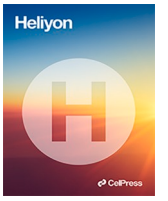
Heliyon 11 (2025) e41941

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Classification of short-term flood events using stochastic variable selection and Gaussian Naïve Bayes classifier: A case study of Sirajganj district, Bangladesh

Chandan Mondal [a,b], Md Jahir Uddin [a,*]

[a] Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna, 9203, Bangladesh
[b] Office of Planning and Development, Rabindra University, Bangladesh

ARTICLE INFO

Keywords:
Gaussian Naïve Bayes
Mutual information
Flood events
Stochastic variable
Predictor

ABSTRACT

Around the world, catastrophes caused by flooding are occurring naturally that cause a great deal of fatalities and financial loss. The loss of life and property can be considerably reduced with precise flood forecasts. The complexity of many flood predicting techniques makes the results difficult to interpret, compromising the process's core goal. This study uses a quick and flexible Gaussian Naïve Bayes (GNB) classifier to categorize eight different years as flooded or non-flooded based on predictor variables obtained via the Mutual Information (MI) technique. During the search, all-sky surface shortwave downward irradiance is identified as the optimum predictor variable out of nineteen stochastic variables, with the highest sensitivity for model accuracy. The model is then validated using four iterations derived from the MAPE of the GNB classification method for Twenty-five percent mean error rates from 4-fold cross-validation indicate that this classification model is suitable for flood forecasting. This high rate of mean error is caused by the short amount of data utilized as training data, as GNB requires huge data records to get effective results. This research could aid in the development and evaluation of hydrological projects in the Sirajganj district.

1. Introduction

Around the world, floods are the most frequent disaster that cause significant financial damages. Normally dry places may be completely or partially submerged as a result of tidal or inland water overflow or the quick formation of runoff [1]. One such hazard is flooding, which is becoming more prevalent worldwide and currently causes 40 billion US dollars in damage annually, affecting around 250 million people [2]. Flood forecasting plays a critical role in lowering flood-related hazards by providing timely hazard information to practitioners, government decision-makers, and citizens who are at risk [3]. There are three main categories of flood forecasting techniques: data-driven models, physically-based models, and hybrid models, which combine data-driven and physically-based models. It has always been challenging to create a physically based hydrological model for flood prediction in catchments with little or no data [4]. By using sophisticated data-driven machine learning approaches, it is possible to overcome the constraints of traditional hydrological models in capturing the intricate, non-linear interactions present in flood dynamics and create more accurate and dependable flood estimation models [5]. Different researchers employ various data-driven methods for flood

* Corresponding author.
E-mail addresses: chandance2k7@gmail.com (C. Mondal), jahirce99@gmail.com (M.J. Uddin).

https://doi.org/10.1016/j.heliyon.2025.e41941
Received 14 October 2024; Received in revised form 12 January 2025; Accepted 13 January 2025
Available online 13 January 2025

forecasting utilizing historical data, including Artificial Neural Network (ANN) [6], Non-parametric and Linear approaches [7], Random Forest (RF) [8], Exponential Smoothing-Long-Short Term Memory (ES-LSTM) [9], Bagging [10], SMOreg [11], Multilayer Perceptron (MLP) [12], Adaptive Neuro-Fuzzy Inference System (ANFIS) [5], and Gaussian Naïve Bayes approaches [13], among others. Data-driven flood classification forecasting can increase the prediction performance effectively by determining predictor variables [14,15]. Alike to prior researchers, the economy and locality of Sirajganj district, Bangladesh, can be maintained with greater success by using data-driven flood classification forecasting.

Sirajganj, an economically significant but disaster-prone district situated alongside the Jamuna River in northwestern Bangladesh, is the locality of several groups of people staying in the low-lying, insecure valleys [16]. According to researcher [17], this district is one of the most flooded districts in Bangladesh, with residents struggling to cope with the natural calamity. This is supported by research conducted by different researchers [16,17], who note that the district experienced flooding in 1949, 1956, 1961, 1962, 1966, 1968, 1974, 1979, 1987, 1988, 1996, 1998, 2002, 2004, 2008, 2016, 2018, and 2020. Some researchers [16–20] recently showed their interest in the flooding scenario of Sirajganj District. A group of researchers [17] concentrate on flood frequency studies utilizing water level data, flood inundation map generation for various return times, flood depth vs. damage curve plotting to address sensitive locations and damages, and so on in this district. Scientist [18] demonstrates a keenness in the application of geographic information systems to examine a portion of the district, where they focus on flood frequency analyses and flood inundation mapping. Another scholar [19] is interested in impact-based forecasting and warning services; thus, they studied a section of Sirajganj and validated their technique through a limited number of focus group interviews at the community level. Researchers [16] conducts statistical analyses to investigate flood propagation, water level variation in relation to the Jamuna River stage in the floodplain, and the development of a hydrodynamic model of the flood spread route in a village in Sirajganj District. Besides, scholar [20] makes a flood susceptibility valuation using the flood vulnerability index technique in Sirajganj Sadar Upazila.

As disasters like floods are common in Sirajganj, flood research, particularly flood forecasting, may help to develop a more viable economy and community in this region. However, research on floods is extremely rare in this area, and research on flood forecasting is nearly nonexistent. The forecasting of floods in this region is of little interest to researchers [19]. Lack of appropriate data that may support this kind of study may be the root of the problem. Therefore, data-driven modeling's ability to identify the best predictor among the available stochastic variables can open up opportunities for this district's flood forecasting study. The present study uses the Gaussian Naïve Bayes (GNB) classifier to classify flood events using an optimum predictor, while the Mutual Information (MI) approach is employed to find the sensitivity of the stochastic variable from which the predictor is chosen. Naïve Bayes classification is a statistical approach for classification and supervised learning that is derived from the Bayes theorem. Made out of directed acyclic graphs, Naïve Bayes is a very simple Bayesian network. It has been verified that Naïve Bayes outperforms well-known complex classification techniques, despite its simplicity and straightforwardness. The GNB approach is used, founded on the idea that the parameters follow a Gaussian distribution [13]. The years 1988, 1998, 2008, and 2018 were selected for the research work as flooded years, according to the study by some researchers [16,17], where a ten-year interval is tried to be maintained. The years 1993, 2003, 2014, and 2023 are selected as non-flooded years, where a ten-year interval is also tried to be maintained in the same manner. Due to the lack of appropriate data in 2013, 2014 has been utilized rather than 2013. This data-driven classification approach is then validated using the K-fold cross-validation methodology, which is one of the most often used methods by researchers for model selection and classifier error assessment [21]. The primary aim and objectives of this study are listed as: i. Identifying new stochastic variables that are appropriate for better flood event classification, ii. Evaluating the performance of data-driven models using limited data sets,

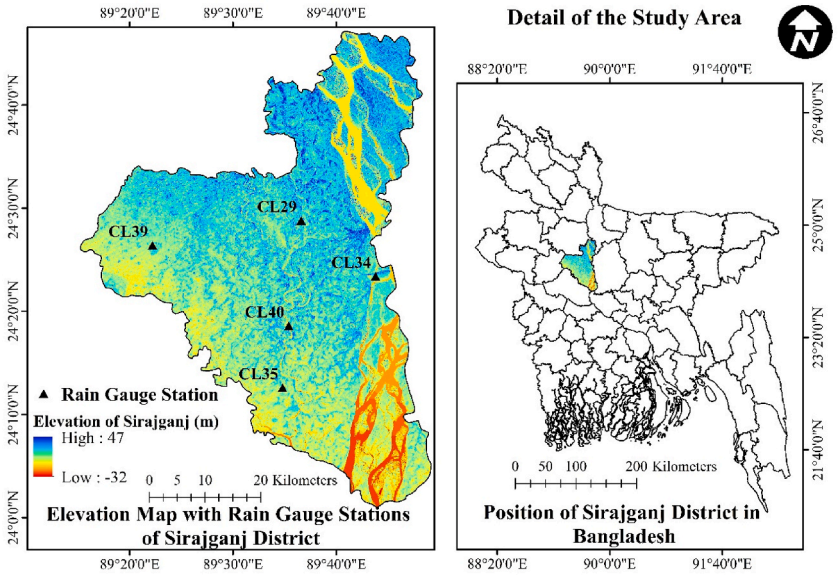


Fig. 1. Elevation and position of study area.

and iii. Suggesting an innovative classification method that can more accurately forecast flood conditions.

This data-driven flood classification forecasting technique is simple, rapid, and inexpensive, and it explains the year's flood conditions based on environmental factors prior to the rainy season. The proposed GNB classification technique can be utilized to produce short-term predictions of the year's flood using the optimum stochastic variable known as a predictor.

2. Methodology

2.1. Study area

The selected study area is Sirajganj, an economically important but disaster-prone district of the Rajshahi division of Bangladesh. Sirajganj district is an agriculturally dominant area in this division, located in the active Brahmaputra-Jamuna floodplain [18]. This district covers an area of 2497.92 square kilometers and is located between latitudes of 24°01′ and 24°47′ North and longitudes of 89°15′ and 89°59′ East. Its elevation lies between −32 m and 47 m above mean sea level. In comparison to other parts of the district, the western and southern provinces of Sirajganj have much lower elevations relative to the mean sea level (Fig. 1). It is surrounded on the north by the Bogra district, on the south by the Pabna and Manikganj districts, on the east by the Tangail and Jamalpur districts, and on the west by the Pabna, Natore, and Bogra districts. Jamuna, Baral, Ichamati, Hurasagar, Chalanbil, etc. are the main rivers and reservoirs of the district. 10 % of Tarash Upazila of the Sirajganj district is covered by the Chalan bill [22]. The climate of this district is usually denoted as tropical monsoon climate. A tropical monsoon weather is distinguished through strong heat, humidity, pressure from the wind, moderate temperatures, and considerable seasonal variations. The most noticeable aspect of the research area's environment is the varied winds that distinguish the summer and winter seasons and play a vital role in the South Asian subcontinental flow structure. Although it is generally assumed that this region has six different seasons, based on rainfall and temperature, three separate seasons can be recognized: the dry winter (November to February), the pre-seasonal monsoonal summer (March to May), and the rainy season (June to October) [23].

2.2. Data

The Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global Digital Elevation Model (DEM) and satellite images from the years 1988, 1993, 1998, 2003, 2008, 2014, 2018, and 2023 are collected from the USGS EarthExplorer website (https://earthexplorer.usgs.gov/). As a fact-finding organization within the US government, the USGS gathers, tracks, evaluates, and disseminates scientific data regarding the conditions and problems of natural resources [24]. Among all the above-mentioned years, 1988, 1993, 1998, and 2008 have thematic mapper (TM) images from the Landsat-5 mission. The year 2003 has both thematic mapper (TM) and enhanced thematic mapper (ETM) images from the Landsat-5 and Landsat-7 missions, respectively, and 2014, 2018, and 2023 have operational land imager (OLI) and thermal infrared sensor (TIRS) images from the Landsat-8 mission. The basic information about the remote sensing data used for this study is summarized in Table 1. Earth Skin Temperature data is obtained from the website https://power.larc.nasa.gov/of NASA POWER. NASA POWER offers current and upcoming climate services for the energy, agricultural, and sustainable buildings communities using NASA data [25]. Five rain gauge stations (Table 2) collect January, February, and March

Table 1
Basic information of remotely sensed data used for the study.

Year	Date of Acquired	Spacecraft ID	Sensor ID	WRS Path/WRS Row	Datum & Ellipsod	UTM Zone	Sun Elevation	Earth Sun Distance	Cloud Cover
1988	10.02.1988	LANDSAT 5	TM	138/43	WGS84	45	38.45828592	0.9867072	0.00
	08.11.1988	LANDSAT 5	TM	138/43	WGS84	45	41.39668051	0.9906479	0.00
1993	22.01.1993	LANDSAT 5	TM	138/43	WGS84	45	33.77229022	0.9841958	0.00
	08.12.1993	LANDSAT 5	TM	138/43	WGS84	45	34.13981498	0.9850279	0.00
1998	05.02.1998	LANDSAT 5	TM	138/43	WGS84	45	38.44972120	0.9859420	0.00
	04.11.1998	LANDSAT 5	TM	138/43	WGS84	45	43.64241458	0.9917835	0.00
2003	26.01.2003	LANDSAT 7	ETM	138/43	WGS84	45	38.14748184	0.9845552	2.00
	18.11.2003	LANDSAT 5	TM	138/43	WGS84	45	40.06671236	0.9885857	0.00
2008	17.02.2008	LANDSAT 5	TM	138/43	WGS84	45	43.59086106	0.9880054	0.00
	01.12.2008	LANDSAT 5	TM	138/43	WGS84	45	37.89191832	0.9860260	1.00
2014	05.03.2014	LANDSAT 8	OLI, TIRS	138/43	WGS84	45	50.30846027	0.9918062	0.03
	02.12.2014	LANDSAT 8	OLI, TIRS	138/43	WGS84	45	39.45762644	0.9859603	0.01
2018	12.02.2018	LANDSAT 8	OLI, TIRS	138/43	WGS84	45	43.68229753	0.9871717	0.21
	11.11.2018	LANDSAT 8	OLI, TIRS	138/43	WGS84	45	44.07959061	0.9901353	0.03
2023	26.02.2023	LANDSAT 8	OLI, TIRS	138/43	WGS84	45	47.85643812	0.9899798	0.03
	09.11.2023	LANDSAT 8	OLI, TIRS	138/43	WGS84	45	44.69263134	0.9906979	0.01

monthly rainfall data from the Bangladesh Water Development Board (BWDB). Since 1959, the Bangladeshi government has authorized BWDB to gather hydrological data throughout the nation [26].

