Optimal flood susceptibility model based on performance comparisons of LR, EGB, and RF algorithms

Ahmed M. Youssef

Saudi Geological Survey

Ali M. Mahdi

South Valley University

Hamid Reza Pourghasemi (✉ hamidreza.pourghasemi@yahoo.com)

Shiraz University <https://orcid.org/0000-0003-2328-2998>

Research Article

Keywords: Flood hazard, EGB, Sustainability, Sentinel-1 images, Egypt

Posted Date: February 10th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1312161/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The Wadi El-Matulla, located in the eastern desert of Egypt, is the most important water basin. The Qift - Qusayr highway (west-east direction) and the Cairo - Aswan eastern desert highway (north-south) pass through the basin. Many urban areas (villages and industrial areas) and agricultural lands are located at the outlet of these basins. In addition, the basin has a promising potential for future economic and urban development as it is located within the Golden Triangle (governmental megaproject). The current study investigates flood hazard modeling and its impact on the area. To determine the optimal algorithm for flood susceptibility mapping, performance comparisons of three techniques were conducted: logistic regression (LR), extreme gradient boosting (EGB), and random forest (RF). Remote sensing, topographic, geologic, and meteorological data were used with the help of field visits to provide spatial database and inventory datasets required by the models. The models' performances were assessed using five statistical indices, ROC, overall accuracy (OAC), kappa index (Ka), RMSE, and mean absolute error (MAE). Our findings indicate that RF is the optimal algorithm for flood susceptibility mapping, followed by EGB and LR. RF provides high performance with high values of ROC (93%), OAC (88%), and Kappa index (0.85) and the lowest values of RMSE (0.34) and MAE (0.12). Finally, the RF model was verified using sentinel-1 images for real floods in 2016 and 2021, and it provides good agreement. The optimal model could help the decision makers and planners to protect the existing facilities and plan the future projects in non-flood prone areas.

Introduction

Nowadays, floods are a global hazard due to climate change and environmental variability. They can exhibit a significant threat to human life and property and cause damage to the physical environment (agricultural land, urban facilities and infrastructures), breakdown of social ties and economic losses (Thomas, 2017; Waqas et al., 2021). Many factors can contribute to flooding, including heavy rainfall (an important factor), inadequate drainage systems to drain excess water or lack of preventive measures, human activities (land use planning, agricultural practices, and built facilities) (Zhao et al., 2019; Ullah and Zhang, 2020). Gao et al. (2019) pointed out that from 1960 to 2014, many disasters occurred worldwide, including 2,171 floods, droughts, and extreme hydrological events. From 2018 to 2021, about 51% of people were displaced due to floods, 35% due to storms, 12% due to earthquakes and tsunamis, and less than 2% due to other types of hazards (Uddin and Matin, 2021). It is an urgent task to create a safe environment for people and property by minimizing the flood risk of areas. To achieve this, flood susceptibility mapping needs to be done as early as possible, especially for future planning areas (Uddin and Matin, 2021). With the advantages of "geographic information systems (GIS) and remote sensing (RS)", susceptibility maps are recently becoming very popular. Remote sensing helps to collect various types of spatial information showing the characteristics of topographic surfaces, vegetation cover, changes in land use over time, and the effects of climate change (Hussain et al., 2020). In addition, the methods of GIS help to create a spatial database, which is considered as the cornerstone for flood susceptibility analysis and delineating of flood-prone zones (Zhao et al., 2019). Various methods have

been developed to map flood-susceptible areas, including the “analytical hierarchy process (AHP) (Seejata et al., 2018; Das and Gupta, 2021); frequency ratio (FR) and index of entropy (IoE) (Feizizadeh et al., 2021), logistic regression (LR)” (Band et al. 2020; Malik et al. 2020), the extreme gradient boosting (EGB) (Naghibi et al., 2020; Mirzaei et al., 2021), and random forest (RF) (Farhadi and Najafzadeh, 2021).

The study of flood hazards is one of the most important and complex applied geomorphological studies for which various datasets are required. Through the application of RS and GIS, these data are available to perform such flood hazard models with a high degree of accuracy in predicting the flood-prone areas. Egyptian authority has elected the Wadi El-Matulla and its surroundings to be an area for future development (agriculture, industrial, and urban). Various researchers were conducted in this area and its surrounding such as geomorphological studies (El-Shamy, 1985; Embabi, 2004) and hydrological conditions (Abd-Allatief et al., 2012). Recently, Egyptian government has chosen this area to be one of the mega projects named the Golden Triangle, subsequently, several studies have been conducted in this region to investigate its groundwater characteristics (Abdelkareem and El-Baz 2015; Gaber et al, 2018). However, to ensure the sustainability of this project, natural hazard evaluation is a must. Unfortunately, hazard vulnerability mapping in Egypt is still new subject with limited studies have been conducted so far (e.g. Abdrabo et al., 2020; Prama et al., 2020; Saber et al., 2021; Hermas et al., 2021). However, the flood susceptibility assessment is a new subject in Egypt and Middle East regions.

The aim of the current work was the development and application of quantitative analysis techniques with the integration of GIS for mapping flood-susceptible in the Wadi El-Matulla and its surroundings catchments, Egypt. The area is promising for agriculture and new development activities as it is part of the Golden Triangle. In addition, the wadis are adjacent to several urban and industrial areas that could be affected by flooding. In order to avert flood hazards and protect the existing facilities and future developments, flood susceptibility modeling was carried out based on the application of multivariate (LR) and machine learning techniques using “Extreme Gradient Boosting (EGB) and Random Forest (RF)”. Also, the results of different models are evaluated to select the optimal model based on various statistical indices. These results will be useful to planners, investors, researchers and decision makers in impact assessment to predict the susceptible flood zones in the future. They can also suggest various mitigation strategies and contribute to sustainable development of the basin.

Study Area Characteristics

Wadi El-Matulla catchment (length is approximately 130km) and its surrounding are located in the eastern desert of Egypt, starting from the “Red Sea Mountains”, crossing the sandy Ababda plateau. Wadi El-Matulla tributaries (e.g., Wadi El-Hammamat, Wadi Zaidun, Wadi El-Mishash and Wadi El-Muweih) drain the catchment area during rainstorms towards the Nile Valley (westwards) where many urban areas are located such as the Al-Kalahin village and the town of Qift. The area is situated between longitudes of 32°50'31" and 34°06'31" and latitudes of 25°26'34" and 26°19'56". Thus, its total area is almost 7,231 km² (Fig. 1). Also, this wadi is characterized by numerous ancient monuments that located along the

ancient Qift-Qusier road, including numerous remains of ancient wells from the Pharaonic and Roman periods (e.g., the Hammam Cleopatra, Bir Lakita and Bir Fawakhier).

