Modelling flood susceptibility zones using hybrid machine learning models of an agricultural dominant landscape of India

Satish Kumar Saini
  Jawaharlal Nehru University
Susanta Mahato
  Jawaharlal Nehru University
Deep Narayan Pandey (✉ deepudai@gmail.com)
  Jawaharlal Nehru University
Pawan Kumar Joshi
  Jawaharlal Nehru University

Abstract

One of the most destructive natural disasters is flood because it destroys a significant amount of property and infrastructure, and often causes death. Due to complexity and ferocity of severe flooding, predicting flood-prone areas is a difficult task. Each year, flooding results in destruction of agriculture, damage to resources, and fatalities in the Asia and the Pacific. Thus, creating flood susceptibility maps at local level is though challenging but inevitable task. In order to implement a flood management plan for the Balrampur district, an agricultural dominant landscape of India, and strengthen its resilience flood susceptibility modeling and mapping is carried out. In the present study, three hybrid machine learning models namely Fuzzy-ANN (Artificial Neural Network), Fuzzy-RBF (Radial Basis Function) and Fuzzy-SVM (Support Vector Machine) with 12 topographic, hydrological and other flood influencing factors were used to determine flood susceptible zones. To ascertain the relationship between the occurrences and flood influencing factors, Correlation Attributes Evaluation (CAE) and multicollinearity diagnostics tests were used. The predictive power of these models was validated and compared using a variety of statistical techniques, including Wilcoxon signed-rank, t-paired tests, and Receiver Operating Characteristic (ROC) curves. Result shows the Fuzzy-RBF model out performed other hybrid machine learning models for modelling flood susceptibility, followed by Fuzzy-ANN and Fuzzy-SVM. Overall, these models have shown promise in identifying flood-prone areas in the basin and other basins around the world. The outcomes of the work would benefit policymakers and government bodies to capture the flood-affected areas for necessary planning, action and implementation.

1. Introduction

The ecosystems of the natural and human world are greatly impacted by floods, which are frequent catastrophe. The magnitude of floods up to a certain degree is generally beneficial for the ecosystem, such as providing riparian corridor with water and nutrients, removing pollutants from flood plain areas, recharging ground water, and improving soil fertility (Hester et al., 2020; Golden et al., 2019). But beyond the tolerance levels these floods result failure in agricultural fields, lives and properties, residential areas, water resources, and natural habitat, among others (Guzzetti and Tonelli, 2004). According to the United Nations Office for Disaster Risk Reduction (UNDRR), between 1996 and 2015, there were 150,061 deaths worldwide attributable to flood occurrences, or 11.1% of all total deaths from disasters worldwide (Chen et al., 2019). Floods were the most common and often occurring form of disaster (about 47 percent between 1995 to 2015), among all the weather-related disasters, and out of all floods about 95 percent were in Asia (Khattri, 2017). The rate and intensity of floods are expected to intensify by 2050, which might result in enormous damage (about US$ 1 trillion). This would further intensify due to rapidly growing population, changing land usage, and the continuing climate change (Islam and Karim 2019; Armal et al., 2020).

The increased frequency of severe occurrences of floods, including droughts, storms, and catastrophic outcomes, are thought to be mostly caused by poorly anticipated climate change. Highly skewed patterns of monsoon rainfall intensity frequently result in flood occurrences in monsoon climate regions (Ullah et al., 2021; Dewan, 2015). Unfavourable changes in channel morphology, particularly aggradations of the channel bed, falling carrying capacity, man-made drainage modifications such capture, piracy of distributaries, and cross-blocking of rivers are some significant causes behind the increase in flood triggering factors (Mahato et al., 2021). The susceptibility of flood is increased by artificially controlling the flow of water from a dam, releasing it suddenly, and artificially raising the water level to embank a river. It can prevent flooding up to a point, but when an embankment is breached, the flood impact multiplies (Froehlich, 2008; Pierce, 2010). People affected by such breaching episodes not only experience tremendous flow impacts, but also some extra negative effects including long-lasting or permanent sand splay damage to agricultural fields (Brázdil et al., 2006; Ielpi et al., 2018) and several water-borne infections (Davies et al., 2015; Semenza, 2020). The chance of a flood effect increases as a result of wetland reclamation in the flood plain region, which decreases the natural capacity of flood water to be stored. In addition to triggering factors, increased population density in flood plains, urban expansion, expansion of infrastructure agriculture and fishing activities in riparian flood plain areas have increased flood exposure, and consequently, the vulnerability of those areas to flooding (Talukdar and Pal, 2020; Wang et al., 2020). The recent meteorological events have increased the risk of flooding in locations with flood plains due to sudden flooding, faster-moving water, and shorter reaction times (Barrocu and Eslamian, 2022). By 2050, it is anticipated that the danger of fluvial floods would have increased globally by around 187 percent due to changes in the triggering variables and degree of exposure, with Asia likely suffering the most (Arnell and Gosling, 2016).

In order to conserve the natural and human ecology, extensive study is needed to explore the cause-nature-consequence links throughout the globe. Floods are a hazard, and this necessitates a focus on the cause-nature-consequence relationships. Adedeji et al. (2012) and Mind'je et al. (2019) correctly said that mapping flood susceptibility is a prerequisite for sustainable flood control and risk management techniques since it provides a useful database for adopting the necessary adaptation and mitigation methods. The use of several pertinent characteristics to address particular susceptibilities, such as land slip, soil erosion, and habitat change, is a recent research trend known as susceptibility modelling (Dragićević et al., 2015 and Barney et al., 2012). When data are available, direct flood parameters can be used in vulnerability modelling; otherwise, proxy parameters can be used in situations when data are few (Jun et al., 2013). Therefore, this strategy can occasionally close the hydrological data gap and offer an acceptable flood forecast at a spatial scale (Molinari et al., 2020). Numerous studies have taken advantage of this chance to map flood susceptibility using a different set of variables than those listed in Table 1. Effective modelling requires the selection of representative parameters and the use of a strong integrating approach (Mahato and Pal, 2019; Pal et al., 2020; Pal et al., 2021).

Table 1
Literature on selecting conditioning parameters for flood susceptibility modelling

References	Soil type/texture	LULC	Buffer	Curvature	Elevation	Flow Acc	Geomorphology	MNDWI	Rain	Slope	SPI	TWI
Choubin et. al., 2019		√	√	√	√					√		√
Arora et. al., 2019	√	√	√	√	√		√		√	√		√
Das 2019		√	√	√	√	√	√		√	√	√	√
Pham et. al., 2020	√	√	√	√	√					√		
Bui et. al., 2019		√	√	√	√				√	√	√	√
Chowdhuri et. al., 2019	√	√	√		√				√		√	√
Chen et. al., 2019	√	√	√	√	√				√	√	√	√
Hong et. al., 2017	√	√	√	√	√				√	√	√	√
Pandey et. al., 2021	√	√	√	√	√		√		√	√		√
Mahato et. al., 2021		√	√	√	√	√				√	√	√
Sahana and Patel, 2019	√	√	√	√	√				√	√		√
Towfiqul Islam et. al., 2020	√	√	√	√	√				√	√	√	√
Pal and Singha, 2021	√	√	√	√	√	√			√	√	√	√
Kundu and Mahato, 2020	√	√		√	√	√			√	√		√
Gourav et. al., 2020		√	√		√					√		

Various flood susceptibility assessment models, including the Analytic Hierarchy Process (AHP) (Vojtek and Vojteková, 2019; Hammami et al., 2019), frequency ratio (a bivariate statistical model, FR) (Rahmati et al. 2016; Shafapour Tehrany et al., 2019), weighted factor (Termeh et al., 2018), Weights-of-Evidence (WoE) (Tehrany et al., 2014; Rahmati et al., 2016), Random Forest (RF) (Vafakhah et al., 2020; Abedi et al., 2022), Remote Sensing (RS) (Dano et al., 2019), Logistic Regression (LR) (Al-Juaidi et al., 2018; Shafapour Tehrany et al., 2019; Pham et al., 2020), fuzzy logic (Akay, 2021; Bouamrane et al., 2022), and Multivariate Discriminant Analysis (MDA) (Choubin et al., 2019), have been used so far for flood susceptibility mapping. However, there are drawbacks and advances to each model and technique. The reviewed literature reveals that a variety of flood susceptibility assessment models, including Multi-Criteria Decision-Making (MCDM), Multivariate Discriminant Analysis (MDA), HYDROTEL, RBF, Soil Water Assessment Tool (SWAT), wetSpa, AHP, FR, weighted factor, WoE, RF, RS, LR, fuzzy logic, and others, have been used so far.

All techniques have been successfully applied and yielded satisfying results in terms of susceptibility modelling across many disciplines. Statistical techniques were used by Arabameri et al. (2019), Shafapour Tehrany et al. (2019), and Mousavi et al. (2022) to analyse flood susceptibility. The knowledge-based multi-criteria approach was used by Das and Gupta (2021) and Nachappa et al. (2020), but those approaches may encounter problems like bias in the expert assessment of the parameters (Kamali et al., 2017; Das and Gupta, 2021; and Comes et al., 2011). But the first two approaches have some limitations (Arabameri et al., 2019). Traditional hydrological models need intensive work to develop and calibrate, and they aren't very trustworthy. Since bias is frequently present in the parameter selection and weighting processes, data-driven and statistical techniques have subjectivity problems. Non-linear machine learning methods may also produce poor outcomes because of

uneven data sets and wide range of values in specific layers (Papandreou and Ziakopoulos, 2022). However, as researchers have observed through applications, machine learning algorithms may yield trustworthy results if parameter selection is done objectively and scientifically, and if the quality of the data is sufficient.

