Journal of Geovisualization and Spatial Analysis (2024) 8:19
https://doi.org/10.1007/s41651-024-00181-5

Enhancing Flood Risk Analysis in Harris County: Integrating Flood Susceptibility and Social Vulnerability Mapping

Hemal Dey[1,2] · **Wanyun Shao**[1,2] · **Md Munjurul Haque**[1,2] · **Matthew VanDyke**[2,3]

Accepted: 5 May 2024

Abstract
Due to climate change, the frequency and intensity of floods have dramatically increased worldwide. The innate social inequality has been exposed and even exacerbated by increasing flooding. It is imperative to assess flood risk in a comprehensive manner, accounting for both physical exposure and social vulnerability. Harris County in Texas, U.S., is selected as the study area as it has experienced a few devastating floods in recent history, with Hurricane Harvey (2017) being the most impactful. First, this study generates a flood susceptibility map (FSM) by applying a Random Forest (RF) model with 500 flood inventory points and 12 flood conditioning factors. Then, it generates a social vulnerability map (SoVM) by applying Principal Component Analysis (PCA) with ten social variables at the census tract level. Finally, it combines FSM with SoVM to produce a flood risk map (FRM) of Harris County. The findings of this study demonstrate that 9.06% of the area of Harris County has high flood susceptibility and 1.45% of the area has a very high social vulnerability. Combining both flood susceptibility and social vulnerability, this study reveals that 5.59% of the total area has a very high risk for flooding. This study further compares the FRM with the Federal Emergency Management Agency's (FEMA) 100-year floodplain map and notes major differences. The comparison reveals that 76.7% of very high and 81.8% of high-risk areas in FRM are underestimated by the FEMA 100-year floodplain. This study produces a comprehensive FRM, highlighting areas where flooding can exacerbate social inequality and cause higher economic costs. FEMA's 100-year floodplain map underestimates a significant portion of high-risk areas suggesting that current zoning and development policy may fail to consider flood risks adequately.

Keywords Flood susceptibility · Social vulnerability · Flood risk · Machine learning

Introduction

The frequency and intensity of flood hazards have increased due to climate change particularly in the coastal regions that are susceptible to rising sea levels and population growth (Wong et al. 2014). The increasing frequency and severity of floods demand a more comprehensive approach to understanding and mitigating flood risk. Traditional flood risk assessments typically prioritize meteorological and hydrological factors, such as rainfall intensity, topography, and river discharge, as key determinants (Park and Lee 2019). The significance of these factors cannot be denied, yet they only partially explain the complex and dynamic relations to flood risk (Eiser et al. 2012). A wider range of factors, such as flood susceptibility and social vulnerability, must be considered to reach a more comprehensive understanding (Haque et al. 2023). Risk decision-makers (e.g., emergency managers and urban planners) and the public need relevant FRM products that effectively and accurately communicate flood risks in support of risk mitigation and reduction. Research has shown that different types of risk decision-makers have different information needs, often based on their technical knowledge and expertise (VanDyke et al. 2021), and the public may often be disengaged from water-related risks unless the risks are particularly salient, for example, unless a flood crisis is imminent or occurring (VanDyke and King 2018). Therefore, to maximize the opportunity for relevant audiences to engage FRM

✉ Wanyun Shao
wshao1@ua.edu

[1] Department of Geography & the Environment, University of Alabama, Tuscaloosa, Alabama, US

[2] Alabama Water Institute, University of Alabama, Tuscaloosa, Alabama, US

[3] Department of Advertising and Public Relations, University of Alabama, Tuscaloosa, Alabama, US

Published online: 22 May 2024

products, information should strive for a comprehensive, precise assessment of flood risks.

The complex interaction between physical exposure to floods and social vulnerability justifies a comprehensive assessment that incorporates both aspects. Research has extensively explored each dimension separately. The physical dimension of flood hazards, particularly flood susceptibility, has been the focus of a large body of work. Flood susceptibility assessment estimates the likelihood and severity of floods and enables us to identify the spatial distribution of flood susceptibility by using past flood events data (flood inventory map) and flood conditioning factors (Rahman et al. 2019). The flood inventory map is a key element of flood susceptibility assessment that contains historical records of past floods (Merz et al. 2007). The past flood locations in the flood inventory map serve as a dependent variable. Additionally, topographical, geological, climatological, and hydrological factors of flooding, such as elevation, slope, precipitation, soil texture, and distance from rivers play a vital role during flood susceptibility assessment as independent variables (Das 2020; Kia et al. 2012; Sarkar and Mondal 2020). The primary product of flood susceptibility assessment is FSM, considered a crucial tool in flood mitigation planning and implementing effective measures to develop sustainable urban environments in less flood-prone areas (Seydi et al. 2022; Youssef et al. 2022). Meanwhile, the concept of "social vulnerability" describes a person's or a community's natural vulnerability to environmental disruptions (Cutter et al. 2003), while social vulnerability map (SoVM) is a choropleth map that depicts the spatial distribution of social vulnerability. It also plays an important role in assessing community preparedness, response, and recovery capacities at different stages of a natural disaster by identifying highly vulnerable regions and communities prone to adverse impacts (Frigerio and De Amicis 2016). It enables policymakers to effectively target interventions, distribute resources, and build the resilience of these communities by analyzing social vulnerability alongside flood susceptibility. Thus, the integration of both flood susceptibility assessment and social vulnerability assessment is crucial within the framework of flood risk assessment. This research initiative aims to combine flood susceptibility and social vulnerability to offer a comprehensive evaluation of flood risk, using Harris County in Texas, U.S. as a case study for illustration of our approach.

This study reflects a shift towards integrating diverse methodologies to understand and manage flood risks effectively. By combining historical data analysis, geospatial technologies, and advanced modeling techniques like Machine Learning (ML) algorithms, the study aims to enhance our capacity for predicting and mitigating flood impacts in various regions. Specifically, the innovation of this study lies in its comprehensive flood risk assessment approach by integrating flood susceptibility and social vulnerability into a coherent framework. Traditionally, flood risk assessments have largely focused on physical features alone, neglecting the social dimension that would either exacerbate or mitigate the impact of flooding. This study takes flood risk assessment to a higher level by employing a bivariate mapping technique that leverages both an FSM, generated using the RF model, and an SoVM, derived through the PCA technique. By combining these two advanced methodologies, the study offers a comprehensive assessment of flood risk, which is the outcome of both the physical susceptibility to flooding and the social capacity/incapacity to respond to flooding. This innovative approach not only increases the comprehensiveness of flood risk assessments but also provides valuable insights for targeted mitigation and adaptation strategies, helping communities build resilience against increasing flood risks. By synthesizing these datasets and methodologies, the study provides a holistic view of flood risk, highlighting the importance of considering both physical and social factors in flood risk assessment and management. Furthermore, this study discusses the comparison between FRM and FEMA's 100-year floodplain map.

Literature review

Over the past few years, Geographic Information System (GIS) and remote sensing (RS) techniques have been increasingly used to assess the physical dimension of natural hazards (Kadri and Nasrallah 2023; Qin et al. 2024; Tabassum et al. 2023) due to geospatial technological advancements, and the availability of real-time remotely sensed data (Hasan et al. 2023; Joyce et al. 2009). It provides a convenient platform to analyze, manipulate, and visualize potential hazard-related remotely sensed data very quickly and efficiently (Sarkar and Mondal 2020; Rahmati et al. 2016). The combination of GIS and RS has become a vital tool to depict flood susceptibility zones and associated flood damage (Kia et al. 2012). Especially, satellite RS-based flood inundation mapping enables us to gather historical flood data through image classification techniques. Many GIS and RS-based methods have already been developed to map flood susceptibility zones, such as bivariate statistical analysis (BSA), multi-criteria decision-making (MCDM), and ML models (Andaryani et al. 2021; Das 2020; Farhadi and Najafzadeh 2021; Islam et al. 2021; Lee et al. 2017; Rahman et al. 2019; Rahmati et al. 2016; Samanta et al. 2018; Sarkar and Mondal 2020). All these methods use GIS and RS techniques, incorporating geospatial analysis to predict future hazards by considering multiple flood conditioning factors. However, BSA and MCDM methods have some performance limitations. For instance, BSA often oversimplifies the interaction between flood points and flood conditioning factors,

neglects non-flood points, and relies only on sum weight or class weights (Dey et al. 2024). MCDM methods are based on expert opinions and, as a result, can introduce bias and variability in predictions (Shahabi et al. 2020). Recently, ML has emerged as a cutting-edge technology by overcoming these issues and providing higher accuracies for predicting various natural hazards (Nachappa et al. 2020).

Researchers increasingly utilize ML models to handle spatial data (Du et al. 2020; Sotiropoulou and Vavatsikos 2023), generate map images (Courtial et al. 2022), and achieve highly accurate and time-efficient RS image classification results (Dou et al. 2024). This trend of RS image classification using ML models is particularly notable in the applications of land use land cover (LULC) classifications and fine mapping using time series of multispectral (Dou et al. 2021) and hyperspectral images (Dou and Zeng 2020). Another important application of ML models using RS images is assessing a variety of natural hazards (Nachappa et al. 2020). ML models have become increasingly popular for flood hazard prediction as they efficiently use historical data to account for the characteristics of flood nonlinearity without requiring knowledge of physical processes (Mosavi et al. 2018). Additionally, ML models demonstrate better performance and cost-effectiveness in flood susceptibility mapping compared to other traditional methods (Rahmati et al. 2020). To date, researchers have adopted a variety of advanced ML models in flood susceptibility assessment. Some common ML models include CART (classification and regression trees) (Rahman et al. 2021), Logistic Regression (LR) (Rahman et al. 2019), Decision Tree (DT) (Seydi et al. 2022), Support Vector Machine (SVM) (Tehrany et al. 2015), Random Forest (RF) (Lee et al. 2017), Extreme Gradient Boosting (XGBoost) (Abedi et al. 2022), Artificial Neural Network (ANN) (Andaryani et al. 2021) and Convolutional Neural Network (CNN) (Youssef et al. 2022).

Compared to other ML models, the RF model has gained popularity in remotely sensed image classification because of its high predictive accuracy, ability to deal with missing values, and capability to detect outliers (Amare et al. 2021; Farhadi and Najafzadeh 2021). Due to its reliable performance and advantages, many researchers utilized RF models in assessing a variety of natural hazards such as landslide susceptibility (Merghadi et al. 2020), land subsidence and sinkhole susceptibility (Elmahdy et al. 2022), groundwater potentiality (Thanh et al. 2022), gully erosion susceptibility (Amare et al. 2021), seismic vulnerability (Han et al. 2020), and wildfire susceptibility (Iban and Sekertekin 2022). RF is also widely recognized as a leading non-parametric ensemble learning technique and has consistently achieved the highest accuracy for flood susceptibility mapping because of its ability to detect nonlinear trends between flooding and flood conditioning factors (Abedi et al. 2022; Dey et al. 2024; Elmahdy et al. 2022; Nachappa et al. 2020). Thus, this study employed the RF algorithm to conduct flood susceptibility mapping in Harris County by utilizing flood inventory points and 12 flood conditioning factors namely elevation, slope, aspect, curvature, topographic wetness index (TWI), stream power index (SPI), precipitation, LULC, distance to rivers, normalized difference vegetation index (NDVI), drainage density, and soil texture.