Geologically, the rocks in the basin area vary between igneous, metamorphic and sedimentary rocks (Conoco, 1987; Said 1993), where the basin is divided into three terrain zones, namely: the Red Sea mountain range occupying the eastern part of the basin and covering (40% of the basin area). This zone is characterized by many mountains such as mountains Matiq 1,112m above sea level (asl), and Nasib Azraq (1,062 m asl). The area is dissected by many normal faults and joints and is crossed by numerous wadis (valleys). The second zone is the plateau zone (Ababda Plateau), which occupies the central and western part of the basin area covering 50% of the area. This zone is composed of sedimentary rocks, mainly sandstone and chalk (Upper Cretaceous) and limestone (Paleocene, Eocene and Pliocene). The plateau is a semi-flat surface and covered by a few isolated mounds. This plateau dissected by numerous wadis. The third zone is the lowest part of the a basin, which accounts for 10% of its area, consists mainly of the valley floor of the main wadis and the fan deposits of the some tributary basins that flow into this zone (mainly filled by thick sand and gravel deposits). Its height ranges from 76 to 160 m, and its surface is semi-level with very low gradient.

Based on meteorological investigation, the catchment area belongs to arid to hyper-arid zone, which characterized by erotic precipitation with an average annual rainfall of less than 100 mm (Sultan et al. 2008). According to the rainstorms, the catchment experienced various flash flood events (e.g., 2016 and 2021). These flooding events are common due to short duration severe storms caused by the high elevation of the Red Sea Hills.

Materials And Method

Data used

Several datasets were used for the current study, including: 1) Digital Elevation Models (DEM): DEM (12.5 resolution) of PALSAR ALOS was acquired in 2020 (<https://asf.alaska.edu/data-sets>). It was used to extract various important parameters for the flood conditions (geomorphological layers), including eight parameters such as elevation, slope angle, plan and profile curvatures using ArcGIS (v.10.8), as well as topographic position index (TPI), convergence index (CI), slope length (LS), and topographic witness index (TWI) using SAGA (v8.0) software. 2) Topographic and geological maps: In the current study, different topographic maps (Egyptian Survey Authority, 1990) at 1:50,000-scale and geological maps (Conoco, 1987) at 1:250,000-scale were used to extract drainage networks and lithological units. 3) Satellite and Google Earth imagery: The current study relies on three types of remote sensing data, including passive satellite imagery, where Sentinel-2 (<https://earthexplorer.usgs.gov>) was utilized to extract NDVI and land use/land cover with a spatial accuracy of 10m. The other type of remote sensing imagery used in this study is Sentinel-1 (SAR), acquired from Sentinels Scientific Data Hub (<https://apps.sentinel-hub.com>). These images were used to identify flooded areas and create a dataset for the inventory. The third type of data is the use of Google Earth imagery, which was used to map the

various flood locations related to the previous events. 4) Metrological data: it was used to estimate the distribution of rainfall depth in the catchment area, three meteorological stations were identified, namely: Qena, Luxor and Qusayr, which were obtained from the “Water Resources Research Institute of the Ministry of Water Resources and Irrigation”. 5) Anthropogenic parameters: roads were extracted from government data and high resolution Google Earth images. 6) Field study: various field studies were carried out between 2018 and 2021 to study the geographical features of the catchment and verify the accuracy of the results (flooded areas extracted from historical records and from Google Earth and radar imageries).

Seven steps are required for flood susceptibility mapping, as follow: 1) data collection from various sources; 2) production of a flood map (inventory map) based on remote sensing imagery, historical records, and field visits; 3) extraction of various flood-conditioning factors and construction of a database using ArcGIS 10.8; 4) application of multivariate (LR) and machine learning models (EGB and RF); 5) construction of flood susceptibility maps; 6) validation of flood models using ROC-AUC and other statistical parameters; and 7) selection of optimal flood model for flood management analysis (Fig. 2).

Inventory map (dependent factor)

Historical and current flood events are crucial for the preparation of the flood inventory, which serves as the basis for flood susceptibility modeling to predict the potentially vulnerable areas for future floods (Sarkar and Mondal, 2020). With the advantages of remote sensing applications, “C-Band Sentinel-1 Synthetic Aperture Radar (SAR)” has been used to detect flooded areas after flood events. These data are available with no charge, less amount of revisit time, works with all-weather conditions, and sensitive to water bodies and wetland areas (Voormansik et al., 2014; Filion et al., 2016; Anusha and Bharathi 2019). Two images from Sentinel-1 SAR were acquired from Sentinels Scientific Data Hub (<https://apps.sentinel-hub.com>) covering the study area and surrounding areas. These images were acquired on 25 February 2016 (flood event on 14 February 2016) and 19 November 2021 (flood event on 13 November 2021). The data were processed using ENVI v.5.5 and ArcGIS v.10.8 software. The data used in the current study were corrected for speckle and noise effects. Most of the applications of polarimetry in the current study are wetland and flooded area identification. Accordingly, Sentinel-1 imagery was acquired after the 2016 and 2021 flood events. Sentinel-1 is dual polarization system with two transmitting and receiving signals in either horizontal (H) or vertical (V) polarization. Different objects on the ground (forest canopy and ocean surface) can have different characteristics leading to produce distinct polarization signature (such as, different reflection, intensities and exchange polarization between H and V). Polarimetric techniques can be used to separate different scattering leading to provide information about these various objects. In the current study, the Sentinel-1 image of the study area shows the composite RGB (color) image generated using various channels such as VV, VH, and VV/VH ratio for red, green, and blue, respectively. The RGB color composite of Sentinel-1 SAR imageries of the current basin area have been used to highlight certain features of the flooded areas after the flood events (downstream area of the El-Matulla basin (Fig. 3 (a, c)). Additionally, the Google Earth program was applied to detect and visualize past floods. Google Earth has a time-track feature that allows us to go back in time to see the impact of flooding on the affected areas. In the current study, Google Earth has been applied to extract the flooded

locations after the two flood events of 2016 and 2021 (Figure 3 (b, d)), where the signature of flowing water is represented by white color (upstream section of the wadi). Field photos were also taken after the November 13, 2021 flood event (Figure 3 (e-f)). As can be seen from the photos, the flood waters flowed through the villages and isolated them. People were afraid of drowning during the flood, accordingly they were standing above their houses, and it seems that after the flood, the community tried to build some dikes to protect themselves from the rains. Based on all this data obtained from Sentinel-1 imagery, Google Earth, archived surveys, and other historical records, a flood database was created in ArcGIS 10.8. In this work, 480 flood points and non-flood points were prepared (Fig. 2). These data were randomly separated into two datasets called training (70% flood and non-flood) to build the flood susceptibility models. The remaining datasets (30% of flood and non-flood) were used for model testing and to help in selecting the optimal model based on accuracy and performance evaluations (Islam et al., 2021; Tang et al., 2021).

