Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin, Egypt

Bosy A. El-Haddad[1], Ahmed M. Youssef [1, 2], Hamid Reza Pourghasemi[3], Biswajeet Pradhan[4,5], Abdel-Hamid El-Shater[1], Mohamed H. El-Khashab[1]

[1]Sohag University, Faculty of Science, Geology Department, Sohag, 82524 Egypt. E-mail: amyoussef70@yahoo.com

[2]Saudi Geological Survey, Applied Geology Sector, Geological Hazards Department, Jeddah 21514, Kingdom of Saudi Arabia

[3]Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran.

[4]Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Information, Systems and Modelling, Faculty of Engineering and IT, University of Technology Sydney, 2007 NSW, Australia

[5]Department of Energy and Mineral Resources Engineering, Sejong University, Choongmu-gwan, ~~Sejong University,~~ 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea

Abstract

Floods represent catastrophic environmental hazards which have a significant impact on environment and human life and their activities. Environmental and water management in many countries require modeling of flood susceptibility to help in reducing the damages and impact of floods. The objective of the current work is to employ four data mining/machine learning models to generate flood susceptibility maps, namely boosted regression tree (BRT), functional data analysis (FDA), general linear model (GLM), and multivariate discriminant analysis (MDA). This study done in Wadi Qena Basin in Egypt. Flood inundated locations were determined and extracted from the interpretation of different data~~-~~sets, including high-resolution satellite images (sentinel-2 and Astro digital) (after flood events), historical records, and intensive field works. In total, 342 flood inundated locations were mapped using ArcGIS 10.5, which separated into two groups; training (has 239 flood ~~points~~ locations represents 70%) and validating (has 103 flood ~~points~~ locations represents 30%), respectively. Nine themes of flood-influencing factors were prepared, including slope angle, slope length, altitude, distance from rivers, landuse/landcover, lithology, curvature, slope-aspect, and topographic wetness index. The relationships between the flood-influencing factors and the flood inventory map were evaluated using the mentioned models (BRT, FDA, GLM, and MDA). The results were compared with flood inundating locations (validating flood ~~points~~sites), which were not used in constructing the models. The accuracy of the models were calculated through~~,~~ the success (training data) and prediction (validation data) rate curves, according to the receiver operating characteristics (ROC) and the area under the curve (AUC). The results showed that the AUC for success and prediction rates are 0.783, 0.958, 0.816, 0.821 and 0.812, 0.856, 0.862, 0.769 for BRT, FDA, GLM, and MDA models, respectively. Subsequently, flood susceptibility maps were divided into five classes, including very low, low, moderate, high, and very high susceptibility. The results revealed that the BRT, FDA, GLM, and MDA models provide reasonable accuracy in flood susceptibility mapping. The produced susceptibility maps might be vitally important for future development activities in the area, especially in choosing new urban areas, infrastructural activities, and flood mitigation areas.

***Keywords*:** Floods, Remote sensing, Data mining, Modeling, GIS, Susceptibility, Egypt.

1. Introduction

Floods are common catastrophic environmental hazards in different areas all over the world, where many cities, highways, and roads were impacted (Taylor et al. 2011; Dawod et al. 2012). These areas are dissected by many wadis that drain rain water towards low-lying areas. In the major Wadi Basins, the flash floods are suddenly occurred that are initiated by intense precipitation generated in rainstorms. The recent rapid urban growth coupled with climate change have led to many environmental problems (e.g., flooding and associated losses of human lives and property) (Zwenzner and Voigt 2009; Kjeldsen 2010; Karmaoui et al., 2014).