Meanwhile, social vulnerability assessment also requires several underlying factors of vulnerability. In addition to the presence of vulnerable groups like the elderly, young women, children, or people with disabilities, the innate social vulnerability is also affected by factors like income, access to resources, education, and healthcare (Rufat et al. 2015). Thus, it is essential to consider economic and educational factors along with demographic information to extract a more comprehensive understanding of social vulnerability. PCA is a commonly applied method in assessing social vulnerability by condensing vital socio-economic information into fewer factors and combining them into a composite index known as the social vulnerability index (SoVI) (Cutter et al. 2003). Many researchers have assessed social vulnerability by calculating the geographic variations of SoVI at various geographic scales by condensing multiple socio-economic indicators such as population density, age, income, education, race, and so on (Cutter et al. 2003; Dey et al. 2023; Khajehei et al. 2020; Shao et al. 2020). Thus, this study adopted the PCA method to prepare the SoVM of Harris County by utilizing ten socioeconomic variables.

Despite many studies having been conducted on flood susceptibility and social vulnerability assessment separately, there however has been limited research that integrates them to prepare a comprehensive flood risk map. Therefore, this study aims to address this research gap by assessing flood risk through the integration of flood susceptibility and social vulnerability assessments. Through this comprehensive approach, this study will offer valuable insights for developing effective flood risk mitigation strategies and enhancing community resilience in flood-prone areas.

Data and Methods

Study area

This study selects Harris County (corresponding to Houston throughout this paper) in the U.S. Gulf Coast to demonstrate our analytical strategy. It is located on the upper Gulf Coast in the southeast of Texas (Fig. 1). It covers an area of 1,778 square miles, out of which most of the region is typically a coastal plain with low elevation, mostly flat (Mukherjee and Singh 2020).

Harris County, where the city of Houston is located, is the third-most populous county in the country. Out of all

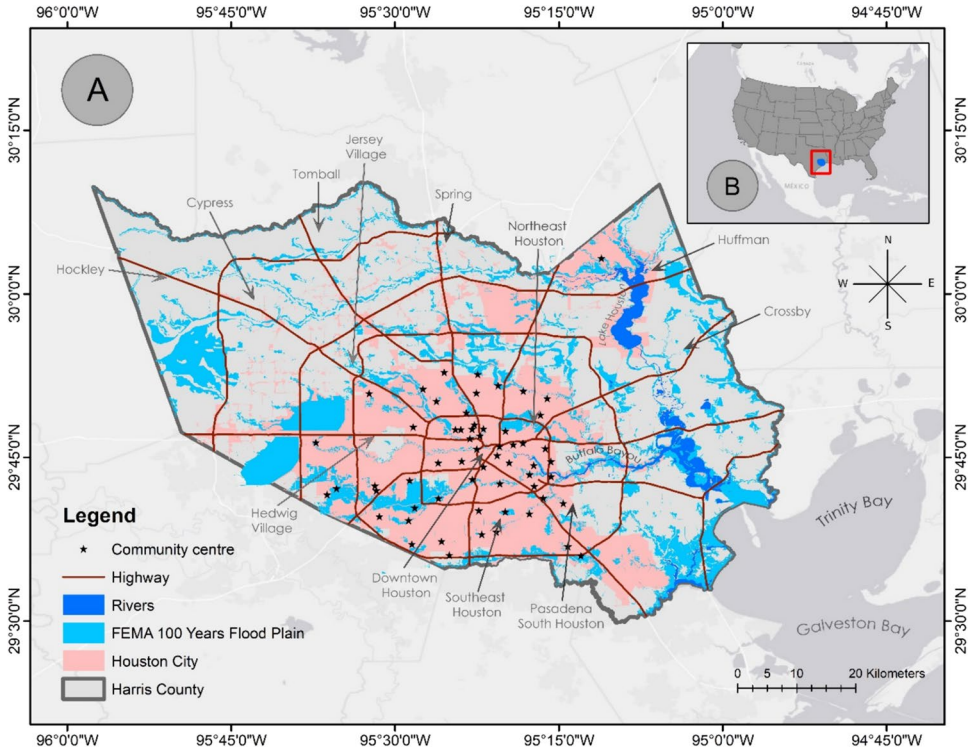

Fig. 1 Study area map (**a**) administrative map of Harris County (**b**) location of Harris County in the context of CONUS

block groups of Harris County, 30.5% of them have high population density. 70% of highly populated block groups are located within the FEMA 100-year and/or 500-year floodplain (Pulcinella et al. 2019). Numerous disastrous floods have occurred in this county in recorded history. Harris County encounters a severe flooding incident every two years (Mukherjee and Singh 2020). The flood caused by Hurricane Harvey was the most devastating in recent history, resulted in a total of 36 deaths. In addition, 154,170 homes and 300,000 vehicles were flooded during the Hurricane Harvey flooding event. Further, neighborhoods with more poor residents, black residents, people with disabilities, and young children faced comparatively higher flood risks during Hurricane Harvey (Chakraborty et al. 2019). Flores et al. (2023) highlighted that the future flood impact in Greater Houston (Harris County) could be catastrophic unless FEMA revise its risk maps by considering social equity. Therefore, due to its recent experience with major floods and revealed social disparities, Harris County serves as an ideal location for demonstrating our conceptual framework and implementing our analytical strategy.

Data sources

Multiple datasets including remotely sensed data and census data are used in this study. The sources of dataset, temporal resolution, spatial resolution, and data output are briefly demonstrated in Table 1. Three datasets are primarily required to prepare the flood inventory map including Sentinel-1 SAR from ESA, flood fatality data from Godfroy and Jonkman (2017), and storm events data from NOAA.

The DEM data are needed to create elevation, slope, aspect, curvature, TWI, SPI, and drainage density layers. The DEM data is obtained from the 3DEP DEM dataset by USGS using the Google Earth Engine (GEE) platform and processed on ArcGIS Pro 2.8. The JRC Global Surface Water dataset by Pekel et al. (2016) is utilized to produce distance from the river layer, downloaded through the GEE platform, and processed on ArcGIS Pro 2.8. The OREGONSTATE/PRISM/Norm91m dataset, a gridded climate dataset for the CONUS, spanning from 1991 to 2020 is downloaded from PRISM/Oregon state dataset for the precipitation layer. The soil texture layer is derived from the Open Land Map Soil Texture Class dataset by Hengl (2018) using GEE. Landsat 8 OLI/TIRS and NLCD datasets are downloaded from USGS to prepare the NDVI and LULC layers using the GEE platform. Finally, this study adopts the census data for 2018 collects ten variables from the Center for Disease Control and Prevention (CDC) at the census tract administrative level and processes them on SPSS and ArcGIS Pro 2.8.

Methodology

This study adopts an interdisciplinary approach to produce FSM, SoVM, and FRM of Harris County. The methodology is visualized as a flowchart in Fig. 2. This methodology can be divided into three major sections: (i) flood susceptibility

Table 1 Details of datasets used in this study

Dataset	Data sources	Temporal resolution	Spatial resolution (m)	Data Output
Sentinel-1 SAR	ESA/Copernicus	2017	10	Flood inventory map
Flood fatality data	Godfroy and Jonkman (2017)	2017	-	Flood inventory map
Strom Events data	NOAA	2017	-	Flood inventory map
JRC Global Surface Water	Pekel et al. (2016)	2020	30	Distance to river
3DEP DEM	USGS	-	10	Elevation, Slope, Aspect, Profile curvature, TWI, SPI, Drainage density
OREGONSTATE/PRISM/Norm91m	PRISM/OREGON STATE	1991–2020	928	Precipitation
Open Land Map Soil Texture Class	Hengl (2018)	2018	250	Soil texture
Landsat 8 OLI/TIRS	USGS	2017	30	NDVI
NLCD	USGS	2016	30	LULC
Census data	CDC/ATSDR	2018	Census Tracts	SoVM

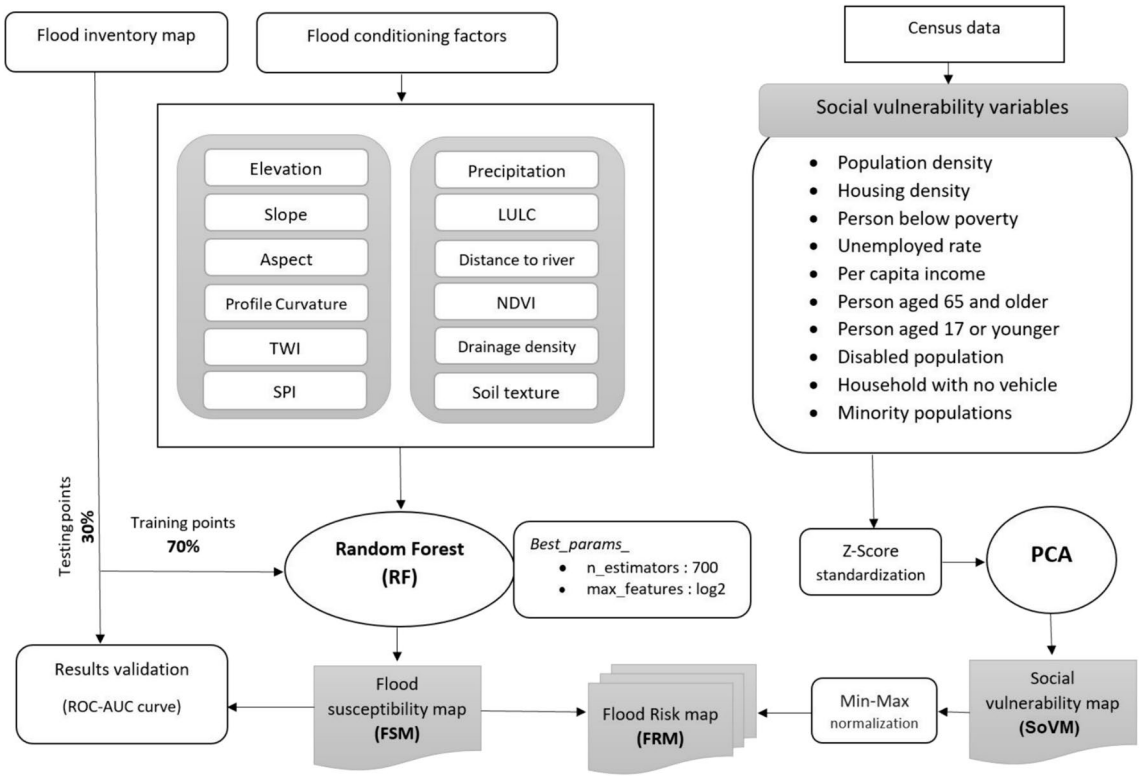

Fig. 2 Methodological framework of this study

assessment, (ii) social vulnerability assessment, and (iii) flood risk assessment. Each section is described below.

Flood susceptibility assessment

Flood inventory map In flood susceptibility analysis, the flood inventory map is the fundamental tool that consists of historical flood information of a particular region affected by flooding (Merz et al. 2007). It plays a vital role in estimating the potentiality of future floods based on previously flooded locations and flood conditioning factors (Tehrany et al. 2015). Since the accuracy of flood susceptibility analysis entirely depends on how precisely past flood information was recorded and used, this study takes meticulous measures to collect flood information from various sources. The most devastating flood, induced by Hurricane Harvey in August

2017 in Harris County, is considered to generate this flood inventory map.