2.3. Analysis technique

2.3.1. Data processing

In this study, Microsoft Excel and various tools from ArcMap, the main application of ArcGIS (version 10.4.1) software, are utilized. Before use, all the satellite images are obtained from Landsat 5, 7, and 8 missions compared with the US_Army_maps_v_17 map using the ArcMap application and Google Earth Pro for geometric corrections [27]. The rainfall time series' missing values are estimated using the Inverse Distance Weighting (IDW) interpolation technique. Many researchers use IDW, a deterministic method for multivariate interpolation using a known distributed set of points [28–31]. Then, double mass analysis is used in order to ensure the consistency of rainfall data [32,33]. For assessing the consistency of hydrological data, double mass analysis is a simple graphical technique. It is a common technique for examining the patterns of records created from meteorological or hydrological data collected at various sites [34].

2.3.2. Land use land cover (LULC) classification

Each satellite image is categorized based on the year that the flood occurred. Digital numbers are converted to spectral radiance, obtained from all types of satellite images captured by the TIRS, OLI, ETM, and TM sensors. The equation below (Eq. (1)) is used to transform digital numbers (DN) in a 1 G product to their original radiance units for TM and ETM [35].

$$L_\lambda = G_{rescale} \times Q_{cal} + B_{rescale} \tag{1}$$

Here L_λ and Q_{cal} means spectral radiance in watts/(m^2.sr.μm) and quantized calibrated pixel value in DN respectively, $G_{rescale}$ and $B_{rescale}$ are band-specific rescaling factors, respectively units of W/(m^2.sr. μm))/DN and W/(m^2.sr. μm). Values of $G_{rescale}$ and $B_{rescale}$ are provided by a scholar [35].

The equation below (Eq. (2)) is employed to transform DNs in a 1 G product into the radiance unit for OLI and TIRS [36].

$$L_\lambda = M_L{}^* Q_{cal} + A_L \tag{2}$$

Here, M_L, Q_{cal} , A_L means radiance multiplicative scaling factor, pixel value in DN and radiance additive scaling factor respectively.

Using spectral radiance from TM, ETM, and OLI images, LULC classifications are established. In order to comply with study area requirements, false-color composite (FCC) images have been generated using layer-stacking radiance taken from four spectral bands (blue, green, red, and infrared) [37]. LULC from false-color composite (FCC) images has been classified unsupervisedly using the Iso Cluster classifier from the ArcMap application of ArcGIS software. To validate the LULC classifications, one hundred reference points are randomly constructed for each time of the year by digitizing one hundred samples from Google Earth for that year [38].

Accuracy assessment LULC can be analyzed (Eq. (3)) by dint of estimating User Accuracy, Producer Accuracy, Overall Accuracy [39], Kappa Coefficient [40].

$$UA = \frac{PCC}{PTR}{}^* 100 \tag{3}$$

Here, *UA*, *PCC* and *PTR* denote user accuracy, number of accurately categorized events in a particular category, total number of categorized events in a particular category respectively (Eq. (4)).

$$PA = \frac{PCC}{PTC}{}^* 100 \tag{4}$$

PA and *PTC* denote producer accuracy and total number of reference events in a particular category respectively (Eq. (5)).

$$OA = \frac{TCS}{TS}{}^* 100 \tag{5}$$

OA, TCS and TS denote overall accuracy, total number of properly categorized events, total sample respectively (Eq. 6)

Table 2
Geographic details of Rain gauge stations provided by Bangladesh Water Development Board (BWDB).

Station ID	Station Name	District	Latitude	Longitude	Elevation (m)(From Mean Sea Level)
CL29	Raiganj	Sirajganj	24.48	89.61	14.33
CL34	Sirajganj	Sirajganj	24.39	89.73	11.28
CL35	Shazadpur	Sirajganj	24.21	89.58	10.06
CL39	Taras	Sirajganj	24.44	89.37	13.11
CL40	Ullapara	Sirajganj	24.31	89.59	8.84

$$\kappa = \frac{TS*TCS - \sum CT*RT}{TS^2 - \sum CT*RT} * 100 \tag{6}$$

where, κ, CT and RT denote Kappa coefficient, column and row total respectively.

2.3.3. Normalized difference vegetation index (NDVI) and normalized difference water index (NDWI) analyses

For NDVI and NDWI analyses of various time periods, planetary reflectance obtained from TM, ETM, and OLI images are utilized. Radiance is converted to spectral reflectance for TM and ETM images. The digital number for OLI images is converted directly to spectral reflectance. In the case of TM and ETM images, combined surface and atmospheric reflectance can be calculated by the equation (Eq. (7)) below [41].

$$\rho_p = \frac{\pi L_\lambda d^2}{ESUN_\lambda \cos\theta_s} \tag{7}$$

here, ρ_p stands for planetary reflectance (unit-less), π is approximately equal to 3.14159, L_λ is spectral radiance on sensor's opening, d and $ESUN_\lambda$ are distance between Earth and Sun (astronomical units) and mean solar exo-atmospheric irradiances respectively, mentioned by a researcher [35], θ_s is solar zenith angle (degrees).

As OLI images, the formula (Eq. (8)) that follows is employed to transform Level 1 DN values to TOA reflectance [36].

$$\rho_p{}' = M_\rho Q_{cal} + A_\rho \tag{8}$$

$\rho_p{}'$ means TOA Planetary Spectral Reflectance, deprived of alteration for solar angle. M_ρ, Q_{cal}, A_ρ are reflectance multiplicative scaling factor, Level 1 pixel value in DN and reflectance additive scaling factor respectively. $\rho_p{}'$ does not represent genuine TOA reflection since it lacks an adjustment for solar elevation angle. When a solar elevation angle is selected, the transformation to real TOA reflectance is as (Eq. (9)) below [36].

$$\rho_p = \frac{\rho_p{}'}{\cos(\theta_{SZ})} = \frac{\rho_p{}'}{\sin(\theta_{SE})} \tag{9}$$

ρ_p, θ_{SZ}, θ_{SE} are TOA planetary reflectance, local sun elevation angle and local solar zenith angle ($\theta_{SE} = 90^\circ$ - θ_{SE}) respectively.

NDVI [42] (Eq. (10)) and NDWI (Eq. (11)) [43] can be determined.

$$NDVI = \frac{NIR - VISR}{NIR + VISR} \tag{10}$$

$$NDWI = \frac{VISG - NIR}{VISG + NIR} \tag{11}$$

Where, *NIR*, *VISR* and *VISG* are planetary reflectance measurements taken from the near infrared, visible red and visible green regions respectively.

The NDVI is a useful indicator that connects vegetation and animal performance, according to recent ecological studies [44]. The balance between the energy that things on Earth receive and emit is measured by the NDVI. When applied to plant communities, this index assigns a value to the area's greenness, which is determined by the amount of vegetation present and its health or growth vigor [45]. The purpose of NDVI is to evaluate the amount of the plant matter, whereas NDWI is created to locate water bodies and saturated water [46]. The usual range for NDVI and NDWI is −1 (low) to 1 (high). Furthermore, high values of NDVI and NDWI represent a dense distribution or high occurrence of the vegetation and water body, respectively, whereas lower values indicate a low existence of the vegetation and water body [47]. Positive NDVI means green vegetation, whereas negative values mean water, clouds, snow, etc. Table 3 shows various NDVI values and how they indicate environmental parameters [48].

Table 4 shows various NDWI values and how they indicate environmental parameters [49].

According to this specification in this study, the total study area is divided into three NDVI classes, such as 0 to −1.0 to denote water bodies, 0 to 0.2 to denote rocks, bare soil, bushes, and pastures, and 0.2 to 1.0 to denote large vegetation. Like the previous index, the total study area is divided into three NDWI classes, such as 0 to −1.0 to denote drought, non-aqueous surfaces, water deficiency, non-aqueous surfaces, and vegetation; 0 to 0.2 to denote flooding; and 0.2 to 1.0 to denote water surface.

Table 3
NDVI values and indication.

NDVI value	Indication
near zero	Rocks and barren soil
0.1 or less	Empty regions of rock, sand or snow
0.2 to 0.3	Bush and meadow
0.6 to 0.8	Temperate and moist dense woodlands

Table 4
NDWI values and indication.

NDWI value	Indication
0.2 to 1.0	Water area
0.0 to 0.2	Flooding and humidity
−0.3 to 0.0	Moderate drought and non-aqueous soils
−0.3 to-1.0	Water scarcity, non-aqueous areas, or plant

2.3.4. *Land surface temperature (LST) analysis*

The proportion of vegetation is determined using the corresponding time period's estimated NDVI. The following equation (Eq. (12)) determines the proportion of vegetation (P_v) [50].

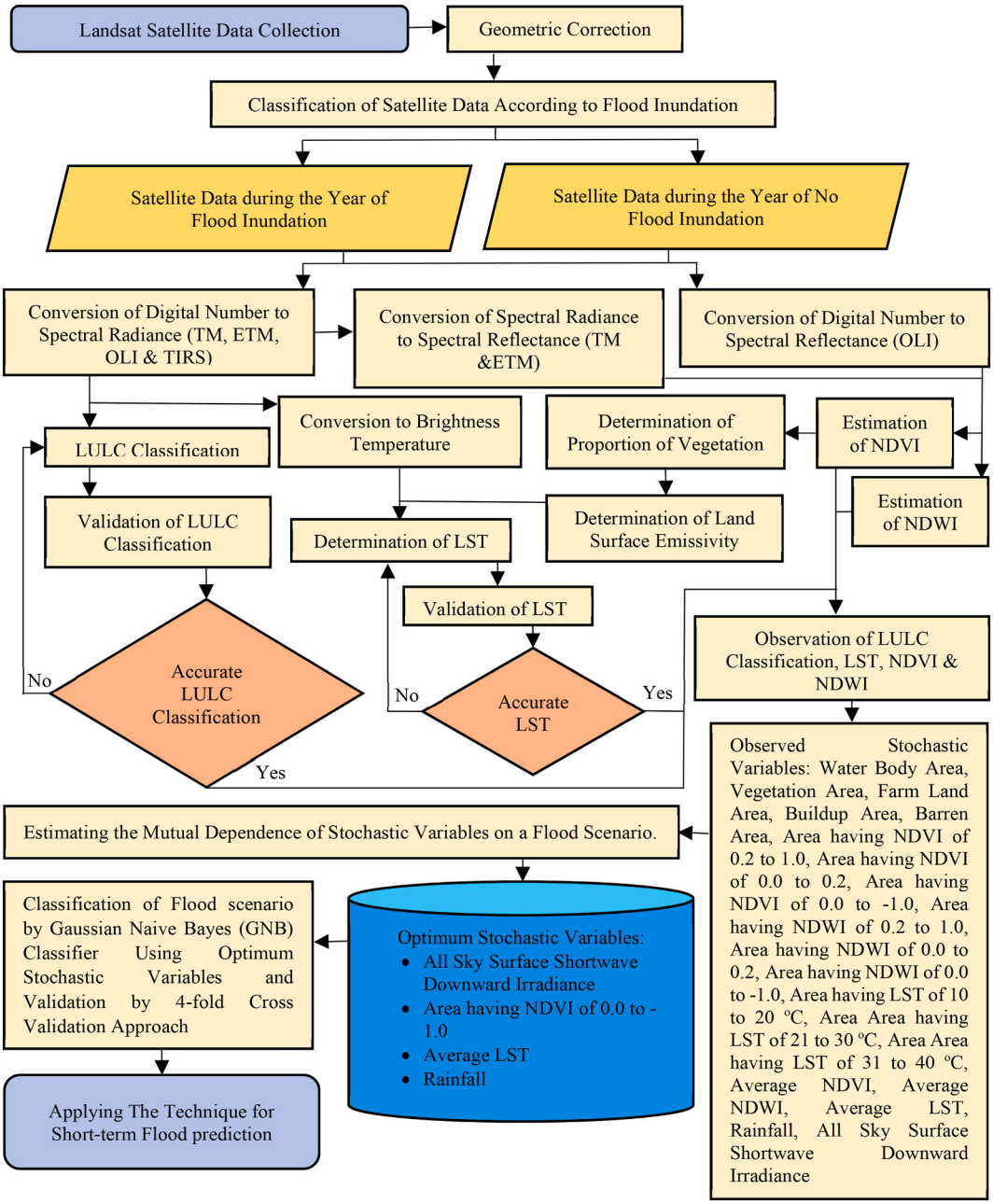


Fig. 2. Flow diagram of the methodology.

$$P_v = \left(\frac{NDVI - NDVI_{MIN}}{NDVI_{MAX} - NDVI_{MIN}}\right)^2 \tag{12}$$

Land surface emissivity e is determined by following formula (Eq. (13)) [50].

$$e = 0.004P_v + 0.986 \tag{13}$$

Further, brightness temperature is obtained from the conversion of spectral radiance obtained from TM, ETM, and TIRS images. The following equation (Eq. (14)) is used to transform the spectral radiance derived from TM, ETM, and TIRS images to brightness temperature [36,41].

$$T = \frac{K2}{\ln\left(\frac{K1}{L_\lambda} + 1\right)} \tag{14}$$

Where, T, $K2$ and $K1$ are brightness temperature (K), calibration constant 2 and calibration constant 1 respectively.

Lastly, an estimate of LST is obtained by integrating brightness temperature and land surface emissivity. LST (^{0}C) is determined by following equation (Eq. (15)) [51].

$$LST = \frac{T}{1 + \frac{wT}{p} \times \ln(e)} - 273.15 \tag{15}$$

Where, w stands for wavelength of emitted radiance (μm), $p = hc/s = 1.438 \times 10^{-2}$ mK, h, s , c stand for Panck's constant (6.626×10^{-34} Js), Boltzmann constant (1.38×10^{-23} J/K) and velocity of light (2.998×10^{-8} m/s).