Flood-conditioning factors (independent factors)

Flood susceptibility model is relying on two main pillars, the first one is the dependent factor (inventory map) and the other one is the independent “flood-conditioning factors” (Rahman et al., 2019; Yagoub et al., 2020; Waqas et al., 2021). In this study, 15 “flood-conditioning factors” have been extracted from various data sources and stored in database layers in the ArcGIS 10.8 with a final pixel size of 12.5-m * 12.5-m. These flood-conditioning factors include; elevation (EL) (Fig. 4a), slope-angle (SA) (Fig. 4b), topographic positioning index (TPI) (Fig. 4c), convergence index (CI) (Fig. 4d), slope length (LS) (Fig. 4e), topographic wetness index (TWI) (Fig. 4f), profile curvature (PrC) (Fig. 4g), plan curvature (PIC) (Fig. 4h), drainage density (DD) (Fig. 4i), geology (GEO) (Fig. 4j), “normalized differential vegetation index (NDVI)” (Fig. 4k), land use/ land cover (LULC) (Fig. 4l), distance to roads (DtR) (Fig. 4m), distance to wadis (DtW) (Fig. 4), and rainfall (RF) (Fig. 4o). In the current work, the weight and the contribution of every flood-related factor is presented in Table (1). The parameters TWI and LS were determined using equations (1 and 2).

$$TWI = \ln \frac{A}{\tan B}$$

1

$$LS = \ln \frac{A}{\tan B}$$

2

where “A is the catchment area, and β is the slope angle (in °)” (Wolock and McCabe, 1995)).

The NDVI index map was produced based on the relationship between band eighth (near-infrared electromagnetic radiation (wave length of NIR = 842nm)) and band fourth (red electromagnetic radiation (wave length of R = 665nm)) (equ. 3). The plant has a very strong near-infrared absorption, and therefore they are used in calculating the NDVI (Rouse et al., 1974).

$$NDVI = \frac{B8 - B4}{B8 + B4}$$

3

Topographic positioning index was calculated based on equation (4) (Guisan et al., 1999)

$$TPI = Z_o - \frac{\sum_{i=1}^n Z_i}{n}$$

4

where Z_o is the height of the desired cell and $\frac{\sum_{i=1}^n Z_i}{n}$ is the average height of neighboring cells.

ArcGIS 10.8 was used to create the stream density layer at 12.5 m resolution using the line density tool, while the distance to roads (DTR) and distance to wadis (DTW) were created using the Euclidean distance tool. Different LULC types in a watershed can significantly affect flood susceptibility and were extracted based on a supervised classification of the Sentinel-2A image. The LULC map was classified into five categories, including agriculture, urban, sparsely tree cover, wadi deposits, and barren rocks. To estimate the amount of rainfall in the 100-year return periods, the rainfall data recorded in the three stations were collected, which are daily data for the period between 1970-2018. Statistical probability distribution methods were applied to select the suitable distribution for the rainfall data of the studied meteorological stations. Our study shows that the generalized extreme value method (GEV) was suitable for the Qena and Qusayr stations; and Person type III was suitable for Qusayr station. The 100-year return period rainfall values were 70, 57, and 45 for Qena, Luxor, and Qusayr stations respectively. The “inverse distance weighting (IDW)” algorithm was used to generate the rain distribution map as it is the most common technique for interpolating scatter points (Kilinc, 2018).

Background of the algorithms

Flood susceptibility mapping is crucial to identify the vulnerable zones of potential flooding and could be created by identifying the relationship between actual flood locations and associated causal factors. Three models were used in this study including logistic regression model (LR), Extreme Gradient Boosting (EGB) and Random Forest (RF).

Logistic Regression (LR).

LR is a common multivariate algorithm that identifies the regression relationship between independent factors (two or more factors) and a dependent variable (one factor) (Liao et al., 1988). It calculates the probability that an event will occur versus the probability that an event will not occur (Sofia et al., 2018). It was used as a proven approach for flood susceptibility mapping and identify the most contribution variables (Wubalem et al., 2020). Equation (5) shows the relationship between flood events and the conditioning factors.

$$P = \frac{1}{(1 + e^{-z})}$$

5

Where P is the probability of flooded or non-flooded area. The coefficients are determined as the best mathematical fit for the specified model. A coefficient indicates the influence of each independent variable on the outcome variable, taking into account all other independent variables. This technique has been adapted as an accurate hazard prediction algorithm that can determine the probability of flooding (Band et al., 2020; Malik et al., 2020). The LR model was performed in the current study using SPSS (v.26) statistical software, with 15 flood-conditioning factors (independent factors) and flood and non-flood points (dependent factor). The predicted value of the model is the sum of the results of the individual products formed by multiplying the coefficient value and the independent variables, as shown in equation (6).

$$z = C + B1X1 + B2X2 + \dots + B15X15$$

6

Where B1 to B15 are the coefficient values and X1 to X15 are the flood related factors. The extracted linear model is LR as a function of the presence or absence of flooding as a function of the values of the independent factors associated with the occurrence of flooding in previous years. The probability of flood susceptibility is expressed by a value ranges from 0 to 1.

Extreme gradient boosting (EGB)

The EGB, a machine learning algorithm, is an advanced supervised algorithm, and was developed by Chen and Guestrin (2016). The major advantages of EGB technique are its ability to generate a strong learner from multiple weak learners outcomes, dealing with missing data in the datasets, and tuning the factors without overfitting the model, and using parallel processes to reduce the computation time (Fan et al., 2018; Naghibi et al., 2020). To conduct EGB model three steps are required as follow: 1) building the learning feature over the entire dataset of factors, 2) constructing the next model over the residuals, and 3) ending the procedure as it reaches "stopping criteria" (Fan et al. 2018). In the current study, the EGB was performed based on the open source code "xgboost" for R-software (Chen et al., 2015).

Random forest (RF)

RF was developed by (Breiman 2001) as an ensemble approach, is generated by "numerous decision trees as predictors and run for classification and regression analyses". Many authors pointed out that RF is a robust and flexible method where random trees are applied by a set of cases through a bootstrapping technique, where the cases not considered in the designed trees are out of the bag (Zhu and Zhang, 2021). RF possess many advantages as follow 1) ability to categorize parameters based on their contribution using "mean decrease accuracy (MDA) and mean decrease Gini (MDG)" (Wang et al. 2016) and the ability to process a large number of datasets and obtain satisfactory results (Rahman et al.

2019). In this study, the flood susceptibility map was created using the random Forest package in R software (Briman and Cutler, 2015). The final map was created and classified using ArcGIS 10.8.

Multicollinearity, models' validation and comparisons

Various methods are used to evaluate factor effectiveness in model construction, such as the multicollinearity method (Yoo and Cho 2019). The “variance inflation factors (VIF) and tolerance (TOL)” were applied to evaluate the effectiveness of factors (equations 7 and 8):

$$\text{TOL} = (1 - R_j^2)$$

7

$$\text{VIF} = \left[\frac{1}{(1 - R_j^2)} \right]$$

8

R_j^2 Represents the regression coefficient of the independent factor J on all other dependent factors. Rahman et al. (2019) indicate that a TOL < 0.10 and a VIF > 5 account for multicollinearity issues.

In addition, reliability of the models' performances can be measured using the “receiver operating characteristics (ROC) and the area under the curves (AUC)” method. It is a crucial step in flood susceptibility modeling (Metz, 1978). The “ROC-AUC” was used in this study to calculate the accuracy of the flood susceptibility maps, which are appropriate indicators and provide valuable information (Liu et al 2021). Cao et al. (2020) classified the AUC values into 4 zones as follow: less than 0.6 which indicates no scientific significance of the model (weak), between 0.6 and 0.7, which shows moderate significance model; from 0.7 and 0.8, represent good model; and above 0.8 indicating a very good model.