Floods often represent the most damaging natural hazards in the low-lying areas of different parts of the world, resulting in loss and injure of human life and properties damage (agricultural and urban areas, bridges, roads, railways, and highways) (Du et al. 2013; Tehrany et al. 2017, 2019; Vojtek and Vojteková 2019). These floods might cause huge economic loss and infected urban areas by microbial development and diseases (Tehrany et al. 2015; Dandapat and Panda 2017). In addition, records of loss of life and damage caused by floods worldwide showed that these have continued to rise steadily during recent years (NFRAG 2008). Many countries all over the world, that located in an arid zone experienced many devastating events of flash floods such as Morocco (1995, 2002, 2008，2014) (Saidi 2010; Echogdali et al., 2018), Algeria (1971, 1974, 1980, 1982, 1984, 2001, 2007, 2008, and 2013) (Kenyon 2007; Warner 2004; Yamani et al., 2016), Chad in 2012 (IRIN, 2013). In Egypt, flash floods frequently occurred in many areas, in 1994, 2010, 2016 (Khidr 1997; Ashour 2002; Moawad 2013; Moawad et al., 2016), and also in Saudi Arabia in different areas (2009, 20011, 2015, 2017, and 2018 in Jeddah) (Youssef et al., 2016). Most flash floods in arid areas are generally unpredictable and infrequent (Reid et al. 1994). Flood frequency and severity in the desert areas vary from year to year (Warner 2004).

There are different flood-influencing factors that could be used to produce the flood susceptibility map for an area. These factors include lithology, slope-angle, slope-aspect, curvature, altitude, distance from main wadis, drain type, slope length, topographic wetness index, and land use/land cover patterns. Many studies have been carried out on flood modeling and susceptibility assessment using hydrological studies, remote-sensing and GIS techniques (e.g., Talei et al. 2010; Kisi et al. 2012; Bubeck et al. 2012; Wanders et al. 2014; Pradhan et al. 2014; Mandal and Chakrabarty 2016; Tehrany et al. 2017; Luu e al., 2018; Mahmoud and Gan 2018; Dano et al., 2019; Kanani-Sadata et al., 2019; Khosravi et al., 2019a; Liu et al., 2019; Wang et al., 2019).

Various modeling approaches were applied to assess flood susceptibility in any specific area which belongs to: (1) heuristic (multi-criterion analysis), (2) inundating analysis, and (3) statistical analysis. Each of them has its own advantages and disadvantages. Heuristic models (such as analytical hierarchy process-AHP) rely mainly on the expert knowledge to assign weights to the various conditioning factors (e.g., Chen et al. 2011; Rozos et al. 2011; Matori 2012; Zou et al. 2013; Sar et al. 2015; Dandapat and Panda 2017; Vojtek and Vojteková 2019; Youssef and Hegab 2019).

The heuristic models are highly subjective and depend on the site itself. Many authors were applying inundating flood models to identify the flood-vulnerable areas (Tsakiris 2014; Pakoksung and Takagi 2016; Pal and Pani 2016; Kumar et al. 2017; Prasad and Pani 2017; Abdelkarim et al. 2019).

Statistical models were also utilized to analyze the flood susceptibility (e.g., artificial neural networks (ANNs), adaptive neuro-fuzzy interface system (ANFIS), weights-of-evidence (WOE), logistic regression (LR), frequency ratio (FR), general linear models (GLMs), decision tree (DT), Shannon's entropy (SE), statistical index (SI), support vector regression (SVR), random forest (RF), boosted regression tree (BRT), classification and regression tree (CART), general linear (GLM), and weighting factor (Wf) (Liao and Carin 2009; Mukerji et al. 2009; Pradhan 2010a; Pradhan and Buchroithner 2010; Kourgialas and Karatzas 2011; Sezer et al. 2011; Kia et al. 2012; Lee et al. 2012; Tehrany et al. 2013, 2014a, b, 2015; Feng et al. 2015; Albers et al. 2016; Gizaw and Gan 2016; Khosravi et al. 2016a,b; Rahmati et al. 2016; Tehrany et al. 2017; Khosravi et al. 2018; Mosavi et al. 2018; Muñoz et al. 2018; Samantal et al. 2018a; Zhao et al. 2018; Choubin et al. 2019; Park et al. 2019).

Many authors indicated that flood susceptibility map could be crucially used to establish an early warning system, emergency plans, reduction and prevention of future floods, and executing of flood management strategies (Bubeck et al. 2012; Mandal and Chakrabarty 2016; Tehrany et al. 2017).

