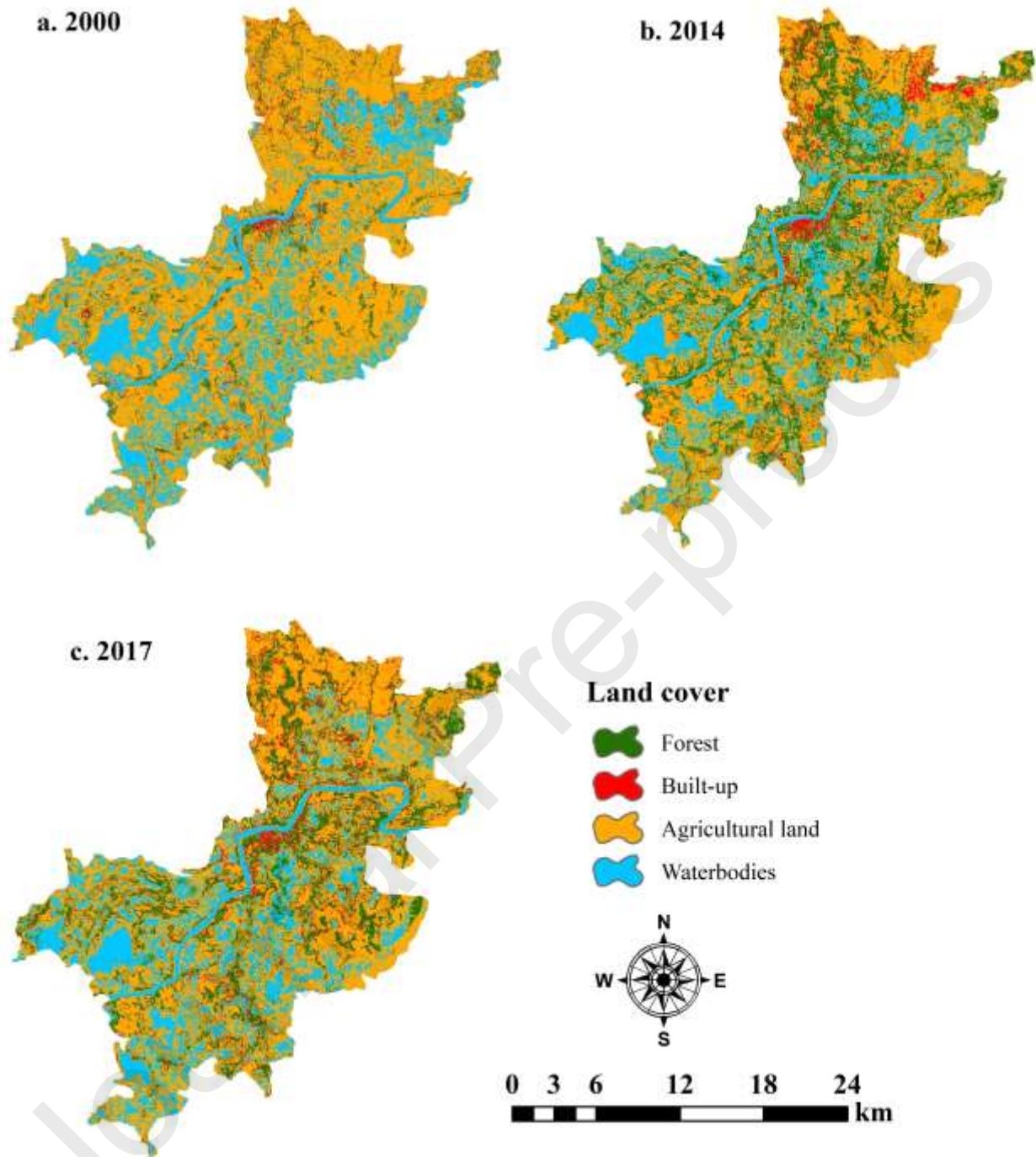
Table S5 Area under the curve (success rate)