To compare between these susceptibility models' various arithmetic indices were applied using 1) Discrimination accuracy measures using (hazard accuracy (HAC), non-hazard accuracy (NHAC), overall hazard accuracy (OHAC), and kappa index (K)) and 2) Reliability assessment using “mean absolute error (MAE) and root mean square error (RMSE)” (Hembram et al., 2021). These indices are shown in equations 9-14. Hazard accuracy (HAC), non-hazard accuracy (NHAC), and overall hazard accuracy (OHAC) equations are determined based on the confusion matrix (Table 2) which depends on flood and non-flood points.

Table 1
Flood conditioning factors characteristics and contributions.

#	Dataset	Data type and scale/resolution	Impacts	Figure/Range
1	EI	Continuous (12.5m)	Elevation is a regulating factor for climatic conditions such as precipitation (Roundy and Chambers, 2021).	Figure 4a (72-1083m)
2	S	Continuous (12.5m)	The slope angle has tremendous effects on runoff, infiltration, and erosion rates. Water generally floods areas with low slope (Khan et al., 2016; Vojtek and Vojteková 2019).	Figure 4b (0-67 deg.)
3	TPI	Continuous (12.5m)	The topographic position index is defined as the height difference between adjacent pixels in the area (Weiss, 2001). If the height of the pixel is higher than the neighboring pixels, then it has a positive value, while a negative pixel means a lower height than the neighboring pixels, which generally represents the flooded areas (De Reu et al., 2013).	Figure 4c (-36.6 - 36.8)
4	CI	Continuous (12.5m)	Convergence index (CI) is a terrain parameter that describes ground topography. Negative values represent low-lying areas (streams), while positive values represent areas with steep slopes (ridges) (Grohmann and Riccomini, 2009). It plays a crucial role in determining the locations with the highest flood potential (Chowdhuri et al., 2020).	Figure 4d (-99.3 - 99.2)
5	LS	Continuous (12.5m)	Slope length (LS) affects water flow and erosion rate (Nguyen et al., 2020; Yariyan et al., 2020).	Figure 4e (0 - 877.3)
6	TWI	Continuous (12.5m)	TWI is a morphometric parameter that affects overland flow and soil moisture and water distribution and is used to delineate flood-prone areas (Mokarrama and Hojati, 2018; Ali et al. 2019).	Figure 4f (2.8 - 26.5)
7	PrC	Continuous (12.5m)	Profile and plan curvature are flood-related variables that indicate the direction of maximum slope of a surface. The surface is linear when the value is close to 0, while the concave surface (convergent) has values below 0 and the convex surface (divergent) has values above 0, which affects the runoff and concentration of water (Popa et al., 2019; Rejith et al., 2019).	Figure 4g, h
8	PIC	Continuous (12.5m)		PrC (-12.2 - 11.1) PIC (-11.1 - 8.6)
9	DD	Continuous (12.5m)	Sangireddy et al. (2016) defined it as the total length of streams in relation to the catchment area. It plays curtail role in flood susceptibility, as higher	Figure 4i (0 –

#	Dataset	Data type and scale/resolution	Impacts drainage density results in higher flood probability (Onuşluoğlu, 2013).	Figure/ Range
10	GEO	Categorical (1:250000)	Geology: influence on hydrological process (Zhao et al. 2019).	Figure 4j 15 units
11	NDVI	Continuous (10m)	Normalized differential vegetation index assesses vegetation distribution and its impact and role in runoff and flood hazard (Ullah and Zhang, 2020).	Figure 4k (-0.51 - 0.79)
12	LULC	Continuous (10m)	LULC have a major influence on inundation potential and affect hydrological processes (Hölting and Coldewey, 2019; Yariyan et al., 2020). Urban expansion due to human activities will increase the rate of runoff (Lei et al., 2021).	Figure 4l 5 units
13	DTR	Polyline (1: 50000)	Roads: Human activities can influence flooding as barriers (Nyssen et al., 2002; Sarkar et al., 2020). Road networks can also increase impervious surfaces that reduce infiltration and increase runoff (Swain et al., 2020).	Figure 4m (0– 58349m)
14	DTW	Polyline (1: 50000)	Floodplains are most susceptible to flooding, so distance from adjacent wadis is critical in flood susceptibility analysis (Mignot et al., 2019; Sarkar et al., 2020).	Figure 4n (0– 12479m)
15	R	Points (Rain stations)	Precipitation is a major factor causing flooding, and many studies have shown that it is a triggering factor that can contribute greatly to flooding potential (Wang et al., 2021). Rainfall is essential factor that cause inundation and many studies indicated that it is a trigger factor that can strongly contribute to the flooding potential (Wang et al., 2021).	Figure 4o (47 – 70mm)

Table 2
Confusion matrix used to quantify different accuracy

Predicted data	Observed data	
	Flood	Non-flood
Flood	True flood (TF)	False non-flood (FNF)
Non-flood	False flood (FF)	True non-flood (TNF)

$$HAC = \frac{(TH)}{(TH + FH)}$$

$$NHAC = \frac{(TNH)}{(FNH + TNH)}$$

10

$$OHAC = \frac{(TH + TNH)}{(TH + TNH + FH + FNH)}$$

11

$$MAE = \frac{1}{n} \sum_{i=1}^{i=n} |X_{ei} - X_{vi}|$$

12

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (X_{ei} - X_{vi})^2}$$

13

$$K = \frac{L_c - L_{exp}}{1 - L_{exp}}$$

14

Where TF = number of correctly classified true-flood pixels, TNF = number of correctly classified true-non-flood pixels, FF = number of false-flood correctly classified, FNH = number of false-non-flood that are incorrectly classified, X_{ei} = values predicted by the model, X_{oi} = values observed by the model, n = the number of data points, L_c = correct flood pixels and correct non-flood pixels, and L_{exp} are the expected matches.

Results And Discussions

Multicollinearity test

The VIF and TOL multicollinearity indicators of the 15 flood-conditioning factors used in the current work is shown in Fig. 5. It was noted that tolerance (TOL) values are acceptable which are above 0.1 (the minimum value is 0.155 of elevation factor) and the VIF values are below 5 (the maximum value is 3.901 of geology factor) except one elevation factor which provide VIF value above 5 (elevation = 6.432). Accordingly, the results showed that there was no multicollinearity among 14 factors which were used in the models and the elevation factor was excluded from model construction in this study.

Factors importance evaluation (FsIE)

The importance of flood-related factors in the current study was determined using RF by calculating the mean Gini decrease. It was presented in Fig. (6) as a percentage (%) after excluding the elevation factor

from the analysis. The results show that three factors are above 10% (distance to wadis has the highest value with 23.9%, followed by TWI with a value of 17.3%, then slope with a value of 13.2%), three factors from 5-10% (drainage density (9.81%), geology (6.17%) and NDVI (5.51%)), and the remaining factors are below 5% (CI has 4.8%, rainfall (4.38%), LULC (4.1%), LS (2.96%), PIC (2.42%), TPI (2.01%), PrC (1.93%) and the lowest value is for distance to road (1.62%)).