One of the most important factors in the physics of land surface processes in all forms, from local to global, is LST. The significance of LST is becoming more widely acknowledged, and there is an intense need to create methods for measuring LST from space [52]. In this research work, the study area is divided into three LST classes, such as 10 to 20 °C, 21 to 30 °C, and 31 to 40 °C. To ensure the accuracy of the LST over various time periods in the study area, a validity analysis for estimated LST stability is performed. Some researchers [52–54] state a standard way to validate LST by comparing data on near-surface temperature obtained from specific ground stations. According to a researcher [53], comparing ground measurements (points) to area-averaged satellite data is useful when the test site's temperature and emissivity are uniform throughout all relevant geographical scales. According to a scientist [55], appropriate places for validation of LST include playas, dry lakes, dense vegetation, and areas with bare soil. LST validation is carried out in this study using the location of the Bangladesh Metrological Department's (BMD) first-class observatory in Tarash, Sirajganj.

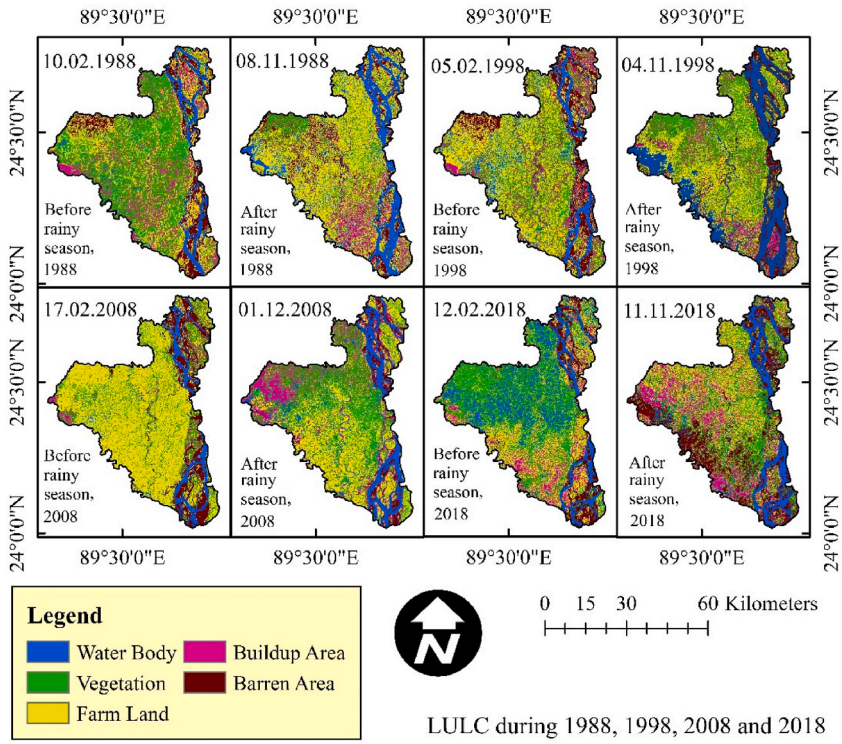


Fig. 3. Spatial distribution of different LULC components before and after rainy season during flooded years (1988, 1998, 2008 and 2018).

2.3.5. Stochastic variables selection

Based on the comparative flood scenario study, some environmental stochastic variables are specified, such as water body area, vegetation area, farmland area, buildup area, barren area, area having NDVI of 0.2–1.0, area having NDVI of 0.0–0.2, area having NDVI of 0.0 to −1.0, area having NDWI of 0.2–1.0, area having NDWI of 0.0–0.2, area having NDWI of 0.0 to −1.0, area having LST of 10–20 °C, area having LST of 21–30 °C, area having LST of 31–40 °C, average NDVI, average NDWI, average LST, rainfall, and all-sky surface shortwave downward irradiance. Here rainfall and all-sky surface shortwave downward irradiance are selected by observing the visual appearance of variation of vegetation, farmland, and barren area before the rainy season of both flooded and non-flooded years (Figs. 3 and 4). An important part of the surface radiation budget (SRB) is the earth's all-sky surface shortwave downward irradiance, which is commonly described as the total of the incoming solar energy over the earth's surface in the shortwave spectrum (0.3–3.0 μm). It is necessary since it is one of the prerequisites that many land models and applications rely on [56]. Climate studies, weather system modeling, watershed runoff analysis, building energy usage modeling, and other applications can benefit from it. The current study discovered that during flooded years, the presence of vegetation and farmland is higher before the rainy season, whereas in non-flooded years, the presence of buildup area is higher at the same time of year. The values of two variables are taken for thirty days, beginning with the date of acquiring a satellite picture prior to the rainy season of the year.

The process for computing the value of the two variables mentioned above is as follows:

- The location of data collection is five rain gauge stations.
- Each location's average variable value over thirty days is calculated.
- This five-location, thirty-day average value is averaged again to provide one value that represents the entire area on a single date.

This is simply a hypothesis for calculating a single value of two variables that represents the entire thirty days of five locations.

2.3.6. Predictor stochastic variables selection

All nineteen stochastic variables are reduced by measuring the amount of information acquired about the flood scenario using the MI approach. Histogram analysis converts the continuous values of stochastic variables to discrete values in order to determine joint and marginal probabilities for MI analysis. The MI values of selected stochastic variables are classified again using histogram analysis to identify predictor stochastic variables. The number of bins can be calculated by the square-root formula (Eq. (16)) as follows [57].

$$k = \lceil \sqrt{n} \rceil \tag{16}$$

And bin width can be calculated by following formula (Eq. (17)) [58].

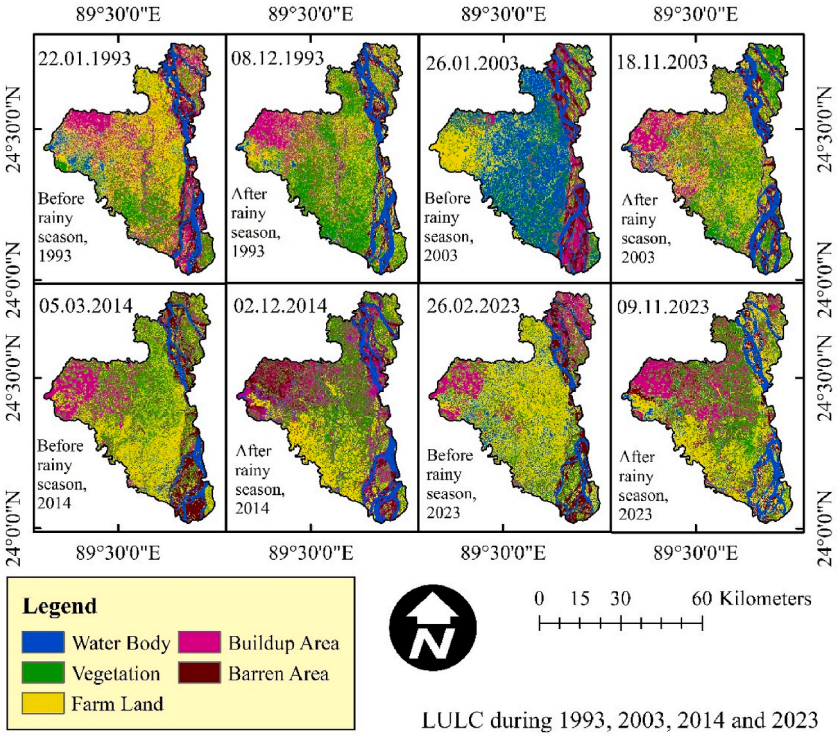


Fig. 4. Spatial distribution of different LULC components before and after rainy season during non-flooded years (1993, 2003, 2014 and 2023).

$$h = \left\lceil \frac{\max x - \min x}{k} \right\rceil \tag{17}$$

Here, n, h, $\max x$ and $\min x$ denote bin number, data point number, bin width and maximum and minimum values of a set of data respectively.

If (X, Y) is a set of discrete random variables having values in the space $\chi \times \gamma$ then the MI can be determined by the following equation (Eq. (18)) [59].

$$I(X;Y) = \sum_{y\in\gamma}\sum_{x\in\chi} P_{(X,Y)}(x,y)\log\left(\frac{P_{(X,Y)}(x,y)}{P_{(X)}(x)P_{(Y)}(y)}\right) \tag{18}$$

Here $P_{(X,Y)}$ denote the joint probability mass function of X and Y. P_X and P_Y are the marginal of probability mass function of X and Y respectively.

MI stands out from other measures of independence between random variables due to its information-theoretic foundation [59]. Unlike the linear correlation coefficient, it is also susceptible to dependences that are not shown by the covariance. Independent component analysis (ICA) is one of the primary sectors in which MI is significant, at least conceptually [60]. Because they are simple to use, crude approximations to MI based on cumulant expansions are common in the ICA literature. However, they only work with distributions that are nearly Gaussian and are only useful for classifying distributions according to their interdependencies; they are not very useful for estimating the true dependencies [61]. MI has the advantage of being able to identify any type of connection [62], whereas correlation is limited to identifying linear associations [63]. It is a concept from information theory that has been utilized a lot recently as a standard for feature selection techniques. Its capacity to represent both linear and non-linear dependency relationships between two variables is the reason behind this [64].

2.3.7. Flood scenario classification and model validation

The GNB classifier is then used to categorize all flood scenarios by calculating the prior probability of each flood scenario class and the likelihood of the predictor stochastic variables. It is a probabilistic classifier that applies the Bayes theorem under the assumption of absolute independence. Naive Bayes classifiers are effective and easy to understand when dealing with multiclass classification. The naive Bayes algorithm makes predictions based on the class and the Bayes theorem, assuming that predictors are conditionally independent [65]. GNB evidence is ignored in this instance because it is common in all of the cases. By Bayes' theorem [66], the conditional probability is decomposed as (Eq. (19)).

$$p(C_k|\mathrm{X}) = \frac{p(C_k)p(\mathrm{X}|C_k)}{p(\mathrm{X})} \tag{19}$$

Here, K is possible outcomes, C_k is a problem instance to be classified, $X = (x_1, \ldots\ldots, x_n)$ encoding some n features.

In general using Bayesian probability terminology, the above equation can be written as (Eq. (20)).

$$\textit{Posterior Probability} = \frac{\textit{Prior Probability} \times \textit{Likelihood}}{\textit{Evidence}} \tag{20}$$

Prior Probability is determined by calculating the probability of an incident happening [67]. And *Likelihood* is calculated by the following equation (Eq. (21)) [68] of normal distribution considering the training data contains a continuous attribute.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{21}$$

Here, x, μ and σ are test sample, mean and is variance.

To create four-fold models, the dataset is divided into four folds. Three of the folds are used as the training set while the remaining fold is utilized as the test set for each iteration by the GNB classifier. A pair of flooded and non-flooded years are included in each fold. Therefore, each year is utilized as training and testing data in four iterations, and the errors of GNB classification are calculated using mean absolute percentage error (MAPE).

The most popular goodness of fit metric is likely the MAPE, which can be calculated as follows (Eq. (22)) [69].

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|\widehat{y}_t - y_t|}{y_t} \times 100 \tag{22}$$

Where, n is sample size, $\widehat{y}_t$ is predicted value by the model for time point t and y_t is observed value at time point t

The model is then validated using mean error values derived from the MAPE of the GNB classification method from four iterations for 4-fold cross validation.

Methodologies adopted in this study are shown by the flow diagram in Fig. 2.

3. Result

3.1. Comparative flood scenario study

In this research work, five LULC classes are assigned: vegetation, farmland, water body, built-up area, and bare soil (Figs. 3–8). The areal distribution of LULC classes is shown in Figs. 3 and 4 for flood-inundated years and non-flood-inundated years, respectively. A visible increase in the distribution of vegetation and farmland is found in the middle of Sirajganj during November and December of the flood inundated years 1988, 1988, and 2008 (Fig. 3). In November and December of those years, there were water bodies in the study area's western corner, particularly in 1998. A significant buildup area is found in November 2018, along with farms and vegetation. The study area's western corner, where a large number of water bodies are last observed in November 1998, is significantly deserted in November 2018. A study by scholar [70] stated that devastating flooding occurred in 1988 and 1998, when seventy percent of Bangladesh was inundated, as visualized in Fig. 3. In December 2008, this topographic feature is also found in a very small and dispersed area. Possibly as a result of flooding, there are significantly barren regions on the west side of the study area in November 2018. The presence of farms and vegetation in February, before the rainy season of flood-affected years, is significant, similar to November and December of those years, even when these months fell within the dry season. But the presence of build-up areas is quite low compared to farmland and vegetation. The number of water bodies in the middle of the area is comparatively higher in February 2018 than it has been in previous years.

Fig. 4 shows that the amount of buildup area is comparatively higher in the four non-flood-inundated years during November and December than in the flood-inundated years. Similar to years when there has been flooding, significant amounts of farmland and vegetation are also discovered. With the exception of January 2003, there is significantly more build-up area in the upper western corner of the study area in the months of January, February, and March, before the rainy season in those years. Figs. 5 and 6 show, respectively, the graphical representations of the area distribution of the LULC classes of years that are flooded and years that aren't. Fig. 5 shows how the area covered by farms and vegetation varies year-round, both before and after the rainy season. This district has no forest or dense vegetation. Part of the vegetation area classified in this study is actually full-grown crops or grassland cultivated for livestock. A characteristic that is common in the years 1988, 1998, and 2008 is an increase in the water body's area after the rainy season. In 2018, the area of the water body reduces, but the amount of barren land increases significantly, which may be a result of flooding. In Fig. 6, LULC classes do not show any specific trend in the years of no flooding. As shown in Fig. 6, areas of farmland and vegetation during the years without flooding are similar to those during the years with flooding. After the rainy season, the area of the water body does not change significantly in the following years, except in 2003, when it actually decreases. Build-up areas are expanded after the rainy season every year, with the exception of 1993.