On the Indian subcontinent, different regions are vulnerable to flooding due to geographical variability. A number of perennial rivers are created by the Himalayan glacier melt, which leads to floods during the monsoon season. In addition to river flooding, other factors that contribute include heavy rainfall, cloud bursts, glacial lake outbursts, and tsunamis (Tripathi, 2015; Singh and Kumar, 2013). India, after Bangladesh, has the second-highest number of floods worldwide (Singh and Kumar, 2013). The proportion of flood-prone land in India is about 12.5%. (Usama, 2015). From 1915 to 2015, there were 649 disasters in India, 302 of which were caused by floods (Tripathi, 2015). According to CRED (Centre for Research on the Epidemiology of Disasters) reports, between 1965 and 1975, there were 1000 flood-related deaths per year, but between 2005 and 2015, these were 1700 (Crunch, 2019). About 2% of India's GDP was lost to flood-related economic losses between 2005 and 2015. After Bengal and Bihar, Uttar Pradesh is the state in India that has experienced the most flooding. The state of Uttar Pradesh has a flood-prone area of about 73,400 km^2 (Planning Commission, 2011). Ghaghara, Sarda, Gandak, Rapti, Gomati, Yamuna, and Ganga are the rivers that cause flooding and damage in Uttar Pradesh (Khatoon, 1994; Usama, 2015). Uttar Pradesh has 294.36 lakh hectares of total land, of which 73.36 lakh hectares are at risk of flooding (Usama, 2015). Every year, floods affect about 27 lakh hectares of land in Uttar Pradesh, causing an estimated loss of 432 crore rupees (Usama, 2015). One of Uttar Pradesh's most severely affected districts by flooding is Balrampur. A flood in 2017 greatly damaged property and claimed many lives (Singh, 2018) in around 300 villages. In October 2022, flood affected about 287 villages (Hindustan times, 2022). Since it is situated in the foothills of the Himalaya, it frequently floods as a result of the Rapti and other rivers. Flood damages property and hinders the socioeconomic and agricultural development of the region. In order to implement a flood management plan for Balrampur district and make this district resilient, this paper captures and reports the susceptible flood zones using hybrid machine learning modeling. The model proposed and used refers to intensive flood inventories. For activities aimed at reducing flood hazards and promoting economic development, the findings of this paper will be helpful to governments and policy makers, specifically to the nations of interests and flood-prone regions. The flood susceptibility map prepared here might serve as a starting point for additional study, such as risk and hazard mapping. It can also help the planners and decision-makers take the right steps to regulate and limit this phenomenon in the research region.

2. Study area

Balrampur is an under developed district located in the eastern part of Uttar Pradesh, at the foothills of the Himalaya (Diwakar, 2008). The geographical extent of the district is between 27°08' to 27°54' North latitude and82°02' to 82°49' East longitude (Kumar et al., 2003), covers total area of 3349 sq km (Diwakar, 2008). The district touches border of Siddharthnagar (in East), Shravasti (in West), and Gonda (in South) in addition to the international border of Nepal on the northern side. It is about 113 m from mean sea level (Sanjay and Prakash, 2020). The region has a sub-humid climate with cold winters and warmer summer (Prakash and Verma, 2020), and three distinct seasons- summer, rainy and winter season. It is located in the Tarai region, where the south-west monsoon brings majority of regional rainfall. The floods in the district are primarily caused by the Rapti River. It moves from north to south-east along the elevation drop (390 m), (109.7 m at Tulsipur and 106.68 m at Balrampur). Some other significant sources of flooding during the south-west monsoon season in the district are BudhiRapti, Suawa, and Kuano rivers, along with the KharjharKatni, Kaktava, Mannar, Varuna, and Ary hilly nullah flow (Census of India, 2011). Due to low level surface, the region is affected by frequent floods, resulting the district unhealthy and malaria prone. The district has a 642 person/sqkm population density, with 92% population dwelling in rural areas (Census of India, 2011). The population distribution indicates higher dependency of community on primary income source like agriculture, forest, etc. Flooding affects highly on agricultural lands and damage crops which causes economic losses and promote migration.

3. Materials and methods

With the aid of prior research and in consideration of the study goal and geographic setting of study area, a range of parameters were chosen for flood susceptibility mapping. Data on the chosen parameters was compiled from a number of sources. Table 2 provides details of the parameters, source of data and data resolution/scale. Data from the United States Geological Survey (USGS) was collected for Landsat 8 and SRTM (Shuttle Radar Topography Mission) DEM (Digital Elevation Model). Data on rainfall was derived from MEERA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2) model. The agricultural contingency plan report for the district of Balrampur served as the source for soil types map. Geomorphology map (Geological Survey of India) was procured from the Bhukosh website.

Table 2
Description of the flood conditioning parameters

Parameter	Source of Data	Resolution
Topographic Wetness Index (TWI)	SRTM DEM	30 m
Land Use Land Cover	Landsat8 (United States Geological Survey, USGS)	30 m
Slope	SRTM DEM	30 m
Rainfall	MERRA-2 modeled data	Interpolation
Elevation	SRTM DEM	30 m
Soil	Agriculture Contingency Plan: Balrampur	Vector
Distance from River	Open Street Map (OSM)	Vector
Curvature	SRTM DEM	30 m
Stream Power Index (SPI)	SRTM DEM	30 m
Geomorphology	Bhukosh (Geological Survey of India)	Vector
MNDWI	Landsat8 (USGS)	30 m
Flow Accumulation	SRTM DEM	30 m

3.1 Selection of Parameters contributes to flood susceptibility

The review of literature revealed that a variety of factors influence flood susceptibility, but due to the variation in local features, some parameters have greater impact than others. So, using the Correlation Attributes Evaluation (CAE) approach, the most appropriate parameters were chosen. This method enables evaluation of conditioning parameters among the factors and classes using Pearson's correlation method (Pal and Singha, 2021). Additionally, to reduce the likelihood that the flood susceptibility models will be incorrect, a multicollinearity test was used in this study to determine whether the attached parameters have any multicollinearity consequences (Pal and Singha, 2021; and Pandey et. al., 2021). Tolerances and Variance Inflation Factor (VIF) are used in identifying the error in this test (Pal and Singha, 2021). Following equation was used to calculate VIF and tolerance values (1,2).

Tolerance = 1- r^2 (1)

VIF = 1/tolerance (2)

VIF value above 10 and tolerance level less than 0.1 shows a multicollinearity issue (Pal and Singha, 2021). In this study factors for flood susceptibility model were included which have VIF value less than 2.5.

3.2 Flood conditioning parameters

3.2.1 Geomorphology

Geomorphology and flood susceptibility are closely related (Pandey et. al., 2022). Through the processes of erosion and deposition, flood changes the topography (Cavalli et al., 2017). Numerous geomorphic features have impact on direction of water flow. Geomorphology indicates indices of previous flood events. The geomorphology of the study area comprises of active flood plain, dams and reservoir, hills and plains, older alluvial plain, older flood plain, piedmont alluvial plain, pond, river, lake and water bodies (Fig. 2a).

3.2.2 Soil

The hydrologic properties of the surface, such as the rate and amount of infiltration, erosion, and runoff are influenced by soil characteristics (Rodríguez-Caballero et al, 2013). Flood susceptibility is negatively correlated with the rate of infiltration and surface runoff (Yang & Zhang, 2011). Lower runoff and infiltration rates reduce flood susceptibility, and vice versa as it determines degree of drainage.

3.2.3 Distance from River

Floods are most likely to occur in areas along both river banks (Buraas et al., 2014). The majority of the flood-affected land in rural areas is found close to river streams (Janizadeh et al., 2021). As a result, flood susceptibility is higher close to rivers and decreases as distance from them grows. This is because water intensity is highest along river banks and in flood plains. Expansion in the flooded area is brought on by

sediment accumulation in the river basin and rising water discharge. The Euclidean map of distance from river source in the district is shown in Fig. 2c.

3.2.4 Curvature

Curvature is a sign of a particular direction's changing slope (Minár et al., 2020). Surfaces with a positive curvature are convex, those with a zero curvature are flat, and those with a negative curvature are concave. Water runoff is associated with regions with negative values that are highly susceptible to flooding, whereas regions with positive curvature are less susceptible to flooding (Zaharia et al., 2017). Curvature helps to define water surface topography relation affecting heterogeneity and hyporheic flow. It helps the model in accurate representation of water velocity which proves beneficial in susceptibility mapping. The curvature map of the site is represented in Fig. 2d.

3.2.5 Elevation

In flood susceptibility modelling, elevation is also a key component that is frequently used. Water flows from high altitude to low altitude, and there is a negative correlation between that and flood susceptibility (Giordano et al., 2007). Because of this, areas with low elevation are more vulnerable to flooding than those with high elevation, and vice versa (Ishaya et al., 2009). The elevation ranges from 79−235 meter (Fig. 2e).

3.2.6 Flow Accumulation

The area where river water is gathered or stored is known as a flow accumulation. The susceptibility to flooding increases with flow accumulation and vice versa (Zingaro et al., 2020). The potential to recharge ground water is greatest in areas where water flow accumulation is high. Data from the SRTM DEM have been used to create a flow accumulation map. The study area has flow accumulation value ranging from zero to 636 (Fig. 2f).

3.2.7 Land Use Land Cover (LULC)

LULC has the ability to modify and examine variables like surface runoff, water infiltration capability, sediment transport, moisture holding capability, heat albedo, and carries a relationship of flood susceptibility in a region (Sugianto et al., 2022). For instance, built-up areas have low infiltration capacities due to the concrete, whereas vegetation has higher infiltration capacities and lower runoff rates, which is why flood events in settlement areas are more intense and frequent than those in vegetated areas. Land use is key determinant of flood susceptibility influencing frequency and spread of flooding (Owrangi et al., 2014). LULC map of the study area comprises of 6 classes: agriculture, stream sediment, settlement, vegetation, wqater bodies and wetland (Fig. 3g).

3.2.8 MNDWI

Normalized Difference Water Index (NDWI) could not distinguish between shallow water regions, so Modified Normalized Difference Water Index (MNDWI) was used to represent the entire water body (Du et al., 2016). For locating open water areas, the MNDWI is superior to the Normalised Difference Water Index (NDWI) (Sahana and Patel, 2019). The formula used for calculation of MNDWI is:

$$MNDWI = \frac{(Green - SWIR)}{(Green + SWIR)}$$

Where, Green refers to green wavelength and SWIR refers to Shortwave Infra-red wavelength in which satellite data is acquired. The value of MNDWI generally ranges between − 1 and 1, however, in this case it is -0.52 to 0.25 (Fig. 3h).

3.2.9 Rainfall

One of the most important parameters for determining flood susceptibility is rainfall. It functions as a fundamental force that causes flooding (Breugem et al., 2020). The main cause of floods in India and the study area is heavy rainfall that occurred in a very short amount of time between June and September. Due to heavy rainfall in a brief period of time, the monsoon season is when India experiences the majority of its floods. Rainfall with a high intensity and frequency causes flooding (Zellou and Rahali, 2019), but it is not clear that increase in what amount of rainfall will result in increasing proportionality of flood. The rainfall distribution map is shown in Fig. 3i.

3.2.10 Slope

The physiographic feature of slope regulates water velocity. High runoff velocity and minimal vertical percolation of water are suitable for high slope gradient slopes. So there is inverse relationship between slope and flood susceptibility (Whipple et al., 1998). Higher the slope angle means lower the flood susceptibility; i.e. at lower slope angle or flat leads to higher flood susceptibility as it regulates surface runoff and infiltration (Santos et al., 2019). In the study area the slope angle ranges from 0º − 24.2º (Fig. 3j).

