Heliyon 10 (2024) e33982

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Forecasting of flash flood susceptibility mapping using random forest regression model and geographic information systems

Mohamed Wahba [a,*], Radwa Essam [b], Mustafa El-Rawy [c,d], Nassir Al-Arifi [e,**], Fathy Abdalla [f], Wael M. Elsadek [g]

[a] Civil Engineering Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt
[b] Mathematics and Engineering Physics Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt
[c] Civil Engineering Department, Faculty of Engineering, Minia University, Minia, 61111, Egypt
[d] Civil Engineering Department, College of Engineering, Shaqra University, Dawadmi, 11911, Saudi Arabia
[e] Chair of Natural Hazards and Mineral Resources, Geology and Geophysics Department, King Saud University, Riyadh, 11451, Saudi Arabia
[f] Deanship of Scientific Research, King Saud University, Riyadh, 145111, Saudi Arabia
[g] Civil Engineering Department, Faculty of Engineering, South Valley University, Qena, 83521, Egypt

ARTICLE INFO

Keywords:
Flash floods
Climate change
Flood susceptibility map
Random forest
ROC

ABSTRACT

Flash floods, rapid and devastating inundations of water, are increasingly linked to the intensifying effects of climate change, posing significant challenges for both vulnerable communities and sustainable environmental management. The primary goal of this research is to investigate and predict a Flood Susceptibility Map (FSM) for the Ibaraki prefecture in Japan. This research utilizes a Random Forest (RF) regression model and GIS, incorporating 11 environmental variables (involving elevation, slope, aspect, distance to stream, distance to river, distance to road, land cover, topographic wetness index, stream power index, and plan and profile curvature), alongside a dataset comprising 224 instances of flooded and non-flooded locations. The data was randomly classified into a 70 % training set for model development, with the remaining 30 % used for model validation through Receiver Operating Characteristics (ROC) curve analysis. The resulting map indicated that approximately two-thirds of the prefecture as exhibiting low to very low flood susceptibility, while approximately one-fifth of the region is categorized as high to very high flood susceptibility. Furthermore, the RF model achieved a noteworthy validation with an area under the ROC curve of 99.56 %. Ultimately, this FSM serves as a crucial tool for policymakers in guiding appropriate spatial planning and flood mitigation strategies.

1. Introduction

Flash floods have emerged as the predominant natural disaster in numerous global regions. These events entail a range of adverse impacts, including disruptions to urban traffic and daily routines, extensive property damage, threats to vital infrastructure, and various forms of pollution [1]. Consequently, the outcomes derived from flood susceptibility maps hold the potential to enhance our understanding of how to effectively mitigate the magnitude of destruction associated with urban flooding. Moreover, the achievement

* Corresponding author.
** Corresponding author.
E-mail addresses: m_wahba@mans.edu.eg (M. Wahba), radwa.essam@mans.edu.eg (R. Essam), mustafa.elrawy@mu.edu.eg (M. El-Rawy), nalarifi@ksu.edu.sa (N. Al-Arifi), fabdalla@ksu.edu.sa (F. Abdalla), wael.elsadek@eng.svu.edu.eg (W.M. Elsadek).

https://doi.org/10.1016/j.heliyon.2024.e33982
Received 19 January 2024; Received in revised form 1 July 2024; Accepted 1 July 2024
Available online 3 July 2024

of effective flood management and prevention necessitates the precise prediction of flood risks [2].

Even today, numerous buildings and infrastructures situated in highly flood-prone areas remain vulnerable to flash floods. This vulnerability persists due to insufficient hazard mapping and a lack of proactive mitigation measures [3,4]. Therefore, it is imperative to ascertain flood-prone regions to mitigate additional devastation and destruction [5,6].

Moreover, the integration of Geographical Information System (GIS) with flood modeling has greatly improved the spatial representation of floods at various return intervals, as exemplified in the research conducted by Refs. [7–10]. Furthermore, the application of advanced GIS technology enables the rapid assessment of flood risks within specific geographic areas [11,12]. It is important to note that the accuracy of input data and the integration of GIS methodologies with machine learning and statistical approaches significantly impact the effectiveness of GIS processing, as highlighted by Ref. [13].

Various methodologies have been utilized to develop spatial flood hazard maps. These methods encompass statistical indices, frequency ratios, Shannon's entropy, generalized linear models, logistic regression, weights-of-evidence, multivariate discriminant analysis, weighting factors, flexible discriminant analysis, generalized additive models, multivariate logistic regression, and other multivariate statistical approaches, as documented in studies by Ref. [14]. Additionally, multi-criteria decision-making analysis has been applied in studies by Refs. [15,16].

Furthermore, it is worth noting that, recently, some studies have employed machine learning techniques to predict flash floods using various environmental factors [72,73]. These studies have employed the combination of geographic information system (GIS) and artificial intelligence to determine the susceptibility to flash flood. Likewise, machine learning techniques, including support vector machines (SVM), artificial neural networks (ANNs), least squares SVM (LSSVM), backpropagation ANNs, classification and regression trees (CART), and random forest (RF), have been leveraged by numerous researchers to develop Flood Susceptibility Maps (FSM), as evidenced in studies by Refs. [17,18].

Recent research has demonstrated the effectiveness of machine learning methods in addressing the challenges associated with hazard map modeling, particularly in regions lacking access to hydraulic and hydrological data. One particularly promising machine learning technique is Random Forest (RF). The RF technique is capable of uncovering intricate patterns by employing multiple tree-based algorithms to generate repeated predictions for various phenomena. This approach offers significant advantages as a spatial modeling tool for evaluating the availability of water and other natural resources. RF stands out due to its exceptional prediction accuracy, reasonable tolerance thresholds, and its ability to mitigate overfitting problems, as highlighted by Refs. [19,20].

Japan has been experiencing a concerning increase in the adverse impacts of flash floods. In September 2015, copious precipitation resulted in the inundation of the Kinu River basin, leading to a subsequent overflow and the breach of the river's embankment in the eastern sector of Joso City, located within Ibaraki Prefecture, Japan [21].

Furthermore, in July 2018, a period of heavy rainfall occurred in Ibaraki Prefecture, driven by a seasonal precipitation front and a concurrent typhoon. This led to flooding and landslides, disrupting transportation systems, causing damage to residential buildings, and resulting in multiple fatalities. Additionally, in 2019, Ibaraki Prefecture experienced a significant flood event triggered by Typhoon Hagibis, causing extensive damage to properties and vital infrastructure and tragically claiming at least one life.

In the year 2020, Typhoon Hagibis emerged as a formidable and expansive tropical cyclone, inducing unprecedented levels of precipitation across extensive regions of Japan. Apart from the inundation-related devastation, this typhoon precipitated a total of 962 landslides, marking the highest incidence of landslides attributed to a typhoon on record. Furthermore, approximately two weeks subsequent to Typhoon Hagibis making landfall in Japan, a low-pressure system situated along the southern coastline of the country, along with the occurrence of Typhoon Bualoi, engendered another historic bout of rainfall within the Ibaraki Prefectures, resulting in successive episodes of flood-related destruction [22].

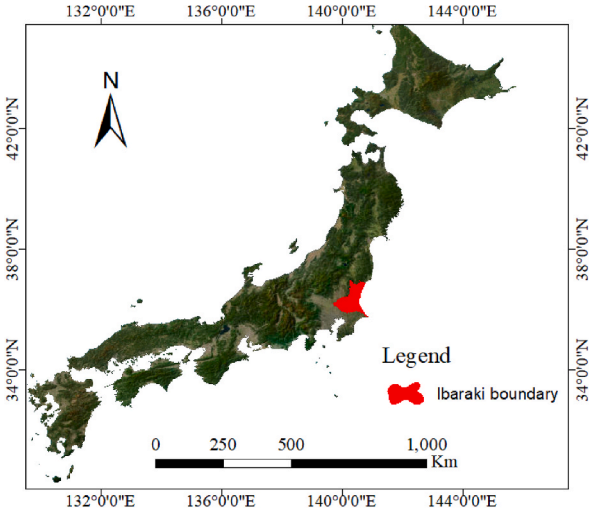

Fig. 1. Ibaraki prefecture location.

The main objective of this research is to assess and predict Flash Flood Hazard Susceptibility Mapping using a Random Forest Regression Model and Geographic Information System (GIS) in the Japanese prefecture of Ibaraki. This particular prefecture was chosen as the focal point of investigation in this research due to its recurring vulnerability to flash flood incidents over the course of several decades. In this particular study area, previous simulations have been conducted to estimate the runoff depth which is not enough to represent the flood susceptibility. That's why this research has focused on most of the causative environmental parameters as input in the Random Forest model. These causative factors represent the main producer of the flood hazard such as elevation, slope, aspect, distance to streams, rivers, and roads, land cover, topographic wetness index, stream power index, plan curvature, and profile curvature. Ultimately, this research produces meaningful insights that can be compared with other machine learning methodologies used for flood hazard mapping in the identical geographic area.

2. Materials and methods

2.1. Study area

Ibaraki Prefecture, located along the eastern coast of Honshu, Japan's main island, is bordered by the Pacific Ocean to the east. Covering an approximate area of 6,100 km^2, the prefecture is home to around 2.87 million residents. Its climate falls into the humid subtropical category, characterized by hot and humid summers and relatively mild winters. Fig. 1 provides a visual representation of the geographical location of Ibaraki Prefecture. Ibaraki Prefecture is experiencing recurring vulnerability to flash flood incidents over the course of several decades.

Significantly, in September 2011, the arrival of Typhoon Talas had a significant impact on Japan, leading to substantial rainfall in various regions, including Ibaraki Prefecture. The aftermath included extensive flooding, landslides, and mudslides. Of particular note was the overflow of the Kinugawa River in Ibaraki, resulting in severe flooding in Joso City. These floods caused extensive damage to homes, critical infrastructure, and agricultural land.