The topographic configuration of the elevation (Fig. 1) and the monthly rainfall pattern of January, February, and March (Table 5) of those years in the study area can be used to explain the distribution of LULC classes before and after the rainy season throughout the flooded and non-flooded years. Fig. 1 shows that the western portion of the study area is lower than the middle, except for the Jamuna River in the east. The western portion is the floodplain of the Baral River, which flows slightly outside the study area's administrative borders. Except for its susceptibility to flooding, this floodplain of lowest elevation is always covered with vegetation and farmland, although over the years. This lowest-elevation floodplain is always covered with farmland and vegetation, in spite of its susceptibility to flooding over time. This is only possible when flooding generated alluvial soil that makes the land fertile and saturated.

Table 5 provides a monthly rainfall data set of January, February, and March from five rain gauge stations provided by BWDB that explains both the common and distinctive characteristics of the LULC classes of flood-inundated and non-inundated years. There is a significant amount of water body area before the rainy season in both January of the non-flooded 2003 and February of the flooded 2018. Although it rains 5.3 mm, 4.60 mm, 8.30 mm, 7.10 mm, and 1.40 mm in CL29, CL34, CL35, CL39, and CL40 in January 2003, respectively, the monthly rainfall for February 2018 at CL34 and CL40 stations is 11.00 and 123.00 mm, respectively. This is a relatively small amount of rainfall for a month, but if it fell in a single day, that is cause for concern. Since these two months fell within the dry season, it appears possible that a similar occurrence might occur in a location outside of the study area. Farmers also begin irrigating their land during this time, which, combined with the phenomenon of rainfall, causes the land to appear as a water body. The

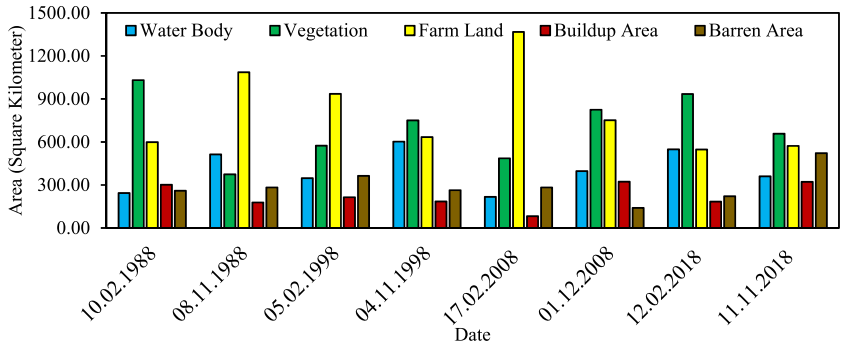


Fig. 5. Area distribution of different LULC components before and after rainy season during flood inundated years.

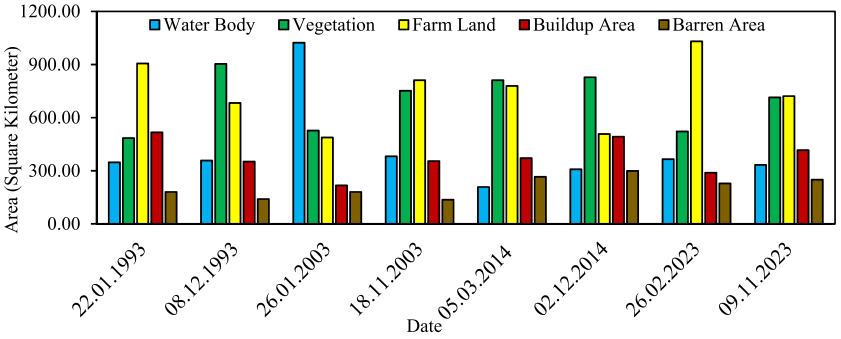


Fig. 6. Area distribution of different LULC components before and after rainy season during non-flood inundated years.

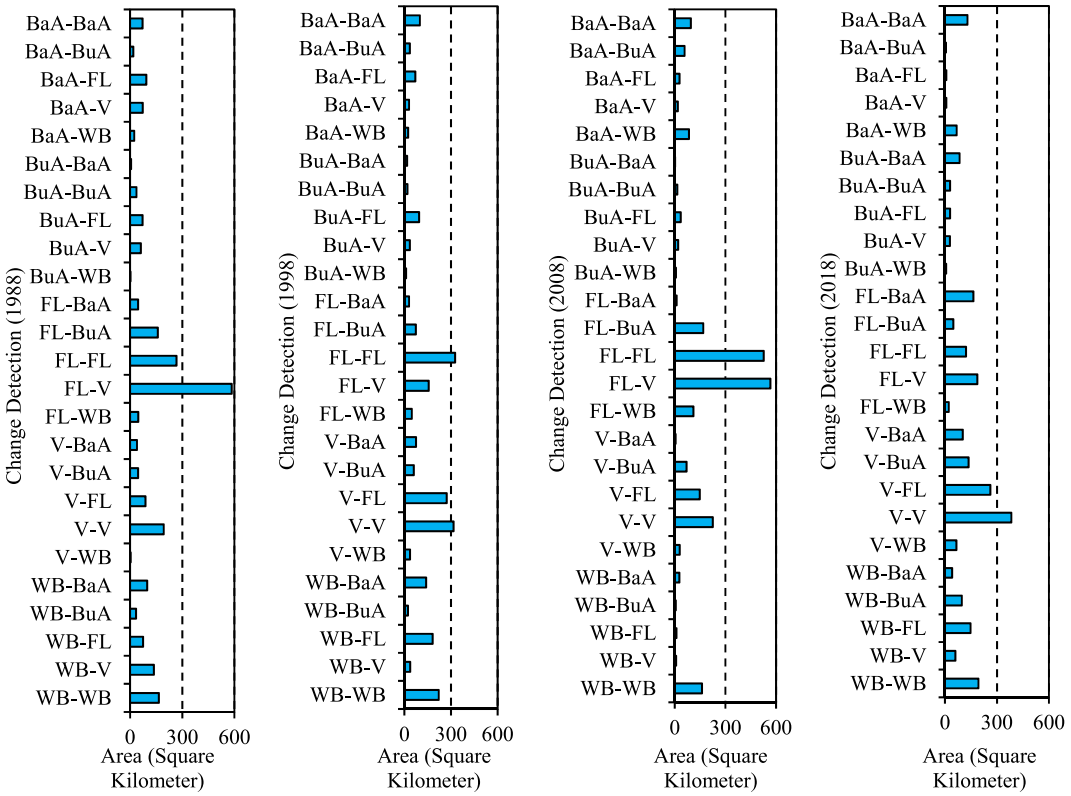


Fig. 7. Detection of LULC component's area interchange before and after rainy season during flood inundated years (WB, V, FL, BuA, BaA stand for water body, vegetation, farm land, buildup area and barren area respectively).

Sirajganj district is primarily a rural, agricultural area without dense vegetation and large buildup areas. Fig. 4 during non-flood-inundated years clearly shows that the majority of people have farmland surrounding their habitat, where vegetation, farmland, and buildup areas were situated together. This build-up area is significantly absent in the flood-inundated years in Fig. 3. Nearly every station shows evidence of rainfall in February of 1988, 1998, 2008, and 2018, with 1988 having the highest amounts of 37.80, 57.70, 69.80, 121.80, and 85.10 mm in CL29, CL34, CL35, CL39, and CL40, respectively. This phenomenon causes a significant portion of the rural buildup area to often become covered in vegetation.

In years when there are floods, the LULC classes exchange mostly agricultural and vegetative areas before and after the rainy season (Fig. 7). Since agriculture dominated the study area, this interchange may be explained by soil properties like saturation and fertility brought about by alluvial soil from flooding. Fig. 8, which shows the interchange of LULC classes before and after the rainy season in years without flooding, highlights a nearly identical pattern, with the exception of 2014. Significant vegetation replacement takes place in the buildup area in 2014. This may have occurred as a result of the unusual irrigation and agricultural practices chosen in the upper western corner of the study area (Fig. 4).

The average, minimum, and maximum NDVI, NDWI, and LST of flood-inundated and non-flood-inundated years are listed in

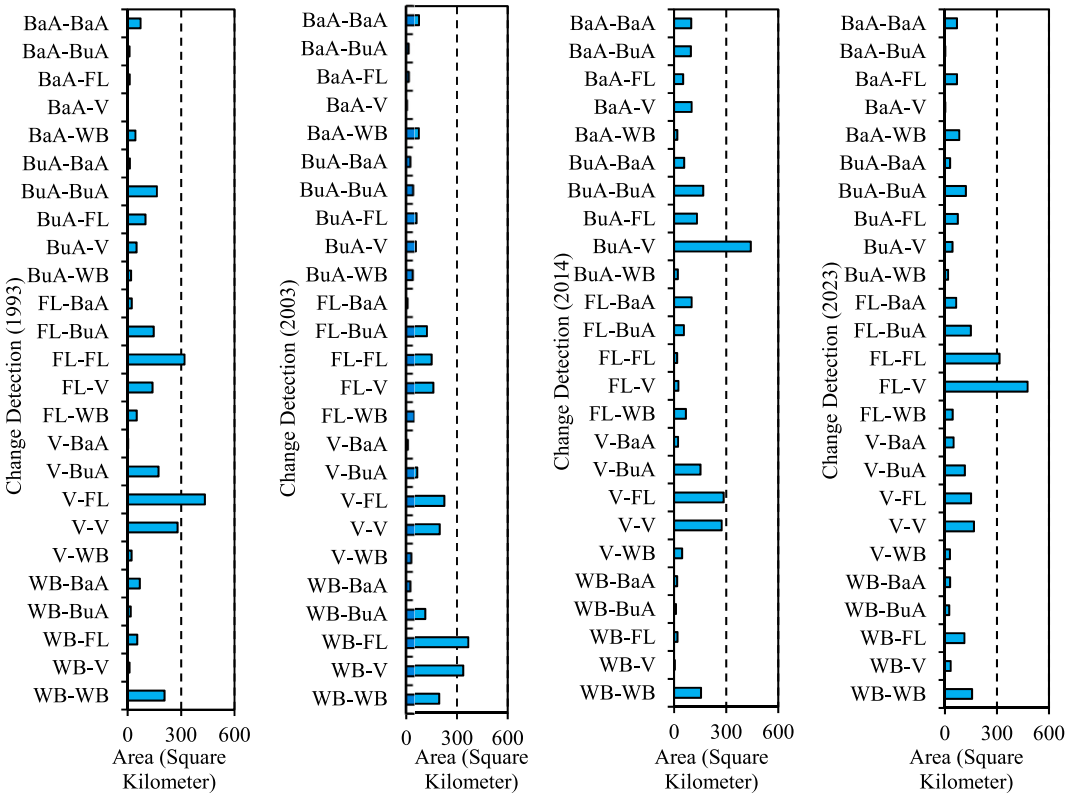


Fig. 8. Detection of LULC component's area interchange before and after rainy season during non-flood inundated years (WB, V, FL, BuA, BaA stand for water body, vegetation, farm land, buildup area and barren area respectively).

Table 6. It is found in flood-inundated years that, except 1998, every year after the rainy season, the average NDVI increases more than before the rainy season. The minimum and maximum NDVI of the study area decrease and increase, respectively, after the rainy season in every flood-inundated year. Also, except in 1998, every year after the rainy season, the average NDWI is lower than before the rainy season. In the case of minimum NDWI, except in 1988, every year it decreases, and maximum NDWI increases every year after the rainy season. But average, minimum, and maximum LST increase after the rainy season in every flood-inundated year. It is very tough to find out any pattern or trend in average, minimum, and maximum NDVI, NDWI, and LST after a rainy season in non-flood inundated years. Average NDVI increases after the rainy season in 1993, 2003, and 2023, but minimum and maximum NDVI do not match any pattern. Every year, average NDWI decreases after the rainy season, but like the NDVI, minimum and maximum NDWI do not match any pattern. In cases of average and maximum LST, except 1988, every year it decreases, which is opposite in nature as compared with flood-inundated years. The minimum LST does not show any pattern at all in this case. Hence, non-flood-inundated years do not show any satisfactory topographic property.

Area coverage at different ranges of NDVI is shown in Fig. 9, where it is very difficult to interpret the nature of the vegetation as this LULC class is easily affected by anthropogenic and natural phenomena. Areas having NDVI of 1.0 to 0.0 in both flooded and non-flooded years increase following the monsoon season. However, this increase is remarkable in 1988 and 1998, more than doubling when compared to the same NDVI range in areas prior to the rainy season. Following the rainy season, the area of the NDVI ranged from −1.0 to 0.0, indicating water bodies and, in fact, evidence of catastrophic flooding that occurred in 1988 and 1998. In Fig. 9, the area having NDVI of 0.2–1.0 and 0.0 to 0.2, which do not interpret any particular pattern in the study area. There are no dense forests or rocky areas in this district, so these two types of NDVI ranges indicate small vegetation, grassland, farmland, riverside barren areas, etc. In every year that is flooded or not, areas with NDVI values between 0.0 and 0.2 decrease somewhat following the monsoon season, with the exception of 2003. The majority of this district is two-crop land, and the different stages of the crop cycle are indicated by the area covered by this range of NDVI. Every year but 2003 experiences an increase in the area with an NDVI in the 0.2 to 1.0 range after the rainy season. This is a cornfield and Napier grassland, part of the crop cycle used as animal feed, because there is no dense vegetation. It may be that this unique NDVI feature prior to the 2003 rainy season is a long-term effect of the 2002 flooding. NDVI analysis (Fig. 9) eliminates doubt regarding the probability that a major portion of the study area would turn into a water body in 2003 before the rainy season (Fig. 6). Fig. 9 analysis reveals that the majority of the area is covered by various forms of vegetation during this period, with the majority falling within the NDVI ranges of 0.0–0.2 (833.43 square kilometers) and 0.2 to 1.0 (1444.65 square kilometers).