This study uses three different data sources to generate this flood inventory map. First, the Sentinel-1 Synthetic Aperture Radar (SAR) dataset is used to extract flood inundated locations during Hurricane Harvey. The Sentinel-1 SAR dataset is a widely preferred dataset for flood inundation mapping because of its reliability, cloud penetration capability, and fine resolution (10 m) (Joyce et al. 2009). In this process, spatial filtering is applied to remove the 'salt and pepper noise' present in SAR data for atmospheric correction. Next, a division band ratio algorithm is applied to pre-flood and post-flood images to extract inundated zones. Later, the Global Surface Water dataset by Pekel et al. (2016) is masked out from that extracted inundated zone to depict actual and more precise flooded areas. This whole process is conducted in the GEE platform using JavaScript API.

Second, the flood fatality data of Hurricane Harvey collected from Godfroy and Jonkman (2017) is considered to ground truth flooded locations. Out of 500 inventory points, 250 points are taken as flood points, with 37 flood points adopted from the flood fatality dataset and the remaining 213 points being randomly taken from flooded areas extracted from Sentinel-1 SAR data. Similarly, 250 non-flood points are also taken randomly from non-flood regions of Harris County and used to prepare the training and testing datasets (Fig. 3). Third, the NOAA storm event dataset is used to confirm the location of all these flood points. Finally, these 500 flood inventory points are split into training (70%) and testing (30%) datasets to train and test the RF model.

Flood Conditioning Factors The occurrence of natural hazards such as floods, landslides, and cyclones is mainly determined by several causal factors. The factors that contribute directly or indirectly to the flood occurrence in a particular area are called flood conditioning factors. Several topographical, climatological, geological, environmental, and anthropogenic factors control flood occurrences (Sarkar and Mondal 2020). Before performing flood susceptibility analysis, it is crucial to select relevant flood conditioning factors for a particular study area (Kia et al. 2012). So, after an in-depth literature review (Das 2020; Kia et al. 2012; Rahmati et al. 2016; Samanta et al. 2018; Sarkar and Mondal 2020; Shahabi et al. 2020; Tehrany et al. 2015; Tehrany and Kumar 2018), a total 12 relevant flood conditioning factors have been meticulously selected for this study. Below, a concise description of these factors is provided.

Elevation The elevation is widely considered as a critical topographic factor in flood susceptibility assessment, as low elevations and flat regions are more susceptible to flooding due to gravitational force directing the flow of surface water in relation to changes in elevation (Das 2020; Tehrany et al. 2015). The elevation of the entire Harris County has a range between -7.33–98.99 m (Fig. 4a).

Slope The topographic slope influences the speed and direction of surface runoff, water accumulation process, and surface infiltration (Rahmati et al. 2016, 2020). Generally, low-lying with a gentle topographic slope region are flood susceptible areas (Kia et al. 2012). The slope of Harris County has a range between 0 to 40.14 degrees (Fig. 4b).

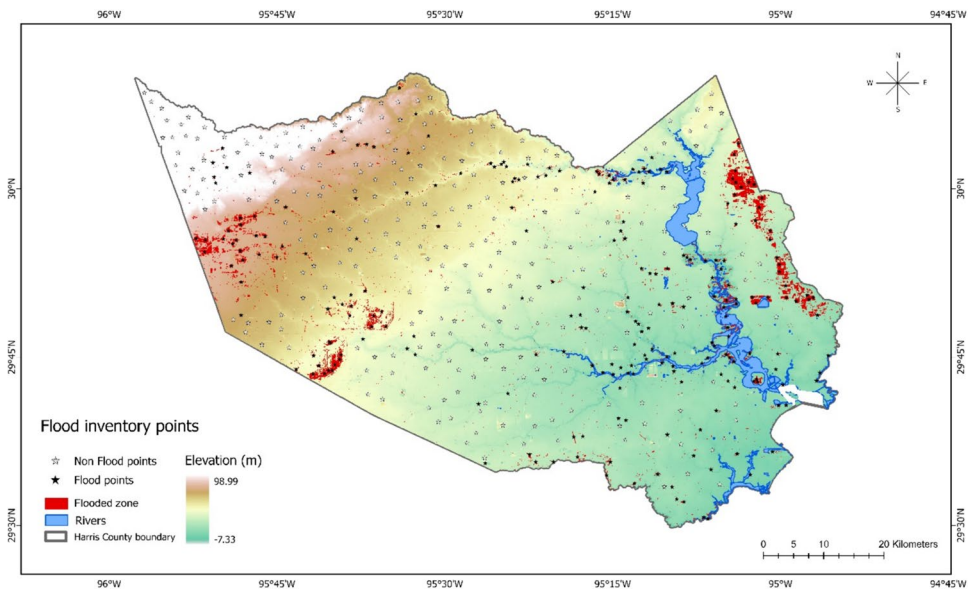

Fig. 3 Flood inventory map of Hurricane Harvey induced flooding in Harris County; white stars and black stars represent non-flood points and flood points respectively; red color represents the extracted flooded zone from Sentinel-1 SAR dataset; and blue color represents rivers of Harris County

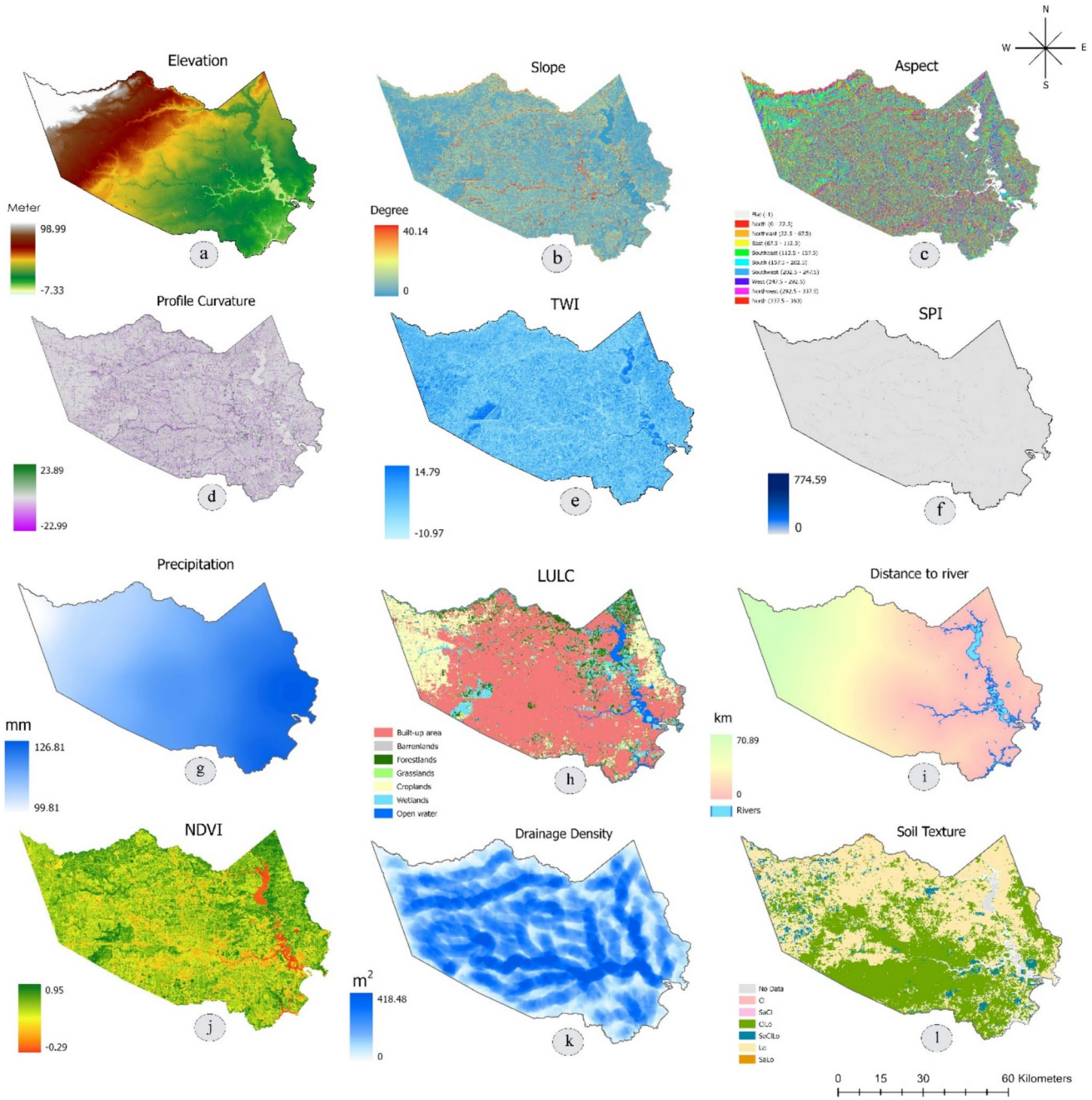

Fig. 4 Maps of 12 flood conditioning factors. (**a**) elevation; (**b**) slope; (**c**) aspect; (**d**) profile curvature; (**e**) TWI; (**f**) SPI; (**g**) precipitation; (**h**) LULC; (**i**) distance to river; (**j**) NDVI; (**k**) drainage density; (**l**) soil texture

Aspect The topographic aspect indicates the direction of a slope or land surface face and plays a significant role in flood susceptibility assessment as it influences precipitation patterns (Farhadi and Najafzadeh 2021). The aspect layer of Harris County has ten different faces (Fig. 4c).

Curvature Topographic curvature represents the physical attributes of a basin in terms of erosion and runoff process (Andaryani et al. 2021). It affects the water budget of floodplains and can differentiate between areas where surface runoff diverges and converges (Islam et al. 2021). The curvature value of Harris County ranges from -22.99 to 23.89 with a majority of the region exhibiting values close to zero (Fig. 4d). Typically, areas with zero curvature values are at higher risk for flooding (Shahabi et al. 2020).

TWI The TWI refers to the spatial distribution of wetness across a landscape that helps to understand the controls of

the overland water flow as wetter areas are more likely to produce runoff (Ali et al. 2019; Samanta et al. 2018). Consequently, a higher TWI value indicates a higher likelihood of flooding. The TWI of Harris County is calculated using Eq. 1.

$$TWI = In\left(A_S/tan\beta\right) \quad (1)$$

where, A_s is the specific catchment area (in m^2/m) and β is the slope (in radians). In Harris County, the TWI value ranges from -10.97 to 14.79 (Fig. 4e).

SPI The SPI helps to understand the dynamics of rivers and streams that are correlated with the erosive power of the overland flow, discharge rate, and soil wetness status of a basin (Tehrany and Kumar 2018). A higher SPI value represents a higher susceptibility to flooding (Seydi et al. 2022). The SPI is computed for the catchment area using Eq. 2.

$$SPI = As * tan\beta \quad (2)$$

Here, A_s is the specific catchment area (in m^2/m) and β is the slope (in radians). The SPI value ranges from 0 to 774.59 (Fig. 4f).

Precipitation Precipitation is considered as a crucial climatological factor and a prime source of water flow during flooding (Rahmati et al. 2020). Heavy precipitation causes rivers to overflow, and consequently inundates the low-lying surrounding floodplain areas (Das 2020). The monthly mean precipitation of Harris County ranges from 99.81 to 126.81 mm (Fig. 4g).