2.2. Data and materials

The digital elevation model (DEM) with resolution 30 m, which contains terrain elevation information, was obtained from Yamazaki-Lab after conducting adjustment through the following website: http://hydro.iis.u-tokyo.ac.jp/~yamadai/JapanDir/index.html. The DEM was utilized to delineate the study area to estimate the basins, flow direction, flow accumulation. This essential morphometric delineation was used to estimate the environmental factors. Additionally, the study area boundaries were delineated using a polygon shapefile obtained from the DIVA-GIS website (https://www.diva-gis.org/Data). The acquired boundaries serve as crucial role for delineating the study area and accurately determining its geographic extent.

Furthermore, land cover data was sourced from the ESRI (Living Atlas) website (https://livingatlas.arcgis.com/). Land cover serves as a pivotal environmental parameter in assessing flood susceptibility, as it characterizes the terrain's surface type. Vegetated areas typically exhibit lower susceptibility to flooding due to their higher infiltration rates, while paved and urban areas tend to display higher susceptibility owing to their reduced capacity for runoff absorption. Shapefiles for road and river maps, essential for calculating distances within the ArcMap software, were obtained from Open Streat Map data hosted at the Geofabrik website (http://download.geofabrik.de/asia/japan.html). This shapefile was utilized to generate the distance to river and road maps which are used as important environmental parameters.

Utilizing geospatial information, the process of discerning inundated and non-inundated locations was executed by cross-referencing a flood hazard survey from the Ministry of Land, Infrastructure, Transport and Tourism of Japan accessible at (https://disaportal.gsi.go.jp/index.html). The integration of this spatial data assumes a crucial significance in the production of the flood inventory map. Table 1 shows the usage and spatial resolution of the utilized data. Furthermore, the produced environmental maps were resampled in ArcMap to obtain the same resolution of 30 m.

2.3. Flood inventory map

The flood inventory map assumes a central and indispensable role in assessing flood vulnerability, as highlighted in prior research [28,29]. Similarly, the precision of the flood hazard map stands to benefit from an increase in the number of inundated locations [30]. The crux of this map resides in its depiction of observed flooded and non-flooded points. In this study area, a total of 224 points were considered, with 112 identified as non-flooded and 112 as flooded. In addition, the flooded points have been extracted from a survey

Table 1
The utilized data classification.

Data type	usage	resolution	year	source
Polygon shapefile	Identification of the Ibaraki prefecture boundary	N/A	2023	[23]
Polygon shapefile	Estimation of distance from roads and rivers	N/A	2023	[24]
Raster DEM	Identification of the terrain elevation	30 m	2022	[25]
Raster land cover	Classification of the spatial land cover	10 m	2023	[26]
Point shapefile	Identification of the flooded and non-flooded zones	N/A	2023	[27]

conducted in https://disaportal.gsi.go.jp/. The flooded pointed were selected from the maximum estimation of runoff that generated form multiply rainfall event. On the other hand, the non-flooded points have been selected randomly from the mentioned website from spots that did not experience flooding. In the context of machine learning techniques, both spatial data pertaining to flooded and non-flooded areas assume paramount significance in training the model to make predictions across the entire study region. Fig. 2 provides a visual representation of the flood inventory map.

2.4. Methodology

In this investigation, three major phases were executed, commencing with preparatory processing, followed by the estimation of environmental factors, culminating in the utilization of the machine learning model (see Fig. 3). In the initial phase of this study, a digital elevation model (DEM) with a spatial resolution of 30 m was utilized and processed using ArcMap 10.8. The initial utilization of this software involves the visual estimation of environmental factors. Following the estimation process, these factors are transformed into ASCI files for exportation into R software, where the machine learning model will be programmed.

In the ArcMap software, the spatial information derived from the elevation values was of paramount importance, as it was subsequently employed for various geospatial tasks, including filling of the DEM, delineation of flow directions, determination of flow accumulation, and identification of drainage basins and sub-basins.

Furthermore, land cover data were extracted and tailored to the specific study area. In addition, the distances to streams were calculated based on the previously obtained flow accumulation and flow direction information. Similarly, the distances to roads and rivers were estimated utilizing region-specific road and river maps. Additionally, key terrain attributes such as aspect, slope, profile curvature, and plan curvature were computed from the DEM. Subsequently, the Stream Power Index (SPI) and Topographic Wetness Index (TWI) were derived from slope and flow accumulation data. All environmental variables were then processed and resampled to a uniform resolution of 30 m.

In the context of the machine learning approach, spatial data points representing flooded and non-flooded areas were exported to the R software. These data points were randomly split into two sets, with 70 % allocated for training the predictive model and the remaining 30 % reserved for assessing model accuracy. These ratios were previously utilized by Ref. [73,74]. The regression Random Forest algorithm was then applied to predict the likelihood of flood occurrence in specific pixels.

To assess the accuracy of the RF regression model, the Receiver Operating Characteristic-Area Under Curve (ROC-AUC) method was employed, which is a widely recognized approach for evaluating the performance of machine learning techniques and addressing criteria selection and interpretive concerns [31].

Following that, satisfactory values of Mean Square Error (MSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) were obtained. The final step involved using the trained model to predict flood hazard potential across all pixels. The resulting

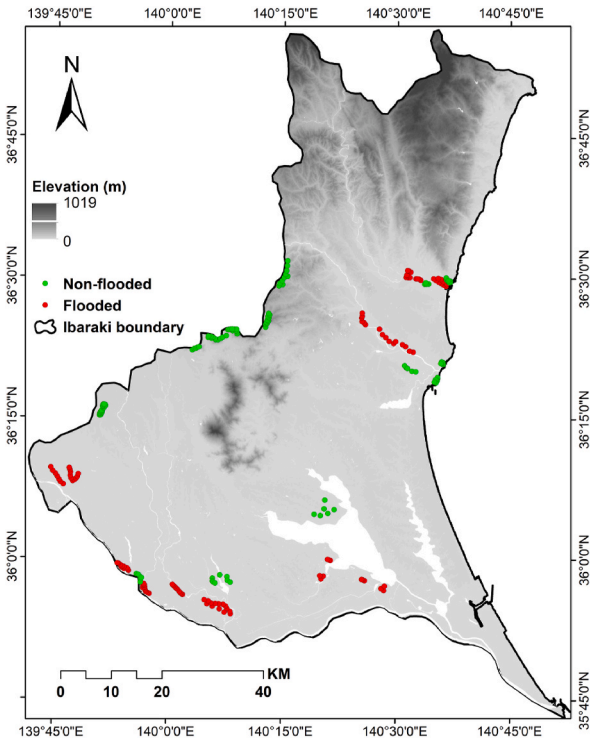

Fig. 2. The flood inventory map.

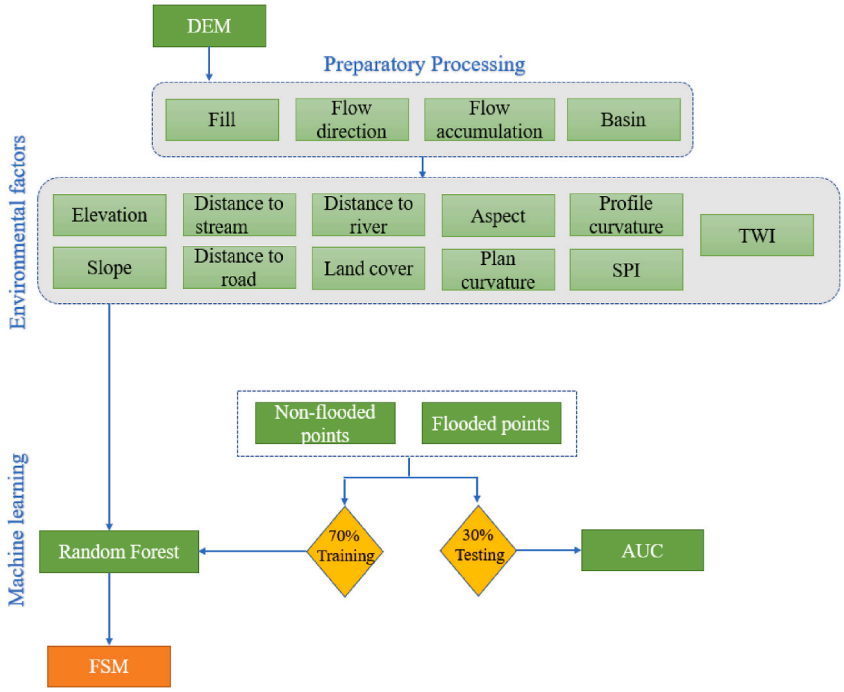

Fig. 3. The methodological framework.

predictions were projected onto a map in Arc Map and classified into five distinct categories, generating the Flood Susceptibility Map (FSM).

2.4.1. Flood environmental factors

Producing a flood hazard map for any given catchment necessitates the critical identification and incorporation of relevant geographic causal factors to ensure efficacy and accuracy [32]. However, within the context of most flood hazard maps, a notable lack of consensus exists concerning specific parameters [33]. In this study, eleven environmental parameters were used to indicate the flood occurrence. Fig. 4(a–k) illustrates the flood environmental factors, encompassing the following components:

Although there is no specific, universally recognized standard for the selection of flood causative factors that influence flood events [34], the interplay between various topographic and environmental factors plays a crucial role in flood risk assessment. Elevation, as highlighted by Samanta et al. [35], is a fundamental meteorological condition that exhibits an inverse relationship with flood severity. Higher elevations are associated with reduced flash flood risk, illustrating the interconnectedness of elevation and flooding. Slope inclination, as noted by Kourgialas and Karatzas [36], contributes significantly to flood dynamics. In the surveyed region, the elevation spans a considerable range, varying from sea level at 0 m to its peak at 1019 m. Additionally, the slope measurements approach a maximum of nearly 60°, predominantly observed in the northern zone of the Ibaraki prefecture. The steeper slopes mitigate flood occurrences due to their impact on flow velocity, creating a link between slope and flood risk. Likewise, Proximity to streams, as emphasized by Opperman et al. [37], is another pivotal factor. Areas near streams are more susceptible to flooding, highlighting the connection between distance to streams and flood vulnerability. In the study area, there are a variety of the distances to stream with maximum value reaches to more than 25 km.