3.2.11 Stream Power Index (SPI)

It is an influential factor for flood susceptibility mapping. SPI is the measurement of sediment transportation and stream bed erosion capacity of a stream. It is calculated using formula:

$$SPI = A_s tan\beta$$

Where, A_s stands for the specific catchment area and β stands for the slope gradient.

SRTM DEM data was used to generate SPI value of the study area map (Fig. 3k). A study show that higher SPI confined channels lead to disastrous transformations of the channel. Streams have loss of water holding capacity due to swallowing of depth increase flooding in the area.

3.2.12 Topographical Wetness Index (TWI)

TWI was proposed by Beven and Kirby (1979) as a symptom of watersheds. TWI is a quantitative index which is used to identify the moisture content of the land surface. It shows the effect of topography on surface water saturation and its runoff. It is expressed as

$$TWI = \ln\left(\frac{A_s}{tan\beta}\right)$$

Where, 'A_s' stands for the cumulative up slope area draining through a point (per unit contour length) and $tan\beta$ stands for the angle of slope at that point. Flood susceptibility is positively correalated with TWI values. The TWI map was prepared from SRTM DEM using Arc-GIS software (Fig. 3l).

3.3 Flood susceptibility models

Figure 4 depicts the paradigm of vulnerability to flooding using three hybrid models. To create hybrid models, fuzzy-logic membership has been used. Fuzzy-ANN (Artificial Neural Network), Fuzzy-RBF (Radial Basis Function) and Fuzzy-SVM (Support Vector Machine) are the three models.

3.3.1 Fuzzy logic

Fuzzy was established by Zadeh in 1965 (Zadeh et al., 1996). It is a composition of non-discrete mathematical equation used for operating various complex problems. Fuzzy is also known specially for its straightforwardness and implementation. The fuzzy membership value, which ranges from 0 to 1, is used to communicate the level of confidence and support for a certain characteristic or variable. The degree of membership of the member can be any value between '0' and '1' if the object is a member of the set, otherwise it has a membership value of '0'.

Fuzzy logic enables more adaptable configurations of weighted map data for GIS modelling. In this fuzzy set technique, the pixel of any causative flood conditioning variable is assumed to be flood-susceptible. The value "1" assigned to this sensitive pixel contrasts with the value "0" assigned to non-vulnerable pixels. However, since many various methods, including the Analytical Hierarchy Process, knowledge-based weighting, and user-defined weighting, have been used to establish the fuzzy membership value, there is no unanimity on how to compute the fuzzy membership function.

3.3.2 ANN (Artificial Neural Network)

The artificial neural network (ANN) is a black box and mathematical model. It has been applied in many fields, such as decision-making, pattern recognition, automated control systems, robotics, and others (Liang et al., 2021). It can handle dynamic, nonlinear, and imbalanced data sets. As a result, it is able to mimic how the human brain functions and is also able to generalise and foresee the results of a wide variety of varied inputs. Because of this, researchers from all over the world are frequently used to answer challenges in a variety of sectors. The ANN model can function as an expert since it can recognise complicated prediction patterns that a non-expert would not be able to. Without going against the assumptions and characteristics of the data, it can act on categorical, continuous, and binary data (Eq. 2).

This study made use of the feed-forward dependent multilayer perceptron (MLP) architecture. An average MLP has three layers, including an input layer, one or more hidden layers, and an output layer for nonlinear activation nodes. Numerous neurons or nodes are present in each layer, and each node in the subsequent layer is given a specific weight. Sharing the data is their responsibility. The MLP trains the network using the back-propagation algorithm until the difference between the expected and output values of the network is as close to zero as possible. The outcomes are therefore produced by the ANN model.

3.3.3 RBF (Radial Basis Function)

Three layers—the input layer, the hidden layer, and the output layer—make up the hidden layer neural network known as the Radial Basis Function Neural Networks (RBF Neural Nets) (Orr, 1996). Each unit in the hidden layer receives a broadcast from the input layer with the input vector's coordinates. Then, depending on the linked RBF, each hidden layer unit creates an activation (Orr, 1996). A linear combination of the activations of the hidden units is then computed by each unit in the output layer.

$$f_i(x) = \sum_{k=1}^{m} w_{ki}\theta(\|x - a_k\|)$$

3

where αis the total number of processing units, are the connecting weights, are the RBF centres or prototypes, and the function $\theta(,)$ is chosen to be a Gaussian function (Du & Swamy, 2014). An unsupervised RBF Neural Nets network uses the training dataset and a k-means algorithm to choose the initial hidden unit centres. Additionally, the initial value of every variance parameter (or parameters) in the network is set to the greatest squared distance between any two cluster centres.

3.3.4 SVM (Support Vector Machine)

SVM is well recognized as a controlled non-parametric quantitative ML technique. Decision planes, also known as the plane of separation of distinct aims or varying class membership, serve as the conceptual foundation for SVM. It can be used with a variety of variables, including continuous, categorical, linear, and non-linear data sets in various members of the class. The segregation of hyper-plane formation in the training sample is the basis of this method. The kernel function is the name of the mathematical operation used to transform data. Which separated the data set between groups for floods and non-floods, which are indicated by 1 and 0 accordingly. The capability of the SVM model is demonstrated by the recognition of correct kernel function. The SVM model's kernel functions were divided into four groups, including the radial basis kernel (RBK) sigmoid kernel (SG), linear kernel (LN), and polynomial kernel (RBF) (PL).PL and RBF kernels have been employed frequently in remote sensing scenarios. RBF kernel is frequently employed in SVM approaches due to the model's superior performance and level of precision compared to other conventional techniques. According to earlier work, the RBF kernel performed better than other kernels in problems involving flood susceptibility. In the most recent work, the RBF kernel function was used to identify the flood probability zones. This approach has been used widely in the context of flood susceptibility. The main issues with SVM modelling are typically related to how challenging it is to capture crucial variables.

3.4 Accuracy assessment and model comparison

Four methods are considered for the accuracy evaluation of the models and model comparison: (1) Area under curve under Receiver operating characteristics (AUC-ROC), (2) Friedmen test, (3) Wilcoxon test, and (4) Correlation coefficient. The true positive rate of several probable cut points of a diagnostic test is plotted against the false positive rate to create the ROC curve. The ROC curve produces the area under the curve, which represents model accuracy. The susceptibility model validity is checked using a non-parametric test. No statistical inference is necessary for non-parametric models (Buchet al., 2017). Despite the fact that the data were normally distributed, a non-parametric test—the Friedman test and paired t test—was used (Martinez-Alvarez et al., 2018). The test was used to compare substantial variations between two or more models (Beasley and Zumbo, 2003). The first implication for carrying out this test is that, at a significance level of 5% ($p < 0.5$), flood model performances are equal (null hypothesis). When comparing two or more models, the outcome was not used if the P value for the Friedman test held true for all models (Mahato et al., 2021). The Wilcoxon signed-rank test was used to examine the statistical significance of systematic pairwise differences among flood models in order to solve this issue. To evaluate the significance of differences across five susceptibility models, the p-value and the z-value were utilised in this case. The null hypothesis was rejected since the z-value was greater than the critical values of z (±1.96) and the p-value was less than 5%; as a result, the susceptibility models performed substantially differently (Islam et al., 2021). Correlation coeeficient computed between flood susceptibilty models using 200 points. A test of significance must be conducted in both circumstances. Highly significant results for the correlation coefficient can be interpreted as indicating the models' applicability in the real world.

4. Results and Discussion

4.1 Determining the influence of flood susceptibility parameters

The multicollinearity test is one of the contemporary methods for removing collinear characteristics and identifying the most important ones for the creation of a flood susceptibility model. A multicollinearity problem among the conditioning Parameters is indicated by a variance inflation factor (VIF) > 5 or a tolerance 0.01, according to the examined literature. To determine the relative importance of the factors influencing flood susceptibility, the CAE rank value of each parameter was calculated using a multicollinearity test. The higher rankings for soil (0.543), MNDWI (0.247), flow accumulation (0.051), geomorphology (0.049), TWI (0.032), aspect (0.027), slope (0.019), and SPI (0.019) indicate that these characteristics have a significant impact on flood susceptibility modelling. However, the indications of low effect flooding susceptibility include the distance from the river (-0.348), height (-0.343), rainfall (-0.025), and curvature (-0.025).

4.2 Flood susceptibility models

Primarily fuzzy membership was given to all the selected parameters to flood susceptibility was measured with the three hybrid models that are Fuzzy-ANN, Fuzzy-RBF and Fuzzy-SVM. have been categorized into five susceptible classes i.e. very low, low, moderate, high and very high, based on the natural break approach. Confluence stretches of the main river and select sections of major tributary connections are where high to extremely high flood sensitive zones are located. Main water conduits, which are classified as having a very high sensitivity to floods, are

specifically the most at risk. Although it is exceedingly challenging to define a precise high flood affected buffer width alongside the river, the confluence stretch has a breadth of around 15 kilometres. This section of the river's bank is occupied, which necessitates the lateral spread of flood water. This conclusion agrees with findings by Mahmoud and Gan (2018), Hong et al (2018), and Yang et al. (2018), who conducted research in the dry regions of the Middle East, Hengfeng area, China, and Hainan, China, respectively. In addition to this, it is discovered that a few isolated areas of topographical depression are extremely vulnerable to floods, even far from the main river (Fig. 4). According to individual estimates from the Fuzzy-ANN, Fuzzy-RBF and Fuzzy-SVM models, there is a prominent geographical proximity between high and very high flood sensitive zones. All hybrid models created using the aforementioned three models additionally contain spatial adjacency (Fig. 4).

Table 3
Areal distribution under different flood susceptibility models

Susceptibility zones	Class	Fuzzy-ANN	Fuzzy-RBF	Fuzzy-SVM
Very Low	Area in sqkm	865.17 (25.83%)	818.52 (24.44%)	521.54 (15.57%)
Low	Area in sqkm	1100.41 (32.86%)	1167.70 (34.87%)	936.55 (27.97%)
Moderate	Area in sqkm	793.77 (23.7%)	665.61 (19.87%)	1031.62 (30.8%)
High	Area in sqkm.	368.56 (11%)	454.69 (13.58%)	478.04 (14.27%)
Very High	Area in sqkm.	221.11 (6.6%)	242.49 (7.24%)	381.26 (11.38%)

Table 3 demonstrates the calculated area of each susceptibility category across all models. In the case of all the models, out of the total area, 16.60%, 20.82%, 25.65% in Fuzzy-ANN, Fuzzy-RBF and Fuzzy-SVM models respectively, fell within extremely high and high flood sensitive zones. In Fuzzy-SVM model, a comparatively larger region is identified as being in very high flood-prone zones. The characteristics of extremely high to high flood sensitive locations are very low elevation (10m), high flow accumulation, concave topography, very low slope, high wetness index, and very close proximity to rivers. They both Fig. 4, Table 3 indicate physical proximity between the calculated regions and sensitive zones, thus if any one of them looks to be genuine, all the models will likely be correct. Only after accuracy evaluation can a final decision regarding the models' validity be drawn. When the flood susceptibility maps are spatially overlaid with the land use/land cover maps, it is evident that 11 percent of agricultural land is located in very high to high susceptible zones, which may experience regular crop loss. 2 percent of the entire built-up area may be very to extremely susceptible to flooding (Table 4). The similar image is shown by field experience. Flooding in September 2022 resulted in significant crop loss and property devastation.