LULC The LULC is an important anthropogenic factor of flood and can impact the potentiality of flooding by affecting the infiltration process and surface runoff (Andaryani et al. 2021). A region with low vegetation (urban land, or barren lands) is at a higher risk of flooding compared to densely vegetated regions (forests or grasslands) due to unhindered water flow (Kia et al. 2012). This layer of Harris County was categorized into seven broad classes including built-up areas, barren lands, forestlands, grasslands, croplands, wetlands, and open water (Fig. 4h).

Distance to the rivers The distance to the river is a key factor in floods because floods generally occur near the riverbank and inundate the surrounding low-lying floodplain land, significantly influencing the extent and severity of floods (Farhadi and Najafzadeh 2021; Samanta et al. 2018). The value range of distance to river layers of Harris County is between 0 to 70.89 km (Fig. 4i).

NDVI NDVI is inversely correlated with flood occurrences; sparse vegetation is more susceptible to flooding due to easy water flow, whereas dense vegetation in forestland impedes the water flow and reduces flood susceptibility (Tehrany et al. 2015). The NDVI value of Harris County ranges from -0.29 to 0.95 (Fig. 4j).

Drainage density Drainage density is defined as the total length of channel segments within a drainage basin per unit area of the basin (Das 2020). The likelihood of flooding increases with higher drainage density (Sarkar and Mondol 2020). The drainage density of Harris County ranges from 0 to 418.48 m/m2 (Fig. 4k).

Soil texture Soil texture is a prime geological factor that significantly influences flooding, by controlling the amount of infiltration, surface runoff, and inundation process (Rahman et al. 2021; Rahmati et al. 2016). The soil types with slow infiltration ability are prone to flooding, causing a higher peak discharge and river flood. Most of Harris County is composed of ClLo and Lo soil texture (Fig. 4l).

Random Forest model The RF model developed by Breiman (2001) is a supervised ensemble algorithm that is widely used to efficiently solve classification, regression, cluster analysis, and anomaly detection problems (Géron 2022). RF offers several benefits including high predictive accuracy, the ability to identify significant features, the ability to reduce the chance of overfitting, the capability to handle large and high-dimensional data, and the ability to deal with missing values and outliers within the predictor variables, etc. (Amare et al. 2021; Elmahdy et al. 2022). The bootstrap aggregating (bagging) sampling method is used in the training process of RF models, where random subsets of data are selected with replacement (Géron 2022). During the training phase of the RF algorithm, multiple decision trees are constructed for bootstrapped samples. Each tree includes root nodes, child nodes, and leaf nodes. The final prediction is then determined by either majority voting (for classification problems) or by considering the mean (for regression problems) of the predictions gathered from the leaf nodes of all trees (Islam et al. 2021; Lee et al. 2017).

In Fig. 5, the flowchart represents the total training dataset as D, which is later divided into multiple subsets (d1, d2, ….., dn) using a bootstrap sampling method. For each subset, one decision tree is constructed. Each decision tree provides individual prediction (P1, P2,……,Pn). Since flood susceptibility assessment is a classification problem, the overall prediction (P) is determined by the majority votes from the decision trees in the forest. In this flood susceptibility assessment, the RF algorithm is trained using flood inventory points (as the target variable) and flood conditioning factors (as feature or predictor variables).

The entire RF classification process is conducted on Jupyter Notebook IDE using Python language. The

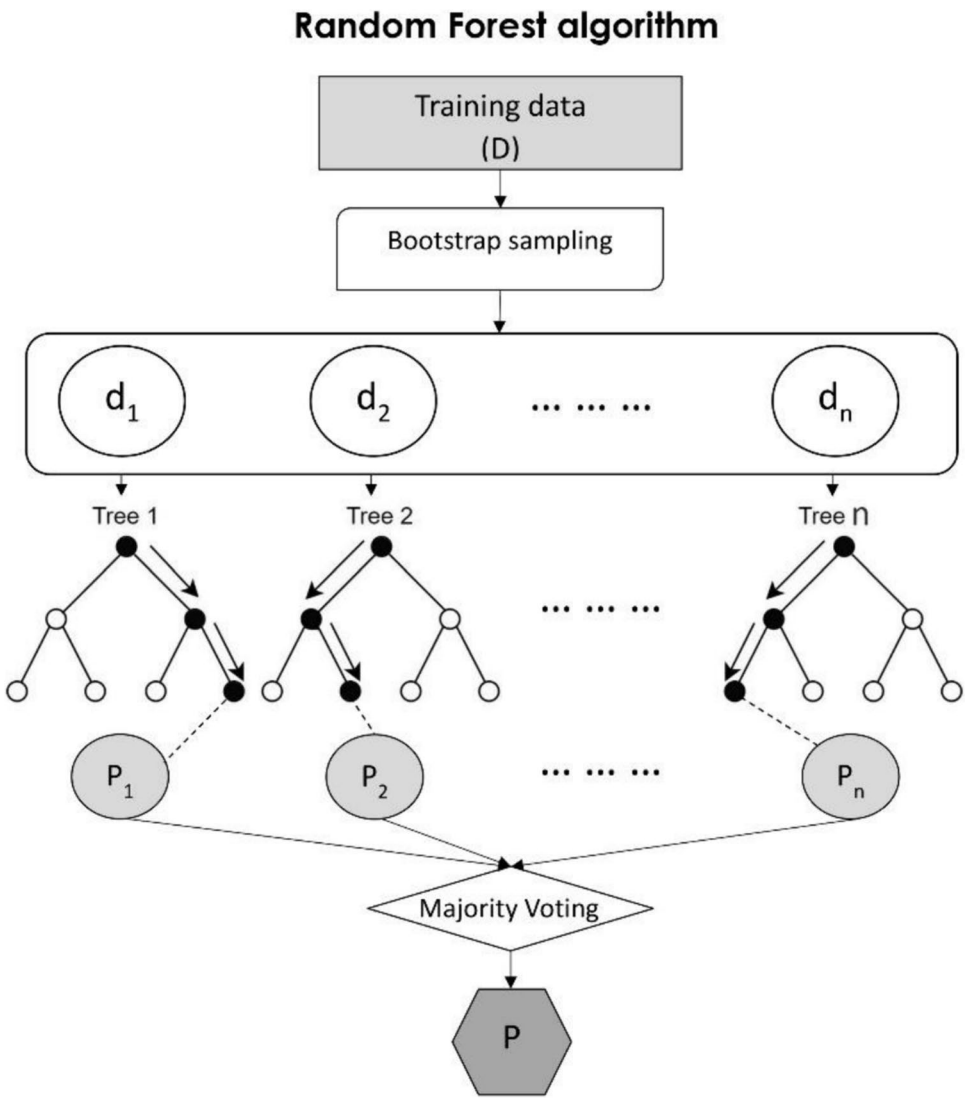

Fig. 5 The classification process of RF algorithm

RandomForestClassifier function is imported from the SciKit Learn package to classify flood susceptibility. To do so, initially, the entire flood inventory dataset (500 inventory points) is split into a 70:30 ratio using the *train_test_split* function. Later, the only training set (70% of the total dataset) is fitted into the RF model. After hyper-parameter tuning, the best optimum parameters are found with *n_estimators = 700, criterion = gini, and max_features = log2*. Following that, the contributions of various flood conditioning factors on flood occurrence are assessed using the *feature_importances_* function. Later, the probability of flood occurrence is estimated by generating a flood susceptibility index (FSI) for each pixel using the *predict_proba* function. The value of FSI ranges from 0 to 1, representing the gradual range of probabilities of flood occurrence. The FSI value of a pixel close to 0 indicates very low susceptibility, and close to 1 indicates very high susceptibility. Equal interval method is employed in ArcGIS Pro 2.8 to categorize FSI into five different classes, namely very low (0–0.2), low (0.2–0.4), moderate (0.4–0.6), high (0.6–0.8), and very high (0.8–1.0). Following that, the area percentage under each category of flood susceptibility is calculated and tabulated.

Social vulnerability assessment

This study adopts a widely recognized formula developed by Cutter et al. (2003) to map social vulnerability by utilizing a very popular data dimension reduction technique known as PCA. To do so, this study initially selects ten variables to create a SoVI of Harris County at the census tract level based on a comprehensive examination of the literature on SoVI assessment (Cutter et al. 2003; Dey et al. 2023; Rufat et al. 2015; Shao et al. 2020). These

variables include population density, housing density, percentage of persons living below the poverty, unemployment rate, per capita income, percentage of persons aged 65 or older, percentage of persons aged 17 or younger, percentage of disabled population, households with no vehicles, and percentages of racial minorities.

First, all the variables are standardized using Z-scores (Eq. 3),

$$z = \frac{(x - \mu)}{\sigma} \tag{3}$$

where z is the normalized value, x is the value of each data variable, μ is the mean value of the variable, and σ is the standard deviation of the variable. After calculating the Z score, PCA is conducted in SPSS using varimax rotation (100 iterations). Later, three-factor components are selected based on eigenvalues (greater than 1.00) (Table 2). These factors are named based on their dominant variables, only those with factor loading values above 0.600 or below -0.600 are considered for naming. The three dominant factors, namely poverty and inaccessibility, income and dependent populations, and population density explain the variance of 36.88%, 22.49%, and 16.39% of the total data, respectively. Cardinality is assigned to each factor depending on their dominant variable's contribution to overall social vulnerability.

In terms of quantifying social vulnerability, there is no established theory guiding the weighting of different factors in a SoVI. Most studies assume all constituent factors are equal in their contributions to the overall social vulnerability and thus assign equal weight to each factor (Rufat et al. 2015). The social vulnerability outcome on the other hand is sensitive to different weighting schemes (e.g., weighted vs. non-weighted) (Reckien 2018). Some studies use the percentage of variance explained to weight each selected constituent factor (de Sherbinin and Bardy 2015; Dey et al. 2023; Shao et al. 2020). The primary objective of mapping the spatial distribution of social vulnerability is to highlight places and areas that display the highest vulnerability across a region. The effort can then inform and guide policymakers to effectively and efficiently allocate resources to the places/areas where they are needed the most. Spatial variations illustrate the disparities of variables/factors across space as well as the extent to which the variable/factor varies. Using ratios of variance explained to the total variance explained by the selected factors to represent contributions to the overall social vulnerability outcome can amplify factors that display the most variances explained in a spatial context. Adopting this reasoning, weights in this study are determined by computing the ratio of the variance explained by each individual factor to the total variance explained by all selected component factors. Next, all these factors are summed up after being multiplied with the corresponding weight to create SoVI (Eq. 4).

$$SoVI = 0.481F1 + 0.296F2 + 0.216F3 \tag{4}$$

After calculating SoVI scores for each census tract, they are joined with Harris County's census tract shapefile using their unique FIPS code. The SoVI scores range from -1.52 to 3.35. The scores are categorized into five different groups including very low (-1.52- -0.58), low (-0.59—-0.11), moderate (-0.12—0.35), high (0.36 -1.01), and very high (1.02–3.35) using natural breaks classification in ArcGIS Pro 2.8.. The final step involves preparing a choropleth map of social vulnerability based on the classified SoVI scores using a diverging color palette.