Flood susceptibility models

In this work, the ArcGIS software 10.8 was used to conduct flood susceptibility maps for wadi El-Matulla basin and its surrounding, Egypt. Natural breaks classified (Nicu, 2018) was applied to divide the flood susceptibility index maps (FSIMs) into five zones (including; “very low, low, moderate, high, and very high”) (Cao et al., 2020; Swain et al., 2020). In the current study, flood susceptibility maps were generated using LR, EGB, and RF (Fig. 7 (a- c)). Each susceptibility map has five categories with different coverage percentages as follow: for LR susceptibility map: VL= 18.7%, L = 32.3%, M = 20.5%, H = 21.5%, and VH = 7.0% (Fig. 7 (a, d)); for EGB susceptibility map VL= 30.1%, L = 45.5%, M = 18.2%, H = 4.4%, and VH = 1.8% (Fig. 7 (b, d)); and for RF susceptibility map VL = 51.7%, L = 23.7%, M = 16.2%, H = 7.1% and VH = 1.3% (Fig. 7 (c, d)).

Models performance assessment and comparisons

Validation datasets (30%), for both “flood and non-flood locations”, which were not applied for model construction, were utilized to identify the model’s performance (prediction rate). Additionally, other statistical parameters such as overall accuracy, Kappa parameter, RMSE, and MSA statistical indices were used to compare between the models (Fig. 8). Our results show that the AUC value of RF is 0.93, followed by 0.86 for EGB and finally 0.80 for LR. The overall accuracy values show that RF has the highest value of 0.88, followed by 0.82 for EGB and 0.76 for LR. The statistical indices Kappa, RMSE and MAE show that RF has the best performance with a Kappa value of 0.85, RMSE value of 0.34 and finally MSA with a value of 0.12, followed by EGB and finally logistic regression with a Kappa value of 0.51, RMSE value of 0.49 and MAE value of 0.24. Based on our results, RF provides the best performance in predicting flood prone areas.

Discussion

Future development of buildability is an essential issue to ensure sustainable urban and infrastructural development worldwide. Accordingly, understanding the extent of flood-prone areas is a critical factor in ensuring a future sustainable environment. In the current work, predictions were made by applying LR (easy to interpret), EGB, and RF (high performance). Based on these models flood susceptibility models were built and followed by selecting the best prediction model based on various statistical indices that can be used for future urban, agriculture, and industrial suitability development in the study area. This will provide important information to decision makers and planners to select less flood-prone areas.

In the present work, LR multivariate statistical analysis was applied and shows acceptable performance with AUC = 80%, which is in agreements with the flood susceptibility work that conducted by Tehrany et al. (2019) using “statistical index (Wi), logistic regression (LR), and frequency ratio (FR)”, with AUC for LR reaching 79.54%. Moreover, EGB is a boosting technique known to be a robust feature in data mining algorithms, resulting in better performance than the model LR with an AUC value of 86%. Our findings are also very close to the results of Abedi et al. (2021). They were used four models named “Classification and Regression Tree (CART), RF, Boosted Regression Trees (BRT), and XGBoost”. They concluded that RF performs the best with (AUC = 95.6%), while “XGBoost” performs well (AUC = 89.2%). Studies indicated that the main advantage of Extreme Gradient Boosting is its strong regularization parameter that overcomes the overfitting problem (Abedi et al., 2021). In addition, Janizadeh et al. (2021) applied Extreme Randomize Tree and Extreme Gradient Boosting for flood susceptibility analysis and found that EGB provides higher model effectiveness (AUC = 91.95%) than ERT, which provides an AUC of 91.37%, which is consistent with our results for EGB, which provides a higher AUC (86.0%) than logistic regression (LR) with an AUC of 80.0%.

Our findings indicated that the RF and EGB algorithms have provided high accuracy than LR model. RF produced high accuracy and performance due to its unique characteristics such as robust in features detection, strong in identifying noise and outliers, working with various inputs and large datasets without removing factors, and more interpretable (Park et al., 2019; Paul et al., 2019). In the current work, the result of RF susceptibility map (Fig. 9a) was verified by comparing with real flooded areas in 2016 and 2021 using Sentinel-1 images (VV-VH-VV /VH, RGB). After special filtering, we were able to highlight the flooded areas in blue color that were inundated during the February 14, 2016 (Fig. 9b) and November 13, 2021 (Fig. 9c) flood events. The result of the RF susceptibility map agreed well with the Sentinel-1 images.

Chen et al. (2020) indicated that RF has its ability to categorize various factors based on their contribution). Accordingly, results concluded that distance to wadis play a crucial role in the model results, the second factor is TWI, then slope, drainage density, geology, and NDVI (Fig. 6). Our findings indicate that distance to wadis has a major contribution to flood susceptibility modelling. In addition, lower slopes are associated with the western area of the study area, which is characterized by high TWI and lower elevation, resulting in higher flood susceptibility. In addition, high slope areas away from wadis and moderate to high elevations are not expected to flood, so future development in these areas is sustainable (see green and yellow areas in Fig. 9a).

Conclusion

Identifying areas at high risk of flooding is a critical step in implementing long-term sustainable projects. However, the lack of high quality data (both spatial and temporal) poses a challenge to the accurate management of flood disasters, especially in developing countries such as Egypt. This study provides an accurate methodology to produce powerful flood susceptibility maps for large regions by applying three models including LR, EGB and RF. The comparison of the models shows that RF and EGB have high

performance and their results have satisfactory efficiency. The EGB model had an AUC value of 0.86, which is considered very good. On the other hand, the RF model had an AUC value of 0.93, which is considered to have excellent predictive ability in evaluating flood susceptibility. Therefore, we highly recommend the use of both RF and EGB approaches for future flood vulnerability assessment. In addition, the results of the significance of the variables showed a high influence of the distance to wadis, the topographic wetness index, slope, drainage density, geology and NDVI in modelling flood susceptibility. This means that topographic parameters play an important role in flood modelling. It was also found that susceptibility maps could be improved if the quality of spatial resolution of digital elevation models is increased. This will improve the factors derived from DEM. Our results indicate that the western part of the study area is most affected by flooding, along the contact between the Nile floodplain and the areas reclaimed from the desert, which include lowland regions. The current work proposes an appropriate method to compare different models in order to select the optimal flood susceptibility model, which could be a cornerstone for the sustainability of future development. Moreover, the flood-prone areas could be saved from flood damage by appropriate preventive measures.