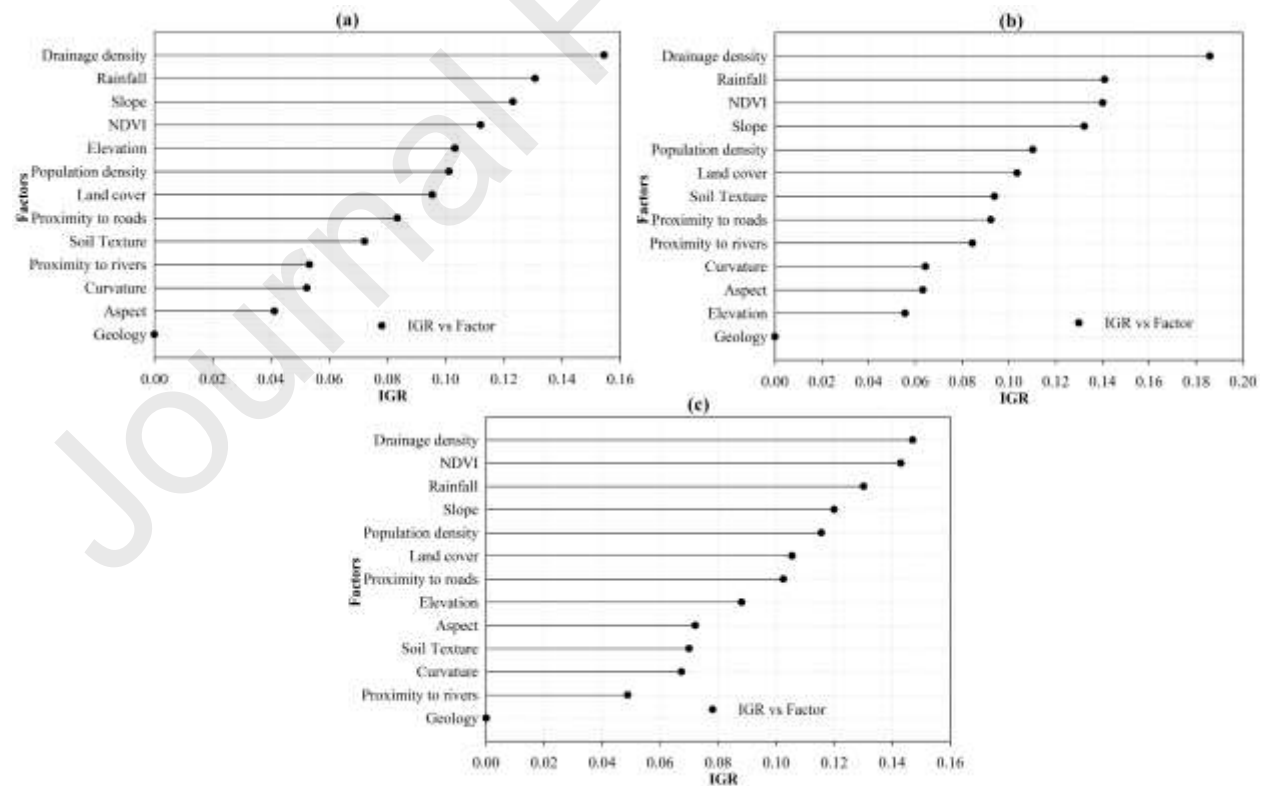
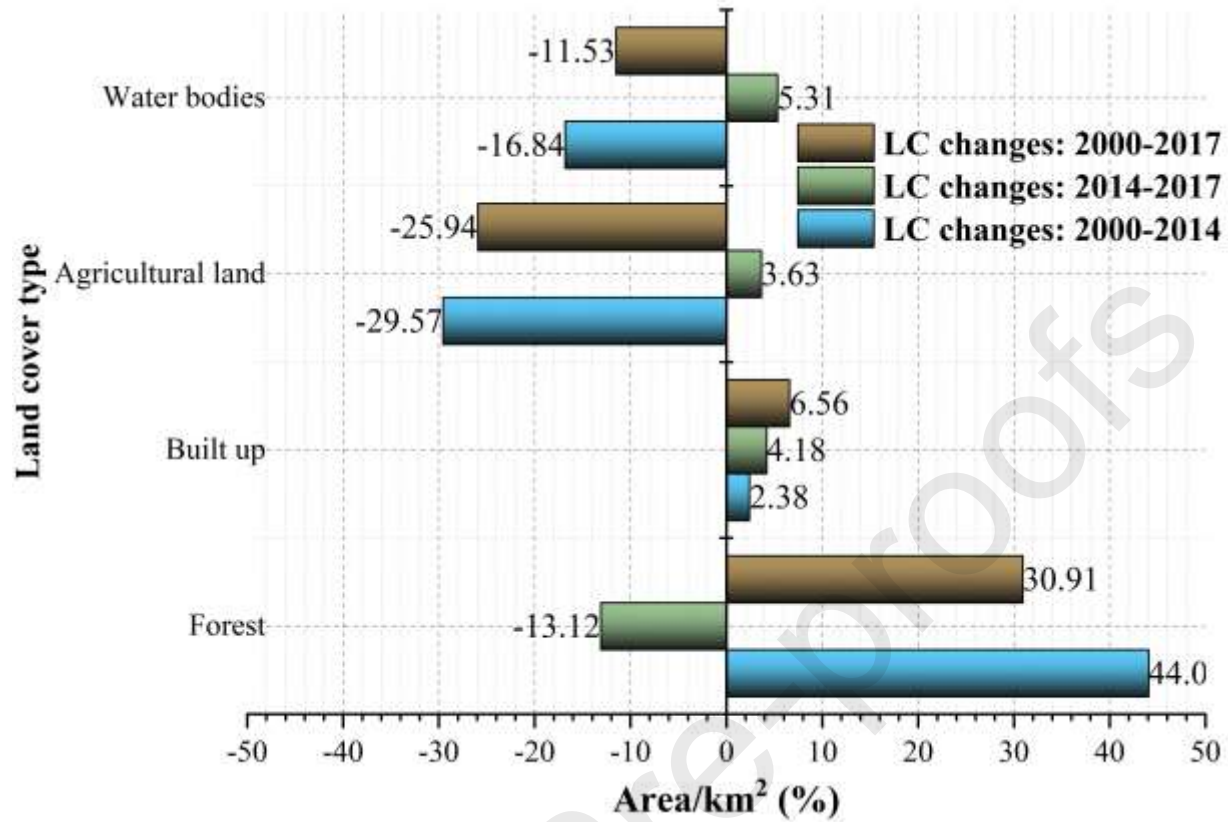
Table S6 Area under the curve (prediction rate)