According to Fig. 10, the majority of the study area before and after the rainy season every year is in the NDWI of 0.0 to −1.0,

Table 5
Monthly rainfall pattern of January, February and March.

Year	January					February					March				
	CL29	CL34	CL35	CL39	CL40	CL29	CL34	CL35	CL39	CL40	CL29	CL34	CL35	CL39	CL40
1988	0.00	0.00	0.00	0.00	0.00	37.80	57.70	69.80	121.80	85.10	78.40	94.60	53.50	13.70	54.40
1998	24.10	12.40	30.00	1.40	10.90	5.00	7.40	10.00	0.00	5.30	67.80	60.90	78.00	30.50	146.70
2008	0.00	39.00	0.00	0.30	29.80	0.50	3.50	0.00	0.00	17.20	0.00	27.50	21.00	0.00	31.00
2018	12.90	28.70	8.00	5.20	22.00	0.00	11.00	0.00	0.00	123.00	9.70	28.20	0.00	2.10	58.60
1993	0.00	13.10	0.00	0.90	5.00	11.20	2.10	10.00	0.00	9.00	72.10	111.90	171.00	10.00	40.10
2003	5.30	4.60	8.30	7.10	1.40	22.84	38.30	5.61	10.99	12.51	53.30	61.00	15.00	0.00	77.40
2014	0.00	1.30	0.00	0.00	0.00	17.07	21.70	18.06	0.00	22.00	4.32	0.00	8.93	0.00	13.20
2023	0.00	0.00	0.00	0.00	5.70	0.00	0.00	0.00	0.00	289.30	11.50	0.00	45.00	1.15	322.60

Table 6
Average, minimum and maximum values of NDVI, NDWI and LST of different years.

Year	Date	NDVI			NDWI			LST		
		Avg	Min	Max	Avg	Min	Max	Avg	Min	Max
1988	10.02.1988	0.1789	−0.2384	0.7588	−0.1331	−0.7288	0.3649	23.7646	19.3337	31.4226
	08.11.1988	0.1845	−0.5655	0.8231	−0.1340	−0.7171	0.6681	26.6057	24.5815	35.0964
1998	05.02.1998	0.2160	−0.4991	0.7657	−0.1662	−0.6666	0.6031	22.3213	18.4135	29.7332
	04.11.1998	0.1973	−0.6158	0.8156	−0.1261	−0.7151	0.6669	26.5724	23.2944	33.4453
2008	17.02.2008	0.1022	−0.2093	0.4458	−0.0380	−0.3510	0.3095	20.6031	16.5721	28.0746
	01.12.2008	0.2857	−0.4567	0.7896	−0.2178	−0.6868	0.5571	23.3828	19.3308	31.0027
2018	12.02.2018	0.2296	−0.4148	0.7511	−0.1621	−0.6379	0.5306	23.1800	18.4040	33.1938
	11.11.2018	0.3606	−0.4674	0.7849	−0.2845	−0.6718	0.5544	25.8710	22.4802	34.6510
1993	22.01.1993	0.1695	−0.2629	0.6282	−0.1210	−0.5424	0.3525	16.8352	13.7270	22.8845
	08.12.1993	0.2323	−0.4925	0.7874	−0.1836	−0.7004	0.5931	22.4870	18.8691	28.4876
2003	26.01.2003	0.2194	−0.3186	0.7147	−0.1504	−0.5936	0.4360	26.1325	20.9374	36.2623
	18.11.2003	0.2800	−0.4470	0.7853	−0.2174	−0.6577	0.5363	24.5868	14.8929	32.5858
2014	05.03.2014	0.3411	−0.3948	0.7751	−0.2497	−0.6699	0.4844	24.9457	20.4591	34.9033
	02.12.2014	0.3225	−0.3581	0.7452	−0.2556	−0.6314	0.4642	23.5485	20.7664	30.8707
2023	26.02.2023	0.2748	−0.2941	0.7243	−0.2031	−0.6133	0.4192	26.9117	22.5045	39.0125
	09.11.2023	0.2767	−0.3222	0.6376	−0.2117	−0.5365	0.3783	26.1261	23.3457	34.0002

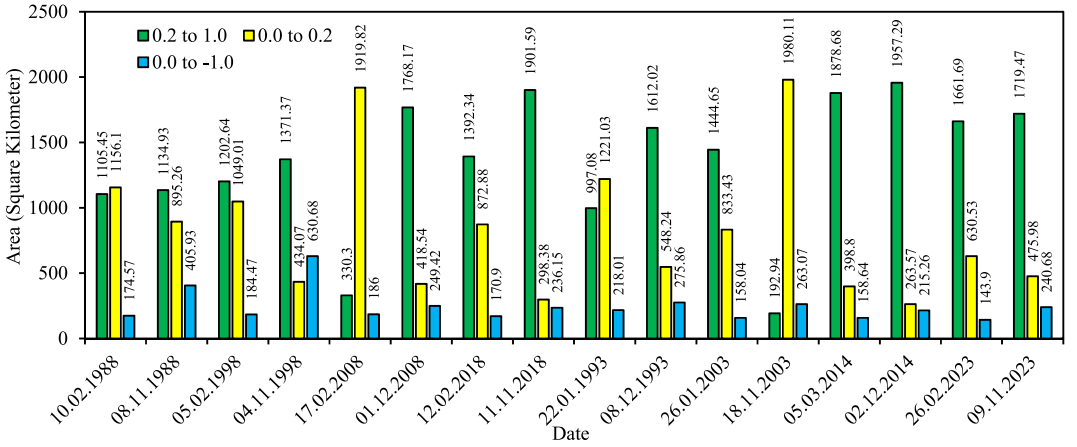


Fig. 9. Area coverage at different ranges of NDVI.

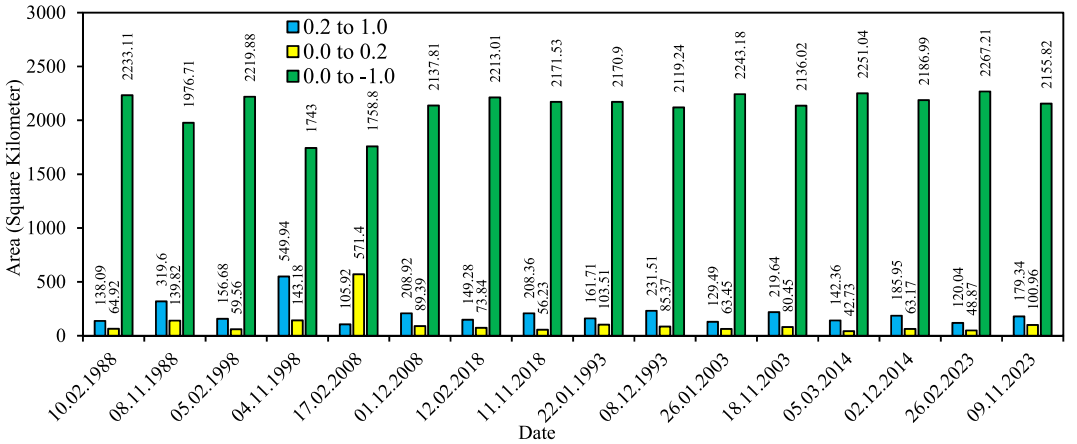


Fig. 10. Area coverage at different range of NDWI.

denoting the non-aqueous surface or vegetation. During the year of flooding, the study area in this NDWI range does not exhibit any particular pattern before or after the rainy season. The amount of non-aquatic surface and vegetation, however, appears to decrease after the rainy season, according to the area with this range of NDWI in 1988 and 1998. This is obviously the effect of massive flooding in 1988 and 1998. The amount of non-aquatic surface and vegetation decreases slightly after the rainy season, according to years without flooding. It is also discovered that there has been significant flooding according to the area having NDWI of 0.2 and 1.0. Every year following the rainy season, the area within this NDWI range increases slightly, with the exception of 1988 and 1998, when the increment is significant. Again, the area having an NDWI of 0.0–0.2 increases above twice after the rainy seasons in 1988 and 1998; however, 2008 and 2018 do not exhibit the same characteristics. During years without floods, the area of this NDWI range does not increase significantly after the rainy season. These analyses indicate that the study area's water surface area increases even through the dry season as a result of significant flooding.

Fig. 11 shows that the majority of the area has an LST of 21–30 °C almost every year, both before and after the rainy season. 2008 is an exception to the convention before the rainy season, during the years of flooding, with a notable portion of the land having LSTs of 10–20 °C. 1993, before the rainy season, is an exception in years without flooding, with the majority of the region having LST between 10 and 20 °C. A portion of the surface in the study area has an LST of 31–40 °C before the rainy season in 2003, 2014, and 2023. In general, the LST analysis of flood-prone and non-flood-prone years does not identify any unique characteristics before or after the rainy season.

3.2. Mutual dependence study

From the preceding comparative flood scenario analysis, certain stochastic variables are chosen that are supposed to be mutually dependent on the flood scenario. These selected variables are water body area, vegetation area, farmland area, buildup area, barren area, area having NDVI of 0.2–1.0, area having NDVI of 0.0–0.2, area having NDVI of 0.0 to −1.0, area having NDWI of 0.2–1.0, area having NDWI of 0.0–0.2, area having NDWI of 0.0 to −1.0, area having LST of 10–20 °C, area having LST of 21–30 °C, area having LST of 31–40 °C, average NDVI, average NDWI, average LST, rainfall, and all-sky surface shortwave downward irradiance. Only the mutual dependency of stochastic variables and the flood scenario prior to the relevant year's rainy season is determined. Maximum and minimum NDVI, NDWI, and LST are not considered because they do not represent the entire study region. Similarly, the LULC component's area interchange for both flooded and non-flooded years is ignored because it expresses very little specific information. Meanwhile, all sky surface shortwave downward irradiance and rainfall are chosen as stochastic variables based on a visual evaluation of Figs. 3 and 4. In Fig. 3, the presence of vegetation and farmland prior to the rainy season in flooded years indicates the potential of rainfall. Whereas in Fig. 4, the presence of a buildup area and a barren area throughout the same period of non-flooded years indicates a lack of rainfall.

Table 7 shows the MI values for the selected stochastic variables that describe the flood scenario. The area having NDVI of 0.0 to −1.0 exhibits the highest (0.69) MI values, while all sky surface shortwave downward irradiance has the second highest (0.52) MI values. Average LST and rainfall are the third highest (0.41), and the area having LST of 31–40 °C has the fourth highest (0.38) MI values, while the area having LST of 21–30 °C has the lowest (0.01) MI values (Table 7). According to MI values, areas having NDVI of 0.0 to −1.0 share the most information to indicate a flood scenario, while areas having LST of 21–30 °C share the least. Fig. 12 shows the classification of nineteen selected stochastic variables into five histogram bins (B1 to B5). Here B5 has the highest MI-valued stochastic variables, B4 contains the second highest MI-valued stochastic variables, and B3 contains three stochastic variables with the third and fourth highest MI values (Table 7). Meanwhile, B1 and B2 consist of six and eight stochastic variables, respectively, with MI values ranging from 0.01 to 0.27. As it is noted, fourteen out of nineteen stochastic variables fall into B1 and B2, with values ranging from 0.01 to 0.15 and 0.15 to 0.29, respectively. This means that the majority of stochastic variables provide relatively little information about the flood scenario.

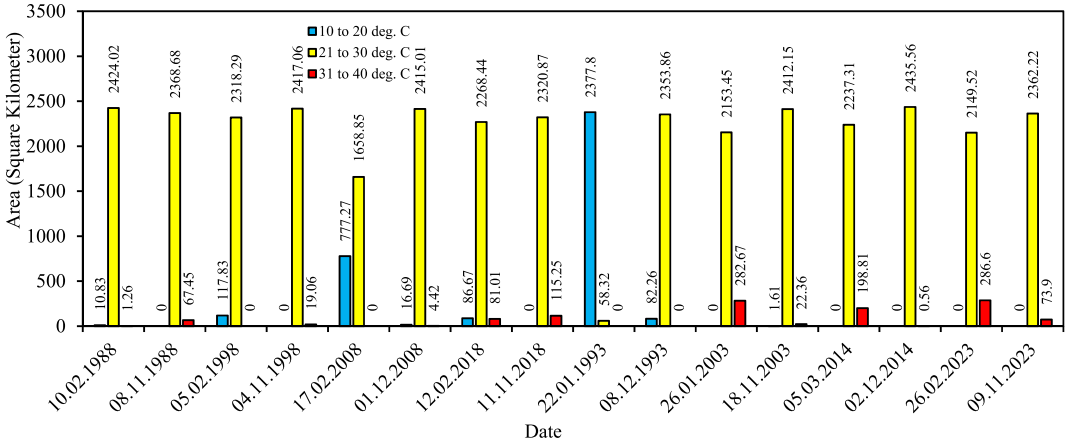


Fig. 11. Area coverage at different range of LST.

Table 7
Mutual Information (MI) value of stochastic variables.