Declarations

Declaration of competing interest

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

References

1. Abedi R, Costache R, Shafizadeh-Moghadam H, Pham QB (2021) Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto International*. DOI: 10.1080/10106049.2021.1920636
2. Abd-Allatief TA, Abd El-Rahman AA, Sabet SH, Saad AM, Ali MZT (2012) Delineating the ground water occurrence using geophysical methods in Wadi El-Mathula area, Qena, Egypt. *Assiut University, Journal of Geology*
3. Abdelkareem M, El-Baz F (2015) Analyses of optical images and radar data reveal structural features and predict groundwater accumulations in the central Eastern Desert of Egypt. *Arab J Geosci* 8:2653–2666. <https://doi.org/10.1007/s12517-014-1434-7>
4. Abdrabo KI, Kantoush SA, Saber M, Sumi T, Habiba OM, Elleithy D, Elboshy B (2020) Integrated methodology for urban flood risk mapping at the microscale in ungauged regions: A case study of Hurgada. *Egypt Remote Sensing* 12(21):3548. <https://doi.org/10.3390/rs12213548>
5. Ali SA, Khatun R, Ahmad A, Ahmad AN (2019) Application of GIS-based analytic hierarchy process and frequency ratio model to flood vulnerable mapping and risk area estimation at Sundarban region, India. *Model Earth Syst Environ*. <https://doi.org/10.1007/s40808-019-00593-z>

6. Anusha N, Bharathi B (2019) Change Detection and Flood Water Mapping Using Sentinel-1A Synthetic Aperture Radar Images. *J Comput Theor Nanosci* 16:3544–3549
7. Band SS, Janizadeh S, Chandra Pal S, Saha A, Chakraborty R, Melesse AM, Mosavi A (2020) Flash Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms. *Remote Sensing* 12(21):3568. <https://doi.org/10.3390/rs12213568>
8. Breiman L (2001) Random forests, *Machine Learning*. 45:5–32. 10.1023/A:1010933404324
9. Briman L, Cutler A (2015) Package ‘randomForest’. pp. 29 (Date/Publication 2015-10-07)
10. Cao Y, Jia H, Xiong J, Cheng W, Li K, Pang Q, Yong Z (2020) Flash Flood Susceptibility Assessment Based on Geodetector, Certainty Factor, and Logistic Regression Analyses in Fujian Province, China. *ISPRS International Journal of Geo-Information*; 9(12):748. <https://doi.org/10.3390/ijgi9120748>
11. Chen T, He T, Benesty M (2015) Xgboost: Extreme Gradient Boosting; R Package Version 0.3-1; Technical Report; 2015; pp. 1–4. Available online: <http://cran.fhcrc.org/web/packages/xgboost/vignettes/xgboost.pdf>
12. Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, ACM*, 785-794
13. Chen J, Li Q, Wang H, Deng M (2020) A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: a case study of the Yangtze River Delta, China *Int. J Environ Res Public Health* 17(1):49
14. Chen RC, Dewi C, Huang SW, Caraka RE (2020) Selecting critical features for data classification based on machine learning methods. *J Big Data* 7:52. <https://doi.org/10.1186/s40537-020-00327-4>
15. Chowdhuri I, Pal SC, Chakraborty R (2020) Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India. *Adv Space Res* 65(5):1466–1489
16. Conoco (1987) Geological Map of Egypt, Scale 1:500,000, Sheet NG36NE Quseir, NG36NW Asyut, NG36SE Gebel Hamata, and NG36SW Luxor, Egypt. Cairo, Egypt, The Egyptian General Petroleum Corporation
17. Das S, Gupta A (2021) Multi-criteria decision based geospatial mapping of flood susceptibility and temporal hydro-geomorphic changes in the Subarnarekha basin, India. *Geosci Front* 12(5):101206. <https://doi.org/10.1016/j.gsf.2021.101206>
18. De Reu J, Bourgeois J, Bats M, Zwertvaegher A, Gelorini V, De Smedt P, Chu W, Antrop M, De Maeyer P, Finke P, Van Meirvenne M, Verniers J, Crombé P (2013) Application of the topographic position index to heterogeneous landscapes. *Geomorphology* 2013, 186, 39–49. <https://doi.org/10.1016/j.geomorph.2012.12.015>
19. Desmet P, Govers G (1996) A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units. *J Soil Water Conserv* 51(5):427–433
20. El-Shamy IZ (1985) Quantitative Geomorphology and Surface Water Conservation in WadiMatula - WadiAbbad Area, Central Eastern Desert, *Annals of the Geological Survey of Egypt*, vol. x

21. Embabi NS (2004) *The Geomorphology of Egypt: Landforms and Evolution*, vol 1. The Nile Valley and Western Desert; The Egyptian Geographical Society, Cairo, Egypt
22. Fan J, Wang X, Wu L, Zhou H, Zhang F, Yu X, Lu X, Xiang Y (2018) Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: a case study in China. *Energy Convers Manag* 164:102–111. <https://doi.org/10.1016/j.enconman.2018.02.087>
23. Farhadi H, Najafzadeh M (2021) *Technique Water* 13(21):3115. <https://doi.org/10.3390/w13213115>. Flood Risk Mapping by Remote Sensing Data and Random Forest
24. Feizizadeh B, Gheshlaghi HA, Bui DT (2021) An integrated approach of GIS and hybrid intelligence techniques applied for flood risk modeling. *J Environ Planning Manage* 64(3):485–516
25. Fillion R, Bernier M, Paniconi C, Chokmani K, Melis M, Soddu A, Talazac M, Lafortune F-X (2016) Remote sensing for mapping soil moisture and drainage potential in semi-arid regions: Applications to the Campidano plain of Sardinia, Italy. *Sci. Total Environ.* 2016, 573, 862–876
26. Gaber A, Abdelkareem M, Abdelsadek IS, Koch M, El-Baz F (2018) Using InSAR Coherence for Investigating the Interplay of Fluvial and Aeolian Features in Arid Lands: Implications for Groundwater Potential in Egypt. *Remote Sensing* 10(6):832. <https://doi.org/10.3390/rs10060832>
27. Gao L, Tao B, Miao Y, Zhang L, Song X, Ren W (2019) A global data set for economic losses of extreme hydrological events during 1960-2014. *Water Resour Res* 55:5165–5175
28. Grohmann CH, Riccomini C (2009) Comparison of roving-window and search-window techniques for characterising landscape morphometry. *Comput Geosci* 35:2164–2169
29. Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modeling of plant species distribution *Plant Ecol* (1999), 10.1023/A:1009841519580
30. Hembram TK, Saha S, Pradhan B, Abdul Maulud KN, Alamri AM (2021) Robustness analysis of machine learning classifiers in predicting spatial gully erosion susceptibility with altered training samples. *Geomatics Natural Hazards and Risk* 12(1):794–828. <https://doi.org/10.1080/19475705.2021.1890644>
31. Hermas E, Gaber A, Mohammed El Bastawesy M (2021) Application of remote sensing and GIS for assessing and proposing mitigation measures in flood-affected urban areas. *Egypt The Egyptian Journal of Remote Sensing and Space Science* 24:119–130. <https://doi.org/10.1016/j.ejrs.2020.03.002>
32. Hölting B, Coldewey WG (2019) In: *Hydrogeology* (ed) Surface water infiltration. Springer, Berlin, Heidelberg, pp 33–37
33. Hussain S, Mubeen M, Akram W, Ahmad A, Habib-ur-Rahman M, Ghaffar A, Amin A, Awais M, Farid HU, Farooq A, Nasim W (2020) Study of land cover/land use changes using RS and GIS: a case study of Multan district, Pakistan. *Environ Monit Assess* 192, 2 (2020). <https://doi.org/10.1007/s10661-019-7959-1>
34. Islam AMdT, Talukdar S, Mahato S, Kundu S, Eibek KU, Pham QB, Kuriqi A, Linh NTT (2021) Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci Front*