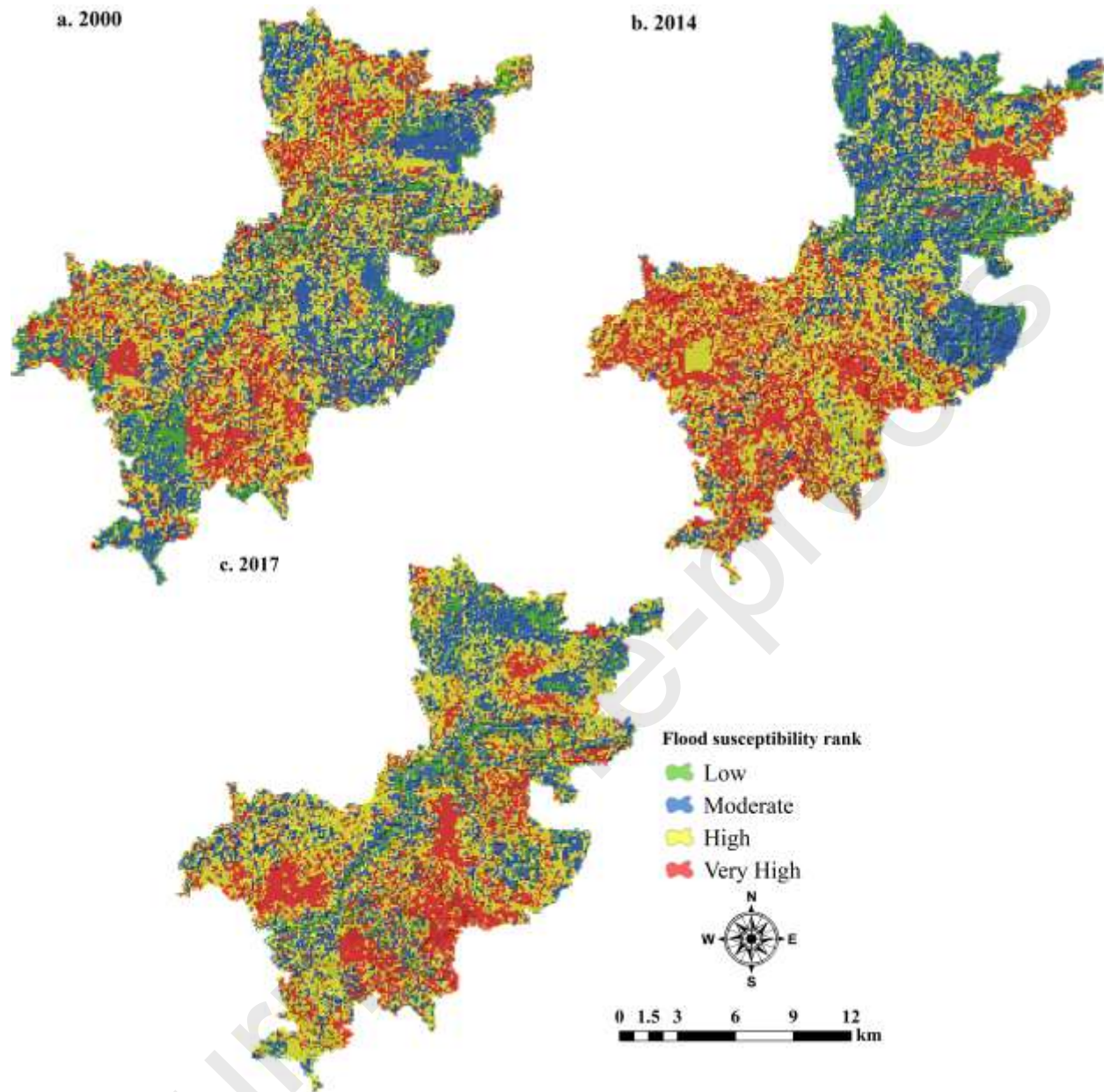
Table S7 Directional semi-variogram values for land cover scenarios with changes in flood areas