The presence of roads and urban development, as discussed by Tehrany et al. [38], introduces artificial obstructions to natural water flow, amplifying flood risks. This underscores the link between distance to roads and flood vulnerability. Furthermore, the proximity to rivers, as elucidated by Khosravi et al. [39], directly influences soil moisture levels and, consequently, flood occurrences. This establishes a clear connection between distance to rivers and flood risk assessment. These two essential types of distances reached to maximum of 0.50 and 3.0 km for roads and rivers respectively. Moreover, land cover, as a major contributor, as mentioned by Komolafe et al. [40], can significantly impact surface runoff, groundwater seepage, and evapotranspiration rates. This interconnectedness underscores the need to consider land cover in comprehensive flood risk assessments. The land cover mapping reveals a greater presence of vegetation in the northern regions, notably in areas of higher elevation. This distribution pattern could result in a reduced likelihood of flood hazards within the northern zones. Aspect, a topographic characteristic, influences climatic elements that, in turn, affect the occurrence of natural phenomena, including floods. The relationship between aspect and climatic conditions further highlights the multifaceted nature of flood risk assessment. Additionally, the terrain's curvature, encompassing both plan curvature and profile curvature, offers insights into flow patterns and the ground's capacity to either facilitate or hinder water movement. Understanding these curvatures establishes a link between topography and flood dynamics. From the estimated curvature maps, it is clearly that the land at the southern area is more likely to be flat, whereas the northern region has the highest curvature values.

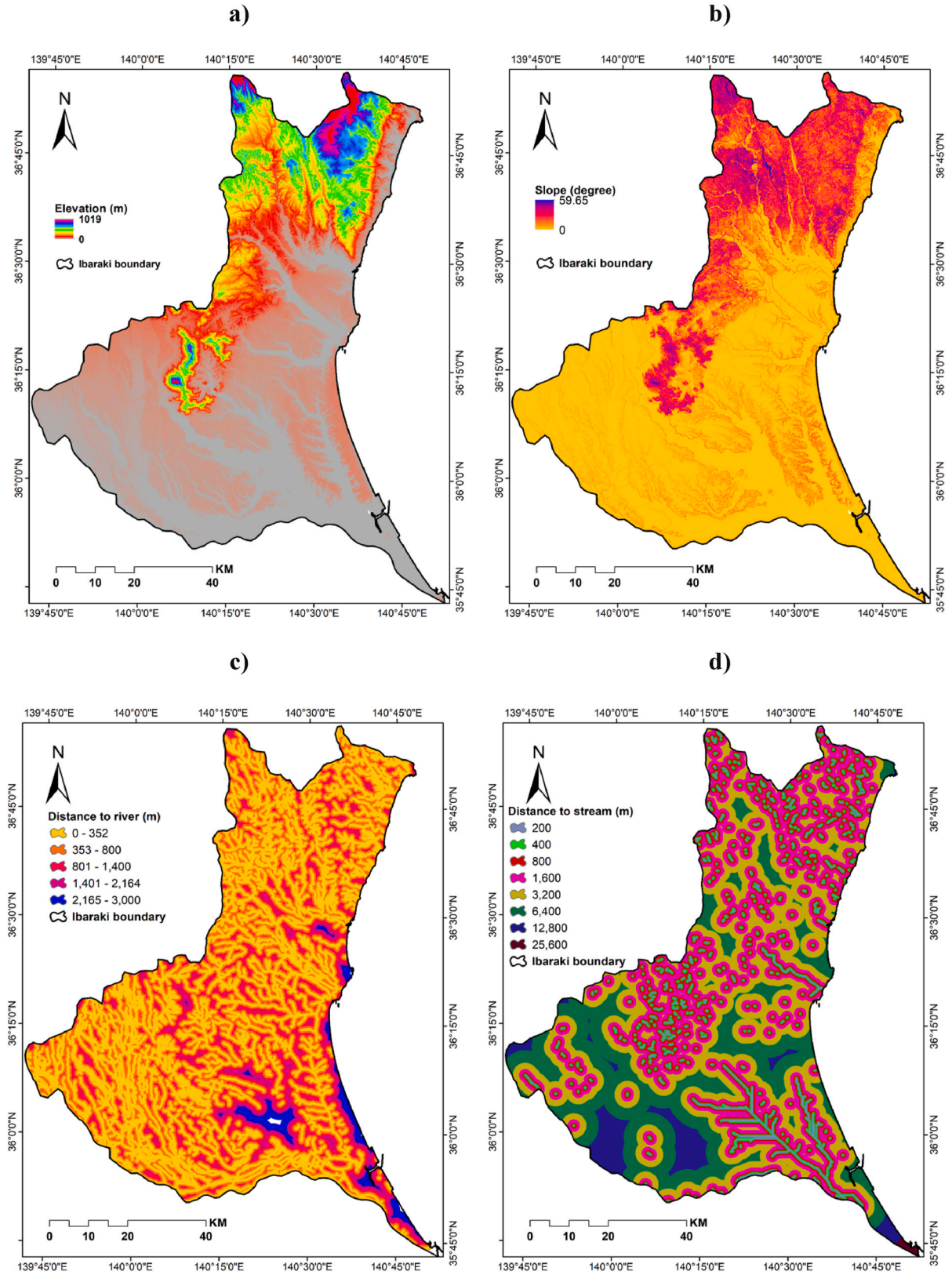

Fig. 4. The flood environmental factors: a) elevation (m); b) slope (degree); c) distance to river (m); d) distance to stream (m); e) distance to road (m); f) land cover; g) aspect (degree); h) TWI; i) SPI; j) profile curvature (degree); k) plan curvature (degree).

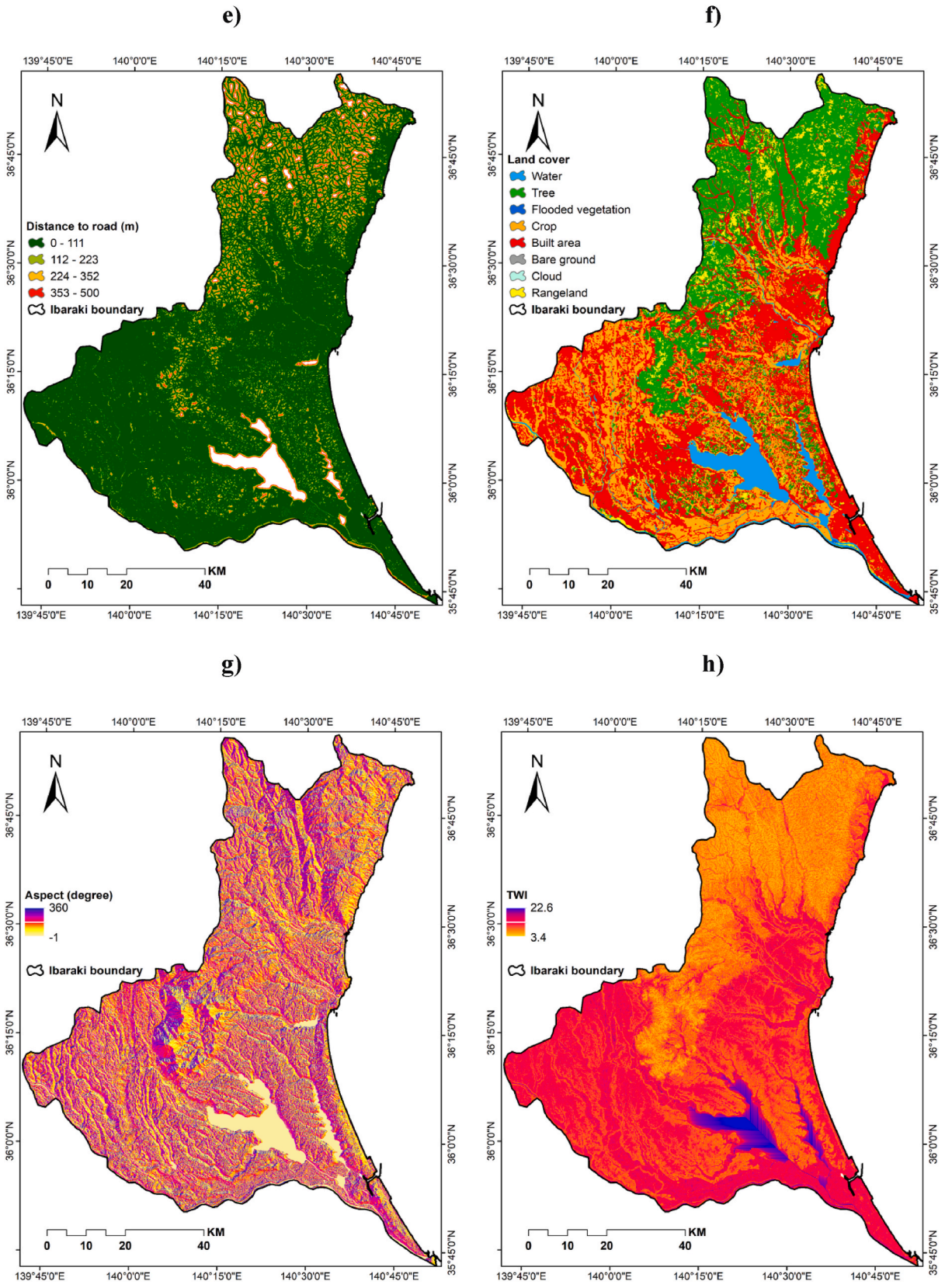

Fig. 4. (*continued*).