Flood risk assessment

The term risk comprises two major aspects: hazard susceptibility and social vulnerability meaning the potentiality of

Table 2 PCA component summary of SoVI variables

Component	Cardinality	Name	% Variance Explained	Dominant variables	Factor loading	Weight
F1	+	Poverty and inaccessibility	36.885	Z_Poverty	0.626	0.481
				Z_Unemployed	0.721	
				Z_Disability	0.792	
				Z_NOVehicle	0.759	
F2	+	Income and Dependents populations	22.491	Z_Income	-0.726	0.296
				Z_Age65	-0.789	
				Z_Age17	0.761	
				Z_Minority	0.702	
F3	+	Population's density	16.393	Z_POPDensity	0.930	0.216
				Z_Housing_Density	0.968	

human life and infrastructure damage when exposed to hazards (Merz et al. 2007; Vojtek and Vojteková 2016). In this study, the FRM of Harris County is created by combining FSM and SoVM. To ensure the consistency and alignment of the value range between FRM and SoVM, initially, SoVI is normalized into the 0–1 range by applying the Min–Max normalization formula. The Min–Max normalization formula is given below in Eq. 5.

$$x\prime = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{5}$$

where x' is the normalized value, x is the value of SoVI of each census tract, max(x) is the maximum value of SoVI, and min(x) is the minimum value of SoVI among all census tracts.

Later the polygon feature of the normalized SoVI is converted into a raster file to ensure the same spatial resolution and geographic coordinate system as the FSM. Normalized SoVM and FSM are then multiplied to create a flood risk index based on the following formula (Eq. 6).

$$FRI = SoVI \times FSI \tag{6}$$

where, FRI represents the flood risk index, SoVI social vulnerability index, and FSI flood susceptibility index. Hence, if the value for a specific region is high in both FSM and SoVM, the corresponding value of FRM would be high as well. The FRI scores range from 0 to 0.79, and are classified into five major categories including very low (0–0.07), low (0.08–0.12), moderate (0.13–0.18), high (0.19–0.27), and very high (0.28–0.79), using natural breaks classification in ArcGIS Pro 2.8. Natural breaks classification is used in classifying both SoVI and FRI because the values of SoVI and FRI are not evenly distributed and contain big leaps. Natural breaks classification is the most suitable when data value is not evenly distributed and there are big leaps in the data (Ayalew and Yamagishi 2005). Later the percentage of areas for each category of flood risk is quantified.

Results

Contributions of flood conditioning factors to flooding in Harris County

The pie chart (Fig. 6) illustrates the contribution of different flood conditioning factors (in percentage) to flooding in the study area. The findings reveal that elevation has the highest impact on flooding, accounting for a 15% contribution. Furthermore, distance to river, precipitation, and LULC are also significant factors with contributions of 13%, 12%, and 12%, respectively.

Other factors including slope, NDVI, drainage density, aspect, curvature, TWI, SPI, and soil texture also contribute to the flooding in Harris County, although to a lesser extent.

To further validate this analysis, this study compared the result of the RF model with that of the XGBoost model. XGBoost, a high-performing ML model, showed a highly similar pattern achieved by the RF model in ranking the impacts of flood conditioning factors in Romania (Abedi et al. 2022). In this study, the XGBoost identified elevation, distance to river, and LULC as dominant factors of flood occurrence which are very similar to what the RF model revealed (see Supplementary Materials).

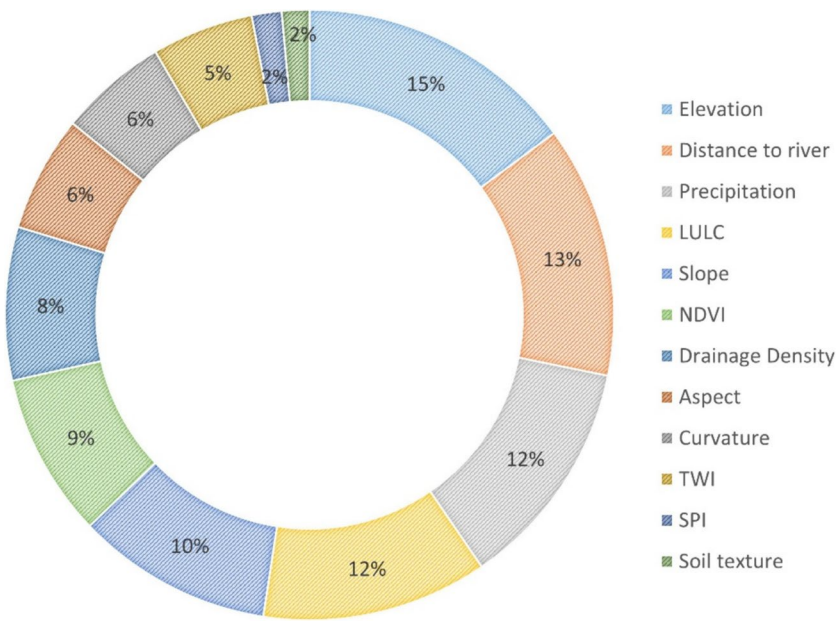

Fig. 6 Impact of flood conditioning factors on flood occurrence in Harris County

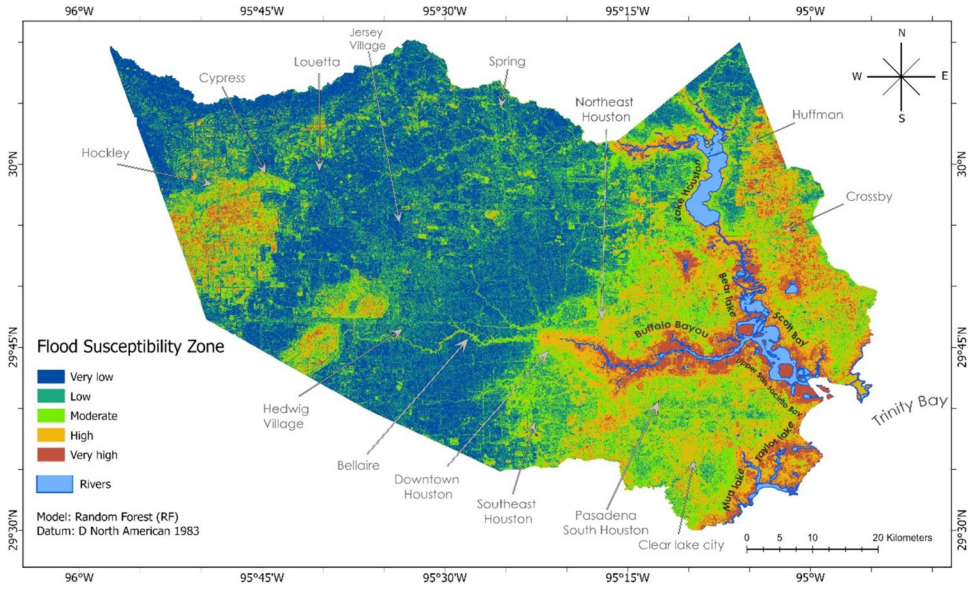

Fig. 7 Flood susceptibility map of Harris County

Flood susceptibility map of Harris County

Figure 7 shows the distribution of different flood susceptible zones throughout Harris County. According to the susceptibility analysis, about 9.06% (388.01km^2) of Harris County is very highly susceptible to flooding, while 17.83% (763.55 km^2) is classified as highly susceptible to flooding. High and very high flood susceptible areas are mainly located in the eastern part. Specifically, most of them are near open water bodies such as Buffalo Bayou River, Scott Bay, Upper San Jacinto Bay, Bear Lake, Mud Lake, and Taylor Lake, and some croplands in Crosby and Huffman region (located at northeastern region of Harris County) and Hockley regions (located at the western side of Harris County) (Fig. 7).

The analysis also reveals that 22.81% (975.81km^2) of Harris County has a moderate susceptibility, while 17.79% (761.76km^2) and 32.52% (1392.44km^2) fall into the low and very low flood susceptibility categories, respectively. The moderate flood-susceptible zone encompasses most of Pasadena, South Houston, and Clear City Lake area, whereas most of Harris County such as Bellaire, Hedwig Village, Jersey Village, Louetta, and Aldine areas were classified as low or very low flood susceptibility zones.

Social vulnerability map of Harris County

Table 3 presents the results of the social vulnerability analysis including the classification of social vulnerability in Harris County, along with the number of census tracts, their estimated area, percentage of area, total population, and percentage of population associated with each category. Table 3 indicates that 37 census tracts are classified as very highly socially vulnerable, accounting for 1.45% (64.15km^2) of the total area. These census tracts consist of 3.52% of the total population. Furthermore, an area of 456.51km^2, which represents 10.34% of the total area, is classified as having a high social vulnerability, consisting of 180 census tracts. A significant percentage of the population (19.53%) lives in these census tracts. As Fig. 8 shows, the areas in Harris County that are considered to be very high and high social vulnerability tend to be concentrated in the downtown areas, such as northeast Houston, Greater Fifth Ward, and Frenchtown, as well as some parts of southeast Houston and Chinatown.

A total of 225 census tracts, covering 27.35% (1207.91km^2) of the total area, are identified as having moderate social vulnerability. This category also includes 30.59% of the total population. Moreover, 31.99% of the total area, estimated as 1412.98km^2 is classified as having a

Table 3 Results of social vulnerability analysis

Vulnerability categories	No. of Census tract	Estimated Area (km^2)	% Of Area	Total Population	% Of Population
Very low	150	1274.89	28.87	875,234	19.01
Low	192	1412.98	31.99	1,258,022	27.33
Moderate	225	1207.91	27.35	1,407,999	30.59
High	180	456.51	10.34	899,161	19.53
Very high	37	64.15	1.45	162,107	3.52

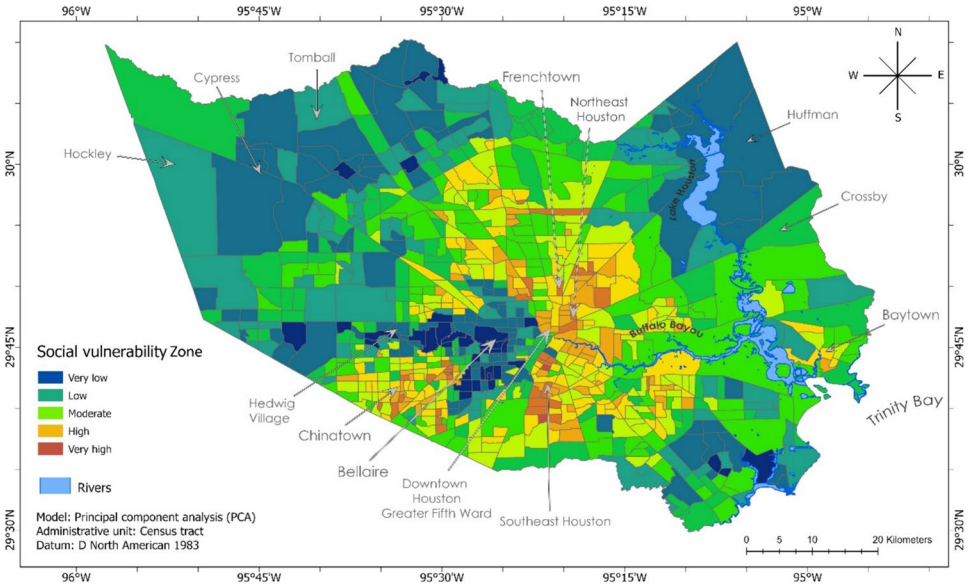

Fig. 8 Social vulnerability map of Harris County

low social vulnerability, where 27.33% of the total population resides. Among all census tracts, 150 of them fall into the very low social vulnerability category, covering 28.87% (1274.89 km^2) of the total area. The population in these census tracts is 875,234, which makes up 19.01% of the total population. Overall, areas outside the downtown area are characterized by moderate, low, and very low social vulnerability, as shown in Fig. 8. Specifically, most of the census tracts close to Bellaire and Hedwig Village fall into this very low socially vulnerable category.