In Egypt, during the last few decades, urban areas and many infrastructures (highways, railways, and roads) are expanding toward the flood-prone areas, and accordingly, floods occur more frequently (Youssef and Hegab 2019). Thus, different locations are often prone to flash floods, which are irregular in time and space since the rainfall differs significantly from north to south. Such events usually lead to severe damages and mortality. Various authors studied floods in Egypt. Foody et al. (2004) predicted the sensitive areas to flash flooding based mainly on land cover distribution and soil properties in the Eastern Desert of Egypt. Milewski et al. (2009) used multiple remote sensing data-sets to identify the relatively larger precipitation events that are more likely to produce runoff and recharge in Sinai Peninsula and the Eastern Desert of Egypt. Moawad (2012) used the hydro-morphometric parameters and soil characteristics to reveal the characteristics of flash floods in Safaga - El Qusier area along the Egyptian Red Sea Coast. Moawad (2013) used the black-box model (BBM) based on the curve number (CN) approach developed by the United States Department of Agriculture, Soil Conservation Service (SCS 1985), and real-time satellite precipitation (HYDIS) to analyze the 18 January 2010 flash flood event in wadi El Arish (Northern Sinai).

In this study, four data mining models were adapted to construct a flood susceptibility map using remote-sensing and GIS tools. These techniques are boosted regression tree (BRT), functional data analysis (FDA), general linear model (GLM), and multivariate discriminant analysis (MDA). These models were selected for a number of reasons, including being newly applied in the field of flood susceptibility in Egypt, adequate for regional- and semi regional-scale applications, and relying mainly on remote-sensing datasets rather than intensive field investigations. We believe that the results obtained from our study provide a considerable contribution to the ~~flood~~ literature dealing with the spatial flood assessment. The flood susceptibility maps can identify and delineate flood-vulnerable areas, so that planners and decision-making can choose favorable locations for future development, such as new urban areas.

2. **Study area**

The study area includes Wadi Qena Basin, covering an area of 14,558 km^2 between latitudes 26°11′44″ and 28°04′42″ N and longitudes 32°15′45″ and 33°37′50″ E (Fig. 1). Wadi Qena is one of the largest basin in Egypt. It belongs to the Great Sahara Desert which considered the world's largest hot desert (covering ten countries such as Mauritania, Morocco, Mali, Algeria, Niger, Tunisia, Libya, Chad, Egypt, and Sudan). The most crucial characteristics of the Sahara Desert are severe aridity, high temperatures, low humidity, and strong winds (Laity 2008).

The Wadi Qena area as part of the Sahara Desert is characterized by the abrupt change of weather patterns that cause~~s~~ most devastating flash floods. The study area receives flash flood water from the mountains and foothills that located to the east, west and north through natural drainage Wadis. Many flood events were occurred in Wadi Qena basin due to intense thunderstorms in the years of 2014-2016, and 2018 causing devastating to the area. Annual average rainfall of the Sahara Desert is less than 100 mm for about 75% of its area, however, less than 20 mm for the remaining area (Warner 2004). Most flash floods in the arid desert (e.g., Sahara Desert) are characterized by high intensity, short duration, fast flowing water, suddenly occurring with little time to respond, and imposing immense risk to people and property (Sene 2013). Most of the arid areas rainfall is variable and spotty (the affected area often limited by the size of the clouds) (Laity 2008). The elevations of the study area range between 113 m and 1,878 m above mean sea level.

The study area stroked by flash many times before. The most catastrophic events were recorded in three consecutive events in 2014- 2016. Most of the damages were occurred along different highways that crossed the area (Fig. 2). In addition to that the Qena City which located at the mouth of Wadi Qena was impacted.

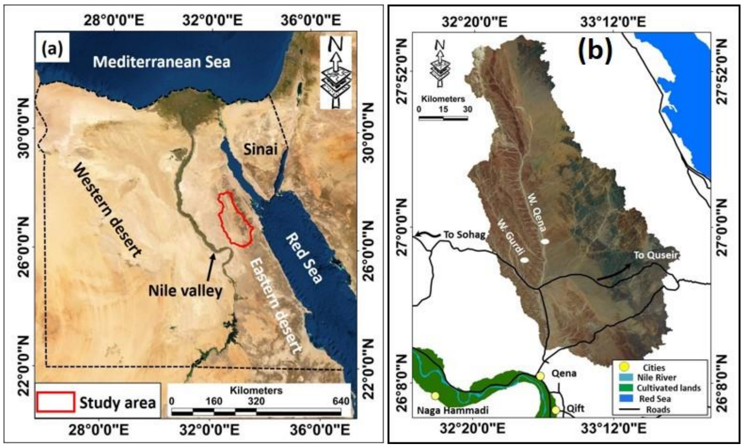

Fig. 1 a) Location of the study area ~~in relation to the surrounding areas~~; b) a zoomed close up view of the study area.