Stochastic Variables	Mutual Information with Flood Scenario	Bin Range	Bin
Area having NDVI of 0.0 to −1.0	0.69	0.57 to 0.71	B5
All sky surface shortwave downward irradiance	0.52	0.43 to 0.57	B4
Average LST	0.41	0.29 to 0.43	B3
Rainfall	0.41	0.29 to 0.43	B3
Area having LST of 31–40 °C	0.38	0.29 to 0.43	B3
Average NDWI	0.27	0.15 to 0.29	B2
Vegetation area	0.27	0.15 to 0.29	B2
Average NDVI	0.22	0.15 to 0.29	B2
Water body area	0.17	0.15 to 0.29	B2
Buildup Area	0.17	0.15 to 0.29	B2
Barren Area	0.17	0.15 to 0.29	B2
Area having NDVI of 0.2–1.0	0.17	0.15 to 0.29	B2
Area having NDVI of 0.0–0.2	0.17	0.15 to 0.29	B2
Farm land area	0.11	0.01 to 0.15	B1
Area having NDWI of 0.0–0.2	0.10	0.01 to 0.15	B1
Area having NDWI of 0.0 to −1.0	0.10	0.01 to 0.15	B1
Area having LST of 10–20 °C	0.10	0.01 to 0.15	B1
Area having NDWI of 0.2–1.0	0.04	0.01 to 0.15	B1
Area having LST of 21–30 °C	0.01	0.01 to 0.15	B1

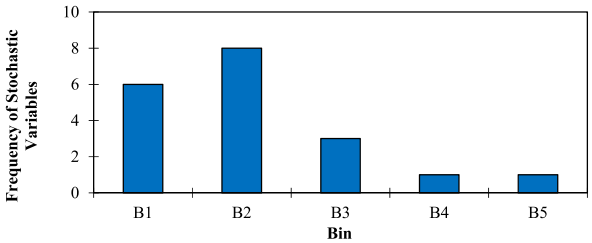


Fig. 12. Histogram of frequency distribution shape of stochastic variables.

For further GNB analysis, five stochastic variables from B3, B4, and B5 are used as predictors. Here MI is a computationally efficient method for global sensitivity analysis (GSA), which evaluates how model outputs are influenced by inputs [71]. Here five stochastic variables from B3, B4, and B5 are more significant for model accuracy (Table 7).

Table 8
Prior probability of each flood scenario class and likelihood of each parameter.

Year	Flood Scenario	Prior Probabilities	Likely hood of Parameters				
			Area (Square Kilometer) having NDVI of 0.0 to −1.0 (P1)	All sky surface shortwave downward irradiance (P2)	Average LST (P3)	Rainfall (P4)	Area (Square Kilometer) having LST of 31–40 °C (P5)
1988	Flood	0.5	0.00575	3.8287E-11	0.14014	0.0618	1.8236E-04
	No Flood	0.5	0.00548	6.256E-124	0.02929	0.0114	1.6136E-04
1998	Flood	0.5	0.00639	2.85884132	0.14065	0.0484	1.8519E-04
	No Flood	0.5	0.00026	1.8897E-28	0.01399	0.0098	1.8502E-05
2008	Flood	0.5	0.00795	2.7418E-49	1.0185E-05	0.0659	1.8519E-04
	No Flood	0.5	0.00026	1.0883E-38	0.01266	1.6251E-07	1.4770E-05
2018	Flood	0.5	0.00996	1.33693473	0.14824	0.0477	0.0000
	No Flood	0.5	0.00034	1.4885E-59	0.01559	0.0091	1.8928E-05
1993	Flood	0.5	0.00498	1.202E-117	0.00250	0.0002	1.8236E-04
	No Flood	0.5	0.00373	4.903E-122	2.5726E-20	0.0118	1.6135E-04
2003	Flood	0.5	0.00613	1.7877E-09	0.06229	0.0172	1.8391E-04
	No Flood	0.5	0.00026	1.11404652	0.01391	0.0100	1.8502E-05
2014	Flood	0.5	0.00757	9.1573E-69	0.00140	0.0935	1.8462E-04
	No Flood	0.5	0.00026	0.75942178	0.01269	3.0240E-10	1.4770E-05
2023	Flood	0.5	0.00640	4.1497E-11	0.02768	0.0228	0.0000
	No Flood	0.5	0.00034	1.3089E-09	0.01538	0.0093	1.8928E-05

3.3. Flood scenario classification by GNB

The GNB classifier posits that each parameter (in this case, predictor stochastic variables) has an independent capability for predicting flood scenarios based on its ability to forecast the likelihood of a circumstance being classed as a perfect flood scenario. Table 8 shows the prior probability of each flood scenario class as well as the likelihood of each parameter given whether it is flooded or not. The values of prior probability are always kept constant in each case by providing the same number of training data in each class of flood scenario, in order to comply with the responsibility of likelihood of each parameter to classify the flood scenario perfectly. Table 9 depicts the various combinations of five parameters (B3, B4, and B5) likelihood with previous probability and their ability to categorize flood scenarios correctly. P2 alone has the best ability to accurately classify flood scenarios (six out of eight), while having the second highest MI value (52). P1 has the greatest MI value (69) but the lowest capacity to correctly categorize flood scenarios; nonetheless, it can be observed that it correctly classifies only actual flood events. P3 and P4 accurately classify five of the eight flood scenarios, whereas P5 correctly classifies four of them. P1 and P5 have the highest and fourth-highest MI values, respectively; however, they both have the same capacity to correctly classify. The main distinction is that P5 correctly classifies one no-flood (2023) incident while P1 exclusively classifies actual flood events. The combination of P1 and P2 has the same categorization capabilities as using simply P1. The seventh, eighth, and ninth parameter combinations (Table 9) demonstrate the same number of valid classification capabilities (five out of eight) but different scenarios in some cases. P2 has a lower capacity for shearing information on floods than P1, but it classifies flood scenarios the best. P1 can only classify true flood scenarios, whereas P2 can classify both types, with the exception of 1993 and 2008. In this study, P2 is the optimum stochastic variable for classifying flood scenarios.

The GNB Classifier uses Bayes' theorem to forecast membership probabilities for each class, such as the likelihood that a given record or data point belongs to a specific category. The class with the biggest potential is deemed extremely likely. The maximum likely class with the highest probability is determined by multiplying the prior probability of every flood scenario type by the likelihood of each parameter. This classifier assumes that all parameters are independent of one another. One parameter's existence or absence has no impact on another's existence or absence. This study uses comparative flood analysis to choose several environmental variables. The MI analysis is then used to establish the superiority of the variables for gathering information about actual flood events before the rainy season begins. In this study, the possibility of a flood occurring each year is characterized by estimating the likelihood of the optimum stochastic variable given whether it is a flood or not. Short-term flood predictions for the current year can be made almost accurately by using previous years' weather and geography data prior to the rainy season as training data and the same sort of data at that time of the current year as testing.

4. Validity and accuracy analysis

4.1. Accuracy assessment of LULC classification

The main goal of the accuracy assessment is to define how closely related the actual value is to the qualitative information from satellite data. Error matrix analysis is a good method for the accuracy calculation of supervised image classification. An error matrix of image classification should be used to evaluate the accuracy of users and producers [38]. In case of accuracy assessment, the Kappa statistic is commonly used to assess interpreter dependability. The significance of rater dependability signifies the degree of data collection in different studies that are exact signs of the variables measured [72]. It is usually regarded as a more relevant measurement than traditional percentage match computations since it accounts for the possibility of matches occurring by chance. Some scholars have suggested that it is theoretically simpler to analyze disagreement between items [73]. Its values range from 0 to 1, with 0 denoting no agreement and 1 denoting perfect agreement [47]. The value of Kappa and their indications are given in Table 10 [72].

From Tables 11 and 12, it can be seen from the Kappa Coefficient values that the measured LULC matches the true data in a moderate to strong manner. Tables 11 and 12, respectively, list the user accuracy, producer accuracy, overall accuracy, and kappa coefficient of various flooded and non-flooded years.

4.2. Validity analysis of LST

The estimated LST of the selected location situated in the study area is compared with the earth skin temperature of the same

Table 9
Likelihood combination with prior probability.

Serial No.	Five parameter's likelihood combination with prior probability	Correct flood scenario classification (Year)
01	P1	1988, 1998, 2008, 2018
02	P2	1988, 1998, 2018, 2003, 2014, 2023
03	P3	1988, 1998, 2018, 2014, 2023
04	P4	1988, 1998, 2008, 2018, 1993
05	P5	1988, 1998, 2008, 2023
06	P1, P2	1988, 1998, 2018, 2003, 2014, 2023
07	P1,P2,P3	1988, 1998, 2018, 2003, 2014
08	P1,P2,P3,P4	1988, 1998, 2018, 2003, 2014
09	P1,P2,P3,P4, P5	1988, 1998, 2003, 2014, 2023

Table 10
Different values of Kappa and their indication.

Kappa value	Indication
0 to 0.20	Degree of acceptance is none
0.21 to 0.39	Degree of acceptance is minimal
0.40 to 0.59	Degree of acceptance is weak
0.60 to 0.79	Degree of acceptance is moderate
0.80 to 0.90	Degree of acceptance is strong
Above 0.90	Degree of acceptance almost perfect

Table 11
User Accuracy, Producer Accuracy, Overall Accuracy and Kappa Coefficient of different LULC components at flooded years.

Year	Date	User Accuracy (%)		Producer Accuracy (%)		Overall Accuracy (%)	Kappa Coefficient (%)
1988	10.02.1988	Water Body	100.00	Water Body	100.00	85	80.57
		Farm Land	100.00	Farm Land	74.29		
		Vegetation	74.19	Vegetation	100.00		
		Buildup Area	56.25	Buildup Area	100.00		
		Barren Area	100.00	Barren Area	75.00		
	08.11.1988	Water Body	100.00	Water Body	100.00	90	86.96
		Farm Land	96.97	Farm Land	91.43		
		Vegetation	86.96	Vegetation	95.24		
		Buildup Area	72.73	Buildup Area	100.00		
		Barren Area	100.00	Barren Area	70.00		
1998	05.02.1998	Water Body	100.00	Water Body	100.00	83	77.12
		Farm Land	97.37	Farm Land	84.09		
		Vegetation	65.00	Vegetation	92.86		
		Buildup Area	64.00	Buildup Area	100.00		
		Barren Area	100.00	Barren Area	47.06		
	04.11.1998	Water Body	100.00	Water Body	92.31	91	87.54
		Farm Land	90.91	Farm Land	93.02		
		Vegetation	80.00	Vegetation	80.00		
		Buildup Area	90.91	Buildup Area	100.00		
		Barren Area	100.00	Barren Area	77.78		
2008	17.02.2008	Water Body	100.00	Water Body	85.71	81	73.95
		Farm Land	83.33	Farm Land	62.50		
		Vegetation	73.33	Vegetation	94.29		
		Buildup Area	78.57	Buildup Area	100.00		
		Barren Area	100.00	Barren Area	62.50		
	01.12.2008	Water Body	93.33	Water Body	90.32	86	81.91
		Farm Land	77.78	Farm Land	80.77		
		Vegetation	85.71	Vegetation	80.00		
		Buildup Area	82.35	Buildup Area	82.35		
		Barren Area	91.67	Barren Area	100.00		
2018	12.02.2018	Water Body	100.00	Water Body	100.00	88	84.59
		Farm Land	96.67	Farm Land	82.86		
		Vegetation	73.91	Vegetation	94.44		
		Buildup Area	77.27	Buildup Area	100.00		
		Barren Area	100.00	Barren Area	70.59		
	11.11.2018	Water Body	91.67	Water Body	100.00	80	73.61
		Farm Land	97.14	Farm Land	79.07		
		Vegetation	63.16	Vegetation	92.31		
		Buildup Area	60.00	Buildup Area	100.00		
		Barren Area	88.89	Barren Area	44.44		

location collected from the website (https://power.larc.nasa.gov/) of NASA POWER. The difference between the estimated LST and earth skin temperature is listed in Table 13.

4.3. Flood event classification model validation

The mean error value obtained from MAPE is 25 % for a 4-fold CV, making the flood classification model reasonable (see Table 14). The MAPE values for the first, second, third, and fourth iterations are 50 %, 0 %, 50 %, and 0 %, respectively, with a mean of 25 %. According to scholar [69], a MAPE error of 20–50 % suggests a reasonable outcome.

5. Discussion

Many professionals utilize LULC analysis to ensure optimal planning and long-term management of land surface. LULC maps of an

Table 12
User Accuracy, Producer Accuracy, Overall Accuracy and Kappa Coefficient of different LULC components at non-flooded years.

Year	Date	User Accuracy (%)		Producer Accuracy (%)		Overall Accuracy (%)	Kappa Coefficient (%)
1993	22.01.1993	Water Body	100.00	Water Body	85.71	90	86.53
		Farm Land	85.71	Farm Land	94.74		
		Vegetation	86.67	Vegetation	81.25		
		Buildup Area	95.45	Buildup Area	95.45		
		Barren Area	88.89	Barren Area	80.00		
	08.12.1993	Water Body	88.89	Water Body	100.00	87	82.95
		Farm Land	100.00	Farm Land	70.97		
		Vegetation	74.36	Vegetation	100.00		
		Buildup Area	92.86	Buildup Area	86.67		
		Barren Area	100.00	Barren Area	77.78		
2003	26.01.2003	Water Body	100.00	Water Body	84.62	79	71.53
		Farm Land	73.91	Farm Land	60.71		
		Vegetation	75.56	Vegetation	94.44		
		Buildup Area	84.62	Buildup Area	84.62		
		Barren Area	100.00	Barren Area	75.00		
	18.11.2003	Water Body	100.00	Water Body	100.00	93	90.70
		Farm Land	100.00	Farm Land	85.71		
		Vegetation	86.84	Vegetation	97.06		
		Buildup Area	85.71	Buildup Area	92.31		
		Barren Area	100.00	Barren Area	88.89		
2014	05.03.2014	Water Body	100.00	Water Body	87.50	81	74.58
		Farm Land	75.00	Farm Land	62.07		
		Vegetation	73.81	Vegetation	93.94		
		Buildup Area	81.82	Buildup Area	81.82		
		Barren Area	100.00	Barren Area	81.82		
	02.12.2014	Water Body	100.00	Water Body	92.86	91	88.00
		Farm Land	87.50	Farm Land	94.59		
		Vegetation	88.24	Vegetation	83.33		
		Buildup Area	95.24	Buildup Area	95.24		
		Barren Area	88.89	Barren Area	80.00		
2023	26.02.2023	Water Body	88.24	Water Body	93.75	94	92.07
		Farm Land	100.00	Farm Land	89.74		
		Vegetation	87.50	Vegetation	100.00		
		Buildup Area	95.45	Buildup Area	95.45		
		Barren Area	90.00	Barren Area	100.00		
	09.11.2023	Water Body	87.50	Water Body	93.33	88	84.03
		Farm Land	89.74	Farm Land	92.11		
		Vegetation	80.00	Vegetation	80.00		
		Buildup Area	95.00	Buildup Area	86.36		
		Barren Area	80.00	Barren Area	80.00		

Table 13
Comparison of calculated LST with measured earth skin temperature at the position of first class observatory of Bangladesh Metrological Department in Tarash, Sirajganj.