Flood risk map of Harris County

Figure 9 demonstrates the spatial distribution of flood risk zones throughout Harris County. Regions close to Buffalo Bayou River are at very high risk of flooding. Specifically, the areas close to Galena Park, Denver Harbor, Greater Fifth Ward, Second Ward, and some parts of northern downtown Houston are categorized as being at high or very high risk of flooding. Additionally, a few regions in Bay Town are also classified as high or very high flood risk zones. The flood risk analysis reveals that 5.59% (239.52km^2) of the entire area is at very high risk of flooding and 12.25% (524.59km^2) of the total area is at high risk of flooding in Harris County.

On the other hand, 945.34 km^2 (equivalent to 22.07% of the total area), 1220.92 km^2 (equivalent to 28.51% of the total area), and 1352.11 km^2 (equivalent to 31.57% of the total area) of Harris County are categorized as moderate, low, and very low flood risk zones, respectively.

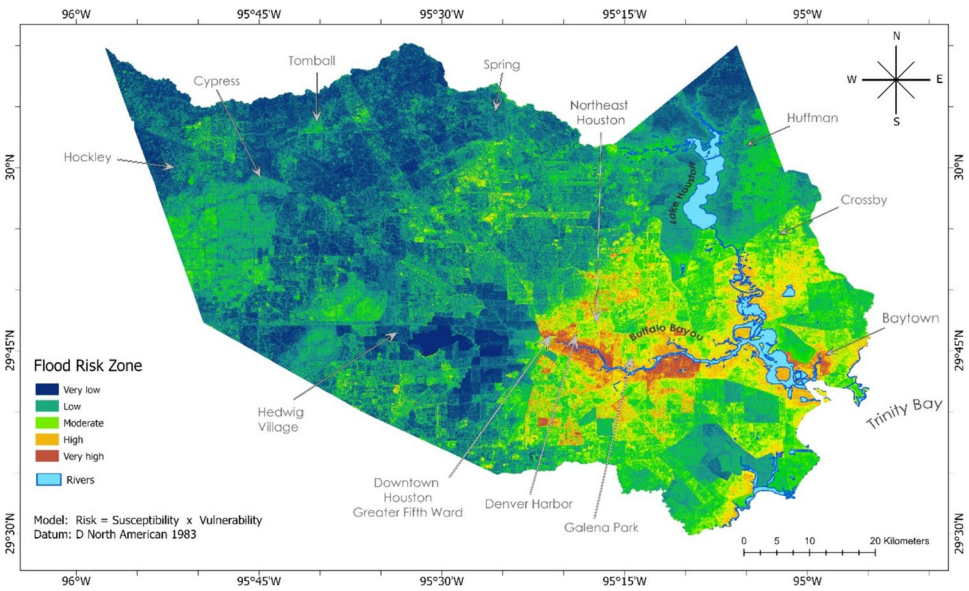

Fig. 9 Flood risk map of Harris County

Validation of the RF model

The ROC-AUC curve is a widely used tool in ML to evaluate prediction model accuracy due to its easy-to-understand and clear depiction of accuracy (Youssef et al. 2022). In this study, the performance of the RF model is evaluated by the ROC-AUC curve. A maximum AUC value (close to 1) indicates the very good prediction capability of an ML model. Figure 10 shows that the AUC-ROC value for the generated RF model is 0.92, indicating very good prediction accuracy.

Moreover, the RF model achieves an overall accuracy of 0.83, a precision score of 0.82, a recall score of 0.84, and an F1 score of 0.83 on the test dataset.

Discussion

Effective flood management necessitates continuous and rigorous flood risk assessment, monitoring, and forecasting. An FRM generated from comprehensive flood risk assessment can be used as an effective and powerful tool in flood risk management. This study combines two approaches: an FSM based on the physical aspects of flooding and an SoVM that is generated from social factors. By integrating these two maps, the study aims to create a comprehensive FRM for Harris County to achieve better flood risk management. This study identified major contributing factors of flooding in Harris County and the spatial distribution of flood risk zones. Most importantly, this study reveals high flood risk regions that are overlooked by FEMA's 100-year floodplain map.

In the United States, the FEMA 100-year floodplain, also known as Special Flood Hazard Areas (SFHAs), denotes areas with a 1% annual chance of being flooded, guiding federal flood insurance and mitigation (Highfield et al. 2013). It is a key indicator for determining flood risk and flood insurance in the U.S., and helps local decision-makers with safety assessments for further development and building infrastructures (Highfield et al. 2013). The boundary of SFHAs is usually prepared by calculating the 100-year return periods of flood water discharge by analyzing historical gauge station data (e.g. stream flow discharge and runoff) on a hydraulic model (Vojtek and Vojteková 2016). Next, base flood elevation (the water surface elevations of the 1% annual chance flood) is generated and overlaid onto topographic data such as DEM to delineate a 100-year flood plain (Crowell et al. 2010).

One major limitation of the FEMA 100-year floodplain is its lack of comprehensiveness, resulting in inconsistencies with real flood damages and excluding areas with substantial populations vulnerable to flooding (Huang and Wang 2020). The probable reason behind this is the FEMA flood risk assessment has been conducted on a large geographic scale and modeling which sometimes overlooks populations at risk (Flores et al. 2023; Highfield et al. 2013). In addition, FEMA 100-year flood map only focuses on physical factors (topography, rainfall, and river discharge) but disregards social factors. Political factors can also affect the remapping process, resulting in less effective responses from individuals and groups, according to studies. This emphasizes the need for unbiased and thorough flood risk mapping to reduce flooding impacts. (Pralle 2019). Furthermore, FEMA's 100-year floodplain map is binary, indicating whether a region is in or beyond the floodplain (Fig. 1), which might have substantial behavioral ramifications. Residents within FEMA's floodplain are more likely to buy flood insurance voluntarily (Shao et al. 2017), while those outsiders might overlook flood consequences (Shao et al. 2019). Research

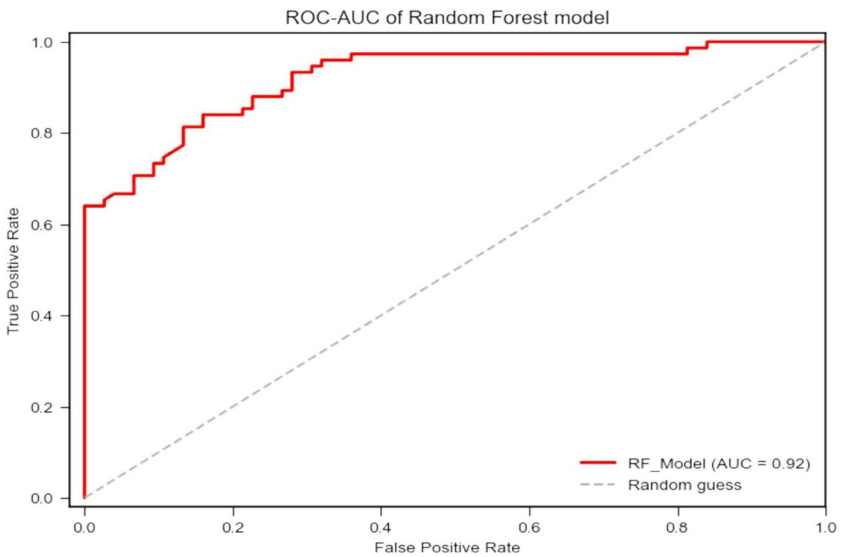

Fig. 10 ROC-AUC curve of RF model

demonstrates that including social vulnerability in flood control project assessment and planning can reduce the negative effects of floods on social stability (Greene et al. 2015).

A comprehensive study should include multiple components that strengthen its findings and boost its credibility to aid long-term planning and policymaking. By incorporating a wide array of influential elements, we were able to develop a more comprehensive assessment and FRM, categorizing flood risk into five distinct levels, spanning from a very low to very high levels (Fig. 9). In addition, FRM is a more dependable tool for assessing flood risk because it considers both physical and social factors. It considers flood risk as a product of both hazards (the physical effects of actual flooding) and vulnerability (potentially exposed populations and social infrastructure) (Vojtek and Vojteková 2016). Furthermore, upon comparing the two maps, it becomes evident that there are areas of high susceptibility, in the vicinity of Hockley, Huffman, and Crosby regions, that are overlapped with the FEMA 100-year floodplain (Fig. 11a). The contrast also underscores FEMA's tendency to underestimate the level of risk in the vicinity of Bays and overestimate the risk level near Buffalo Bayou River, a characteristic that is more accurately portrayed by the FRM (Fig. 11b).

Table 4 displays regions that have been categorized as very high or high in terms of FSM and FRM but are situated beyond the FEMA 100-year floodplain boundary. The table shows that the FEMA 100-year floodplain map underestimated 62.6% of the very highly susceptible areas and 77.9% of the highly susceptible areas in FSM. In addition, 76.7% of the regions classified as very high risk and 81.8% of the areas classified as highly risky in FRM are located outside the FEMA 100-year floodplain map. This finding is highly pertinent to the research conducted by Flores et al. (2023), which revealed that around 1 million individuals, accounting for 16% of the overall population, reside in areas of Greater Houston that are highly susceptible to flooding. However, these areas were not included in the FEMA 100-year floodplain map.

Table 4 Percentage of underestimated flood risk areas by FEMA 100-year floodplain map

Risk level	Underestimated area in FSM (km^2)	% Of area in FSM	Underestimated areas in FRM (km^2)	% Of area in FRM
Very high	242.8	62.6	183.3	76.7
High	594.7	77.9	428.6	81.8

The RF model also demonstrates that Harris County is prone to flooding near Galveston and Trinity Bay, and it identifies many contributing factors. Flood susceptibility is significantly affected by elevation, river proximity, precipitation, and LULC (Fig. 6). Numerous flood risk management studies have confirmed the significance of those factors in flood susceptibility (Rahmati et al. 2016; Lee et al. 2017). Figure 4a depicts how gravitational forces lead water to flow towards Galveston and Trinity Bay as elevation decreases. The high precipitation in these bay areas (Fig. 4g) and proximity to river channels (Fig. 4i) increase the risk of flooding. Heavy precipitation combined with proximity to a river or shore can cause compound flooding, which is often more severe than individual impacts (Bevacqua et al. 2019). This underscores the necessity of understanding the connections between flood risk management strategy components. The built-up land in these areas, with impermeable surfaces that limit water infiltration and accelerate surface runoff, makes them more vulnerable to flooding (Fig. 4h). Thus, Harris County's Galveston and Trinity Bay regions are especially prone to flooding due to their low elevation, proximity to rivers, high precipitation, and urbanization.