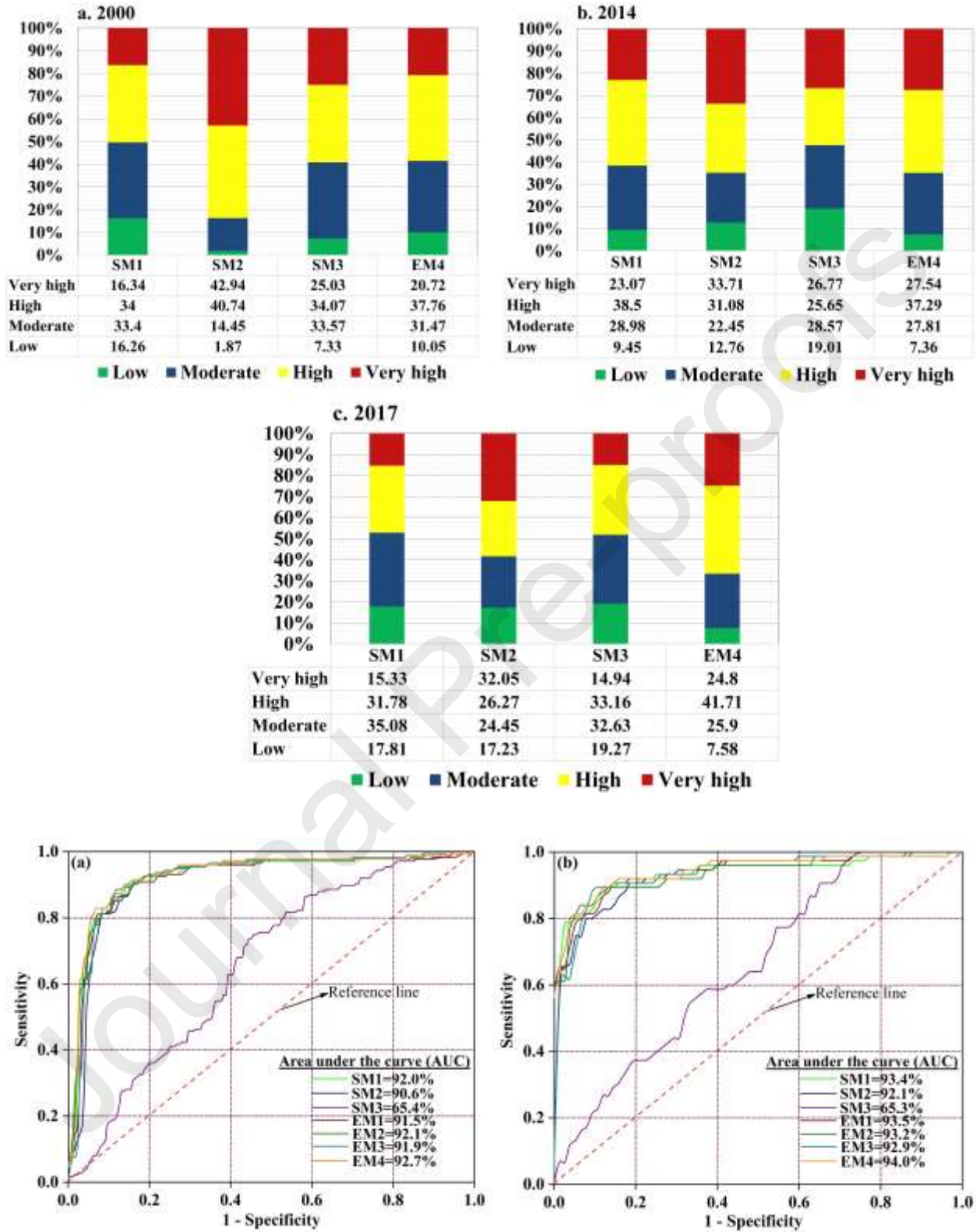


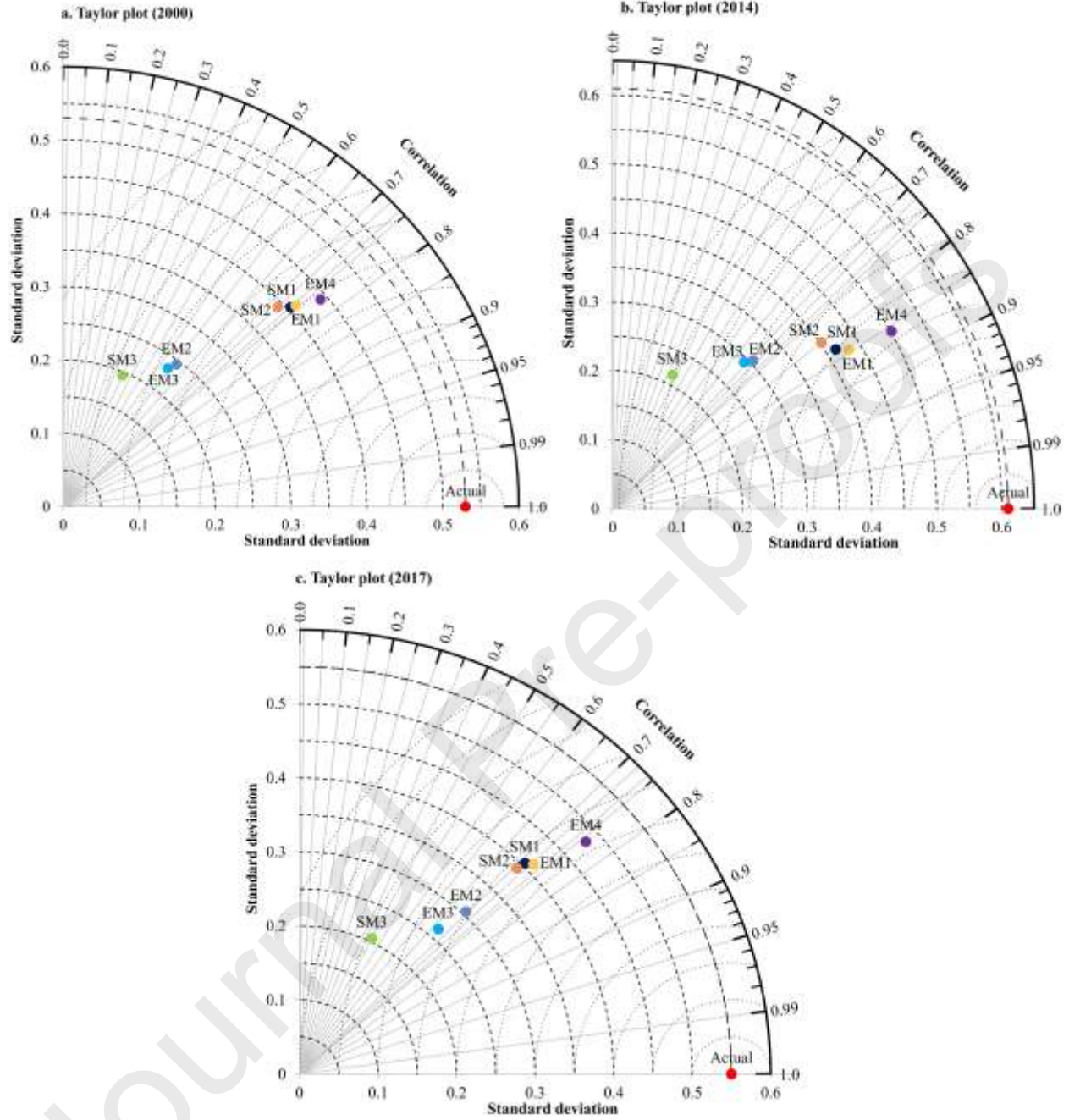


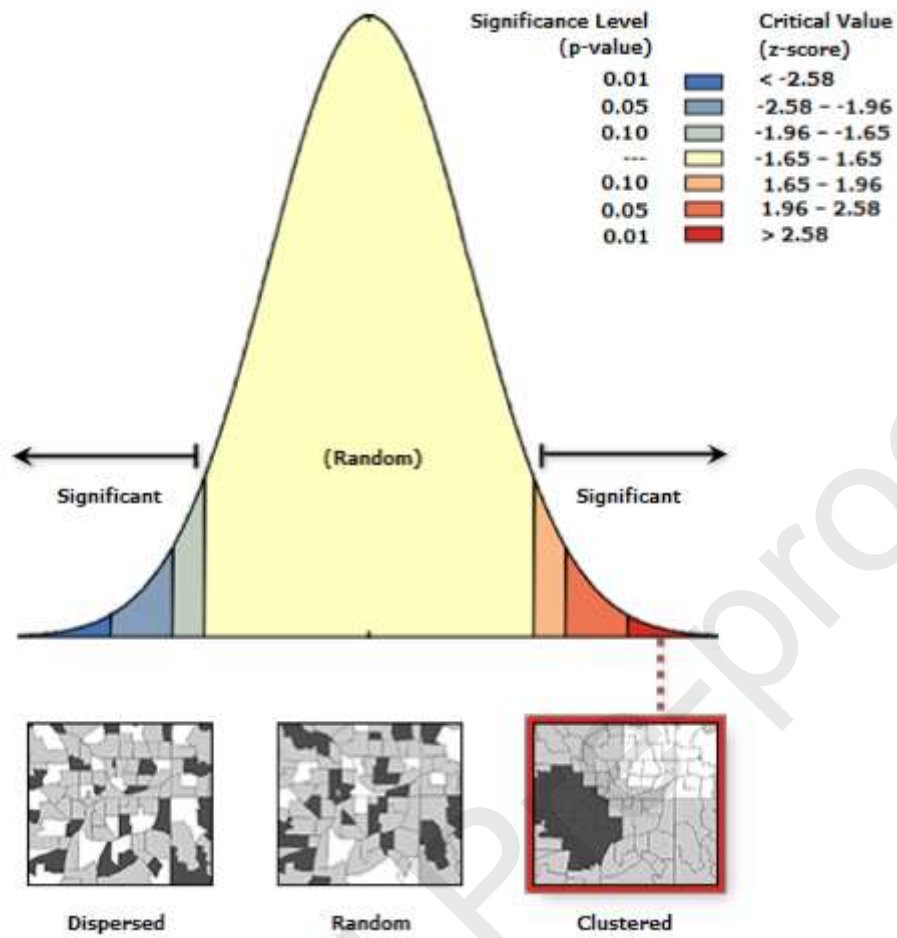