Fig. 2 Different photographs that were captured for various flood events hit the area; a and b) were taken in 2014 event; c and d) were taken in 2015 event; and e and f) were taken in 2016 event.

3. **Data and methodology**

The current work illustrates the utilization of various datasets to be applied in flood susceptibility mapping. Many stages of methodologies were used in this research including preparation of various datasets extracted from different sources and types (remote sensing images and geological and topographical data), establish a flood inventory map, models construction, and finally checking the models validation (Fig. 3).

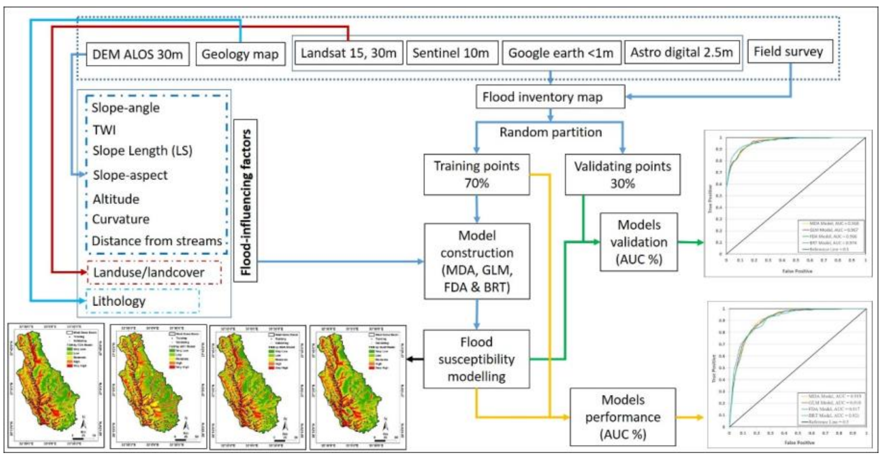

Fig. 3 ~~a f~~Flowchart showing the data and modeling steps used to produce a reliable flood susceptibility map.

3.1. Stage I: Data and inventory map preparation

Different data sources and types were extracted and used in this research (Table 1). Many field investigations have been carried out for the study area to collect data related to the existing impact of the flooded areas at different times, to get information from the local people in the area related to previous, current, and future problems, and to take photographs to document different situations. Other data types including historical reports which were collected from different sources such as the civil defense authority, and from the department of transportation. According to these historical records, the frequency of flood events could be identified, especially those affected urban and infrastructure areas. In addition to that satellite images were acquired including, Landsat 8, Operational Land Imager (OLI) images with 30-m resolution acquired in 2018 for the study area which were obtained from Earth explorer website (https://earthexplorer.usgs.gov). Landsat 8 data consists of eleven bands; a layer stacking was conducted for bands (1-7) to create an image mosaic with 30-m spatial resolution, followed by image fusion with band 8 (panchromatic 15-m resolution) to create a final mosaic with 15-m spatial resolution. Also, a high-resolution image was used (Google Earth images, DigitalGlobe). A Digital Elevation Model (DEM) 30-m spatial resolution was acquired from ALOS Global Digital Surface Model (ALOS World 3D-30m) which used to extract different data sets such as stream networks, slope-angle, slope-aspect, curvature, LS (slope length), TWI (topographic wetness index), and altitude. The topographic maps of 1:50,000-scale was used to verify the main Wadis that were extracted from DEM. Finally, the Geological map 1:250,000-scale was used to map the lithology units. All the datasets used in the current study are in a digital format with a unified projection (UTM-Zone 36, WGS84 datum).

Table 2 Data sources and datasets used in the current study.