Table 4
Land use specific flood area detection from Fuzzy-RBF model

LULC/ Flood	Very low	low	Moderate	High	Very High
Waterbodies	0.004	0.027	0.022	0.070	1.372
Wetland	0.016	0.110	0.130	0.434	1.830
River sediment	0.162	0.533	0.523	0.620	0.367
Vegetation	7.424	8.695	3.254	3.127	0.309
Agriculture	16.221	23.975	15.276	8.219	2.916
Settlement	0.646	1.476	0.688	1.103	0.453

4.3 Accuracy assessment using ROC curve

The Receiver Operating Characteristic (ROC) Area Under Curve (AUC) values for all applied hybrid models range from 0.776 to 0.872, demonstrating very good to excellent agreement between Flood Susceptibility Models (FSMs) and the actual ground situation. So, all of the models could be considered to be reliable. The Fuzzy-RBF was found to be the most indicative of the three hybrid models used, with an AUC value of (0.872). Fuzzy-ANN was second, with (0.831), and Fuzzy-SVM was third (0.776). The best model for predicting flood susceptibility in the study area among the three hybrid models is fuzzy-RBF. Flood inventory map was used to validate the hybrid models of flood susceptibility (Fig. 5).

Table 5
Mean rank of flood susceptible models from Friedman test

Models	Mean Rank	Chi-square	Monte Carlo Sig. at 99% confidence interval
Fuzzy-ANN	2.05	589.039	0.00
Fuzzy-RBF	2.71		
Fuzzy-SVM	1.24		

4.4 Model comparison

For the purposes of validating and contrasting different flood susceptibility models, flood inventory data sets were used. The Friedman's test revealed that all hybrid models' average rank was notable at a value of 0.05, indicating a notable difference in the performance of flood susceptibility models (Table 5). The analysis demonstrates that the Fuzzy-RBF model's higher mean rank value (2.71) reflects its greater accuracy (Table 5). There are significant differences between all of the models, as indicated by the calculated chi-square value (589.039) and the degree of freedom. All models could not be compared as comparable models with satisfactory performance. The top three models are fuzzy-RBF, fuzzy- ANN, and fuzzy- SVM. The Friedman test is applied to the entire set of model outputs; pair-wise comparisons are not included in its scope. In order to compare the flood susceptibility models on a pair-by-pair basis, Wilcoxon Signed Rank tests were used. According to the results of the Wilcoxon Signed Rank test, which were calculated in Table 6, there were significant variations in the flood susceptibility models (P value < 0.05 and Z value > critical value).

Table 6
Wilcoxon signed rank test for flood susceptible models

Models		N	Mean Rank	Sum of Ranks	Z value	Monte Carlo Sig. at 99% confidence interval	Significance
Fuzzy-RBF vs Fuzzy-ANN	Negative Ranks	0[a]	0.00	0.00	-14.45	0	Yes
	Positive Ranks	265[b]	133.00	35245.00			
	Ties	135[c]					
Fuzzy-SVM vs Fuzzy-RBF	Negative Ranks	306[d]	154.50	47277.00	-15.194	0	Yes
	Positive Ranks	1[e]	1.00	1.00			
	Ties	93[f]					
Fuzzy-ANN vsFuzzy-SVM	Negative Ranks	1[g]	1.00	1.00	-15.202	0	Yes
	Positive Ranks	306[h]	154.50	47277.00			
	Ties	93[i]					
a. Fuzzy-RBF < Fuzzy-ANN	b. Fuzzy-RBF > Fuzzy-ANN	c. Fuzzy-RBF = Fuzzy-ANN	d. Fuzzy-SVM < Fuzzy-RBF	e. Fuzzy-SVM > Fuzzy-RBF			
f. Fuzzy-SVM = Fuzzy-RBF	g. Fuzzy-ANN < Fuzzy-SVM	h. Fuzzy-ANN > Fuzzy-SVM	i. Fuzzy-ANN = Fuzzy-SVM				

4.5 Significance of the study

This year (2022) also due to persistently severe rains in the region, a flood devastated many parts of Balrampur, Uttar Pradesh, posing challenges for the local villagers. The water level of rapti river remained above the danger level and flowing 130 cm above the danger mark. The flood that devastated more than 200 villages in the Balrampur area resulted in lengthy traffic lines at the National Highway-730 for vehicles. Water has risen 3 feet above ground level on the NH-730 as a result of the flooding. Following this, the district administration halted allowing vehicles to proceed, resulting in a lengthy line-up of trucks and other vehicles on the roadway. Thousands of people have been impacted by the flood in more than 200 villages so far. Jagtapur, Panditpurva, Jhovahna, Kalandarpur, Gangapur, Kodari, Lalpur, Phagunia, Jogiya Kalan, Lal Nagar, Durgapur, and Sherpur are just a few of the villages that was impacted (Fig. 7). The flood has spread beyond rural regions to urban areas as well. Many residential areas in urban area were submerged in water. Many colonies, including Shyam Bihar and the Civil Line of the One India Mall, was also flooded. Water levels on the Balrampur-Bahraich Marg have risen to two feet, forcing many locals to utilise boats to transport goods out of their stores. The water level in Gilauli Bhaga village at its peak flood stage is shown in Fig. 7a, which was taken just before the

monsoon. The state of the research area in flood in 2022 is depicted in Fig. 7. It is obvious that a more accurate and reliable map of flood vulnerability can reduce the expense and harm caused by floods. The generated maps can help governments, planners, managers, and decision-makers come up with better ways to limit additional urbanisation in the vulnerable areas in order to limit harm. The generated map might serve as a starting point for additional study, such as risk and hazard mapping. It can also help the planners and decision-makers take the right steps to regulate and limit this phenomenon in the research region.

5. Conclusions

Present study finds some areas of depressed land with centripetal-type flow convergence and accumulation along the river's edge, as well as the low-lying areas there, are more susceptible to flooding. It is also found to be more successful to predict flood vulnerability using an ensemble modelling approach as opposed to a single machine learning model. However, the individual model's performance in the current study is also satisfactory. The ensemble model's accuracy level is strongly influenced by the accuracy of the combined models. Only use highly ground truth representative models for ensemble modelling, it is advised. A significant portion of formed agricultural land is extremely vulnerable to flooding. The idea of leaving important floodways unusable due to the severe population pressure is untenable. Two methods should be employed to lessen the effects of a flood. A flood must first have a clear flow path in order to move swiftly and reduce both flood depth and stagnation time. flood depth and lateral extent could be reduced by storing flood water in natural reservoirs. Reclamation of wetlands must stop immediately in this case. The next step is to increase locals' ability to withstand flooding. It is obvious that homes constructed of mud must be replaced with elevated-base concrete structures. It is necessary to build crop granaries. When selecting crops during the monsoon, water-resistant varieties should be prioritised. The significance of flood control plans has increased recently, especially in India where flooding occurs yearly. However, the effectiveness of other regions' flood mitigation measures has not yet been examined. Thus, the current study, which was carried out in the Balrampur district of Uttar Pradesh using hybrid machine learning algorithms, will provide crucial information about the approaches to be used to support the local government and other parties in creating efficient flood mitigation strategies and planning land-use policy not only in India but also in other parts of the world.

Declarations

Ethical approval and consent to participate: Not Applicable

Consent to publish: All the co-authors agreed to publish the manuscript

Author's contributions: Satish Kumar Saini collected data, prepared maps, framework of the study and wrote part of manuscript **Susanta Mahato** prepared maps and wrote part of manuscript; **Deep Narayan Pandey** supervised the manuscript, correction and editing of manuscript; **Pawan Kumar Joshi** visualized the work, correction and editing of manuscript. All authors read and approved the final manuscript.

Competing interest: The authors declare that they have no conflict of interest.

Availability of data and materials: All the data and materials related to the manuscript are published with the paper, and available from the corresponding author upon request.

Funding:

"The authors declare that no funds, grants, or other support were received during the preparation of this

manuscript."

Declaration of Interest Statement

"The authors have no relevant financial or non-financial interests to disclose."

Conflict of interest: No potential conflict of interest was reported by the authors.

References

1. Abedi, R., Costache, R., Shafizadeh-Moghadam, H. and Pham, Q.B., 2022. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto International*, 37(19), pp.5479-5496.
2. Adedeji, O.H., Odufuwa, B.O. and Adebayo, O.H., 2012. Building capabilities for flood disaster and hazard preparedness and risk reduction in Nigeria: need for spatial planning and land management. *Journal of Sustainable Development in Africa*, 14(1), pp.45-58.
3. Akay, H., 2021. Flood hazards susceptibility mapping using statistical, fuzzy logic, and MCDM methods. *Soft Computing*, 25(14), pp.9325-9346.

4. Al-Juaidi, A.E., Nassar, A.M. and Al-Juaidi, O.E., 2018. Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. *Arabian Journal of Geosciences*, 11(24), pp.1-10.

5. Arabameri, A., Rezaei, K., Cerdà, A., Conoscenti, C. and Kalantari, Z., 2019. A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. *Science of the Total Environment*, 660, pp.443-458.

6. Arabameri, A., Rezaei, K., Cerdà, A., Conoscenti, C. and Kalantari, Z., 2019. A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. *Science of the Total Environment*, 660, pp.443-458.

7. Armal, S., Porter, J.R., Lingle, B., Chu, Z., Marston, M.L. and Wing, O.E., 2020. Assessing property level economic impacts of climate in the US, new insights and evidence from a comprehensive flood risk assessment tool. *Climate*, 8(10), p.116.

8. Arnell, N.W. and Gosling, S.N., 2016. The impacts of climate change on river flood risk at the global scale. *Climatic Change*, 134(3), pp.387-401.

9. Barney, J.N., Mann, J.J., Kyser, G.B. and DiTomaso, J.M., 2012. Assessing habitat susceptibility and resistance to invasion by the bioenergy crops switchgrass and Miscanthus× giganteus in California. *Biomass and Bioenergy*, 40, pp.143-154.

10. Barrocu, G. and Eslamian, S., 2022. Geomorphology and Flooding. In : *Flood Handbook* (pp. 23-54). CRC Press.

11. Beasley, T.M. and Zumbo, B.D., 2003. Comparison of aligned Friedman rank and parametric methods for testing interactions in split-plot designs. *Computational Statistics & Data Analysis*, 42(4), pp.569-593.