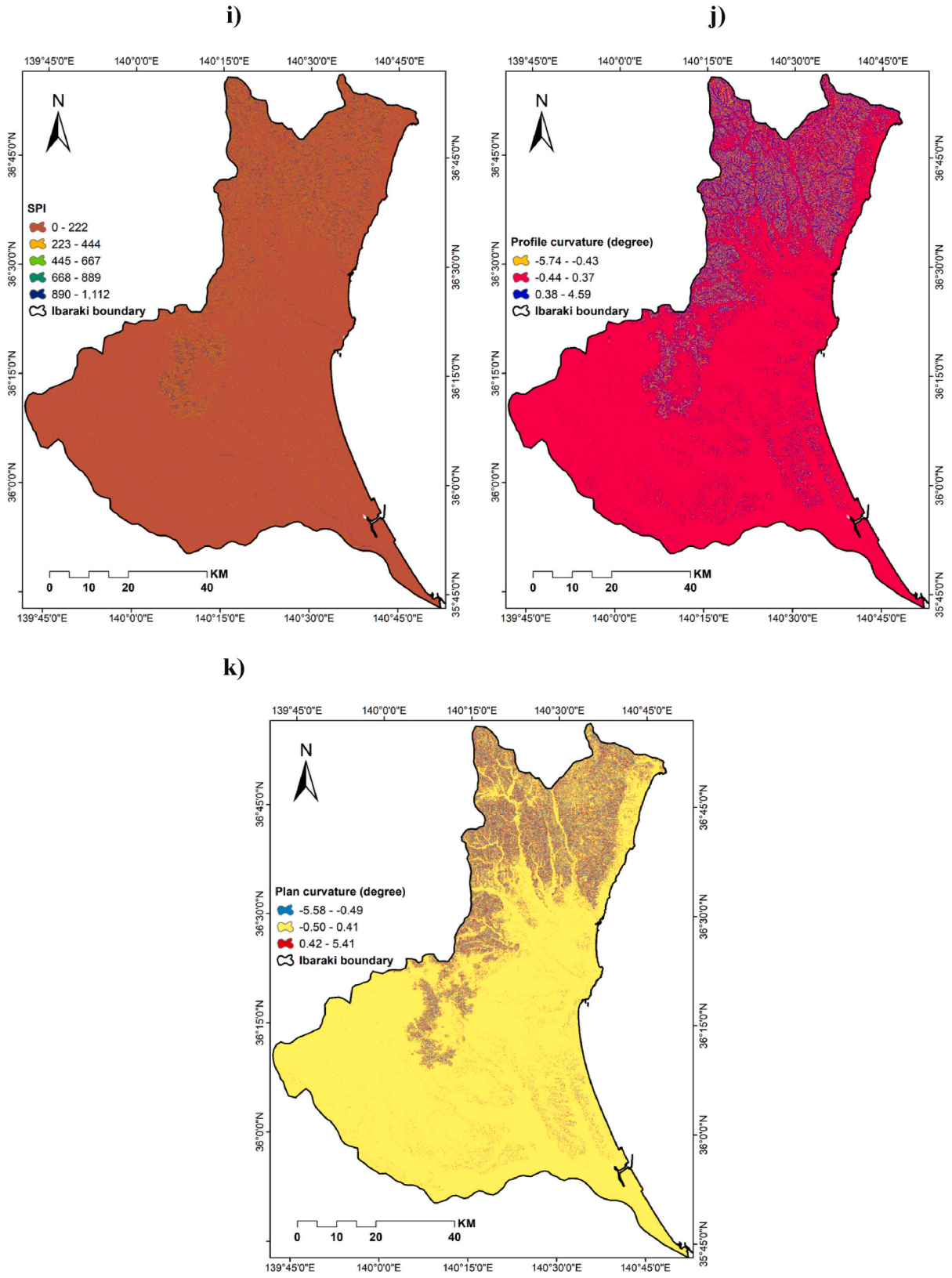

Fig. 4. (*continued*).

Finally, specialized indices like the Stream Power Index (SPI) and Topographic Wetness Index (TWI) provide quantitative tools to evaluate flood-related factors. In the study area, it appears that the low values of TWI exist mostly in the northern region. The SPI's calculation, as defined by Huisman and de By Ref. [40], directly ties together flow accumulation and slope, aiding in identifying soil conservation measures. Similarly, the TWI, as detailed by Mudashiru et al. [41], sheds light on the hydrological properties of a watershed, bridging the gap between topography and downstream water movement. These indices serve as critical connectors in the comprehensive assessment of flood risk. SPI and TWI can be calculated using equations (1) and (2).

$$SPI = (\varphi_s)(\tan \alpha) \tag{1}$$

$$TWI = \ln\left(\frac{\varphi_s}{\tan \alpha}\right) \tag{2}$$

where φ_s is the flow accumulation in a particular watershed area, and α is the slope in degree.

2.4.2. Random forest model

The Random Forest (RF) algorithm stands as a prominent ensemble technique in supervised machine learning, initially introduced by Leo Breiman in 2001 [42]. It finds application in both classification and regression tasks. Known for its reliability in making predictions, the RF algorithm remains effective even with sizable datasets, demonstrating resilience against overfitting. Notably, it accommodates numerous input features without disregarding any and possesses the capability to evaluate the predictive power of all variables [42,43].

RF functions through a collection of diverse trees, each derived from bootstrap samples [44,45]. These data are divided into two categories: "in bag" data, used for model training, and "out-of-bag" data, allocated for model validation. Typically, around two-thirds of the bootstrap samples constitute the training data. Enhancing prediction accuracy and improving model efficacy, according to Taalab et al. [46], involves optimizing two factors: the quantity of trees (ntree) generated within the model and the number of available splitable parameters at each tree node (mtry).

Numerous theoretical and empirical research endeavors have extensively outlined the myriad benefits of Random Forest (RF) models, encompassing their notable strengths such as precise forecast accuracy, resilience in accommodating outliers and noise, and the straightforward avoidance of overfitting concerns. Given this extensive knowledge base, RF theoretically holds significant promise for assessing flood hazard risks, particularly in addressing complex multi-variable and non-linear challenges [47]. In contrast to conventional methods like bagging that utilize an ensemble approach, the structure of RF can decrease the attributes within each Decision Tree (DT), thus augmenting variability between DTs and mitigating issues related to the curse of dimensionality [48].

RF has found successful applications across various domains, including earthquake damage classification [49], anticipation of rock burst occurrences [50], genetic data analysis [51], and computer-aided diagnosis [52]. Conversely, employing RF for flood hazard prediction offers several advantages [53]. Firstly, RF adeptly models the complexities inherent in flood hazard prediction, capturing non-linear dependencies and interactions among multiple factors crucial in these scenarios. Secondly, employing ensemble learning through multiple decision trees allows RF to combine knowledge, enhancing prediction accuracy while minimizing bias and variance. Lastly, RF demonstrates exceptional scalability, efficiently handling extensive datasets, a crucial aspect in flood hazard prediction involving vast spatial and temporal information.

Regarding the RF process, an elementary decision tree structure initiates from a sample input (variable) and branches out according to predetermined criteria. If the criteria are met ('yes'), the decision tree progresses along the outlined route; if not ('no'), it follows an alternative route. This iterative process continues until the decision tree arrives at a terminal node, determining the final outcome (Fig. 5).

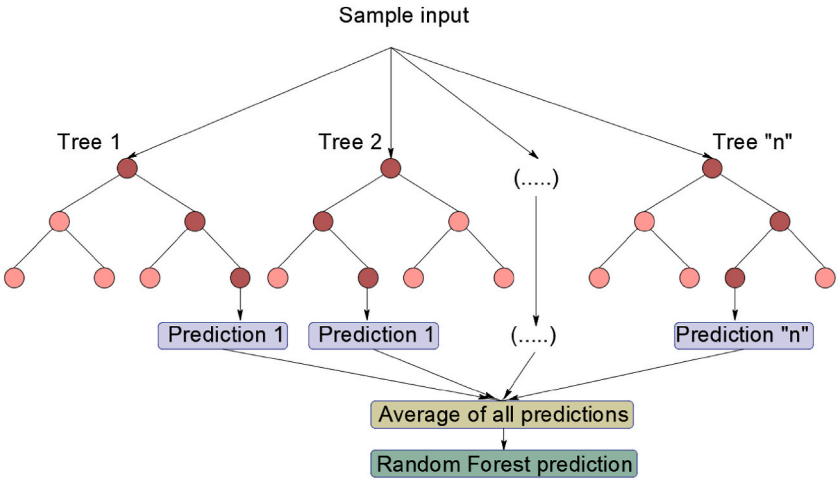

Fig. 5. Schematic structure of the random forest regression model.

In this research endeavor, R-software (version R-4.2.1) has been employed to construct the Random Forest (RF) regression model. Initially, the eleven environmental factors were resampled at a spatial resolution of 30 m using ArcMap. Subsequently, the data contained within each pixel were exported to the R-software environment. Simultaneously, the corresponding data pertaining to flood occurrences were integrated into the model. This flood occurrence dataset was arbitrarily split into two subsets, with 70 % of the data allocated for training the model, while the remaining portion was set aside for model validation.

Furthermore, the model was fine-tuned to determine the optimal values for two key factors: the number of trees (ntree) and the maximum number of nodes in each tree (maxnodes). Fig. 6 visually illustrates the relationship between the root mean square error (RMSE) and variations in both ntree and maxnodes. Based on the estimations, the optimal number of trees has been identified to be 900, with each tree having a maximum node count of 70. The RMSE was estimated via equation (3).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{m} (\rho_i - \sigma_i)^2}{m}} \tag{3}$$

where, ρ_i = prediction, σ_i = actual value, m = total count of data.