Dataset No.	**Data Source**	**Data Type**	**Resolution & Scale**	**Extracted Data**
1	Satellite Imageries	OLI 2014, 2015, 2016, 2018 Sentinel-2 2015, 2016 Astro digital Google Earth 2014, 2015, & 2016	30, 15m 10m 2.5m <1m	- LULC mapping, extracting inundating areas after the flood events in 2014, 2015, & 2016 - Mapping inundating areas after flood events in 2016 - Verify the flood locations after the events 2014, 2015, 2016
2	Geological Data	Quadrangle 1985	1: 250,000	- Lithology units
3	Digital Elevation Model	Grid	30 m	- Slope-aspect, Slope-angle, altitude, TWI, LS, curvature, and main Wadis
4	Field Investigation	Information on the inundated and damaged areas by flood events in 2014, 2015 & 2016	Field trips	- Inundated and damages areas in 2014, 2015, 2016 events

Formatted: Font: Bold

The flood locations were mapped according to previous inundated areas. It is known that recent floods are more likely to happen under the same conditions of the previous floods (Akgun et al. 2012; Tehrany et al. 2013, 2014a, b; Fotovatikhah et al. 2018). Inventory map considers a crucial part for hazard susceptibility modeling (landslides and floods) where the relationship between the existing hazard areas and the factors controlling this hazard is an essential requirement for susceptibility mapping (Petley 2008; Rahmati et al., 2016). In the current study, a flood inventory map was generated according to the integration of different data sources such as historical records, field surveys, and satellite images interpretation. The ~~-~~Flood hazards inventory map shows the spatial distribution of flood hazards in the study area. Different datasets were used to prepare the flood inventory map ~~-~~as shown in (Table 1). The historic flood data was collected from the analysis and interpretation of high~~-~~-resolution images (Google Earth and Astro digital images) from 2006 ~~till~~ to 2016 and medium resolution images (Landsat OLI 2014, 2015, and 2016) and Sentinel-2 images (2015 and 2016). In addition to that more data related to recent flood events (flood occurrences) 2014, 2015, and 2016 were collected from field surveys. The flood hazard locations were identified according to detailed field surveys. Collapses, erosions, and inundated areas caused by flooding were identified through the field surveys (Fig. 2). ~~-~~Other data collected from the civil defense department and previous reports of flash flood for the past 20 years. To extract the real flood areas using high resolution remote sensing images, two time span imageries of Astro digital data, with a special resolution of 2.5 m, were used. The first one was acquired on October 13, 2016, before the flood event that was occurred on October 18, 2016. The second imagery was acquired on November 5, 2016 - after the same flood event. These dates were characterized by a cloud free and covering the whole Wadi basin. In the current research, a true color imagery (band 1, 2, 3 in RGB) was used for these time spans. The Environment for Visualizing Images (ENVI v. 5.4) software was used to extract the inundated areas from the Astro digital image after the flood event. Visual inspection was carried out to compare the areas before and after the flood events. Analysis of these images indicated that inundated areas can be easily detected on the imagery acquired after the flood event. In addition to that imaging enhancing processing method (slicing classification technique) was applied to extract the inundated areas from the image acquired after the flood event (Fig. 4a). The slicing results were verified by filed ~~investigation for some flooded areas.~~ survey data and the data collected from civil defense. Subsequently, critical flooded areas were identified and digitized on the slicing map as point features. ~~Finally, flood locations were collected and digitized as point features.~~ All the different data sources (point features of flood locations)~~These data~~ were collected and assembled together to create the flood inventory map (Fig. 4b). A total of 342 flood locations were identified and mapped in the study area. These flood locations represent~~s~~ the inundated areas after heavy rainstorms that stroked the area previously (areas were highly impacted by flood events). Using R statistical software, the data points were randomly partitioned. According to Naimi and Araújo (2016), the random partition method is a splitting technique in which the flood points randomly separated into training and validating datasets. According to the literature, the percentages commonly applied to split the inventory dataset are 70% and 30% for the training and validating datasets, respectively (Abdulwahid and Pradhan 2017; Chen et al., 2019). In the current work, 239 flood locations (70 % of the sites) were randomly selected for training datasets and the remaining 103 flood locations (30 % of the sites) were used as validating datasets for verification purposes (Fig. 4b). Field surveys indicated that all these locations were previously inundated by floods.