12(3):1674–9871. <https://doi.org/10.1016/j.gsf.2020.09.006>

35. Janizadeh S, Vafakhah M, Kapelan Z, Dinan NM (2021) Hybrid XGboost model with various Bayesian hyperparameter optimization algorithms for flood hazard susceptibility modeling, *Geocarto International*. 10.1080/10106049.2021.1996641
36. Khan MN, Gong Y, Hu T, Lal R, Zheng J, Justine MF, Azhar M, Che M, Zhang H (2016) Effect of Slope, Rainfall Intensity and Mulch on Erosion and Infiltration under Simulated Rain on Purple Soil of South-Western Sichuan Province, China. *Water* 8:528. <https://doi.org/10.3390/w8110528>
37. Kilinc H (2018) Estimation of rainfall distribution map of Turkey by IDW and kriging interpolation method. *American Journal of Engineering Research (AJER)* 7(6):238–241
38. Lanfredi Sofia C, Oliveira SC, Pereira S, Zêzere JL, Corsini A (2018) A comparison between bivariate and multivariate methods to assess susceptibility to liquefaction related coseismic surface effects in the Po Plain (Northern Italy), *Geomatics, Natural Hazards and Risk*. 9:108–126. 10.1080/19475705.2017.1411401. 1
39. Lei X, Chen W, Panahi M, Falah F, Rahmati O, Uuemaa E, Kalantari Z, Ferreira CSS, Rezaie F, Tiefenbacher JP, Lee S, Bian H (2021) Urban flood modeling using deep-learning approaches in Seoul, South Korea. *J Hydrol* 601:126684. <https://doi.org/10.1016/j.jhydrol.2021.126684>
40. Liao SSC, Veneziano D, Whitman RV (1988) Regression models for evaluating liquefaction probability. *J Geotech Eng* 114:389–411
41. Liu J, Wang J, Xiong J, Cheng W, Sun H, Yong Z, Wang N (2021) Hybrid Models Incorporating Bivariate Statistics and Machine Learning Methods for Flash Flood Susceptibility Assessment Based on Remote Sensing Datasets. *Remote Sens.* 2021, 13, 4945. <https://doi.org/10.3390/rs13234945>
42. Malik S, Pal SC, Chowdhuri I, Chakraborty R, Roy P, Das B (2020) Prediction of highly flood prone areas by GIS based heuristic and statistical model in a monsoon dominated region of Bengal Basin. *Remote Sens Appl: Soc Environ* 19:100343
43. Metz CE (1978) Basic principles of ROC analysis. *Semin Nucl Med* 8:283–298
44. Mignot E, Li X, Dewals B (2019) Experimental modelling of urban flooding. A review *J Hydrol* 568:334–342
45. Mirzaei S, Vafakhah M, Pradhan B, Alavi SJ (2021) Flood susceptibility assessment using extreme gradient boosting (EGB), Iran. *Earth Sci Inform* 14, 51–67 (2021). <https://doi.org/10.1007/s12145-020-00530-0>
46. Mokarrama M, Hojati M (2018) Landform classification using a sub-pixel spatial attraction model to increase spatial resolution of digital elevation model (DEM). *Egypt J Remote Sens Sp Sci*. doi:10.1016/j.ejrs.2016.11.005
47. Naghibi SA, Hashemi H, Berndtsson R, Lee S (2020) Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors *Journal of Hydrology* 125197
48. Nicu IC (2018) Application of analytic hierarchy process, frequency ratio, and statistical index to land slide susceptibility: an approach to endangered cultural heritage. *Environ Earth Sci* 77(3):79.

<https://doi.org/10.1007/s12665-018-7261-5>

49. Nguyen B, Minh D, Ahmad A, Nguyen Q (2020) The Role Of Relative Slope Length In Flood Hazard Mapping Using Ahp And Gis (Case Study: Lam River Basin, Vietnam). *GEOGRAPHY, ENVIRONMENT, SUSTAINABILITY*;13(2):115-123. <https://doi.org/10.24057/2071-9388-2020-48>
50. Nyssen J, Poesen J, Moeyersons J, Luyten E, Veyret-Picot M, Deckers J, Haile M, Govers G (2002) Impact of road building on gully erosion risk: A case study from the northern Ethiopian highlands. *Earth Surf. Process. Landf. J. Br. Geomorphol. Res. Group* 2002, 27, 1267–1283
51. Onuşluel Gül G (2013) Estimating flood exposure potentials in Turkish catchments through index-based flood mapping. *Nat Hazards*. doi:10.1007/s11069-013-0717-8
52. Park S, Hamm S-Y, Kim J (2019) Modeling Sustainability 11(20):5659. <https://doi.org/10.3390/su11205659>. Performance Evaluation of the GIS-Based Data-Mining Techniques Decision Tree, Random Forest, and Rotation Forest for Landslide Susceptibility
53. Paul GC, Saha S, Hembram TK (2019) Application of the GIS-Based Probabilistic Models for Mapping the Flood Susceptibility in Bansloi Sub-basin of Ganga-Bhagirathi River and Their Comparison. *Remote Sens Earth Syst Sci* 2:120–146. <https://doi.org/10.1007/s41976-019-00018-6>
54. Popa MC, Peptenatu D, Drăghici CC, Diaconu DC (2019) Flood Hazard Mapping Using the Flood and Flash-Flood Potential Index in the Buzău River Catchment, Romania. *Water*. 2019; 11(10):2116. <https://doi.org/10.3390/w11102116>
55. Prama M, Omran A, Schröder D, Abouelmagd A (2020) Vulnerability assessment of flash floods in Wadi Dahab Basin, Egypt. *Environ Earth Sci* 79:114. <https://doi.org/10.1007/s12665-020-8860-5>
56. 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, 585–601 (2019). <https://doi.org/10.1007/s41748-019-00123-y>
57. Rahman R, Dhruba SR, Ghosh S, Pal R (2019) Functional random forest with applications in dose-response predictions *Sci. Rep* 9(1):1–14
58. Rejith RG, Anirudhan S, Sundararajan M (2019) Delineation of groundwater potential zones in hard rock terrain using integrated remote sensing, GIS and MCDM techniques: A case study from vamanapuram river basin, Kerala, India, in: *GIS and Geostatistical Techniques for Groundwater Science*. doi:10.1016/B978-0-12-815413-7.00025-0
59. Roundy BA, Chambers JC (2021) Effects of elevation and selective disturbance on soil climate and vegetation in big sagebrush communities. *Ecosphere* 12:e03377. <https://doi.org/10.1002/ecs2.3377>
60. Rouse JrJW, Haas R, Schell J, Deering D (1974) Monitoring vegetation systems in the great plains with erts. *NASA Special Publication* 351:309
61. Saber M, Abdrabo KI, Habiba OM, Kantosh SA, Sumi T (2020) Impacts of Triple Factors on Flash Flood Vulnerability in Egypt: Urban Growth, Extreme Climate, and Mismanagement. *Geosciences*. 2020; 10(1):24. <https://doi.org/10.3390/geosciences10010024>
62. Said R (1993) *The River Nile; geology, hydrology and utilization*. Pergamon Press, Oxford