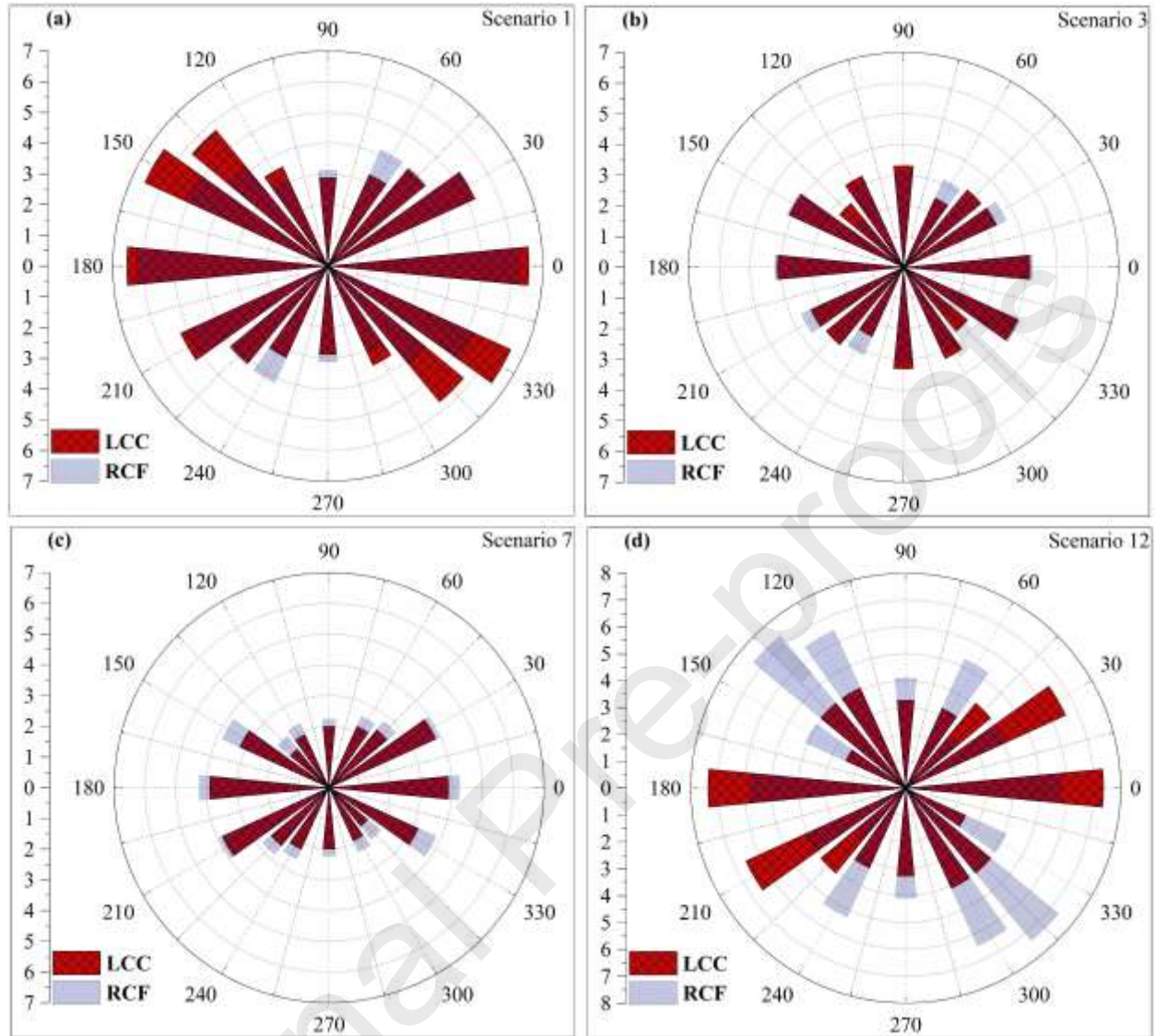


Table 1 Flood influencing factors.