Note that high flood susceptibility does not always imply high flood risk. The interaction of flood susceptibility and social vulnerability generates flood risk. Hockley in western

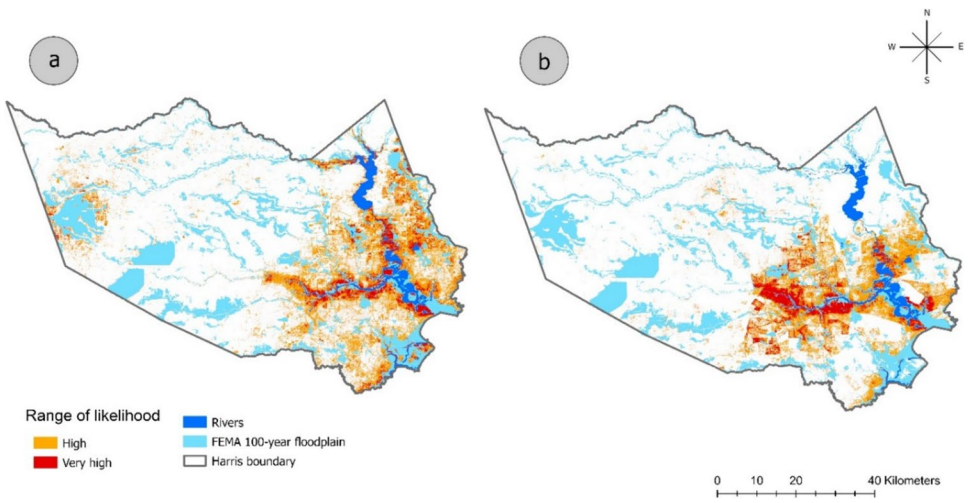

Fig. 11 Comparison between FEMA 100-year floodplain map with FSM and FRM (**a**) Underestimated flood susceptible areas by FEMA 100-year floodplain map (**b**) Underestimated flood risk areas by FEMA 100-year floodplain map; Red color represents underestimated very high categorized zones, Orange color represents underestimated high categorized zones

Harris County is very highly susceptible to floods (Fig. 7), but its low flood risk (Fig. 9) is linked to low social vulnerability (Fig. 8), which is largely attributable to a small population. Hoffman has high flood susceptibility (Fig. 7), but low flood risk (Fig. 9). Our findings revealed that some places might be significantly vulnerable and resilient at the same time due to some other factors and such findings are consistent with others (Haque et al. 2022; Shah et al. 2018) Despite significant socioeconomic vulnerability (Fig. 8), the areas surrounding northeastern Houston and Chinatown pose negligible flood risk (Fig. 9). Figures 7 and 8 illustrate that regions near the Buffalo Bayou riverbank are both flood-susceptible and socially vulnerable, resulting in high flood-risk zones (Fig. 9).

In summary, this study aims to complement the FEMA 100-year floodplain map for better flood risk management. The utilization of ML algorithms in the creation of flood susceptibility maps signifies a notable progression in the field of disaster management, providing enhanced effectiveness and accuracy in forecasting regions susceptible to flooding. Through the utilization of data from many sources, ML algorithms can effectively manage the intricate connections between flood conditioning factors and flood occurrences and determine the most influential aspects. The combination of social vulnerability and flood susceptibility within a unified framework for flood risk assessment provides a more holistic comprehension of flood hazards, encompassing both the probability of flooding occurrences and the possible consequences for impacted areas. The present methodology for creating an FRM can be extended to encompass the generation of real-time and high-resolution FRM for additional cities globally that are susceptible to flooding. This can be achieved by using current census data and pertinent data that encompass a range of flood conditioning factors. Furthermore, this methodology also facilitates the identification of the key elements that contribute to flood events, as well as the delineation of flood-prone areas as an outcome. The utilization of integrated assessment can facilitate the development of focused and fair policy interventions that prioritize the distribution of resources to the most susceptible people, thereby enhancing the overall resilience of the community. This research offers significant insights into the communication of flood risks and the development of sustainable solutions for long-term mitigation.

Conclusion

A comprehensive flood risk assessment is vital for sustainable urban development. In this study, we use advanced ML techniques, specifically the RF algorithm, to construct an FSM and integrate it with a SoVM to produce a comprehensive FRM for Harris County. The analysis identifies key factors like elevation, river proximity, precipitation, and LULC as major flood drivers. This study also revealed that almost 18% of Harris County is at very high or high flood risk and most of these regions are overlooked by the FEMA 100-year floodplain map. The comprehensiveness of our FRM offers a deeper insight into flood risks than FEMA's 100-year floodplain map and can guide disaster management efforts more effectively.

Nonetheless, this study does have certain limitations that warrant consideration in future research. These limitations include a lack of validation for the FRM and the arbitrary decisions on the selection and weighting of social vulnerability variables due to the fact that no consensus exists. Future research should validate assessments with historical data, explore different ML algorithms, and involve local stakeholders for more context-specific approaches. The study's focus on Harris County also limits its geographic scope, suggesting a need for future research on larger areas, like the Southeastern US, for broader understanding. Despite these limitations, the study underscores the value of combining flood susceptibility and social vulnerability in flood risk management and urban planning.

Acknowledgements Funding for this project was provided by the National Oceanic and Atmospheric Administration (NOAA), awarded to the Cooperative Institute for Research on Hydrology (CIROH) through the NOAA Cooperative Agreement with The University of Alabama, NA22NWS4320003. The authors declare they have no financial interests. On behalf of all authors, the corresponding author states that there is no conflict of interest. The research was conducted in compliance with ethical standards. All data come from publicly available sources. Thus, no ethical approval or informed consent was needed during the data compilation process.

Data Availability Data will be available based on reasonable requests.

Declarations

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Abedi R, Costache R, Shafizadeh-Moghadam H, Pham QB (2022) Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. Geocarto Int 37(19):5479–5496. https://doi.org/10.1080/10106049.2021.1920636

Amare S, Langendoen E, Keesstra S, Ploeg MVD, Gelagay H, Lemma H, van der Zee SE (2021) Susceptibility to gully erosion: applying random forest (RF) and frequency ratio (FR) approaches to a small catchment in Ethiopia. Water 13(2):216. https://doi.org/10.3390/w13020216

Andaryani S, Nourani V, Haghighi AT, Keesstra S (2021) Integration of hard and soft supervised machine learning for flood susceptibility mapping. J Environ Manage 291:112731. https://doi.org/10.1016/j.jenvman.2021.112731

Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the

Kakuda-Yahiko mountains central Japan. Geomorphology 65(1–2):15–31
Bevacqua E, Maraun D, Vousdoukas MI, Voukouvalas E, Vrac M, Mentaschi L, Widmann M (2019) Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change. Sci Adv 5(9):eaaw5531. https://doi.org/10.1126/sciadv.aaw5531
Breiman L (2001) Random Forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324
Chakraborty J, Collins TW, Grineski SE (2019) Exploring the Environmental Justice Implications of Hurricane Harvey Flooding in Greater Houston, Texas. Am J Public Health 109:244–250. https://doi.org/10.2105/AJPH.2018.304846
Courtial A, Touya G, Zhang X (2022) Constraint-Based Evaluation of Map Images Generalized by Deep Learning. J Geovisualiz Spat Anal 6:13. https://doi.org/10.1007/s41651-022-00104-2
Crowell M, Coulton K, Johnson C, Westcott J, Bellomo D, Edelman S, Hirsch E (2010) An estimate of the US population living in 100-year coastal flood hazard areas. J Coastal Res 26(2):201–211. https://doi.org/10.2112/JCOASTRES-D-09-00076.1
Cutter SL, Boruff BJ, Shirley WL (2003) Social vulnerability to environmental hazards. Soc Sci Q 84(2):242–261
Das S (2020) Flood susceptibility mapping of the Western Ghat coastal belt using multi-source geospatial data and analytical hierarchy process (AHP). Remote Sens Appl : Soc Environ 20:100379. https://doi.org/10.1016/j.rsase.2020.100379
Dey H, Shao W, Moradkhani H et al (2024) Urban flood susceptibility mapping using frequency ratio and multiple decision tree-based machine learning models. Nat Hazards 1-29. https://doi.org/10.1007/s11069-024-06609-x
Dey H, Shao W, Pan S, Tian H (2023) The spatiotemporal patterns of community vulnerability in the U.S. Mobile Bay from 2000 - 2020. Appl Spat Anal Pol 1–22. https://doi.org/10.1007/s12061-023-09549-4
Dou P, Zeng C (2020) Hyperspectral image classification using feature relations map learning. Remote Sens 12(18):2956. https://doi.org/10.3390/rs12182956
Dou P, Shen H, Li Z, Guan X (2021) Time series remote sensing image classification framework using combination of deep learning and multiple classifiers system. Int J Appl Earth Obs Geoinf 103:102477. https://doi.org/10.1016/j.jag.2021.102477
Dou P, Huang C, Han W, Hou J, Zhang Y, Gu J (2024) Remote sensing image classification using an ensemble framework without multiple classifiers. ISPRS J Photogramm Remote Sens 208:190–209. https://doi.org/10.1016/j.isprsjprs.2023.12.012
Du P, Bai X, Tan K et al (2020) Advances of Four Machine Learning Methods for Spatial Data Handling: a Review. J Geovisualiz Spat Anal 4:13. https://doi.org/10.1007/s41651-020-00048-5
Eiser JR, Bostrom A, Burton I et al (2012) Risk interpretation and action: A conceptual framework for responses to natural hazards. Intl J Disaster Risk Reduct 1:5–16. https://doi.org/10.1016/j.ijdrr.2012.05.002
Elmahdy SI, Mohamed MM, Ali TA, Abdalla JED, Abouleish M (2022) Land subsidence and sinkholes susceptibility mapping and analysis using random forest and frequency ratio models in Al Ain. UAE Geocarto Intl 37(1):315–331. https://doi.org/10.1080/10106049.2020.1716398
Farhadi H, Najafzadeh M (2021) Flood risk mapping by remote sensing data and random forest technique. Water 13(21):3115. https://doi.org/10.3390/w13213115
Flores AB, Collins TW, Grineski SE, Amodeo M, Porter JR, Sampson CC, Wing O (2023) Federally overlooked flood risk inequities in Houston, Texas: Novel insights based on dasymetric mapping and state-of-the-Art flood modeling. Ann Am Assoc Geogr 113(1):240–260. https://doi.org/10.1080/24694452.2022.2085656
Frigerio I, De Amicis M (2016) Mapping social vulnerability to natural hazards in Italy: A suitable tool for risk mitigation strategies. Environ Sci Policy 63:187–196. https://doi.org/10.1016/j.envsci.2016.06.001
Géron A (2022) Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.".
Godfroy M, Jonkman B (2017) Fatalities due to hurricane Harvey (2017). 4TU.ResearchData. Dataset. https://doi.org/10.4121/uuid:95690fdd-b13f-4bf9-a28d-c9b924696a96
Greene G, Paranjothy S, Palmer SR (2015) Resilience and vulnerability to the psychological harm from flooding: The role of social cohesion. Am J Public Health 105(9):1792–1795
Han J, Kim J, Park S, Son S, Ryu M (2020) Seismic vulnerability assessment and mapping of Gyeongju, South Korea using frequency ratio, decision tree, and random forest. Sustainability 12(18):7787. https://doi.org/10.3390/su12187787
Haque MM, Islam S, Sikder MB, Islam MS (2022) Community flood resilience assessment in Jamuna floodplain: A case study in Jamalpur District Bangladesh. Intl J Disaster Risk Reduct 72:102861. https://doi.org/10.1016/j.ijdrr.2022.102861
Haque MM, Islam S, Sikder MB, Islam MS, Tabassum A (2023) Assessment of flood vulnerability in Jamuna floodplain: a case study in Jamalpur district. Bangladesh Nat Hazards 116(1):341–363. https://doi.org/10.1007/s11069-022-05677-1
Hasan MA, Mia MB, Khan MR, Alam MJ, Chowdury T, Al Amin M, Ahmed KMU (2023) Temporal changes in land cover, land surface temperature, soil moisture, and evapotranspiration using remote sensing techniques—a case study of Kutupalong Rohingya Refugee Camp in Bangladesh. J Geovisualiz Spat Anal 7(1):11. https://doi.org/10.1007/s41651-023-00140-6
Hengl T (2018) Soil texture classes (USDA system) for 6 soil depths (0, 10, 30, 60, 100 and 200 cm) at 250 m (Version v02). Zenodo. https://doi.org/10.5281/zenodo.1475451
Highfield WE, Norman SA, Brody SD (2013) Examining the 100-year floodplain as a metric of risk, loss, and household adjustment. Risk Anal : an Intl J 33(2):186–191. https://doi.org/10.1111/j.1539-6924.2012.01840.x
Huang X, Wang C (2020) Estimates of exposure to the 100-year floods in the conterminous United States using national building footprints. Intl J Disaster Risk Reduct 50:101731. https://doi.org/10.1016/j.ijdrr.2020.101731
Iban MC, Sekertekin A (2022) Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces. Turkey Ecol Inform 69:101647. https://doi.org/10.1016/j.ecoinf.2022.101647
Islam ARMT, Talukdar S, Mahato S et al (2021) Flood susceptibility modelling using advanced ensemble machine learning models. Geosci Front 12(3):101075. https://doi.org/10.1016/j.gsf.2020.09.006
Joyce KE, Belliss SE, Samsonov SV, McNeill SJ, Glassey PJ (2009) A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. Prog Phys Geogr 33(2):183–207. https://doi.org/10.1177/0309133309339563
Kadri CB, Nasrallah Y (2023) GIS-based AHP technique for assessment of desertification in western highlands of Algeria. J Geovisualiz Spat Anal 7(2):18. https://doi.org/10.1007/s41651-023-00147-z
Khajehei S, Ahmadalipour A, Shao W, Moradkhani H (2020) A place-based assessment of flash flood hazard and vulnerability in the contiguous United States. Sci Rep 10:448. https://doi.org/10.1038/s41598-019-57349-z
Kia MB, Pirasteh S, Pradhan B, Mahmud AR, Sulaiman WNA, Moradi A (2012) An artificial neural network model for flood simulation using GIS: Johor River Basin. Malaysia Environ Earth Sci 67(1):251–264. https://doi.org/10.1007/s12665-011-1504-z