To implement the Random Forest Regression algorithm in R software, there are 8 essential steps as following:

1) Data Preparation: Collect and preprocess the flood and non-flood data with their corresponding environmental parameters.
2) Loading Data into R: loading the preprocessed data to the dataset into R environment.
3) Splitting Data into Training and Testing Sets: Split the dataset into training and testing sets.
4) Model Training: Train the Random Forest regression model using the "randomForest" package in R.
5) Model Evaluation: Evaluate the model's performance using metrics like MSE, RMSE, or R-squared on the testing dataset.
6) Model Tuning: Fine-tune model parameters such as the number of trees or tree depth to optimize performance.
7) Prediction: generate the predictions on new data using the trained model with the 'predict' function.
8) Visualization and Interpretation: Visualize model outputs, such as feature importance or predicted hazard maps, to gain insights and aid interpretation.

2.4.3. Model performance

Numerous techniques are available for assessing the efficacy of machine learning models. In this study, we utilize the Mean Absolute Error (MAE), Mean Square Error (MSE), and R-squared (R^2) metrics to evaluate the performance of both classification and regression models. The mathematical expressions for these metrics are presented in equations (4)–(6), respectively.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{m} |\rho_i - \sigma_i|}{m} \tag{4}$$

$$\mathrm{MSE} = \frac{1}{m} \sum_{i=1}^{m} (\sigma_i - \rho_i)^2 \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (\rho_i - \sigma_i)^2}{\sum_{i=1}^{m} (Z - \sigma_i)^2} \tag{6}$$

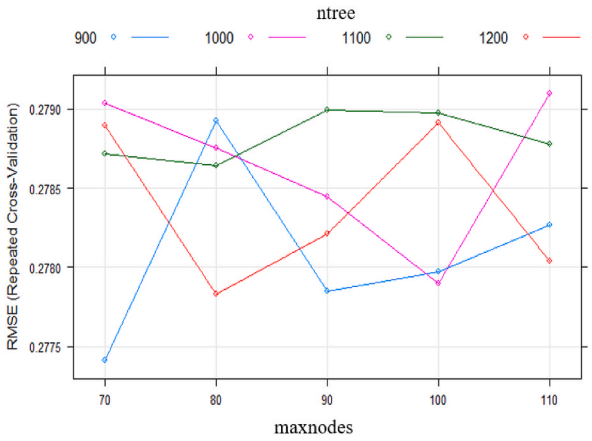

Fig. 6. Tuning the random forest model.

Since, $z = \frac{1}{m}\sum_{i=1}^{m}\sigma_i$.where, z = the mean of actual values,

Furthermore, the Mean Absolute Error (MAE) serves as a quantitative measure in assessing models by computing the mean of the absolute differences between predicted and actual values. This metric finds utility in the evaluation of models where substantial errors are considered undesirable. Conversely, the Mean Squared Error (MSE) estimates the mean of the squared differences between predicted and actual values, assigning greater significance to larger errors. MSE proves valuable in evaluating models that prioritize accurate prediction of extreme values. In a related vein, the Root Mean Squared Error (RMSE) emerges as the square root of MSE, employed to convey error in units congruent with the target variable. RMSE offers insight into the extent of deviation between predictions and actual values. Lastly, the R-squared metric quantifies the ration of variance in the target variable elucidated by the model, taking values within the range of 0–1. Higher R-squared values denote a superior fit of the model to the data. These four metrics have been computed and presented in Table 2.

The MAE states that, on average, the model's predictions deviate from the actual values by approximately 0.137 units. A lower MAE indicates that the model's predictions are closer to the actual values, implying a good fit of the model to the data. On the other hand, 0.038 for the MSE implies better model performance. MSE places more weight on larger errors than MAE, which means the model is generally providing reasonably accurate predictions with minimal outliers. Moreover, R^2 value of 0.885 indicates that the model accounts for approximately 88.5 % of the variability in the target variable. This is a relatively high R^2 value, suggesting that the model is performing well in explaining and predicting the variation in the data. In essence, about 88.5 % of the variance in the dependent variable can be attributed to the independent variables included in the model.

3. Results and discussions

3.1. Flood susceptibility map (FSM) generation

The Flood Susceptibility Map (FSM) represents a crucial tool for spatially assessing the risk of flash floods. This map was generated using a Random Forest (RF) regression model, illustrating a systematic approach for flood hazard evaluation within the study area. The process of hazard prediction involved initial execution within the Python programming environment, followed by the transfer of prediction data to ArcMap for cartographic representation, leading to the creation of the Flood Susceptibility Map (FSM) as depicted in Fig. 7. Subsequently, the forecasted values were stratified into five distinct hazard categories, namely "very low," "low," "moderate," "high," and "very high," employing the equal interval classification tool available within the ArcMap.

The resulting flood susceptibility map reveals a noteworthy concentration of areas classified as having very low and low susceptibility in the northwest portion of the study area. Conversely, the remaining three susceptibility categories exhibit a significant distribution in the southern region. This spatial pattern suggests that areas with lower slope values in low-lying regions are at a heightened risk of flash floods, consistent with prior research findings [54,55]. Additionally, the proximity to rivers plays a significant role in flood occurrence, particularly in areas closer to riverbanks [38]. This observation is consistent with the generated FSM, which consistently indicates that areas near rivers and roads are more vulnerable to flooding. Furthermore, it is worth noting that urbanization can potentially exacerbate the risk of floods, as areas undergoing urban development with reduced infiltration rates are particularly susceptible to flood hazards, as exemplified in the study by Ref. [56]. This vulnerability arises from the capacity of rivers to overflow during floods and the limited infiltration capabilities of roads, causing water accumulation and increasing the risk of flash floods.

Furthermore, Wahba et al. [57] previously produced a flood hazard map, identifying areas with gentle gradients and low elevations as more susceptible to flooding in two basins in the Yamaguchi and Shimane prefectures, Japan, a finding supported by this study's methodology. The influence of land cover on hazard classification within the FSM is also notable, with forested areas predominantly exhibiting low hazard levels due to their high permeability, which mitigates surface runoff accumulation. Conversely, areas characterized as having high and very high hazard levels are mainly associated with built-up regions, underscoring the role of urbanization in exacerbating flood hazards. This underscores the potential for urban areas to amplify flood vulnerability. However, well-planned urban areas that undergo hydrological studies and incorporate appropriately designed drainage systems capable of handling maximum water accumulation during flood events are less likely to experience flooding [15,58,59].

From another perspective, the proportions of hazard levels were determined based on the study region's area, as illustrated in Fig. 8. It's evident that a significant portion of the study area, 37.71 %, is categorized as having a "Very Low" flood hazard level. This suggests that a substantial part of the region is at a relatively low risk of flooding. This information can be reassuring for residents and businesses located in these areas, as it indicates a reduced likelihood of being affected by floods. Similarly, nearly a third of the area falls under the "Low" flood hazard category. While this is slightly higher than the "Very Low" category, it still implies that a considerable portion of the region faces relatively low flood risks. These regions might be suitable for various types of development, including residential and commercial, with a lower risk of flood-related damage.

On the other hand, 8.31 % of the area is categorized as having a "Moderate" flood hazard level. This category suggests that specific

Table 2
MAE, MSE, and R^2 for the study region.

	MAE	MSE	R^2
Value	0.137	0.038	0.885

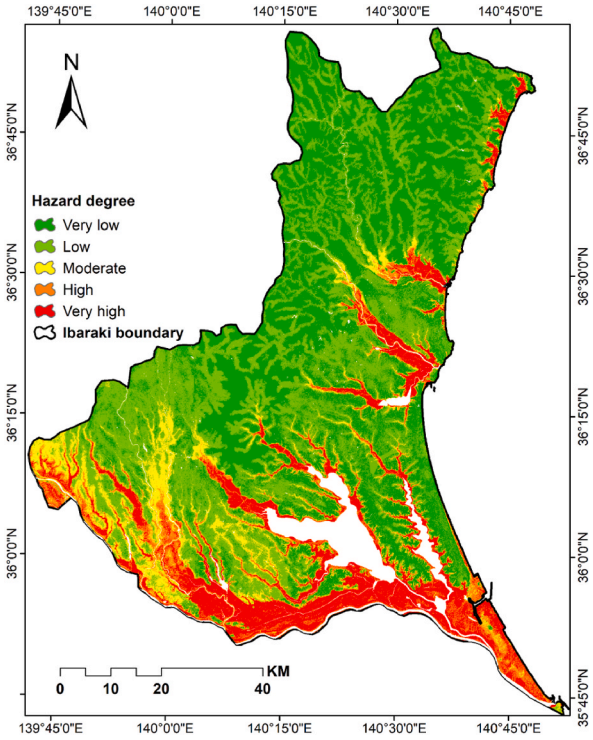

Fig. 7. The flood susceptibility map.

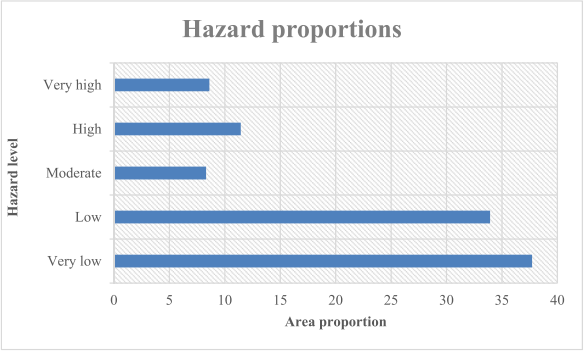

Fig. 8. Proportion of study area according to the hazard level.

precautions and mitigation measures may be necessary to manage flood risks effectively. It's essential for local authorities and residents in these areas to be aware of the potential hazards and take appropriate steps to reduce vulnerability.