Factor	Association between influencing factors and flood susceptibility	Class	Classification method
Elevation (m)	<p>Elevation is one of the most essential factors in the assessment of flood susceptibility (Choubin et al., 2019)</p> <p>Generally, flood and elevation are inversely proportional, which indicates that relatively more elevated areas are relatively less susceptible to floods; however, flat areas are extremely prone to flooding</p> <p>An elevation map was prepared from Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) with a resolution of 30 m (acquired from https://earthexplorer.usgs.gov/)</p>	(i) 0–4; (ii) > 4–8; (iii) > 8–12; (iv) > 12–16; and (v) > 16	Manual (Fig. S2-d)

Factor	Association between influencing factors and flood susceptibility	Class	Classification method
Slope (degree)	The slope map (30 m × 30 m cell size) was created from DEM data Comprises the effect of gravity on the formation of surface runoff and its velocity (Khosravi et al., 2019). It is a significant factor in flood susceptibility analysis	(i) 0–2; (ii) > 2–4; (iii) > 4–6; (iv) > 6–8; and (v) > 8	Manual (Fig. S2 (e))
NDVI ²⁰⁰⁰ NDVI ²⁰¹⁴ NDVI ²⁰¹⁷	NDVI is another important factor, affecting floods (Bui et al., 2019) NDVI map was prepared using Landsat 7 (2000) and Landsat 8 (2014 and 2017) images via the GEE interface. The mathematical equation for calculating NDVI is as follows: Landsat ETM+: $NDVI (2000) = \frac{Band\ 4 - Band\ 3}{Band\ 4 + Band\ 3}$ Landsat 8: $NDVI (2014, 2017) = \frac{Band\ 5 - Band\ 4}{Band\ 5 + Band\ 4}$	(i) Below 0; (ii) 0–0.075; (iii) 0.075–0.1; and (iv) > 0.1	Manual (Fig. S2-g, h, and i)
Aspect	Aspect is considered to be an important factor in flood susceptibility assessment and aids in the calculation of the water flow direction (Tehrany and Kumar, 2018) The aspect map (30 m × 30 m cell size) was prepared from DEM data	(i) Flat; (ii) North; (iii) Northeast; (iv) East; (v) Southeast; (vi) South; (vii) Southwest; (viii) West; (ix) Northwest; and (x) North	Manual (Fig. S2-f)
Plan curvature	The plan curvature of the study area aids in the acquisition of information on geomorphologic characteristics of the given area (Costache and Bui, 2019) The DEM was considered to create the plan curvature map using a cell size of 30 m × 30 m	(i) Concave (-1); (ii) Flat (0); and (iii) Convex (+1)	Manual (Fig. S2-j)
Mean monsoon rainfall (mm) ²⁰⁰⁰ Mean monsoon rainfall (mm) ²⁰¹⁴ Mean monsoon rainfall (mm) ²⁰¹⁷	Rainfall is the main factor contributing to flooding in the study area. Most of the rainfall occurs in the monsoon season (May to September) The monsoon rainfall data was collected from https://www.hydrology.bwdb.gov.bd/ for the years 2000, 2014, and 2017 In this study, the rainfall thematic map (30 m × 30 m cell size) was prepared using mean monsoon rainfall data which were collected from eight rainfall stations using the inverse distance weighting (IDW) interpolation method (Khosravi et al., 2019)	(i) 791.28–809.31; (ii) 809.31–820.81; (iii) 820.81–828.14; (iv) 828.14–832.82; and (v) 832.82–835.80 (i) 833.40–837.45; (ii) 837.45–845.40; (iii) 845.40–861.02; (iv) 861.02–891.67; and (v) 891.67–951.85 (i) 864.60–890.13; (ii) 890.13–907.94; (iii) 907.94–920.37; (iv) 920.37–929.05; and (v) 929.05–935.10	Geometric interval (Fig. S2-a, b, and c)
Drainage density (km/km ²)	Drainage density is an important factor affecting flood susceptibility because it impacts peak discharge (Sajedi-Hosseini et al., 2018) The higher the drainage density, the higher the possibility of flooding (Bui et al., 2019) In this study, the drainage density map was generated from DEM data with a resolution of 30 m	(i) 0.02–0.29; (ii) 0.29–0.52; (iii) 0.52–0.72; (iv) 0.72–0.92; and (v) 0.92–1.14	Natural breaks (Fig. S2-m)