Station	Location		Year	Date	LST	Earth Skin Temperature	Difference
	Latitude	Longitude					
Tarash, Sirajganj	24.43	89.38	1988	10.02.1988	23.41	22.53	0.88
				08.11.1988	27.18	27.62	−0.44
			1998	05.02.1998	19.96	19.21	0.75
				04.11.1998	24.20	23.40	0.80
			2008	17.02.2008	17.21	16.83	0.38
				01.12.2008	21.69	21.27	0.42
			2018	12.02.2018	18.98	17.94	1.04
				11.11.2018	22.87	22.03	0.84
			1993	22.01.1993	16.52	15.95	0.57
				08.12.1993	23.71	22.86	0.85
			2003	26.01.2003	24.30	23.84	0.46
				18.11.2003	21.56	20.87	0.69
			2014	05.03.2014	23.58	23.42	0.16
				02.12.2014	20.92	19.69	1.23
			2023	26.02.2023	24.66	23.92	0.74
				09.11.2023	23.46	22.05	1.41

Table 14
4-fold cross-validation.

Iteration	Folds				MAPE (%)	Mean MAPE (%)
1st iteration	Fold 1 (1988, 1993)	Fold 2 (1998, 2003)	Fold 3 (2008, 2014)	Fold 4 (2018, 2023)	E1	E= (E1+ E2+ E3+ E4)/4
2nd iteration	Fold 1 (1988, 1993)	Fold 2 (1998, 2003)	Fold 3 (2008, 2014)	Fold 4 (2018, 2023)	E2	
3rd iteration	Fold 1 (1988, 1993)	Fold 2 (1998, 2003)	Fold 3 (2008, 2014)	Fold 4 (2018, 2023)	E3	
4th iteration	Fold 1 (1988, 1993)	Fold 2 (1998, 2003)	Fold 3 (2008, 2014)	Fold 4 (2018, 2023)	E4	
Testing data fold				Training data folds		

area provide information to assist users in understanding the current terrain. It allows users to explore, map, and monitor the landscape, as well as better comprehend the effects of natural hazards and human activities. While NDWI can be used to discover water bodies, NDVI can be used to monitor plant health over large areas by detecting changes in vegetation cover. On the other hand, LST is an important environmental metric with applications in meteorology, climatology, agriculture, urban planning, and environmental monitoring. All of these analytical investigations provide a thorough grasp of the effects of numerous natural disasters on environmental factors. These parameters can aid in the short-term prediction of a natural disaster by providing stochastic variables. MI serves as a gateway to identifying these stochastic variables by assessing the stochastic dependency between the variable and the related disaster. Meanwhile, short-term disaster prediction refers to the technique of anticipating what will happen in the near future. The Gaussian Naïve Bayes (GNB) classifier has the ability to anticipate a disaster extremely quickly by calculating the likelihood of observing each environmental parameter associated with the disaster.

In this study, comparative LULC analysis demonstrates that the flood primarily affected vegetation, farmland, buildup areas, and barren areas. Even before the rainy season, certain signs of upcoming flooding can be observed in the form of altered vegetation, farmland, and barren areas. The presence of farmland and vegetation in February, before the rainy season in flooded years, was significant, as are November and December in those years, even though these months fell during the dry season. However, the presence of built-up areas is quite small when compared to farmland and vegetation. In February 2018, there are more water bodies in the middle of the area than in prior years. Except for January 2003, non-flooded years have much more buildup area in the upper western corner of the study region in the months of January, February, and March, prior to the rainy season. Rainfall variability is possible, which has an impact on vegetation, farmland, and built-up areas. In their research [74], show the positive influence of rainfall on vegetation and farmland propagation in China, whereas [75] illustrate the effect of drought on barren areas in Brazil. And these variations in rainfall suggest changes in cloud cover, which affects incoming solar energy. In this analysis, all-sky shortwave downward irradiance is believed to represent incoming solar energy. It is widely accepted to be the total amount of solar energy that reaches the surface of the earth in the shortwave spectrum (0.3–3.0 μm) [56]. Clouds control how much solar radiation a planet absorbs and how much light reaches its surface. Generally speaking, poorer solar energy absorption and a higher albedo are correlated with more cloud cover. Clouds control how much solar radiation a planet absorbs and how much light reaches its surface. Typically, better albedo and decreased solar energy absorption are correlated with greater cloud cover [76]. Low, dense clouds reflect solar radiation and chill the surface of Earth [77]. When clouds gather at low altitudes, it usually indicates that rain or other forms of precipitation may fall [78]. This effect was also observed in average LST and rainfall before the rainy season, when LST values are typically lower and rainfall amounts are higher in flooded years than in non-flooded years (Table 15). Only average NDVI, NDWI, and LST are considered for further MI investigation because maximum and minimum NDVI, NDWI, and LST often represent very small areas. Also, area coverage of different ranges of NDVI, NDWI, and LST is also kept to observe their mutual dependency on floods. A total of nineteen stochastic variables were chosen from the comparative scenario analysis, and five of them were shortlisted as predictors by quantifying mutual dependency on the flood scenario. It has been discovered that areas having NDVI of 0.0 to −1.0, which denotes water bodies, have the

Table 15
Values of Predictor stochastic variables.

`	All Sky Surface Shortwave Downward Irradiance	Area having NDVI of 0.0 to −1.0 (Square Kilometer)	Average LST	Rainfall	Area having LST of 31–40 °C (Square Kilometer)
10.02.1988	3.81	174.57	23.76	19.64	1.26
	3.46	184.47	22.32	15.76	0.00
05.02.1998					
	3.17	186	20.60	14.02	0.00
17.02.2008					
	3.63	170.9	23.18	17.66	81.01
12.02.2018					
	4.68	218.01	16.84	3.8	0.00
22.01.1993					
	4.25	158.04	26.13	5.34	282.67
26.01.2003					
	4.38	158.64	24.95	15.8	198.81
05.03.2014					
	4.11	143.9	26.91	6.36	286.60
26.02.2023					

highest (0.69) MI values and reveal most about flooding. Table 15 also shows that flooded years had a higher amount of area having NDVI of 0.0 to −1.0 before the rainy season. This indicates that this stochastic variable is likely to display symptoms of a flood in the near future. Similarly, the MI value (0.52) of all sky surface shortwave downward irradiance and its value (Table 15) during flooded years as compared to non-flooded years reveal a similar but opposite character to the prior variable. Researcher [79] conducted a study to anticipate rainfall by selecting explanatory variables using MI. They employ this approach to anticipate rainfall using artificial neural networks. Their explanatory factors include rainfall, atmospheric pressure, dry bulb temperature, relative humidity, and wind speed, with the highest MI value of 0.215 for dry bulb temperature. The variables included in this study come from a comparative flood study; thus, they are fairly uncommon. The only variable in common between the current study and the study by that group of researchers [79] is rainfall, which has a MI value of 0.41 in the present study and 0.159 in their study [79]. Further research is required to realize the capability of stochastic factors to shear flood information, as described in the current work.

For short-term predictions using the Gaussian Naïve Bayes (GNB) classifier, the stochastic variable of all sky surface shortwave downward irradiance can correctly classify six out of eight flood scenarios. Combining this variable with an area having NDVI of 0.0 to −1.0 can also achieve similar results. Other variables and their combinations have poor classification properties when predicting the flood scenario. This indicates that all sky surface shortwave downward irradiance has the best potential to anticipate flood scenarios. Researchers [13] conducted a study in which they used the GNB algorithm to predict droughts and floods using monthly rainfall, maximum and minimum temperature, relative humidity, wind speed, sunshine duration, potential evapotranspiration, and the Standardized Precipitation Index (SPI) as predictor variables. They discovered that SPI values performed better in prediction. Discovering more optimum stochastic variables, such as all-sky surface shortwave downward irradiance, can improve prediction capability. The use of this type of variable to make short-term predictions is uncommon, so more research is needed. All sky surface shortwave downward irradiance data utilized in this study are average values determined by averaging individual values from five locations. And the values at each location are derived using the prior thirty-day average from the day the satellite image was acquired. In this study, the approach is also used to compute rainfall. It is a hypothesis to test the accuracy of these variables as predictors, and the results are better. From the 4-fold cross-validation procedure, a 25 % mean MAPE value shows that the model is reasonable. This is due to the short amount of training data, whereas GNB performs best with a big number of training data. Using more flooded and non-flooded years as training data, this approach can improve the results.

More research is needed to identify such optimum stochastic variables by alternating the number of data points and days. Increasing the number of training data points can also lead to increased accuracy during testing.

6. Conclusion

One of the most dangerous natural disasters is flooding because it primarily alters the pattern of land cover. The longest period of flooding has the most detrimental effects on various environmental elements and land use practices. A clear picture of how flood events regulate land cover dynamics and land use practices in the Sirajganj district is provided by this study's LULC analysis, which was conducted using the Iso Cluster unsupervised classification method, NDVI, NDWI, and LST analysis. And some parameters from this study are chosen as stochastic variables, which are then evaluated using MI to determine the mutual dependence on flood. The Gaussian Naïve Bayes classifier predicts the correct flood scenario by employing shortened stochastic variables as predictors from the mutual information study. The optimum stochastic variable is selected as all-sky surface shortwave downward irradiance. This type of study is uncommon, but in the case of Sirajganj, it is the first. The important feature of this study is that it is more accurate, inexpensive, simple, and straightforward to implement. This model can be highly precise if more training data is added, making the classification process effective. Additionally, discovering more predictor factors can improve the classification procedure's accuracy. Consequently, investors seeking to reduce the risk of natural disasters would find this well-established data-intelligent analytic study useful. This study concludes that the variable selection strategy via mutual information is preferable. The short-term prediction accuracy of the Gaussian Naïve Bayes classifier in forecasting flood scenarios is sensitive to the predictor stochastic variables used.

CRediT authorship contribution statement

Chandan Mondal: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation. **Md Jahir Uddin:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization.

Data availability statement

The data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to express their deep gratitude to the Bangladesh Water Development Board (BWDB) for supporting and

providing rainfall data for the study area. The authors are also very thankful to the reviewers for their valuable comments and suggestions.