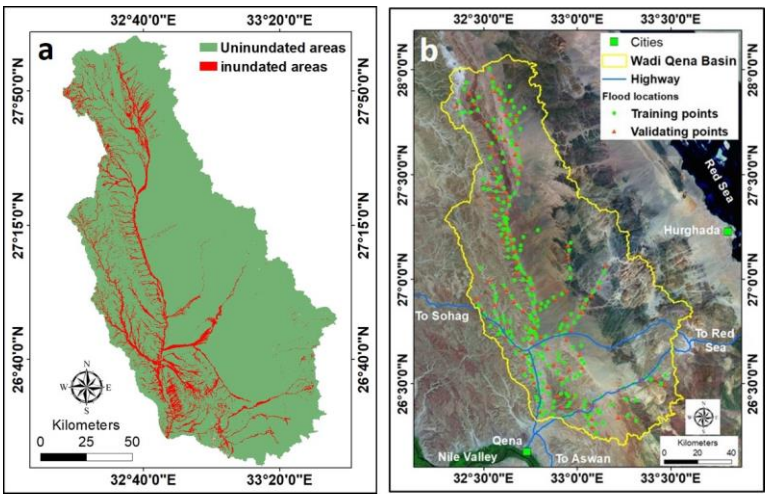

Fig. 4 a) ~~The slicing map extracted from satellite image after the f~~Flood event showing the distribution of inundated areas along Wadi Qena basin;~~.~~ b) flood inventory data used to test and validate the models.

3.2. Stage II: Generating the flood-influencing variables

In terms of flood-influencing factors, the selection of the most influential parameters is ~~vitally~~ important for flood susceptibility analysis. Floods are initiated by rainfall, the most significant variable in the occurrence of floods. However, many other influential factors are involved (Lawal et al., 2012). Hölting and Coldewey, (2019) indicated that during precipitation in a drainage catchment, the runoff depends on the condition of the catchment, for example, catchment area, topography, and LULC types. ~~Determining the flood-influencing variables is vitally important for flood susceptibility analysis~~~~.~~ Different flood-influencing variables have been selected in the current work according to previous literatures (Pradhan 2010a; Kia et al. 2012; Lee et al. 2012; Tehrany et al. 2014a, b; Rahmati et al., 2016; Khosravi et al., 2016a; Al-Juaidi et al., 2018; Luu et al., 2018; Mahmoud and Gan 2018; Samanta et al., 2018b; Dano et al., 2019; Kanani-Sadata et al., 2019; Khosravi et al., 2019a; Liu et al., 2019; Mind'je, et al., 2019; Wang et al., 2019; Vojtek and Vojteková 2019). ~~In the current research, n~~Nine flood-influencing variables were used, which generated and stored in a database folder in a Geographic Information System (GIS) for data interpretation and analysis. These variables include distance from wadi, landuse/landcover (LULC), lithology, slope-angle, TWI, altitude, slope length (LS), curvature, and slope-aspect (Fig. 5). All layers were converted into a grid spatial database by 30×30-m pixel size which have UTM coordinate system zone 36 with a datum of WGS 84. Seven themes were extracted from DEM (five layers including slope-aspect, slope-angle, altitude, distance from main wadis, and curvature, were extracted using ArcGIS 10.5 software and two layers including topographic wetness index and slope length, were extracted using SAGA software).

The main Wadis consider the pathways for runoff waters where the nearby areas are vulnerable to flooding (Opperman et al., 2009). The shorter the distance from the main wadis, the higher the probability of flooding, especially where the wadis have a low storage capacity (Predick et al. 2008). In this study, main Wadis were extracted from the DEM and verified using the topographic map (1:50,000). Distance from main Wadis was calculated using the Euclidean tool in ArcGIS 10.5 environment (Fig. 5a). This map was categorized into 5 classes from 0-100m, 100-200m, 200-300m, 300-400m, and >400m.

Landuse/landcover in any area has a crucial impact in runoff velocity, interception, percolation, and evapo-transportation (Yalcin et al., 2011). Different soil characteristics can impact the extent of runoff in the basin area. Some soil types has greater infiltration of rainfall compared to others, which leads to a smaller runoff volume (Tehrany et al. (2019). Many studies indicated that LULC map is vitally important in identifying of flood-prone areas (Karlsson et al., 2017; Komolafe et al., 2018). The landuse/landcover map was prepared from the interpretation of Landsat satellite images (OLI) acquired in 2018. Four LULC types were extracted, including bare rock, bare soil, rainfed less tree crop, and grass~~-~~land (Fig. 5b).