Factor	Association between influencing factors and flood susceptibility	Class	Classification method
Proximity to rivers (m)	Areas close to main channels are more likely to be flooded, thus, proximity to rivers plays a significant role in the assessment of flood susceptibility (Chowdhuri et al., 2020) The proximity river map was created by applying the Euclidian distance method based on the digital stream order (Mukherjee and Singh, 2020)	(i) 0–500; (ii) > 500–1000; (iii) > 1000–1500; (iv) > 1500–2000; and (v) > 2000	Manual (Fig. S2-l)
Proximity to roads (m)	Impermeable surfaces (roadways, pavements, and parking spaces) increase the rainfall-runoff process (Mukherjee and Singh, 2020). Hence, proximity to roads is another factor affecting flood susceptibility The road network layer was obtained from the Local Government Engineering Department (LGED) at a scale of 1:800,000 The proximity to roads was calculated using the Euclidean distance method (Mukherjee and Singh, 2020)	(i) 0–500; (ii) > 500–1000; (iii) > 1000–1500; (iv) > 1500–2000; and (v) > 2000	Manual (Fig. S2-k)
Geology	The geology greatly influences the structure of the drainage pattern, contributing to floodplain growth (Islam and Sado, 2000) A geological map with a scale of 1:100,000 was obtained from the Soil Resource Development Institute (SRDI) (https://www.barc.gov.bd/)	(i) Dihing and Dupi Tila formations; (ii) Young gravelly sand; (iii) Marsh clay and peat; and (iv) Waterbodies	Geology unit (Fig. S2-o)
Soil texture	The soil texture is a significant factor influencing floods because it controls the rate of infiltration and surface runoff The soil texture thematic layer was collected (scale 1:100,000) from the SRDI (https://www.barc.gov.bd/)	(i) Mixed sand and silt loam; (ii) Predominantly loam; (iii) Mostly clay; (iv) Mixed clay and silt loam; (v) Mixed silt loam and silty clay loam; and (vi) Waterbodies	Soil unit (Fig. S2-n)
Population density ²⁰⁰⁰	An increase in population increases the risk of flooding (Li et al., 2019) More densely populated areas are often more susceptible to floods	(i) 54.67–214.12; (ii) 214.12–373.58; (iii) 373.58–533.03; (iv) 533.03–692.48; and (v) 692.48–851.94	Equal interval (Fig. S2-p, q, and r)
Population density ²⁰¹⁴		(i) 68.32–263.47; (ii) 263.47–458.63; (iii) 458.63–653.79; (iv) 653.79–848.95; and (v) 848.95–1044.11	
Population density ²⁰¹⁷		(i) 76.37–292.27; (ii) 292.27–508.17; (iii) 508.17–724.08; (iv) 724.08–939.98; and (v) 939.98–1155.88	
Land cover ²⁰⁰⁰ Land cover ²⁰¹⁴ Land cover ²⁰¹⁷	Land cover is a crucial factor in the assessment of flood susceptibility because areas with relatively less vegetation are relatively more prone to flooding However, cities and towns are covered with impervious surfaces and barren lands, which increase the flow of surface runoff (Mojaddadi et al., 2017)	(i) Forest; (ii) Built-up; (iii) Agricultural land; and (v) Waterbodies	CART classifier in GEE (Fig. 3)

Factor	Association between influencing factors and flood susceptibility	Class	Classification method
	Land cover maps were generated from Landsat 7 (2000) and Landsat 8 (2014 and 2017) data in GEE		

Table 2 Accuracy of land cover maps

LC 2017	Ground truth point				Grand Total (user)	Accuracy (user)	Overall Accuracy	Kappa coefficient
	Forest	Built-up	Agricultural land	Water bodies				
Forest	293	0	6	9	308	95.13		
Built-up	0	88	1	0	89	98.88		
Agricultural land	5	0	1206	306	1517	79.50		
Water bodies	2	0	120	3839	3961	96.92	92.36	0.83
Grand Total (producer)	300	88	1333	4154	5875			
Accuracy (producer)	97.67	100.00	90.47	92.42				
LC 2014								
Forest	276	1	28	3	308	89.61		
Built-up	0	89	0	0	89	100.00		
Agricultural land	12	0	1458	47	1517	96.11		
Water bodies	1	0	40	3920	3961	98.96	97.75	0.95
Grand Total (producer)	289	90	1526	3970	5875			
Accuracy (producer)	95.50	98.89	95.54	98.74				
LC 2000								
Forest	217	4	23	64	308	70.45		
Built-up	7	75	5	2	89	84.27		
Agricultural land	11	1	1071	434	1517	70.60		
Water bodies	3	0	59	3899	3961	98.43	89.57	0.76
Grand Total (producer)	238	80	1158	4399	5875			
Accuracy (producer)	91.18	93.75	92.49	88.63				

Table 3 Land cover area transition matrix from 2000 to 2017

LC class/Year		2014				
		Agricultural land	Built-up	Forest	Water bodies	Total (looses)
2000	Agricultural land	106.76	7.59	40.45	22.89	177.69
	Built-up	3.05	0.84	2.57	0.85	7.31
	Forest	5.79	0.82	10.17	2.07	18.85
	Water bodies	32.54	0.43	9.65	32.59	75.22
	Total (gain)	148.14	9.68	62.84	58.41	279.07
LC class/Year		2017				
		Agricultural land	Built up	Forest	Water bodies	Total (looses)
2014	Agricultural land	98.81	6.79	16.71	25.84	148.16
	Built-up	6.08	1.64	1.66	0.30	9.69
	Forest	25.36	4.29	25.43	7.77	62.85
	Water bodies	21.53	1.15	5.93	29.81	58.42
	Total (gain)	151.79	13.87	49.73	63.72	279.11
LC class/Year		2017				
		Agricultural land	Built up	Forest	Water bodies	Total (looses)
2000	Agricultural land	111.26	10.84	32.37	23.20	177.70
	Built-up	3.60	0.85	1.87	0.99	7.31
	Forest	6.68	1.03	9.57	1.58	18.85
	Water bodies	30.23	1.13	5.91	37.95	75.22
	Total (gain)	151.77	13.85	49.72	63.72	279.08

Table 4 Dynamic index of LC changes (%) from 2000 to 2017 in the study area

↓LC type/years→	2000–2014		2014–2017		2000–2017	
	SLCDI	BLCDI	SLCDI	BLCDI	SLCDI	BLCDI
Agricultural land	-1.11	12.21	1.22	101.12	-0.858	10.90
Built up	2.17	15.48	21.55	121.41	5.27	17.01
Forest	15.56	28.86	-10.43	89.49	9.64	21.38
Water bodies	-1.49	11.83	4.54	104.48	-0.90	10.86
IBLCDI	13.32		99.91		11.75	

Table 5 Tolerance (TOL) and variance of inflation (VIF) for the flood influencing factor