From another perspective, the proportions of hazard levels were determined based on the study region's area, as illustrated in Fig. 7. It's evident that a significant portion of the study area, 37.71 %, is categorized as having a "Very Low" flood hazard level. This suggests that a substantial part of the region is at a relatively low risk of flooding. This information can be reassuring for residents and businesses located in these areas, as it indicates a reduced likelihood of being affected by floods. Similarly, nearly a third of the area falls under the "Low" flood hazard category. While this is slightly higher than the "Very Low" category, it still implies that a considerable portion of the region faces relatively low flood risks. These regions might be suitable for various types of development, including residential and commercial, with a lower risk of flood-related damage.

Moving towards higher risk categories, 11.44 % is categorized as having a "High" flood hazard level, signifying a notable risk of flooding. Consequently, both short-term and long-term strategies for flood preparedness and mitigation should take precedence in these areas. These strategies may encompass enhancements to infrastructure, the implementation of flood barriers, establishment of early warning systems, and the enforcement of stricter zoning regulations. Additionally, approximately 8.61 % of the region is designated with a "Very High" flood hazard level, representing the utmost risk within the area. These areas demand immediate and comprehensive flood risk management measures, including the development of advanced warning systems, well-defined evacuation

procedures, and potential restrictions on new construction projects to address the heightened vulnerability.

3.2. Spatial planning and mitigation

Utilizing FSM for spatial planning and mitigation involves a comprehensive approach that integrates geographic information to inform decision-making processes. Obtaining the generated FSM is crucial, with an emphasis on representing different levels of flood risk. This map is then incorporated into spatial planning frameworks, guiding zoning regulations, land-use plans, and building codes. In spatial planning, these maps serve as valuable tools for identifying high-risk zones, enabling planners to make informed decisions about land use and infrastructure development. By avoiding construction in high-risk areas or implementing resilient design strategies, spatial planners can reduce the potential impact of floods on communities.

Mitigation efforts also benefit significantly from classified flood hazard maps. These maps enable authorities to prioritize areas that require immediate attention and investment in mitigation measures. For instance, regions identified as high-risk on the maps may need the implementation of flood barriers, levees, or other protective infrastructure. Additionally, the information provided by the maps helps in developing early warning systems and emergency response plans, allowing communities to prepare for and respond to flooding events more effectively. By integrating these maps into spatial planning and mitigation strategies, decision-makers can enhance the overall resilience of communities, reduce property damage, and safeguard lives during flood events.

3.3. Model validation

To assess the precision of the Random Forest algorithm, the Receiver Operating Characteristic-Area Under Curve (ROC-AUC) was applied. This technique is often used to evaluate the accuracy of machine learning methods, as it helps identify interpretive and criterion selection issues [60]. By plotting the True Positive Rate (TPR), also known as sensitivity, against the False Positive Rate (FPR), denoted as specificity, along the y and x axes, respectively, ROC curves are produced. FPR measures the proportion of negative instances or non-events that are erroneously classified as positive or events, whereas TPR determines the proportion of existing positives that are accurately detected [61]. Meanwhile, it signifies the frequency with which the model erroneously forecasts a favorable result despite the fact that the true outcome is unfavorable [62]. The calculations of the TPR and FPR can be conducted through equations (7) and (8).

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

where:

TP is the number of True Positives (correctly predicted positive instances).

FN is the number of False Negatives (actual positive instances incorrectly predicted as negative).

FP is the number of False Positives (actual negative instances incorrectly predicted as positive).

TN is the number of True Negatives (correctly predicted negative instances).

Following the computation of TPR and FPR, the ROC curve has been visually represented in Fig. 9. An AUC value of 50 % indicates that the estimation is not discriminatory [63]. Nonetheless, when the AUC surpasses 90 %, the model can be deemed to have outstanding leverage [64].

The proportions of AUC were estimated as 99.56 %. This output provides greater confidence in the effectiveness of the machine learning technique particularly RF regression model. This transcendence of the adopted machine learning technique comes through a powerful stability of the model and the dependency on both flooded and non-flooded points.

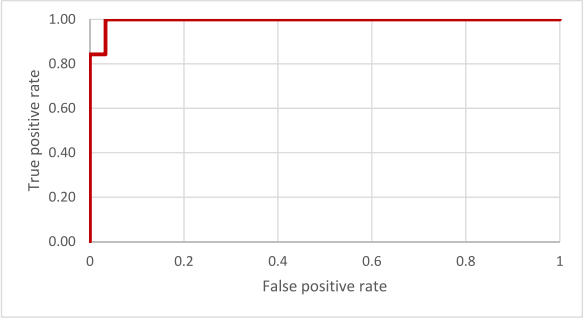

Fig. 9. Receiver operating characteristic curve.

4. Suggested mitigation measures

Ibaraki Prefecture is confronted with the imminent threat of flooding, primarily attributed to its geographical proximity to rivers and coastal regions. In light of this vulnerability, the imperative of instituting robust flood mitigation strategies is very crucial. The safeguarding of both human lives and property hinges upon the judicious implementation of a multifaceted set of measures tailored to the region's specific needs.

One suggested flood mitigation involves the construction and fortification of flood barriers, embankments, and levees, predominantly situated along the courses of rivers and coastal perimeters, with the aim of curtailing the ingress of water into densely inhabited locales [65] Equally pivotal is the institution of stringent floodplain zoning regulations, thereby averting construction activities in flood-prone areas [66] and ensuring that newly erected structures adhere to stringent flood-resistant building codes [67]. Furthermore, the development and enhancement of early warning systems [68] equipped with cutting-edge meteorological monitoring technologies, holds paramount significance, as it enables the timely dissemination of alerts to residents and pertinent authorities, thereby affording them crucial preparatory time in the face of impending inundations. Additionally, ensuring the proficiency of emergency responders and the general populace through recurrent training and drills is essential to foster effective flood event management.

Concomitantly, the construction of reservoirs and detention basins assumes prominence [69], providing a temporary reservoir for excess rainfall, with controlled discharge to mitigate downstream flooding risks. River channelization, involving channel modifications [70] and continuous upkeep, serves to optimize water flow dynamics and reduce flooding perils, mindful of its ecological ramifications. The elevated construction of critical infrastructure, such as hospitals and emergency shelters, contributes to safeguarding them during flood events. Moreover, upgrading and sustaining drainage systems to facilitate efficient rainwater diversion away from inhabited zones is pivotal. Incorporating green infrastructure solutions such as permeable pavements and green roofs, enables stormwater absorption and management [71]. Encouraging or mandating the use of flood-resistant building materials and techniques in new constructions, encompassing elevated foundations and flood barriers, is another vital facet. Public awareness campaigns, community workshops, and educational programs in schools play an instrumental role in enhancing flood risk awareness and safety measures. In parallel, devising comprehensive climate adaptation plans that factor in the potential influence of climate change on flood patterns and integrating strategies to mitigate these effects is indispensable. Additionally, the restoration and protection of natural riparian zones is conducive to erosion reduction [72], water absorption, and the augmentation of the natural flood control capacity of the landscape. Moreover, it is crucial to pinpoint the primary locations where runoff accumulates significantly in order to alleviate the adverse effects of sudden floods [58]. Furthermore, regions proximate to coastlines may employ artificial channels to direct excess runoff into the sea or ocean, as described by Ref. [75]. Lastly, forging collaborative ties with neighboring regions and countries for transboundary flood management, along with the exchange of information and resources, embodies a strategy of international cooperation. Ensuring the efficacious execution of these measures necessitates active engagement with local communities, government agencies, and environmental experts, thereby assuring their resilience and functionality in the face of evolving climate dynamics. Regular maintenance and updates to these measures remain pivotal, aligning with the imperative of adaptability in a climate-altered landscape.

5. Conclusions

In this research, an investigation was undertaken to develop a Flood Susceptibility Map (FSM). The investigation employed a Random Forest (RF) regression model that incorporated 11 environmental variables (including factors such as elevation, slope, distance to streams, rivers, and roads, land cover, stream power index, topographic wetness index, plan curvature, profile curvature, and aspect). This modeling approach was applied to a dataset comprising 224 instances representing both flooded and non-flooded locations. The dataset was divided randomly, with 70 % allocated to a training set for model development, and the remaining 30 % used for model validation through Receiver Operating Characteristics (ROC) curve analysis.

Based on the FSM results, the northwestern region shows a predominance of areas classified as "very low" and "low" susceptibility, whereas the southern part of the study area prominently displays the other three susceptibility categories. This spatial pattern suggests that areas with lower slope values in low-lying terrain are at higher risk of flash floods. The study explores the application of FSM in spatial planning and mitigation strategies. Employing FSM, the research demonstrates its role in informing zoning regulations, land-use plans, and building codes. By identifying high-risk zones, FSM facilitates informed decision-making in spatial planning, enabling the implementation of resilient design strategies and prioritization of mitigation efforts to enhance community resilience and minimize property damage during flood events.

The performance of the machine learning model was assessed by evaluating the Area Under Curve (AUC), yielding an estimated value of 99.56 %. This result instills greater confidence in the efficacy of the machine learning approach, particularly the RF regression model employed. The utilized approach can be applied in alternative topographic regions characterized by diverse climatic conditions to produce the FSM. Finally, the generated FSM serves as a pivotal instrument for policymakers, offering guidance for the implementation of appropriate flood mitigation strategies in the Ibaraki prefecture.

Regarding the limitation in this study, the research introduces FSM within Ibaraki prefecture employing the RF model, marking the initial instance of employing RF to create the FSM within this specific case study. Although employing RF presents potential opportunities for creating flood susceptibility maps, it encounters certain hurdles related to the adequacy and quantity of available data. A larger dataset comprising both flooded and non-flooded points is crucial for enhancing the accuracy of predictive maps generated by the algorithm. Furthermore, this study's implementation within Ibaraki prefecture implies that the model's accuracy might exhibit slight variations when applied to different case studies. Moreover, this investigation represents a crucial initial phase in assessing

mitigation strategies within areas susceptible to flash floods, serving as a foundation for future examinations.