Lee S, Kim JC, Jung HS, Lee MJ, Lee S (2017) Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. Geomat Nat Haz Risk 8(2):1185–1203. https://doi.org/10.1080/19475705.2017.1308971

Merghadi A, Yunus AP, Dou J et al (2020) Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth Sci Rev 207:103225. https://doi.org/10.1016/j.earscirev.2020.103225

Merz B, Thieken AH, Gocht M (2007) Flood risk mapping at the local scale: concepts and challenges. Flood risk management in Europe: innovation in policy and practice 231–251. https://doi.org/10.1007/978-1-4020-4200-3_13

Mosavi A, Ozturk P, Chau KW (2018) Flood prediction using machine learning models: Literature review. Water 10(11):1536. https://doi.org/10.3390/w10111536

Mukherjee F, Singh D (2020) Detecting flood prone areas in Harris County: a GIS based analysis. GeoJournal 85(3):647–663. https://doi.org/10.1007/s10708-019-09984-2

Nachappa TG, Piralilou ST, Gholamnia K, Ghorbanzadeh O, Rahmati O, Blaschke T (2020) Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. J Hydrol 590:125275. https://doi.org/10.1016/j.jhydrol.2020.125275

Park K, Lee MH (2019) The development and application of the urban flood risk assessment model for reflecting upon urban planning elements. Water 11(5):920. https://doi.org/10.3390/w11050920

Pekel JF, Cottam A, Gorelick N, Belward AS (2016) High-resolution mapping of global surface water and its long-term changes. Nature 540(7633):418–422. https://doi.org/10.1038/nature20584

Pralle S (2019) Drawing lines: FEMA and the politics of mapping flood zones. Clim Change 152(2):227–237. https://doi.org/10.1007/s10584-018-2287-y

Pulcinella JA, Winguth AM, Allen DJ, DasaGangadhar N (2019) Analysis of flood vulnerability and transit availability with a changing climate in Harris County. Texas Transp Res Rec 2673(6):258–266. https://doi.org/10.1177/0361198119839346

Qin H, Wang J, Mao X et al (2024) An Improved Faster R-CNN Method for Landslide Detection in Remote Sensing Images. Journal of Geovisualization and Spatial Analysis 8:2. https://doi.org/10.1007/s41651-023-00163-z

Rahman M, Ningsheng C, Islam MM, Dewan A, Iqbal J, Washakh RMA, Shufeng T (2019) Flood susceptibility assessment in Bangladesh using machine learning and multi-criteria decision analysis. Earth Syst Environ 3(3):585–601. https://doi.org/10.1007/s41748-019-00123-y

Rahman M, Ningsheng C, Mahmud GI et al (2021) Flooding and its relationship with land cover change, population growth, and road density. Geosci Front 12(6):101224. https://doi.org/10.1016/j.gsf.2021.101224

Rahmati O, Pourghasemi HR, Zeinivand H (2016) Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province. Iran Geocarto Intl 31(1):42–70. https://doi.org/10.1080/10106049.2015.1041559

Rahmati O, Darabi H, Panahi M et al (2020) Development of novel hybridized models for urban flood susceptibility mapping. Sci Rep 10(1):12937. https://doi.org/10.1038/s41598-020-69703-7

Reckien D (2018) What is in an index? Construction method, data metric, and weighting scheme determine the outcome of composite social vulnerability indices in New York City. Reg Environ Change 18:1439–1451. https://doi.org/10.1007/s10113-017-1273-7

Rufat S, Tate E, Burton CG, Maroof AS (2015) Social vulnerability to floods: Review of case studies and implications for measurement. Intl J Disaster Risk Red 14:470–486. https://doi.org/10.1016/j.ijdrr.2015.09.013

Samanta S, Pal DK, Palsamanta B (2018) Flood susceptibility analysis through remote sensing, GIS and frequency ratio model. Appl Water Sci 8(2):66. https://doi.org/10.1007/s13201-018-0710-1

Sarkar D, Mondal P (2020) Flood vulnerability mapping using frequency ratio (FR) model: a case study on Kulik river basin. Indo-Bangladesh Barind Region Appl Water Sci 10(1):1–13. https://doi.org/10.1007/s13201-019-1102-x

Seydi ST, Kanani-Sadat Y, Hasanlou M, Sahraei R, Chanussot J, Amani M (2022) Comparison of Machine Learning Algorithms for Flood Susceptibility Mapping. Remote Sensing 15(1):192. https://doi.org/10.3390/rs15010192

Shah AA, Ye J, Abid M, Khan J, Amir SM (2018) Flood hazards: household vulnerability and resilience in disaster-prone districts of Khyber Pakhtunkhwa province, Pakistan. Nat Hazards 93:147–165. https://doi.org/10.1007/s11069-018-3293-0

Shahabi H, Shirzadi A, Ghaderi K et al (2020) Flood detection and susceptibility mapping using sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on k-nearest neighbor classifier. Remote Sens 12(2):266. https://doi.org/10.3390/rs12020266

Shao W, Xian S, Lin N, Kunreuther H, Jackson N, Goidel K (2017) Understanding the effects of past flood events, perceived and estimated flood risks on individuals' voluntary flood insurance purchase behaviors. Water Res 108:391–400. https://doi.org/10.1016/j.watres.2016.11.021

Shao W, Feng K, Lin N (2019) Predicting support for flood mitigation based on flood insurance purchase behavior. Environ Res Lett 14(5):054014. https://doi.org/10.1088/1748-9326/ab195a

Shao W, Jackson NP, Ha H, Winemiller T (2020) Assessing community vulnerability to floods and hurricanes along the Gulf Coast of the United States. Disasters 44(3):518–547. https://doi.org/10.1111/disa.12383

de Sherbinin A, Bardy G (2015) Social vulnerability to floods in two coastal megacities: New York City and Mumbai. Vienna yearbook of population research 131–165. https://www.jstor.org/stable/24770028

Sotiropoulou KF, Vavatsikos AP (2023) A decision-making framework for spatial multicriteria suitability analysis using PROMETHEE II and k nearest neighbor machine learning models. J Geovisualiz Spat Anal 7(2):20. https://doi.org/10.1007/s41651-023-00151-3

Tabassum A, Basak R, Shao W, Haque MM, Chowdhury TA, Dey H (2023) Exploring the Relationship Between Land Use Land Cover and Land Surface Temperature: a Case Study in Bangladesh and the Policy Implications for the Global South. J Geovisualiz Spat Anal 7(2):25. https://doi.org/10.1007/s41651-023-00155-z

Tehrany MS, Kumar L (2018) The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. Environ Earth Sci 77:1–24. https://doi.org/10.1007/s12665-018-7667-0

Tehrany MS, Pradhan B, Jebur MN (2015) Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stoch Env Res Risk Assess 29(4):1149–1165. https://doi.org/10.1007/s00477-015-1021-9

Thanh NN, Chotpantarat S, Trung NH, Ngu NH (2022) Mapping groundwater potential zones in Kanchanaburi Province, Thailand by integrating of analytic hierarchy process, frequency ratio, and random forest. Ecol Ind 145:109591. https://doi.org/10.1016/j.ecolind.2022.109591

VanDyke MS, King AJ (2018) Using the CAUSE model to understand public communication about water risks: Perspectives from Texas groundwater district officials on drought and availability. Risk Anal 38(7):1378–1389. https://doi.org/10.1111/risa.12950

VanDyke MS, Armstrong CL, Bareford K (2021) How risk decision-makers interpret and use flood forecast information: Assessing the Mississippi River outlook email product. J Risk Res 24(10):1239–1250. https://doi.org/10.1080/13669877.2020.1819390

Vojtek M, Vojteková J (2016) Flood hazard and flood risk assessment at the local spatial scale: a case study. Geomat Nat Haz Risk 7(6):1973–1992. https://doi.org/10.1080/19475705.2016.1166874

Wong PP, Losada IJ, Gattuso JP et al (2014) Coastal systems and low-lying areas. Clim Change 2104:361–409

Youssef AM, Pradhan B, Dikshit A, Mahdi AM (2022) Comparative study of convolutional neural network (CNN) and support vector machine (SVM) for flood susceptibility mapping: a case study at Ras Gharib, Red Sea, Egypt. Geocarto Intl 1–28. https://doi.org/10.1080/10106049.2022.2046866

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com