12. Bhukosh, Geological Survey of India. https://bhukosh.gsi.gov.in/Bhukosh/MapViewer.aspx. (Accessed on 15 January 2023)

13. Bouamrane, A., Derdous, O., Dahri, N., Tachi, S.E., Boutebba, K. and Bouziane, M.T., 2022. A comparison of the analytical hierarchy process and the fuzzy logic approach for flood susceptibility mapping in a semi-arid ungauged basin (Biskra basin: Algeria). *International Journal of River Basin Management*, 20(2), pp.203-213.

14. Brázdil, R., Kundzewicz, Z.W. and Benito, G., 2006. Historical hydrology for studying flood risk in Europe. *Hydrological Sciences Journal*, 51(5), pp.739-764.

15. Breugem, A.J., Wesseling, J.G., Oostindie, K. and Ritsema, C.J., 2020. Meteorological aspects of heavy precipitation in relation to floods–an overview. *Earth-Science Reviews*, 204, p.103171.

16. Buch, H., Trivedi, I.N. and Jangir, P., 2017. Moth flame optimization to solve optimal power flow with non-parametric statistical evaluation validation. *Cogent Engineering*, 4(1), p.1286731.

17. Buraas, E.M., Renshaw, C.E., Magilligan, F.J. and Dade, W.B., 2014. Impact of reach geometry on stream channel sensitivity to extreme floods. Earth *Surface Processes and Landforms*, 39(13), pp.1778-1789.

18. Cavalli, M., Goldin, B., Comiti, F., Brardinoni, F. and Marchi, L., 2017. Assessment of erosion and deposition in steep mountain basins by differencing sequential digital terrain models. *Geomorphology*, 291, pp.4-16.

19. Census 2011 (July 01). District Census Handbook – Uttar Pradesh. Retrieved July 09, 2018, from Office of the Registrar General & Census Commissioner, India

20. Chen, W., Hong, H., Li, S., Shahabi, H., Wang, Y., Wang, X. and Ahmad, B.B., 2019. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *Journal of Hydrology*, 575, pp.864-873.

21. Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F. and Mosavi, A., 2019. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment*, 651, pp.2087-2096.

22. Comes, T., Hiete, M., Wijngaards, N. and Schultmann, F., 2011. Decision maps: A framework for multi-criteria decision support under severe uncertainty. *Decision Support Systems*, 52(1), pp.108-118.

23. Crunch, C.R.E.D., 2020. 58—Disaster Year in Review (2019). CRED: Brussels, Belgium.

24. Dano, U.L., Balogun, A.L., Matori, A.N., Wan Yusouf, K., Abubakar, I.R., Said Mohamed, M.A., Aina, Y.A. and Pradhan, B., 2019. Flood susceptibility mapping using GIS-based analytic network process: A case study of Perlis, Malaysia. *Water*, 11(3), p.615.

25. Das, S. and Gupta, A., 2021. Multi-criteria decision based geospatial mapping of flood susceptibility and temporal hydro-geomorphic changes in the Subarnarekha basin, India. *Geoscience Frontiers*, 12(5), p.101206.

26. Das, S. and Gupta, A., 2021. Multi-criteria decision based geospatial mapping of flood susceptibility and temporal hydro-geomorphic changes in the Subarnarekha basin, India. *Geoscience Frontiers*, 12(5), p.101206.

27. Davies, G.I., McIver, L., Kim, Y., Hashizume, M., Iddings, S. and Chan, V., 2015. Water-borne diseases and extreme weather events in Cambodia: Review of impacts and implications of climate change. *International Journal of Environmental Research and Public Health*, 12(1), pp.191-213.

28. Dewan, T.H., 2015. Societal impacts and vulnerability to floods in Bangladesh and Nepal. *Weather and Climate Extremes*, 7, pp.36-42.

29. District Census Handbook   (DCHB, Census of India, 2011), Balrampur, Uttar Pradesh.

30. Diwakar, D. M., 2008. A Report On Base Line Survey of Minority Concentrated Balrampur District in Uttar Pradesh", Giri Institute of Development Studies Sector O, Aliganj Housing Scheme Lucknow, Uttar Pradesh.

31. Dragićević, S., Lai, T. and Balram, S., 2015. GIS-based multicriteria evaluation with multiscale analysis to characterize urban landslide susceptibility in data-scarce environments. *Habitat International*, 45, pp.114-125.

32. Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W. and Li, X., 2016. Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing*, 8(4), p.354.

33. Froehlich, D.C., 2008. Embankment dam breach parameters and their uncertainties. *Journal of Hydraulic Engineering*, 134(12), pp.1708-1721.

34. Giordano, A.R., Ridenhour, B.J. and Storfer, A., 2007. The influence of altitude and topography on genetic structure in the long-toed salamander (Ambystoma macrodactulym). *Molecular Ecology*, 16(8), pp.1625-1637.

35. Golden, H.E., Rajib, A., Lane, C.R., Christensen, J.R., Wu, Q. and Mengistu, S., 2019. Non-floodplain wetlands affect watershed nutrient dynamics: A critical review. *Environmental Science & Technology*, 53(13), pp.7203-7214.

36. Guzzetti, F. and Tonelli, G., 2004. Information system on hydrological and geomorphological catastrophes in Italy (SICI): a tool for managing landslide and flood hazards. *Natural Hazards and Earth System Sciences*, 4(2), pp.213-232.

37. Hammami, S., Zouhri, L., Souissi, D., Souei, A., Zghibi, A., Marzougui, A. and Dlala, M., 2019. Application of the GIS based multi-criteria decision analysis and analytical hierarchy process (AHP) in the flood susceptibility mapping (Tunisia). *Arabian Journal of Geosciences*, 12(21), pp.1-16.

38. Hester, E.T., Lin, A.Y.C. and Tsai, C.W., 2020. Effect of floodplain restoration on photolytic removal of pharmaceuticals. *Environmental Science & Technology*, 54(6), pp.3278-3287.

39. Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.X. and Chen, W., 2018. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. *Science of the Total Environment*, 625, pp.575-588.

40. Ielpi, A., Fralick, P., Ventra, D., Ghinassi, M., Lebeau, L.E., Marconato, A., Meek, R. and Rainbird, R.H., 2018. Fluvial floodplains prior to greening of the continents: Stratigraphic record, geodynamic setting, and modern analogues. *Sedimentary Geology*, 372, pp.140-172.

41. Ishaya, S., Ifatimehin, O.O. and Abaje, I.B., 2009. Mapping flood vulnerable areas in a developing urban centre of Nigeria. *Journal of Sustainable Development in Africa*, *11*(4), pp.180-194.

42. Islam, A.R.M.T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K.U., Pham, Q.B., Kuriqi, A. and Linh, N.T.T., 2021. Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12(3), p.101075.

43. Islam, S.M.F. and Karim, Z., 2019. World's demand for food and water: The consequences of climate change. *Desalination Challenges and Opportunities*, pp.57-84.

44. Janizadeh, S., Vafakhah, M., Kapelan, Z. and Dinan, N.M., 2021. Novel Bayesian additive regression tree methodology for flood susceptibility modeling. *Water Resources Management*, *35*(13), pp.4621-4646.

45. Jun, K.S., Chung, E.S., Kim, Y.G. and Kim, Y., 2013. A fuzzy multi-criteria approach to flood risk vulnerability in South Korea by considering climate change impacts. *Expert Systems with Applications*, 40(4), pp.1003-1013.

46. Kamali, F.P., Borges, J.A., Meuwissen, M.P., de Boer, I.J. and Lansink, A.G.O., 2017. Sustainability assessment of agricultural systems: The validity of expert opinion and robustness of a multi-criteria analysis. *Agricultural Systems*, 157, pp.118-128.

47. Khatoon, R., (1994) Floods in Uttar Pradesh: A geographical perspective, *Ph.D. thesis*, Department of Geography, Aligarh Muslim University.

48. Khattri, P., 2017. Rural Livelihoods and Natural Disasters: Observations from Flood Affected Bahraich District of Uttar Pradesh. *Indian Anthropologist*, 47(2), pp.17-33.

49. Kumar, A., Tewari, D.D. and Pande, Y.N., 2003. Ethno-Phytotherapeutics among Tharus of Beerpur-Semra forest range of Balrampur district (UP). *Journal of Economic and Taxonomic Botany*, 27(4), pp.839-844.

50. Liang, Y., Li, S., Yan, C., Li, M. and Jiang, C., 2021. Explaining the black-box model: A survey of local interpretation methods for deep neural networks. *Neurocomputing*, 419, pp.168-182.

51. Mahato, S. and Pal, S., 2019. Groundwater potential mapping in a rural river basin by union (OR) and intersection (AND) of four multi-criteria decision-making models. *Natural Resources Research*, 28(2), pp.523-545.

52. Mahato, S., Pal, S., Talukdar, S., Saha, T.K. and Mandal, P., 2021. Field based index of flood vulnerability (IFV): A new validation technique for flood susceptible models. *Geoscience Frontiers*, 12(5), p.101175.

53. Mahmoud, S.H. and Gan, T.Y., 2018. Urbanization and climate change implications in flood risk management: Developing an efficient decision support system for flood susceptibility mapping. *Science of the Total Environment*, 636, pp.152-167.

54. Martínez-Álvarez, F., Schmutz, A., Asencio-Cortés, G. and Jacques, J., 2018. A novel hybrid algorithm to forecast functional time series based on pattern sequence similarity with application to electricity demand. *Energies*, 12(1), p.94.

55. Minár, J., Evans, I.S. and Jenčo, M., 2020. A comprehensive system of definitions of land surface (topographic) curvatures, with implications for their application in geoscience modelling and prediction. *Earth-Science Reviews*, 211, p.103414.

56. Mind'je, R., Li, L., Amanambu, A.C., Nahayo, L., Nsengiyumva, J.B., Gasirabo, A. and Mindje, M., 2019. Flood susceptibility modeling and hazard perception in Rwanda. *International Journal of Disaster Risk Reduction*, 38, p.101211.

57. Molinari, D., Scorzini, A.R., Arrighi, C., Carisi, F., Castelli, F., Domeneghetti, A., Gallazzi, A., Galliani, M., Grelot, F., Kellermann, P. and Kreibich, H., 2020. Are flood damage models converging to "reality"? Lessons learnt from a blind test. *Natural Hazards and Earth System Sciences*, 20(11), pp.2997-3017.

58. Mousavi, S.M., Ataie-Ashtiani, B. and Hosseini, S.M., 2022. Comparison of statistical and mcdm approaches for flood susceptibility mapping in northern Iran. *Journal of Hydrology*, 612, p.128072.