Data availability statement

Some or all data, models, or code that support the findings of this study are available from the first author upon reasonable request.

CRediT authorship contribution statement

Mohamed Wahba: Writing – review & editing, Writing – original draft, Validation, Software, Investigation, Data curation, Conceptualization. **Radwa Essam:** Writing – original draft, Validation, Formal analysis, Data curation, Conceptualization. **Mustafa El-Rawy:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Nassir Al-Arifi:** Visualization, Validation, Software, Investigation, Funding acquisition, Conceptualization. **Fathy Abdalla:** Writing – review & editing, Writing – original draft, Validation, Funding acquisition, Data curation, Conceptualization. **Wael M. Elsadek:** Visualization, Validation, Supervision, Formal analysis, Conceptualization.

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.

Acknowledgments

The authors are grateful to the Deanship of Scientific Research, King Saud University for funding through Vice Deanship of Scientific Research Chairs.

References

[1] M. Pregnolato, A. Ford, S.M. Wilkinson, R.J. Dawson, The impact of flooding on road transport: a depth-disruption function, Transp. Res. D Transp. Environ. 55 (2017) 67–81, https://doi.org/10.1016/j.trd.2017.06.020.
[2] N. Vivekanandan, Comparison of probability distributions in extreme value analysis of rainfall and temperature data, Environ. Earth Sci. 77 (2018) 5, https://doi.org/10.1007/s12665-018-7356-z.
[3] Y. Huang, A. Bárdossy, K. Zhang, Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data, Hydrol. Earth Syst. Sci. 23 (2019) 2647–2663, https://doi.org/10.5194/hess-23-2647-2019.
[4] S.S. Wahla, J.H. Kazmi, A. Sharifi, S.A. Shirazi, A. Tariq, S.H. Joyell, Assessing spatio-temporal mapping and monitoring of climatic variability using SPEI and RF machine learning models, Geocarto Int. 37 (2022) 14963–14982, https://doi.org/10.1080/10106049.2022.2093411.
[5] M. Wahba, H. Mahmoud, W.M. Elsadek, S. Kanae, H.S. Hassan, Alleviation approach for flash flood risk reduction in urban dwellings: a case study of Fifth District, Egypt, Urban Clim. 42 (2022), https://doi.org/10.1016/j.uclim.2022.101130.
[6] M. Wahba, H.S. Hassan, W.M. Elsadek, S. Kanae, M. Sharaan, Prediction of flood susceptibility using frequency ratio method: a case study of fifth district, Egypt, 14th International Conference on Hydroscience & Engineering (2022).
[7] M. Liu, N. Chen, Y. Zhang, M. Deng, Glacial Lake inventory and lake outburst flood/debris flow hazard assessment after the Gorkha earthquake in the Bhote Koshi basin, Water 12 (2020) 464, https://doi.org/10.3390/w12020464.
[8] M. Zhao, Y. Zhou, X. Li, C. Zhou, W. Cheng, M. Li, K. Huang, Building a series of consistent night-time light data (1992–2018) in Southeast Asia by integrating DMSP-OLS and NPP-VIIRS, IEEE Trans. Geosci. Rem. Sens. 58 (2019) 1843–1856, https://doi.org/10.1109/TGRS.2019.2949797.
[9] M.M. Mansour, M.G. Ibrahim, M. Fujii, M. Nasr, Sustainable development goals (SDGs) associated with flash flood hazard mapping and management measures through morphometric evaluation, Geocarto Int. 37 (2022) 11116–11133, https://doi.org/10.1080/10106049.2022.2046868.
[10] M. Wahba, M. Sharaan, W.M. Elsadek, S. Kanae, H.S. Hassan, Building information modeling integrated with environmental flood hazard to assess the building vulnerability to flash floods, Stoch. Environ. Res. Risk Assess. (2024) 1–21.
[11] Abdelkarim, A.F.D. Gaber, Flood risk assessment of the Wadi Nu'man basin, Mecca, Saudi Arabia (during the period, 1988–2019) based on the integration of geomatics and hydraulic modeling: a case study, Water 11 (2019) 1887, https://doi.org/10.3390/w11091887.
[12] S.S. Alarifi, M. Abdelkareem, F. Abdalla, M. Alotaibi, Flash flood hazard mapping using remote sensing and GIS techniques in Southwestern Saudi Arabia, Sustainability 14 (2022) 14145, https://doi.org/10.3390/su142114145.
[13] R. Costache, Flood susceptibility assessment by using bivariate statistics and machine learning models - a useful tool for flood risk management, Water Resour. Manag. 33 (2019) 239–3256, https://doi.org/10.1007/s11269-019-02301-z.
[14] M. El-Rawy, W.M. Elsadek, F. De Smedt, Flash flood susceptibility mapping in Sinai, Egypt using hydromorphic data, principal component analysis and logistic regression, Water 14 (2022) 2434, https://doi.org/10.3390/w14152434.
[15] M. El-Rawy, W.M. Elsadek, F. De Smedt, Flood hazard assessment and mitigation using a multi-criteria approach in the Sinai Peninsula, Egypt, Nat. Hazards 115 (2023) 215–236, https://doi.org/10.1007/s11069-022-05551-0.
[16] Y. Kanani-Sadat, R. Arabsheibani, F. Karimipour, M. Nasseri, A new approach to flood susceptibility assessment in data-scarce and ungauged regions based on GIS-based hybrid multi criteria decision-making method, J. Hydrol. 572 (2019) 17–31, https://doi.org/10.1016/j.jhydrol.2019.02.034.
[17] H. Hong, M. Panahi, A. Shirzadi, T. Ma, J. Liu, A.X. Zhu, W. Chen, I. Kougias, N. Kazakis, Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution, Sci. Total Environ. 621 (2018) 1124–1141, https://doi.org/10.1016/j.scitotenv.2017.10.114.
[18] H. Darabi, B. Choubin, O. Rahmati, H.A. Torabi, B. Pradhan, B. Kløve, Urban flood risk mapping using the GARP and QUEST models: a comparative study of machine learning techniques, J. Hydrol. 569 (2019) 142–154, https://doi.org/10.1016/j.jhydrol.2018.12.002.
[19] D.J. Booker, T.H. Snelder, Comparing methods for estimating flow duration curves at ungauged sites, J. Hydrol. 434–435 (2012) 78–94, https://doi.org/10.1016/j.jhydrol.2012.02.031.
[20] O. Rahmati, H.R. Pourghasemi, H. Zeinivand, Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran, Geocarto Int. 31 (2016) 42–70, https://doi.org/10.1080/10106049.2015.1041559.
[21] N. Nagumo, M. Ohara, D. Kuribayashi, H. Sawano, The 2015 flood impact due to the overflow and Dike breach of Kinu River in Joso City, Japan, J. Disaster Res. 11 (2016) 1112–1127, https://doi.org/10.20965/jdr.2016.p1112.