Factors	2000		2014		2017	
	TOL	VIF	TOL	VIF	TOL	VIF
Elevation	0.552	1.811	0.650	1.539	0.586	1.705
Slope	0.892	1.121	0.919	1.088	0.944	1.059
NDVI	0.564	1.773	0.770	1.298	0.758	1.319
Aspect	0.873	1.145	0.858	1.165	0.884	1.131

Factors	2000		2014		2017	
	TOL	VIF	TOL	VIF	TOL	VIF
Rainfall	0.607	1.647	0.679	1.472	0.675	1.482
Land cover	0.847	1.181	0.861	1.161	0.790	1.266
Curvature	0.884	1.131	0.887	1.127	0.951	1.052
Soil Texture	0.481	2.078	0.410	2.438	0.603	1.659
Drainage density	0.412	2.428	0.324	3.091	0.452	2.213
Proximity to rivers	0.670	1.493	0.584	1.712	0.612	1.634
Proximity to roads	0.778	1.285	0.689	1.451	0.737	1.357
Population density	0.663	1.508	0.653	1.532	0.769	1.300

Table 6 Spatial correlation metrics

Variables (in pair)	Moran's, I index	Expected index	Variance	Critical value (z)
RCF and LC	0.164958	-0.000049	0.000009	54.425943
RCF and PG	0.194871	-0.000049	0.000009	64.313122
RCF and RD	0.239414	-0.000049	0.000009	79.328751

Note: LC-Land cover, PG-Population density growth rate and RD-Road density

Table 7 Relationships between relative changes in flood areas and different land cover change scenarios.

Scenario (Sc.)	Change class (2000–2017)	Standardized coefficients*
1	Agricultural land to built-up	0.707
2	Agricultural land to forest	0.172
3	Agricultural land to waterbodies	0.500
4	Built-up to agricultural land	-0.010
5	Built-up to forest	-0.011
6	Built-up to waterbodies	0.017
7	Forest to agricultural land	0.408
8	Forest to built-up	0.290
9	Forest to waterbodies	-0.012
10	Waterbodies to agricultural land	0.156
11	Waterbodies to built-up	0.169
12	Waterbodies to forest	-0.001

*Note: 95% confidence interval

Table 8 Flood susceptible areas according to land cover, 2000-2017

Susceptibility rank	Area (km ²)				% change between	
	LC 2000	LC 2014	LC 2017	LC (2000-2014)	LC (2014-2017)	LC (2000-2017)
Low	28.78	21.07	21.72	-26.80	3.10	-24.54
Moderate	90.11	79.65	76.17	-11.61	-4.37	-15.47
High	108.15	106.78	117.45	-1.26	11.87	10.46
Very High	59.32	78.86	71.02	32.94	-9.94	19.72

Table 9 Population growth (PG) between 2000 and 2017

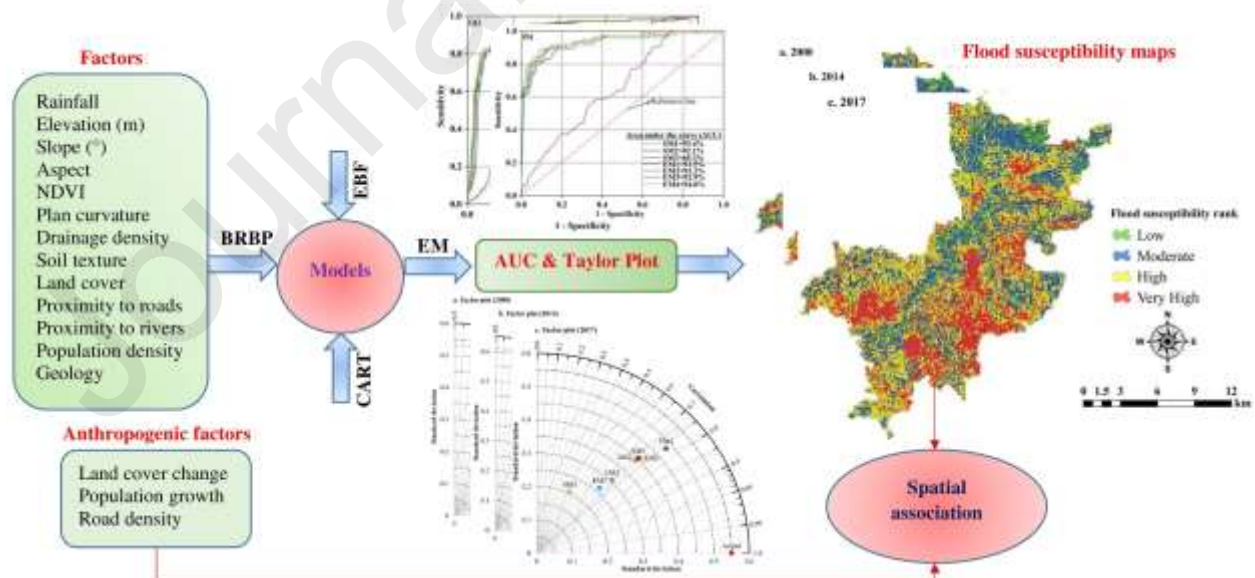
Susceptibility rank	% of PG		
	2000-2014	2014-2017	2000-2017
Low	-19.71	-1.47	-20.89
Moderate	5.83	2.33	8.29
High	13.58	10.89	25.95
Very High	60.68	-5.60	51.68

Table 10 Correlation between the relative change in flooding (RCF) areas and road density (RD), 2000-2017

		RD	RCF
RD	Pearson correlation	1	0.496**
	Sig.		0.000
RCF	Pearson correlation	0.496**	1
	Sig.	0.000	

**Correlation is significant at the 0.01 level (2-tailed)

Graphical abstract



Highlights

- Flood susceptibility was mapped using standalone and ensemble models
- New ensemble models had higher accuracy (AUC > 90%) than standalone ones
- Land cover, population growth, road density, and flooding changes are correlated