59. Nachappa, T.G., Piralilou, S.T., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O. and Blaschke, T., 2020. Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. *Journal of Hydrology*, 590, p.125275.

60. NIDM. (2015). National Institute of Disaster Management. Retrieved from http://www.nidm.gov.in/ (Accessed on 15 January 2023)

61. Orr, M.J., 1996. Introduction to radial basis function networks.

62. Owrangi, A.M., Lannigan, R. and Simonovic, S.P., 2014. Interaction between land-use change, flooding and human health in Metro Vancouver, Canada. *Natural hazards*, *72*(2), pp.1219-1230.

63. Pal, S. and Singha, P., 2022. Analyzing sensitivity of flood susceptible model in a flood plain river basin. *Geocarto International*, 37(24), pp.7186-7219.

64. Pal, S., Khatun, S. and Mahato, S., 2021. Charland mapping and analyzing suitability for settlement and agriculture in ganga river within the stretch of West Bengal, India. *Geocarto International*, pp.1-24.

65. Pal, S., Kundu, S. and Mahato, S., 2020. Groundwater potential zones for sustainable management plans in a river basin of India and Bangladesh. *Journal of Cleaner Production*, 257, p.120311.

66. Pandey, H.K., Singh, V.K. and Singh, S.K., 2022. Multi-criteria decision making and Dempster-Shafer model–based delineation of groundwater prospect zones from a semi-arid environment. *Environmental Science and Pollution Research*, pp.1-19.

67. Pandey, M., Arora, A., Arabameri, A., Costache, R., Kumar, N., Mishra, V.N., Nguyen10, H., Mishra12, J., Siddiqui, M.A., Ray14, Y. and Soni, S., 2022. Flood Susceptibility Modeling in a Subtropical Humid Low-Relief Alluvial Plain Environment: Application of Novel Ensemble Machine Learning Approach. Geospace Observation of Natural Hazards.

68. Pandey, P., Chauhan, P., Bhatt, C.M., Thakur, P.K., Kannaujia, S., Dhote, P.R., Roy, A., Kumar, S., Chopra, S., Bhardwaj, A. and Aggrawal, S.P., 2021. Cause and process mechanism of rockslide triggered flood event in Rishiganga and Dhauliganga River Valleys, Chamoli, Uttarakhand, India using satellite remote sensing and in situ observations. *Journal of the Indian Society of Remote Sensing*, 49(5), pp.1011-1024.

69. Papandreou, C. and Ziakopoulos, A., 2022. Predicting VLCC fuel consumption with machine learning using operationally available sensor data. *Ocean Engineering*, 243, p.110321.

70. Pham, B.T., Phong, T.V., Nguyen, H.D., Qi, C., Al-Ansari, N., Amini, A., Ho, L.S., Tuyen, T.T., Yen, H.P.H., Ly, H.B. and Prakash, I., 2020. A comparative study of kernel logistic regression, radial basis function classifier, multinomial naïve bayes, and logistic model tree for flash flood susceptibility mapping. *Water*, 12(1), p.239.

71. Pierce, M.W., Thornton, C.I. and Abt, S.R., 2010. Predicting peak outflow from breached embankment dams. *Journal of Hydrologic Engineering*, 15(5), p.338.

72. Planning Commission (2011) Report of working group on food management and region-specific issues for XII plan. Govt of India, New Delhi.

73. Planning Commission (2011) Report of working group on food management and region-specific issues for XII plan. Govt of India, New Delhi.

74. Prakash, S.A.D.G.U.R.U. and Verma, A.K., 2020. Seasonal Variations in Prevalence of Ectoparasitic Infestation in Indian Major Carps at Balrampur, Up, India. *Journal of Zoology*, 41(10), pp.121-127.

75. Rahmati, O., Pourghasemi, H.R. and Zeinivand, H., 2016. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto International*, 31(1), pp.42-70.

76. Rahmati, O., Pourghasemi, H.R. and Zeinivand, H., 2016. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto International*, 31(1), pp.42-70.

77. Rodríguez-Caballero, E., Cantón, Y., Chamizo, S., Lázaro, R. and Escudero, A., 2013. Soil loss and runoff in semiarid ecosystems: a complex interaction between biological soil crusts, micro-topography, and hydrological drivers. *Ecosystems*, 16(4), pp.529-546.

78. Sahana, M. and Patel, P.P., 2019. A comparison of frequency ratio and fuzzy logic models for flood susceptibility assessment of the lower Kosi River Basin in India. *Environmental Earth Sciences*, 78(10), pp.1-27.

79. Sanjay, M.C. and Prakash, S., 2020. Ichthyofaunal diversity of rapti river flowing through shravasti and balrampur districts of Uttar Pradesh (India), *Bulletin of Pure and Applied Sciences*, *39*(2).

80. Santos, P.P., Reis, E., Pereira, S. and Santos, M., 2019. A flood susceptibility model at the national scale based on multicriteria analysis. *Science of The Total Environment*, *667*, pp.325-337.

81. Semenza, J.C., 2020. Cascading risks of waterborne diseases from climate change. Nature Immunology, 21(5), pp.484-487.

82. ShafapourTehrany, M., Kumar, L., NeamahJebur, M. and Shabani, F., 2019. Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. *Geomatics, Natural Hazards and Risk*, 10(1), pp.79-101.

83. ShafapourTehrany, M., Kumar, L., NeamahJebur, M. and Shabani, F., 2019. Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. *Geomatics, Natural Hazards and Risk*, 10(1), pp.79-101.

84. ShafapourTehrany, M., Kumar, L., NeamahJebur, M. and Shabani, F., 2019. Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. *Geomatics, Natural Hazards and Risk*, 10(1), pp.79-101.

85. Singh, O., & Kumar, M. (2013). Flood events, fatalities and damages in India from 1978 to 2006. *Natural Hazards*, 69(3), 1815-1834.

86. Singh, P. (2018). Flood Study in Balrampur District of Uttar Pradesh India. *International Journal of Science and Research*, 9 (6), pp. 62-65.

87. Sugianto, S., Deli, A., Miswar, E., Rusdi, M. and Irham, M., 2022. The Effect of Land Use and Land Cover Changes on Flood Occurrence in Teunom Watershed, Aceh Jaya. *Land*, 11(8), p.1271.

88. Talukdar, S. and Pal, S., 2020. Modeling flood plain wetland transformation in consequences of flow alteration in Punarbhaba river in India and Bangladesh. *Journal of Cleaner Production*, 261, p.120767.

89. Talukdar, S., Ghose, B., Salam, R., Mahato, S., Pham, Q.B., Linh, N.T.T., Costache, R. and Avand, M., 2020. Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. Stochastic *Environmental Research and Risk Assessment*, 34(12), pp.2277-2300.

90. Tehrany, M.S., Pradhan, B. and Jebur, M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology*, 512, pp.332-343.

91. Termeh, S.V.R., Kornejady, A., Pourghasemi, H.R. and Keesstra, S., 2018. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Science of the Total Environment*, 615, pp.438-451.

92. Tripathi, P. (2015). Flood disaster in India: an analysis of trend and preparedness. *Interdisciplinary Journal of Contemporary Research*, 2(4), 91-98.

93. Ullah, W., Wang, G., Lou, D., Ullah, S., Bhatti, A.S., Ullah, S., Karim, A., Hagan, D.F.T. and Ali, G., 2021. Large-scale atmospheric circulation patterns associated with extreme monsoon precipitation in Pakistan during 1981–2018. *Atmospheric Research*, 253, p.105489.

94. United States Geological Survey (USGS). Accessed through https://earthexplorer.usgs.gov/.

95. Usama, M. (2015). Management of Floods in Flood Prone Regions of Eastern Uttar Pradesh. *Management*, 2(11).

96. Vafakhah, M., Mohammad HasaniLoor, S., Pourghasemi, H. and Katebikord, A., 2020. Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping. *Arabian Journal of Geosciences*, 13(11), pp.1-16.

97. Vojtek, M. and Vojteková, J., 2019. Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. *Water*, 11(2), p.364.

98. Wang, Y., Fang, Z., Hong, H. and Peng, L., 2020. Flood susceptibility mapping using convolutional neural network frameworks. Journal of Hydrology, 582, p.124482.

99. Whipple, K.X., Parker, G., Paola, C. and Mohrig, D., 1998. Channel dynamics, sediment transport, and the slope of alluvial fans: experimental study. The Journal of geology, 106(6), pp.677-694.

100. Yang, J.L. and Zhang, G.L., 2011. Water infiltration in urban soils and its effects on the quantity and quality of runoff. *Journal of soils and sediments*, *11*(5), pp.751-761.

101. Yang, Y., Guo, H., Wang, D., Ke, X., Li, S. and Huang, S., 2021. Flood vulnerability and resilience assessment in China based on super-efficiency DEA and SBM-DEA methods. Journal of Hydrology, 600, p.126470.

102. Zadeh, L.A., Klir, G.J. and Yuan, B., 1996. Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers (Vol. 6). World Scientific.

103. Zaharia, L., Costache, R., Prăvălie, R. and Ioana-Toroimac, G., 2017. Mapping flood and flooding potential indices: a methodological approach to identifying areas susceptible to flood and flooding risk. Case study: the Prahova catchment (Romania). *Frontiers of Earth Science*, *11*(2), pp.229-247.

104. Zellou, B. and Rahali, H., 2019. Assessment of the joint impact of extreme rainfall and storm surge on the risk of flooding in a coastal area. *Journal of Hydrology*, *569*, pp.647-665.