[22] H. Yokoki, T. Uchida, Y. Akamatsu, S. Seto, S. Onda, T. Yamada, S. Nishimura, T. Tebakari, C. Fujiyama, H. Sakakibara, Y. Shigihara, S. Yokojima, Editorial-special issue on the storm and flood damage in Japan, J. JSCE 10 (2022) 1–7, https://doi.org/10.2208/journalofjsce.10.1_1.
[23] Diva-gis, https://www.diva-gis.org/Data. Accessed on 10 October 2023.
[24] Geofabrik, http://download.geofabrik.de/index.html. Accessed on 10 October 2023.
[25] Yamazaki Lab, http://hydro.iis.u-tokyo.ac.jp/~yamadai/JapanDir/index.html. Accessed on 10 October 2023.
[26] https://livingatlas.arcgis.com/. Accessed on 10 October 2023.
[27] https://disaportal.gsi.go.jp/index.html. Accessed on 10 October 2023.
[28] O. Rahmati, H.R. Pourghasemi, H. Zeinivand, Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran, Geocarto Int. 31 (2016) 42–70, https://doi.org/10.1080/10106049.2015.1041559.
[29] M.S. Tehrany, B. Pradhan, M.N. Jebur, Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS, J. Hydrol. 512 (2014) 332–343, https://doi.org/10.1016/j.jhydrol.2014.03.008.
[30] D. Tien Bui, N.-D. Hoang, A Bayesian framework based on a Gaussian mixture model and radial-basis-function Fisher discriminant analysis (BayGmmKda V1.1) for spatial prediction of floods, Geosci. Model Dev. (GMD) 10 (2017) 3391–3409, https://doi.org/10.5194/gmd-10-3391-2017.
[31] P.V. Gorsevski, P. Gessler, R.B. Foltz, Spatial prediction of landslide hazard using logistic regression and GIS, 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4) 110 (2020) 4–5.
[32] M.B. Kia, S. Pirasteh, B. Pradhan, A.R. Mahmud, W.N.A. Sulaiman, A. Moradi, An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia, Environ. Earth Sci. 67 (2012) 251–264, https://doi.org/10.1007/s12665-011-1504-z.
[33] M.S. Tehrany, B. Pradhan, M.N. Jebur, Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS, J. Hydrol. 512 (2014) 332–343, https://doi.org/10.1016/j.jhydrol.2014.03.008.
[34] W.M. Elsadek, M. Wahba, N. Al-Arifi, S. Kanae, M. El-Rawy, Scrutinizing the performance of GIS-based analytical Hierarchical process approach and frequency ratio model in flood prediction – case study of Kakegawa, Japan, Ain Shams Eng. J. (2023) 102453, https://doi.org/10.1016/j.asej.2023.102453.
[35] S. Samanta, D.K. Pal, D. Lohar, B. Pal, Interpolation of climate variables and temperature modeling, Theor. Appl. Climatol. 107 (2012) 35–45, https://doi.org/10.1007/s00704-011-0455-3.
[36] N.N. Kourgialas, G.P. Karatzas, A flood risk decision making approach for Mediterranean tree crops using GIS; climate change effects and flood-tolerant species, Environ. Sci. Pol. 63 (2016) 132–142, https://doi.org/10.1016/j.envsci.2016.05.020.
[37] J.J. Opperman, G.E. Galloway, J. Fargione, J.F. Mount, B.D. Richter, S. Secchi, Sustainable floodplains through large-scale reconnection to rivers, Science 326 (2009) 1487–1488, https://doi.org/10.1126/science.1178256.
[38] M.S. Tehrany, S. Jones, F. Shabani, Identifying the essential flood conditioning factors for flood-prone area mapping using machine learning techniques, Catena 175 (2019) 174–192, https://doi.org/10.1016/j.catena.2018.12.011.
[39] K. Khosravi, M. Panahi, A. Golkarian, S.D. Keesstra, P.M. Saco, D.T. Bui, S. Lee, Convolutional neural network approach for spatial prediction of flood hazard at the national scale of Iran, J. Hydrol. 591 (2020), https://doi.org/10.1016/j.jhydrol.2020.125552.
[40] A.A. Komolafe, S. Herath, R. Avtar, Methodology to assess potential flood damages in urban areas under the influence of climate change, Nat. Hazards Rev. 19 (2018), https://doi.org/10.1061/(ASCE)NH.1527-6996.00002.
[41] O. Huisman, R.A. de By, Principles of geographic information systems, ITC Educational Textbook Series 1 (2009) 17.
[42] R.B. Mudashiru, N. Sabtu, R. Abdullah, A. Saleh, I. Abustan, Optimality of flood influencing factors for flood hazard mapping: an evaluation of two multi-criteria decision-making methods, J. Hydrol. 612 (2022) 128055, https://doi.org/10.1016/j.jhydrol.2022.128055.
[43] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[44] Trevor Hastie, Robert Tibshirani, Jerome H. Friedman, Jerome H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2, springer, New York, 2009.
[45] A. Cutler, D.R. Cutler, J.R. Stevens, in: C. Zhang, Y.Q. Ma (Eds.), Random Forests, Ensemble Machine Learning, New York, 2012, pp. 157–175, https://doi.org/10.1007/978-1-4419-9326-7_5. Springer.
[46] M. Belgiu, L. Drăguţ, Random forest in remote sensing: a review of applications and future directions, ISPRS J. Photogrammetry Remote Sens. 114 (2016) 24–31, https://doi.org/10.1016/j.isprsjprs.2016.01.011.
[47] Z. Wang, C. Lai, X. Chen, B. Yang, S. Zhao, X. Bai, Flood hazard risk assessment model based on random forest, J. Hydrol. 527 (2015) 1130–1141, https://doi.org/10.1016/j.jhydrol.2015.05.017.
[48] A. Prinzie, D. Van den Poel, Random multiclass classification: generalizing random forests to random MNL and random NB, in: International Conference on Database and Expert Systems Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 349–358.
[49] S. Tesfamariam, Z. Liu, Earthquake induced damage classification for reinforced concrete buildings, Struct. Saf. 32 (2) (2010) 154–164, https://doi.org/10.1016/j.strusafe.2009.09.006.
[50] L. Dong, X. Li, P.E.N.G. Kang, Prediction of rockburst classification using Random Forest, Trans. Nonferrous Metals Soc. China 23 (2) (2013) 472–477, https://doi.org/10.1016/S1003-6326(13)62509-2.
[51] X. Chen, H. Ishwaran, Random forests for genomic data analysis, Genomics 99 (6) (2012) 323–329, https://doi.org/10.1016/j.ygeno.2012.04.003.
[52] D.M. Mihailescu, V. Gui, C.I. Toma, A. Popescu, I. Sporea, Computer-aided diagnosis method for steatosis rating in ultrasound images using random forests, Medical Ultrasonography 15 (3) (2013) 184–190, https://doi.org/10.11152/mu.2013.2066.153.dmm1.
[53] H. Farhadi, M. Najafzadeh, Flood risk mapping by remote sensing data and random forest technique, Water 13 (21) (2021) 3115, https://doi.org/10.3390/w13213115.
[54] C. Cao, P. Xu, Y. Wang, J. Chen, L. Zheng, C. Niu, Flash flood hazard susceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas, Sustainability 8 (2016) 948, https://doi.org/10.3390/su8090948.
[55] N. Kazakis, I. Kougias, T. Patsialis, Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: application in Rhodope-Evros region, Greece, Sci. Total Environ. 538 (2015) 555–563, https://doi.org/10.1016/j.scitotenv.2015.08.055.
[56] M. Wahba, H.S. Hassan, W.M. Elsadek, S. Kanae, M. Sharaan, Novel utilization of simulated runoff as causative parameter to predict the hazard of flash floods, Environ. Earth Sci. 82 (2023) 333, https://doi.org/10.1007/s12665-023-11007-w.
[57] M. Wahba, M. El-Rawy, N. Al-Arifi, M. Mansour, A novel estimation of the composite hazard of landslides and flash floods utilizing an artificial intelligence approach, Water 15 (23) (2023) 4138, https://doi.org/10.3390/w15234138.
[58] M. Wahba, S. Mahmoud, W.M. Elsadek, S. Kanae, H.S. Hassan, Categorization of urban basin according to the runoff depth: case study of Katsushika ward and Edogawa City basin, Japan, in: Asia Conference on Environment and Sustainable Development, Springer Nature Singapore, 2022, pp. 131–142.
[59] A.N.I. Obi, O. Ubani, Analysis of flood risk management in Nigerian urban environment, J. Environ. Earth Sci. 4 (6) (2014) 2224–3216.
[60] P.V. Gorsevski, P. Gessler, R.B. Foltz, Spatial prediction of landslide hazard using logistic regression and GIS, in: 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4), 2000.
[61] H. Wang, H. Zheng, True positive rate, in: Encyclopedia of Systems Biology, Springer, New York, 2013, pp. 2302–2303, https://doi.org/10.1007/978-1-4419-9863-7_255.
[62] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (2006) 861–874, https://doi.org/10.1016/j.patrec.2005.10.010.
[63] J. Fan, S. Upadhye, A. Worster, Understanding receiver operating characteristic (ROC) curves, CJEM 8 (2006) 19–20, https://doi.org/10.1017/S1481803500013336.
[64] B. Choubin, M. Borji, A. Mosavi, F. Sajedi-Hosseini, V.P. Singh, S. Shamshirband, Snow avalanche hazard prediction using machine learning methods, J. Hydrol. 577 (2019), https://doi.org/10.1016/j.jhydrol.2019.123929.
[65] T. Tingsanchali, Urban Flood Disaster Management. Procedia Engineering, Elsevier Ltd, 2012, pp. 25–37, https://doi.org/10.1016/j.proeng.2012.01.1233.
[66] H. Kreibich, P. Bubeck, M. Van Vliet, H. De Moel, A review of damage-reducing measures to manage fluvial flood risks in a changing climate, Mitig. Adapt. Strategies Glob. Change 20 (2015) 967–989, https://doi.org/10.1007/s11027-014-9629-5.

[67] B. Merz, A.H. Thieken, M. Gocht, Flood risk mapping at the local scale: concepts and challenges, Flood risk management in Europe: innovation in policy and practice (2007) 231–251, https://doi.org/10.1007/978-1-4020-4200-3_13.
[68] G. El Afandi, M. Morsy, Developing an early warning system for flash flood in Egypt: case study Sinai Peninsula, in: A. Negm (Ed.), Flash Floods in Egypt. Advances in Science, Technology & Innovation, Springer, Cham, 2020, https://doi.org/10.1007/978-3-030-29635-3_4.
[69] A. Bhusal, A. Kalra, S. Shin, Resilience effect of decentralized detention system to extreme flooding events, J. Hydroinf. 25 (2023) 971–988, https://doi.org/10.2166/hydro.2023.176.
[70] C. Hauer, P. Flödl, H. Habersack, U. Pulg, Critical flows in semi-alluvial channels during extraordinarily high discharges: implications for flood risk management, J. Flood Risk Manag. 14 (2021), https://doi.org/10.1111/jfr3.12741.
[71] A.R. McFarland, L. Larsen, K. Yeshitela, A.N. Engida, N.G. Love, Guide for using green infrastructure in urban environments for stormwater management, Environ. Sci.: Water Res. Techno. 5 (2019) 643–659, https://doi.org/10.1039/c8ew00498f.
[72] R.R. Rudra, S.K. Sarkar, Artificial neural network for flood susceptibility mapping in Bangladesh, Heliyon 9 (6) (2023 Jun 1) e16459, https://doi.org/10.1016/j.heliyon.2023.e16459.
[73] M. Avand, H. Moradi, Spatial modeling of flood probability using geo-environmental variables and machine learning models, case study: Tajan watershed, Iran, Adv. Space Res. 67 (10) (2021 15) 3169–3186, https://doi.org/10.1016/j.asr.2021.02.011.
[74] M. El-Rawy, M. Wahba, H. Fathi, F. Alshehri, F. Abdalla, R.M. El Attar, Assessment of groundwater quality in arid regions utilizing principal component analysis, GIS, and machine learning techniques, Mar. Pollut. Bull. 205 (2024 1) 116645, https://doi.org/10.1016/j.marpolbul.2024.116645.
[75] M. Wahba, M. El-Rawy, N. Al-Arifi, Integrating geographic information systems and hydrometric analysis for assessing and mitigating building vulnerability to flash flood risks, Water 16 (3) (2024) 434, https://doi.org/10.3390/w16030434, 29.