63. Sangireddy H, Carothers RA, Stark CP, Passalacqua P (2016) Controls of climate, topography, vegetation, and lithology on drainage density extracted from high resolution topography data. *J Hydrol* 537:271–282. doi:10.1016/j.jhydrol.2016.02.051
64. 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* 10:17. <https://doi.org/10.1007/s13201-019-1102-x>
65. Seejata K, Yodying A, Wongthadam T, Mahavik N, Tantanee S (2018) Assessment of flood hazard areas using Analytical Hierarchy Process over the Lower Yom Basin, Sukhothai Province. *Procedia Eng* 212:340–347. <https://doi.org/10.1016/j.proeng.2018.01.044>
66. Sentinel-1 User Handbook (https://sentinel.esa.int/documents/247904/685163/Sentinel-1_User_Handbook)
67. Sultan M, Wagdy A, Manocha N, Sauck W, Abdel Gelil K, Youssef AF, Becker R, Milewski A, El Alfy Z, Jones C (2008) An integrated approach for identifying aquifers in transcurrent fault systems: the Najd shear system of the Arabian Nubian shield. *J Hydrol* 349:475–488
68. Swain KC, Singha C, Nayak L (2020) Flood Susceptibility Mapping through the GIS-AHP Technique Using the Cloud. *ISPRS International Journal of Geo-Information*; 9(12):720. <https://doi.org/10.3390/ijgi9120720>
69. Tang X, Machimura T, Liu W, Li J, Hong H (2021) A novel index to evaluate discretization methods: A case study of flood susceptibility assessment based on random forest. *Geosci Front* 12:101253. <https://doi.org/10.1016/j.gsf.2021.101253>
70. Tehrany MS, Jones S, Shabani F (2019) Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *CATENA* 175:174–192. doi: 10.1016/j.catena.2018.12.011
71. Thomas V (2017) Climate Change and Natural Disasters. *Clim. Chang. Nat. Disasters* 2017, 8, 81–94
72. Uddin K, Matin MA (2021) Potential flood hazard zonation and flood shelter suitability mapping for disaster risk mitigation in Bangladesh using geospatial technology. *Progress in Disaster Science* 11:100185. <https://doi.org/10.1016/j.pdisas.2021.100185>
73. Ullah K, Zhang J (2020) GIS-based flood hazard mapping using relative frequency ratio method: A case study of Panjkora River Basin, eastern Hindu Kush, Pakistan. *PLoS ONE* 15(3):e0229153. <https://doi.org/10.1371/journal.pone.0229153>
74. Vojtek M, Vojteková J (2019) Flood Susceptibility Mapping on a National Scale in Slovakia Using the Analytical Hierarchy Process. *Water* 11(2):364. <https://doi.org/10.3390/w11020364>
75. Voormansik K, Praks J, Antropov O, Jagomagi J, Zalite K (2014) Flood mapping with TerraSAR-X in forested regions in Estonia. *IEEE J Select Top Appl Earth Observ Remote Sens* 7:562–577
76. Wang H, Yang F, Luo Z (2016) An experimental study of the intrinsic stability of random forest variable importance measures. *BMC Bioinformatics* 17:60, 26842629. doi: 10.1186/s12859-016-0900-5

77. Wang Y, Fang Z, Hong H, Costache R, Tang X (2021) Flood susceptibility mapping by integrating frequency ratio and index of entropy with multilayer perceptron and classification and regression tree. *J Environ Manage* 289:112449. <https://doi.org/10.1016/j.jenvman.2021.112449>
78. Waqas H, Lu L, Tariq A, Li Q, Baqa MF, Xing J, Sajjad A (2021) Flash Flood Susceptibility Assessment and Zonation Using an Integrating Analytic Hierarchy Process and Frequency Ratio Model for the Chitral District. *Khyber Pakhtunkhwa Pakistan Water* 13:1650. <https://doi.org/10.3390/w13121650>
79. Weiss A (2001) Topographic position and landforms analysis. In *Proceedings of the Poster Presentation, ESRI User Conference, San Diego, CA, USA, 9 July 2001; Volume 200*
80. Wolock DM, McCabe GJ (1995) Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. *Water Resour Res* 31(5):1315–1324. <http://onlinelibrary.wiley.com/doi/10.1029/95WR00471/full>
81. Wubalem A, Tesfaw G, Dawit Z, Getahun B, Mekuria T, Jothimani M (2020) Comparison of statistical and analytical hierarchy process methods on flood susceptibility mapping: in a case study of Tana sub-basin in northwestern Ethiopia, *Nat. Hazards Earth Syst. Sci. Discuss.* <https://doi.org/10.5194/nhess-2020-332>, 2020
82. Yagoub MM, Alsereidi AA, Mohamed EA, Periyasamy P, Alameri R, Aldarmaki S, Alhashmi Y (2020) Newspapers as a validation proxy for GIS modeling in Fujairah, United Arab Emirates: identifying flood-prone areas. *Nat Hazards* 104, 111–141 (2020). <https://doi.org/10.1007/s11069-020-04161-y>
83. Yariyan P, Avand M, Abbaspour RA, Haghighi AT, Costache R, Ghorbanzadeh O, Janizadeh S, Blaschke T (2020) Flood susceptibility mapping using an improved analytic network process with statistical models, *Geomatics. Natural Hazards and Risk* 11(1):2282–2314. <https://doi.org/10.1080/19475705.2020.1836036>
84. Yoo C, Cho E (2019) Effect of Multicollinearity on the Bivariate Frequency Analysis of Annual Maximum Rainfall Events. *Water* 11:905. <https://doi.org/10.3390/w11050905>
85. Zhao G, Pang B, Xu Z, Peng D, Xu L (2019) Assessment of urban flood susceptibility using semi-supervised machine learning model. *Sci Total Environ* 659:940–949
86. Zhu Z, Zhang Y (2021) Flood disaster risk assessment based on random forest algorithm. *Neural Comput & Applic.* <https://doi.org/10.1007/s00521-021-05757-6>

Figures

Figure 1

a) The location map of El-Matulla wadi and its surrounding, Egypt. b) Geological map overlayed by flood inventory data, road and wadis.

Figure 2

Schematic diagram showing applied methodology in the current study

Figure 3

a, c) Sentinel-1 images VV-VH-VV/VH, RGB taken after two flood events. Image (a) taken on 25 February 2016 and image (c) taken on 19 November 2021. Images b and c, on the other hand, are Google Earth images showing the signature of flowing water in the main wadis after the 2016 and 2021 floods,

respectively. Pictures e and f show the flooding after the 2021 event, where villages and roads were inundated with water (1 m depth). Note that people are standing above the roofs of buildings to avoid the flowing water.

Figure 4

Flood-conditioning factors: a) EL, b) S, c) TPI, d) CI, e) LS, f) TWI, g) PrC, h) PIC, i) DD, j) GEO, k) NDVI, l) LULC, m) DTR, n) DTW, and o) R

Figure 5

Multicollinearity results of flood-causative factors.

Figure 6

Factors importance histogram in percent

Figure 7

FSMs for the Wadi El-Matulla basin and its surroundings using LR, EGB and RF models and classes percent

Figure 8

Showing various techniques in model comparisons (ROC, OAC, Kappa, RMSE, and MAE.

Figure 9

Validation of RF model result (a), with real flood inundated areas in 2016 and 2021 using sentinel-1 images (b) on 25 February, 2016, and (c) on 19 November 2021.