105. Zingaro, M., Refice, A., D'Addabbo, A., Hostache, R., Chini, M. and Capolongo, D., 2020. Experimental application of sediment flow connectivity index (SCI) in flood monitoring. *Water*, 12(7), p.1857.
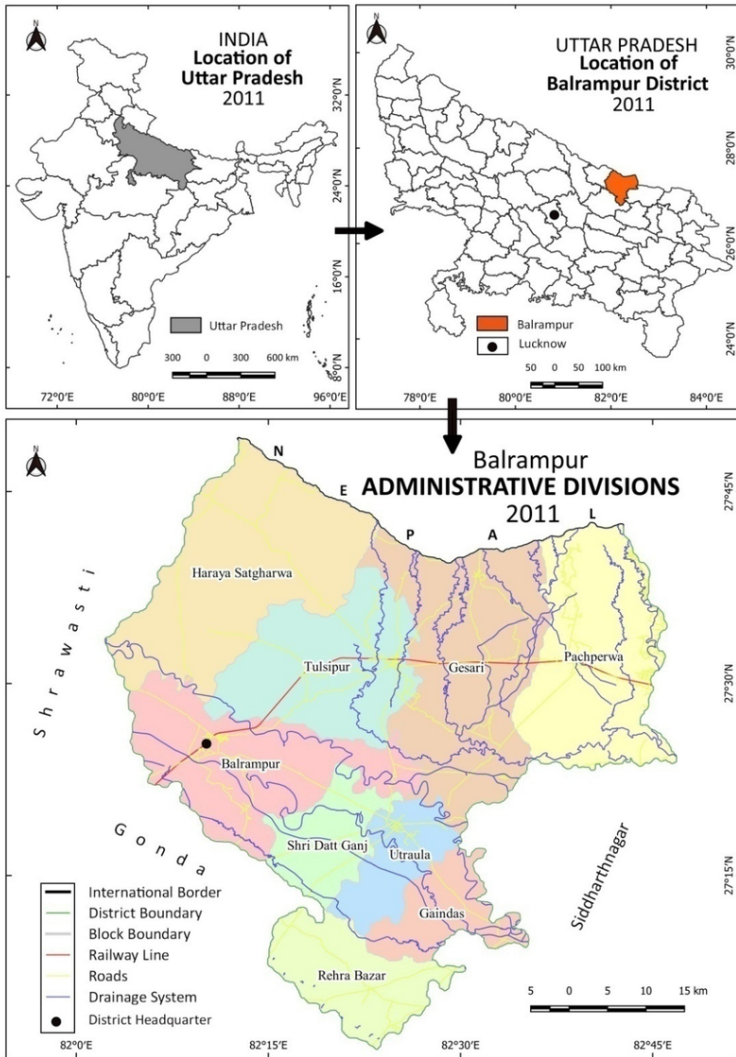
Figures

Figure 1

Map showing the location of study area in India and the drainage network, rail and road infrastructure along with sub-district units (blocks)
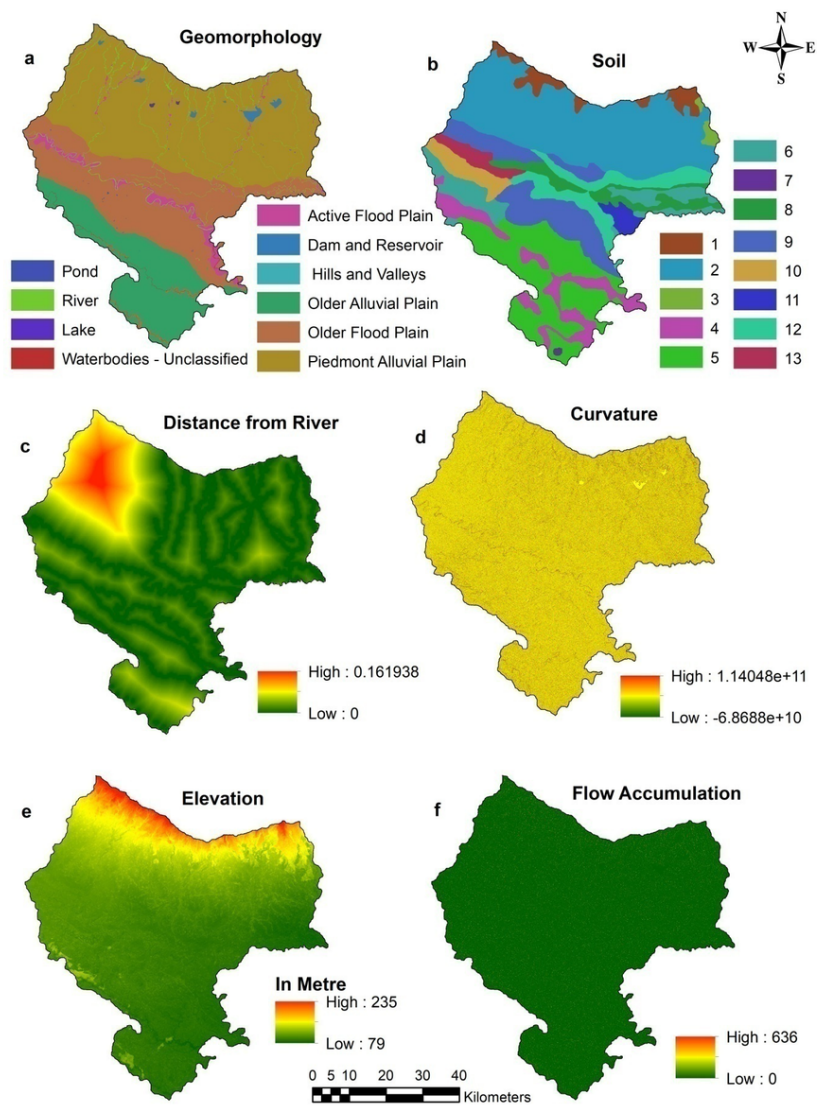
Figure 2

Flood conditioning parameters- a. Geomorphology; b. Soil; c. Distance from River; d. Curvature; e. Elevation and f. Flow Accumulation
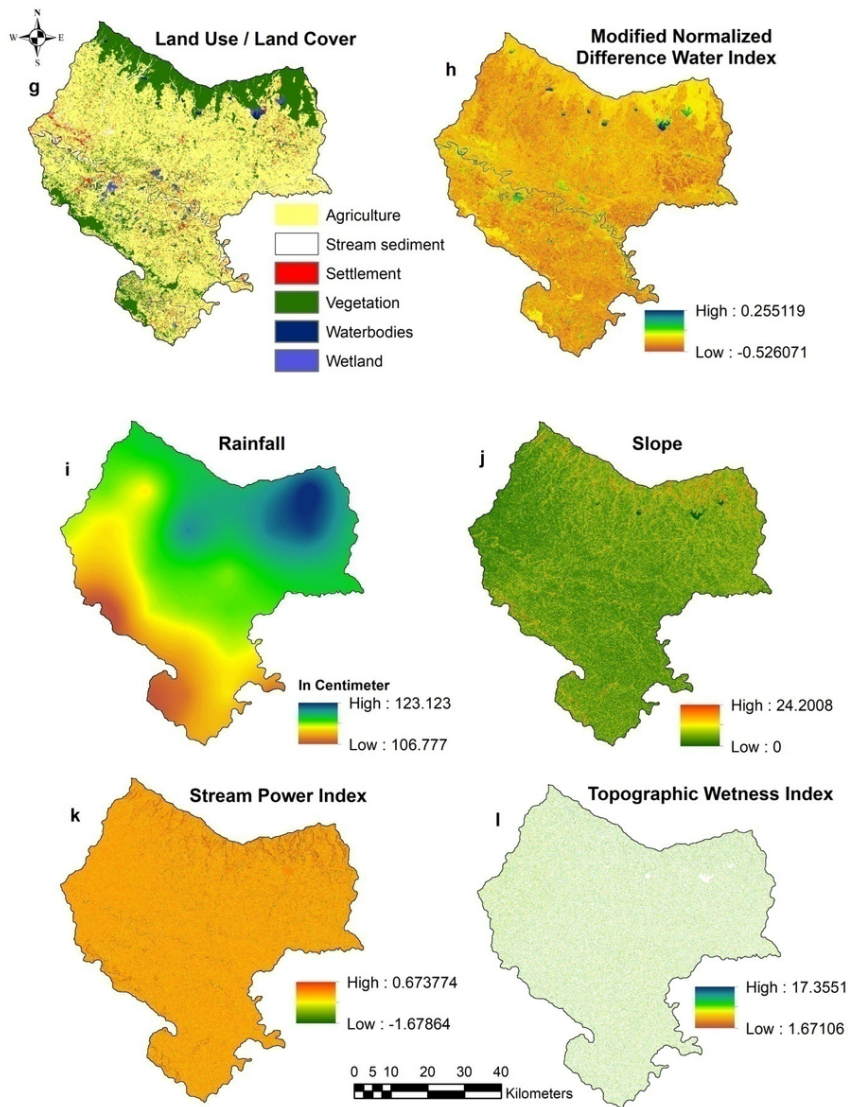
Figure 3

Flood conditioning parameters- g. Land Use / Land Cover; h. Modified Normalized Difference Water Index; i. Rainfall; j. Slope; k. Stream Power Index; l. Topographic Wetness Index
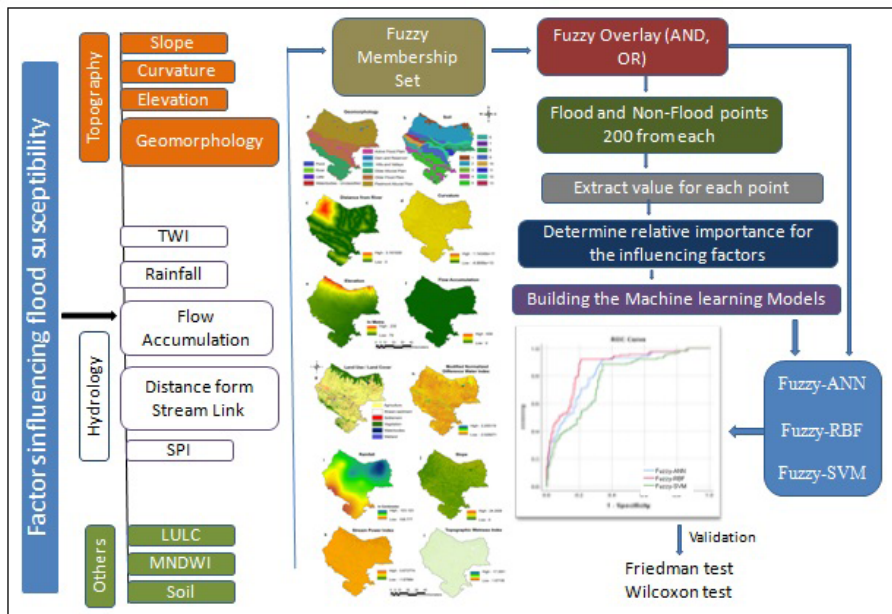
Figure 4
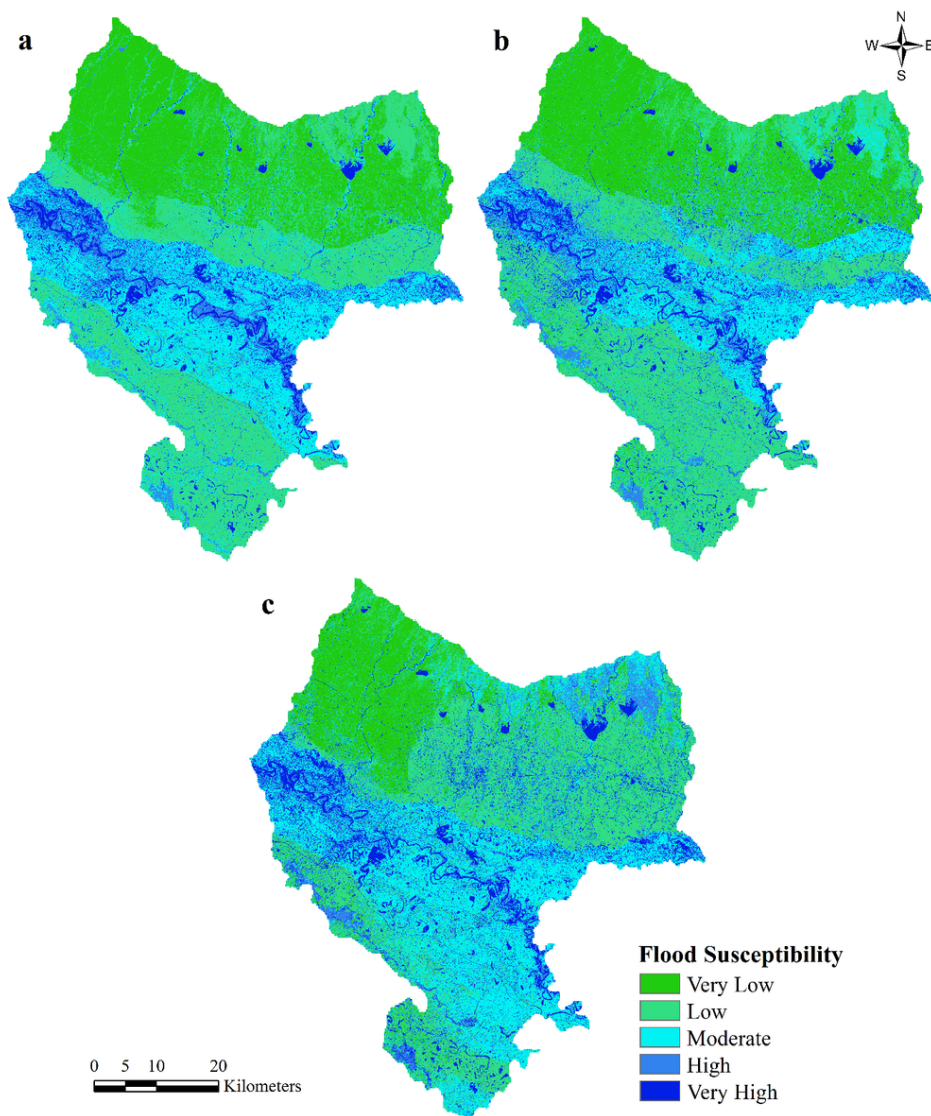
Methodological framework of the study