References

[1] T.D. Gizaw, S. Baye, Flood effects on household livelihoods and it's controlling strategies in gelana woreda, oromia, Ethiopia, Nat. Hazards (2024) 1–32.
[2] H.D. Nguyen, Q.H. Nguyen, D.K. Dang, C.P. Van, Q.H. Truong, S.D. Pham, A.I. Petrisor, A novel flood risk management approach based on future climate and land use change scenarios, Sci. Total Environ. 921 (2024) 171204.
[3] Q. Xu, Y. Shi, J.L. Bamber, C. Ouyang, X.X. Zhu, Large-scale flood modeling and forecasting with FloodCast, Water Res. 264 (2024) 122162.
[4] Y. Wang, J. Liu, C. Li, Y. Liu, L. Xu, F. Yu, A data-driven approach for flood prediction using grid-based meteorological data, Hydrol. Process. 37 (3) (2023) e14837.
[5] A.S. Chafjiri, M. Gheibi, B. Chahkandi, H. Eghbalian, S. Waclawek, A.M. Fathollahi-Fard, K. Behzadian, Enhancing flood risk mitigation by advanced data-driven approach, Heliyon 10 (18) (2024) e37758.
[6] I.M. Hayder, T.A. Al-Amiedy, W. Ghaban, F. Saeed, M. Nasser, G.A. Al-Ali, H.A. Younis, An intelligent early flood forecasting and prediction leveraging machine and deep learning algorithms with advanced alert system, Processes 11 (2) (2023) 481.
[7] C. Mondal, M.J. Uddin, Assessment of climate change induced rainfall trend and variability with non-parametric and linear approach for Sirajganj district, Bangladesh, Heliyon 10 (10) (2024) e31151.
[8] O.O. Aiyelokun, O.D. Aiyelokun, O.A. Agbede, Application of random forest (RF) for flood levels prediction in Lower Ogun Basin, Nigeria, Nat. Hazards 119 (3) (2023) 2179–2195.
[9] A. Msigwa, A.S. Makinde, Application of machine learning techniques in hydrometeorological event prediction. Modeling and Monitoring Extreme Hydrometeorological Events, 2024, pp. 135–161.
[10] J. Sarwar, S.A. Khan, M. Azmat, F. Khan, An application of hybrid bagging-boosting decision trees ensemble model for riverine flood susceptibility mapping and regional risk delineation, Water Resour. Manag. (2024) 1–31.
[11] D. Sabanci, K. Yurekli, M.M. Comert, S. Kilicarslan, M. Erdogan, Predicting reference evapotranspiration based on hydro-climatic variables: comparison of different machine learning models, Hydrol. Sci. J. 68 (7) (2023) 1050–1063.
[12] M.M. Hasan, M.S.M. Nilay, N.H. Jibon, R.M. Rahman, LULC changes to riverine flooding: a case study on the Jamuna River, Bangladesh using the multilayer perceptron model, Results in Engineering 18 (2023) 101079.
[13] O. Aiyelokun, G. Ogunsanwo, A. Ojelabi, O. Agbede, Gaussian Naïve Bayes classification algorithm for drought and flood risk reduction. Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation: Theory and Practice of Hazard Mitigation, 2021, pp. 49–62.
[14] H. Caihong, Z. Xueli, L. Changqing, L. Chengshuai, W. Jinxing, J. Shengqi, Real-time flood classification forecasting based on k-means++ clustering and neural network, Water Resour. Manag. (2022) 1–15.
[15] G. Wei, W. Ding, G. Liang, B. He, J. Wu, R. Zhang, H. Zhou, A new framework based on data-based mechanistic model and forgetting mechanism for flood forecast, Water Resour. Manag. 36 (10) (2022) 3591–3607.
[16] A. Iqbal, M.S. Mondal, M.S.A. Khan, H. Hakvoort, W. Veerbeek, Flood propagation processes in the Jamuna river floodplain in Sirajganj, in: Water Management: A View from Multidisciplinary Perspectives: 8th International Conference on Water and Flood Management, Springer International Publishing, Cham, 2022, pp. 45–67.
[17] M.H. Ali, B. Bhattacharya, A.K.M.S. Islam, G.M.T. Islam, M.S. Hossain, A.S. Khan, Challenges for flood risk management in flood-prone Sirajganj region of Bangladesh, Journal of Flood Risk Management 12 (1) (2019) e12450.
[18] S.R. Bhuiyan, A. Al Baky, Digital elevation based flood hazard and vulnerability study at various return periods in Sirajganj Sadar Upazila, Bangladesh, Int. J. Disaster Risk Reduc. 10 (2014) 48–58.
[19] F. Sai, L. Cumiskey, A. Weerts, B. Bhattacharya, R. Haque Khan, Towards impact-based flood forecasting and warning in Bangladesh: a case study at the local level in Sirajganj district, Natural Hazards and Earth System Sciences Discussions (2018) 1–20.
[20] M.A. Aktar, K. Shohani, M.N. Hasan, M.K. Hasan, Flood vulnerability assessment by flood vulnerability index (FVI) method: a study on Sirajganj Sadar Upazila, International Journal of Disaster Risk Management 3 (1) (2021) 1–13.
[21] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, S. Ridella, The'K'in K-fold Cross validation, ESANN 102 (2012, April) 441–446.
[22] Sirajganj District. (2023). Retrieved from Banglapedia: https://en.banglapedia.org/index.php/Sirajganj_District.
[23] R. Ahmed, S. Karmakar, Arrival and withdrawal dates of the summer monsoon in Bangladesh, Int. J. Climatol. 13 (7) (1993) 727–740.
[24] J.L. Faundeen, R.L. Kanengieter, M.D. Buswell, US geological survey spatial data access, Journal of Geospatial Engineering 4 (2) (2002) 145, 145.
[25] B. Hegyi, P.W. Stackhouse, P. Taylor, F. Patadia, Nasa POWER: providing present and future climate services based on NASA data for the energy, agricultural, and sustainable buildings communities, in: 104th American Meteorological Society (AMS) Annual Meeting, 2024, January.
[26] M.M. Hossain, A. Zahid, Bangladesh Water Development Board: a bank of hydrological data essential for planning and design in water sector, in: Proceedings of the 2nd International Conference on Advances in Civil Engineering 2014, Chittagong University of Engineering and Technology, 2014. Bangladesh, December 2014.
[27] N.G. Kardoulas, A.C. Bird, A.I. Lawan, Geometric correction of SPOT and Landsat imagery: a comparison of map-and GPS-derived control points, Photogramm. Eng. Rem. Sens. 62 (10) (1996) 1173–1177.
[28] L. Ju, S. Guo, X. Ruan, Y. Wang, Improving the mapping accuracy of soil heavy metals through an adaptive multi-fidelity interpolation method, Environ. Pollut. 330 (2023) 121827.
[29] S. Bel-Lahbib, K. Ibno Namr, B. Rerhou, F. Mosseddaq, B. El Bourhrami, L. Moughli, Assessment of soil quality by modeling soil quality index and mapping soil parameters using IDW interpolation in Moroccan semi-arid, Modeling Earth Systems and Environment 9 (4) (2023) 4135–4153.
[30] P.L. Ohlert, M. Bach, L. Breuer, Accuracy assessment of inverse distance weighting interpolation of groundwater nitrate concentrations in Bavaria (Germany), Environ. Sci. Pollut. Control Ser. 30 (4) (2023) 9445–9455.
[31] W. Yang, Y. Zhao, D. Wang, H. Wu, A. Lin, L. He, Using principal components analysis and IDW interpolation to determine spatial and temporal changes of surface water quality of Xin'anjiang river in Huangshan, China, Int. J. Environ. Res. Publ. Health 17 (8) (2020) 2942.
[32] L. Yang, G. Zhao, P. Tian, X. Mu, X. Tian, J. Feng, Y. Bai, Runoff changes in the major river basins of China and their responses to potential driving forces, J. Hydrol. 607 (2022) 127536.
[33] O.M. Baez-Villanueva, M. Zambrano-Bigiarini, H.E. Beck, I. McNamara, L. Ribbe, A. Nauditt, N.X. Thinh, RF-MEP: a novel Random Forest method for merging gridded precipitation products and ground-based measurements, Rem. Sens. Environ. 239 (2020) 111606.
[34] J.A. Burton, The Value of Conserving Genetic ResourcesMargery Oldfield Available from Superintendent of Documents, US Government Printing Office, Washington DC 20402, USA, 1985, p. 55, 1984, $8• 50. Oryx, 19(1), 55.
[35] G. Chander, B. Markham, Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges, IEEE Trans. Geosci. Rem. Sens. 41 (11) (2003) 2674–2677.
[36] V. Ihlen, K. Zanter, Landsat 8 (L8) Data Users Handbook Version 5.0. Department of the Interior, United States Geological Survey, 2019. Sioux Falls, South Dakota.
[37] S. Das, D.P. Angadi, Land use-land cover (LULC) transformation and its relation with land surface temperature changes: a case study of Barrackpore Subdivision, West Bengal, India, Remote Sens. Appl.: Society and Environment 19 (2020) 100322.
[38] M.B. Roy, A. Ghosh, S. Mohinuddin, A. Kumar, P.K. Roy, Analysing the trending nature in land surface temperature on different land use land cover changes in urban lakes, West Bengal, India, Modeling Earth Systems and Environment 8 (4) (2022) 4603–4627.
[39] Accuracy Metrics. (2019). Retrieved from Humboldt State University: http://gsp.humboldt.edu/olm/Courses/GSP_216/lessons/accuracy/metrics.html.

[40] A.J. Tallón-Ballesteros, J.C. Riquelme, Data mining methods applied to a digital forensics task for supervised machine learning. Computational Intelligence in Digital Forensics: Forensic Investigation and Applications, 2014, pp. 413–428.
[41] Landsat 7 Data Users Handbook, Retrieved from USGS, 2019, DECEMBER 5. https://www.usgs.gov/media/files/landsat-7-data-users-handbook.
[42] S. Sruthi, M.M. Aslam, Agricultural drought analysis using the NDVI and land surface temperature data; a case study of Raichur district, Aquatic Procedia 4 (2015) 1258–1264.
[43] S.K. McFeeters, The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, Int. J. Rem. Sens. 17 (7) (1996) 1425–1432.
[44] N. Pettorelli, J.O. Vik, A. Mysterud, J.M. Gaillard, C.J. Tucker, N.C. Stenseth, Using the satellite-derived NDVI to assess ecological responses to environmental change, Trends Ecol. Evol. 20 (9) (2005) 503–510.
[45] C.L. Meneses-Tovar, NDVI as indicator of degradation, Unasylva 62 (238) (2011) 39–46.
[46] S. Szabo, Z. Gácsi, B. Balazs, Specific Features of NDVI, NDWI and MNDWI as Reflected in Land Cover Categories, 2016.
[47] A.M. Saqr, M. Nasr, M. Fujii, C. Yoshimura, M.G. Ibrahim, Monitoring of agricultural expansion using hybrid classification method in southwestern fringes of Wadi El-Natrun, Egypt: an appraisal for sustainable development, in: Asia Conference on Environment and Sustainable Development, Springer Nature Singapore, Singapore, 2022, November, pp. 349–362.
[48] NDVI, Retrieved from earth observing system. https://eos.com/make-an-analysis/ndvi/, 2023.
[49] Retrieved from earth observing system, Normalized Difference Water Index, 2023. https://eos.com/make-an-analysis/ndwi/.
[50] J.A. Sobrino, J.C. Jiménez-Muñoz, L. Paolini, Land surface temperature retrieval from LANDSAT TM 5, Rem. Sens. Environ. 90 (4) (2004) 434–440.
[51] U. Avdan, G. Jovanovska, Algorithm for automated mapping of land surface temperature using LANDSAT 8 satellite data, J. Sens. 2016 (2016) 1–8.
[52] Z.L. Li, B.H. Tang, H. Wu, H. Ren, G. Yan, Z. Wan, J.A. Sobrino, Satellite-derived land surface temperature: current status and perspectives, Rem. Sens. Environ. 131 (2013) 14–37.
[53] P.K. Srivastava, T.J. Majumdar, A.K. Bhattacharya, Surface temperature estimation in Singhbhum Shear Zone of India using Landsat-7 ETM+ thermal infrared data, Adv. Space Res. 43 (10) (2009) 1563–1574.
[54] F. Mukherjee, D. Singh, Assessing land use–land cover change and its impact on land surface temperature using LANDSAT data: a comparison of two urban areas in India, Earth Systems and Environment 4 (2020) 385–407.
[55] C. Coll, V. Caselles, J.M. Galve, E. Valor, R. Niclos, J.M. Sánchez, R. Rivas, Ground measurements for the validation of land surface temperatures derived from AATSR and MODIS data, Rem. Sens. Environ. 97 (3) (2005) 288–300.
[56] Y.C. Yu, J. Shi, T. Wang, H. Letu, C. Zhao, All-sky total and direct surface shortwave downward radiation (SWDR) estimation from satellite: applications to MODIS and Himawari-8, Int. J. Appl. Earth Obs. Geoinf. 102 (2021) 102380.
[57] H.O. Lohaka, Making a Grouped-Data Frequency Table: Development and Examination of the Iteration Algorithm (Doctoral Dissertation, Ohio University, 2007.
[58] B.D. Venables Wnripley, Modern Applied Statistics with S. Statistics and Computing, Springer, New York, 2002.
[59] T.M. Cover, Elements of Information Theory, John Wiley & Sons, 1999.
[60] S. Roberts, R. Everson (Eds.), Independent Component Analysis: Principles and Practice, Cambridge University Press, 2001.
[61] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys. 69 (6) (2004) 066138.
[62] L. Batina, B. Gierlichs, E. Prouff, M. Rivain, F.X. Standaert, N. Veyrat-Charvillon, Mutual information analysis: a comprehensive study, J. Cryptol. 24 (2) (2011) 269–291.
[63] P. Schober, C. Boer, L.A. Schwarte, Correlation coefficients: appropriate use and interpretation, Anesth. Analg. 126 (5) (2018) 1763–1768.
[64] M.A. Sulaiman, J. Labadin, Improved feature selection based on mutual information for regression tasks, Journal of IT in Asia 6 (1) (2016) 11–24.
[65] R.G. Poola, L. Pl, COVID-19 diagnosis: a comprehensive review of pre-trained deep learning models based on feature extraction algorithm, Results in Engineering 18 (2023) 101020.
[66] A. Stuart, K. Ord, Kendall's Advanced Theory of Statistics, Distribution Theory, vol. 1, John Wiley & Sons, 2010.
[67] S.M. Lee, P.A. Abbott, Bayesian networks for knowledge discovery in large datasets: basics for nurse researchers, J. Biomed. Inf. 36 (4–5) (2003) 389–399.
[68] J.S. Rosenthal, First Look at Rigorous Probability Theory, A, World Scientific Publishing Company, 2006.
[69] J.J.M. Moreno, A.P. Pol, A.S. Abad, B.C. Blasco, Using the R-MAPE index as a resistant measure of forecast accuracy, Psicothema 25 (4) (2013) 500–506.
[70] M.M.Q. Mirza, Three recent extreme floods in Bangladesh: a hydro-meteorological analysis. Flood Problem and Management in South Asia, 2003, pp. 35–64.
[71] P. Jiang, K. Son, M.K. Mudunuru, X. Chen, Using mutual information for global sensitivity analysis on watershed modeling, Water Resour. Res. 58 (10) (2022) e2022WR032932.
[72] M.L. McHugh, Interrater reliability: the kappa statistic, Biochem. Med. 22 (3) (2012) 276–282.
[73] Jr R.G. Pontius, M. Millones, Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment, Int. J. Rem. Sens. 32 (15) (2011) 4407–4429.
[74] X. Yue, T. Zhang, X. Zhao, X. Liu, Y. Ma, Effects of rainfall patterns on annual plants in horqin sandy land, inner Mongolia of China, Journal of Arid Land 8 (2016) 389–398.
[75] L. Costa, A.A. Sant'Anna, C.E.F. Young, Barren lives: drought shocks and agricultural vulnerability in the Brazilian Semi-Arid, Environ. Dev. Econ. 28 (6) (2023) 603–623.
[76] R. Mueller, J. Trentmann, C. Träger-Chatterjee, R. Posselt, R. Stöckli, The role of the effective cloud albedo for climate monitoring and analysis, Rem. Sens. 3 (11) (2011) 2305–2320.
[77] S. Graham, NASAEarth observatory, Retrieved from, https://earthobservatory.nasa.gov/features/Clouds#:~:text=Low%2C%20thick%20clouds%20primarily%20reflect,the%20surface%20of%20the%20Earth, 1999, March 1.
[78] B.J. Mason, Clouds, Rain and Rainmaking, Cambridge University Press, 1975.
[79] M.S. Babel, G.B. Badgujar, V.R. Shinde, Using the mutual information technique to select explanatory variables in artificial neural networks for rainfall forecasting, Meteorol. Appl. 22 (3) (2015) 610–616.