Flood Susceptibility
- Very Low
- Low
- Moderate
- High
- Very High

0 5 10   20
Kilometers

Figure 5

Figure 4 Flood Susceptibility maps through; a (Fuzzy-ANN); b (Fuzzy-RBF); c (Fuzzy-SVM)
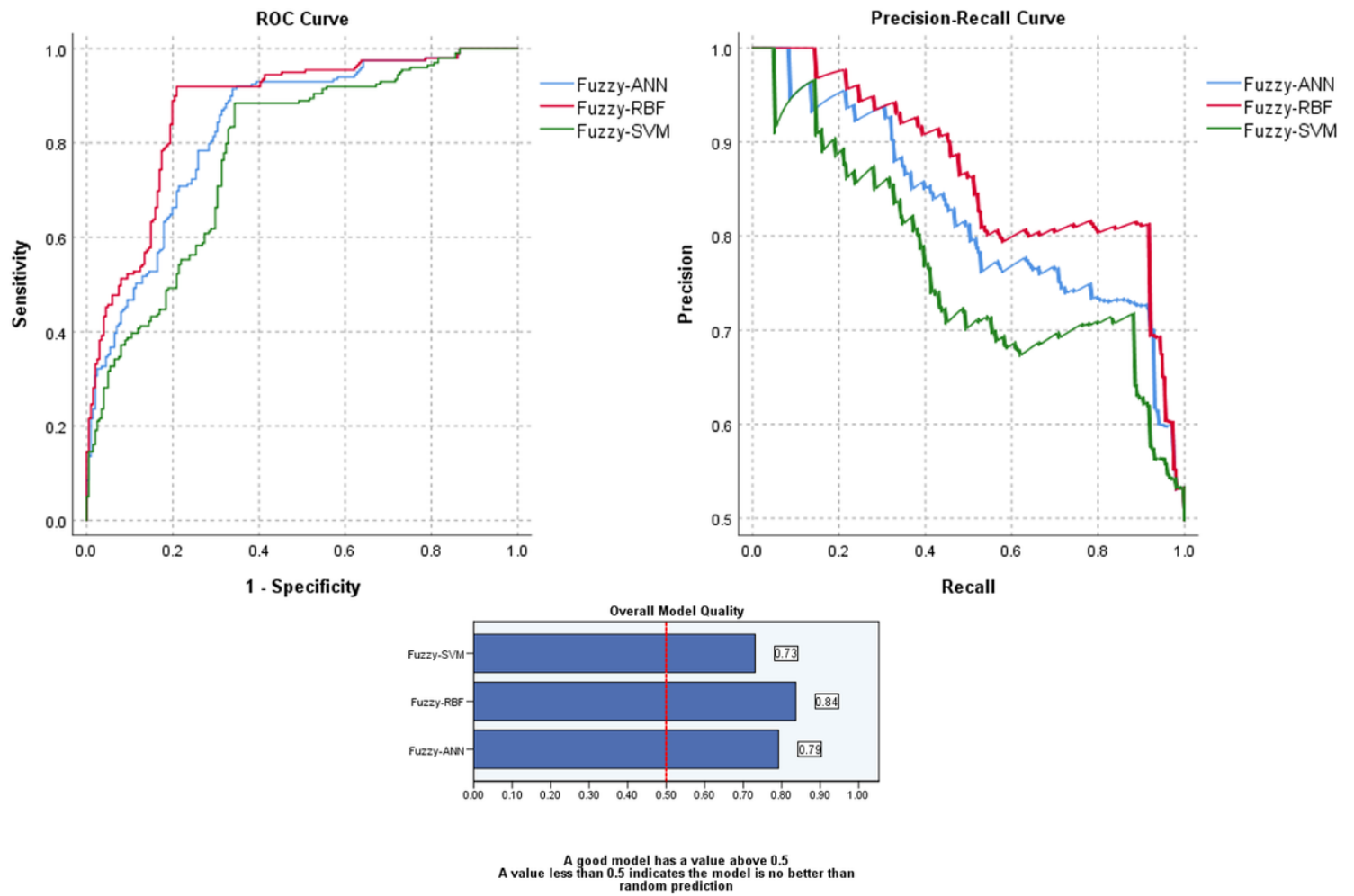

Figure 6

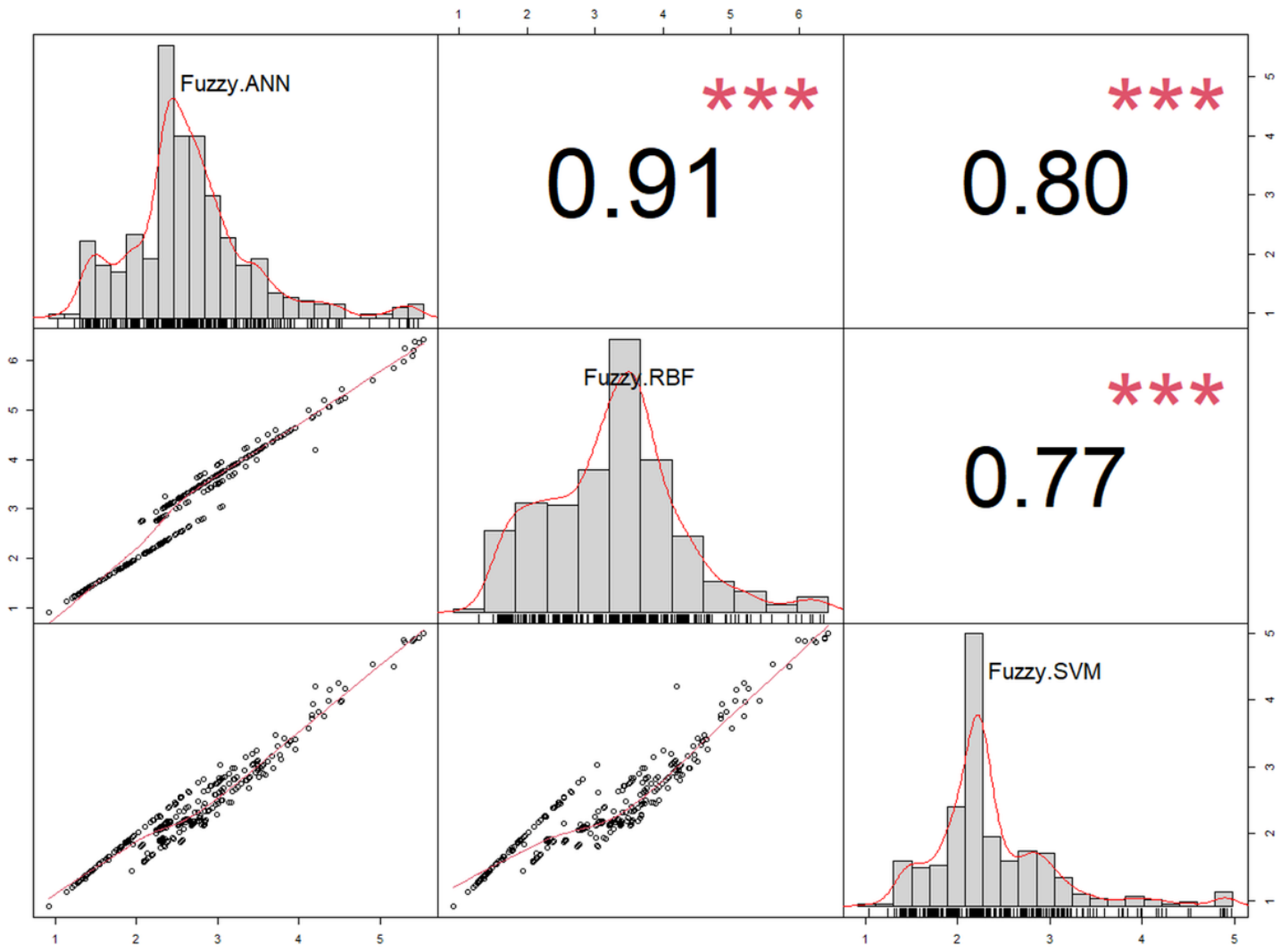Figure 5Model validation using Receiver Operating Characteristic curve (ROC)

Figure 7

Figure 6 The correlation between the flood susceptible models

Figure 8

Figure 7 Severe flooding and water logging problems in various locations throughout the Balrampur district during survey work