Lithological units can affect ~~have an important role on~~the hydrological processes (the amount and speed of water flow) due to the differences in permeability of rocks and sediments in any watershed area (Ward and Robinson, 2000; Regmi et al. 2013; Khosravi et al. 2019b). In this study, lithology units were extracted from the geological database (1:250,000 scale). Four main lithological units were mapped including (1) Wadi deposits, (2) gravel deposits, (3) sedimentary rocks, and (4) Precambrian rocks (Fig. 5c).

Slope-angle considers an important physiographic parameter in flood behavior where the runoff velocity increased in high slope areas and water will inundated low slope areas (Meraj et al., 2015; Tien Bui et al., 2016; Rahmati et al. 2016). Tehrany et al. (2019) mentioned that steep slopes have less time for infiltration, which causes an increase in water flow. The slope-angle map~~,~~ was generated from the DEM layer in ArcGIS environment, ~~. In the study area, the slope angles~~ rangese from 0.0° to 84° (Fig. 5d).

Topographic wetness index (TWI) represents the spatial variations of wetness (amount of water collected) in a watershed area (Gokceoglu et al. 2005; Rahmati et al. 2016). It is applied to measure topographic control on hydrological procedures (Chen and Yu, 2011). TWI is calculated according equation (1):

$$TWI = lin\ \frac{A}{\tan B} \qquad (1)$$

where A is the specific catchment area (m^2) and β is the slope gradient (in degrees), respectively.

TWI shows the water infiltration capability in an area, and subsequently, the regions with potential for floods. In fact, flat area absorbs more water than steep terrain, where the gravity acting increase the water flowing down the hilly slopes towards flat areas (Tehrany et al. (2019). Areas around streams and flat lands (flooded areas) have greater TWI value than that in areas with slopes. In the current study, the TWI was calculated in the SAGA-GIS environment ranging from 2.39 to 24.79 (Fig. 5e).

Flood occurrence is likely affected by altitude where low elevation regions are more prone to floods (Botzen et al. 2013). Runoff moves from the hillsides of mountains (high elevation areas) and reaches the lower ground (lower elevation areas), causing flooding. ~~altitude~~ Altitude is controlled by several geological and geomorphological processes (material types, wind action, rainfalls, and erosions) (Tehrany et al. 2014a,b; Tien Bui et al. 2016; Khosravi et al. 2016a). Kia et al. (2012) indicated that the altitude considers an amplifying factor in the occurrence of floods because it has an influence on the amount and velocity of runoff. Subsequently~~,~~ altitude has a vital role in identifying areas that are susceptible to flooding. Altitude values of the study area range from 113 to 1,878 m (Fig. 5f).

Slope length (LS) is an important factor, which ~~in which soil erosion can be detected (Bohner and Selige 2006).~~ describes soil erosion, represents the combined effects of slope length and steepness, and affects soil particle transport (Bohner and Selige 2006; Park et al. 2019). Bera (2017) indicated that as the slope length increases, the soil erosion due to water also increases as a result of greater accumulation of surface runoff. It was calculated in the SAGA-GIS environment using the universal soil loss equation (USLE) based on slope and specific catchment area. In the current study, slope length (LS) ranges from 0 to 73.6 (Fig. 5g).

Slope-aspect can be defined as the direction of the maximum slope of the earth surface. The slope-aspect map was derived ~~generated~~ from ~~in ArcGIS environment from~~ the DEM map in ArcGIS environment. The slope-aspect layer is shown in classes of flat (−1), North (0°–22.5°; 337.5–360°), North-East (22.5–67.5°), East (67.5–112.5°), South-East (112.5–157.5°), South (157.5–202.5°), South-West (202.5–247.5°), West (247.5– 292.5°), and North-West (292.5–337.5°) (Fig. 5i).

In the current study, the flood-influencing variables were nominal, ordinal, and scale. Some factors are ordinal, such as slope-angle, curvature, distance from main Wadis, TWI, and LS, while altitude was in a ratio scale; however, after classification it transformed to an ordinal scale. In addition, the nominal factors are lithology, LULC, and slope aspect.

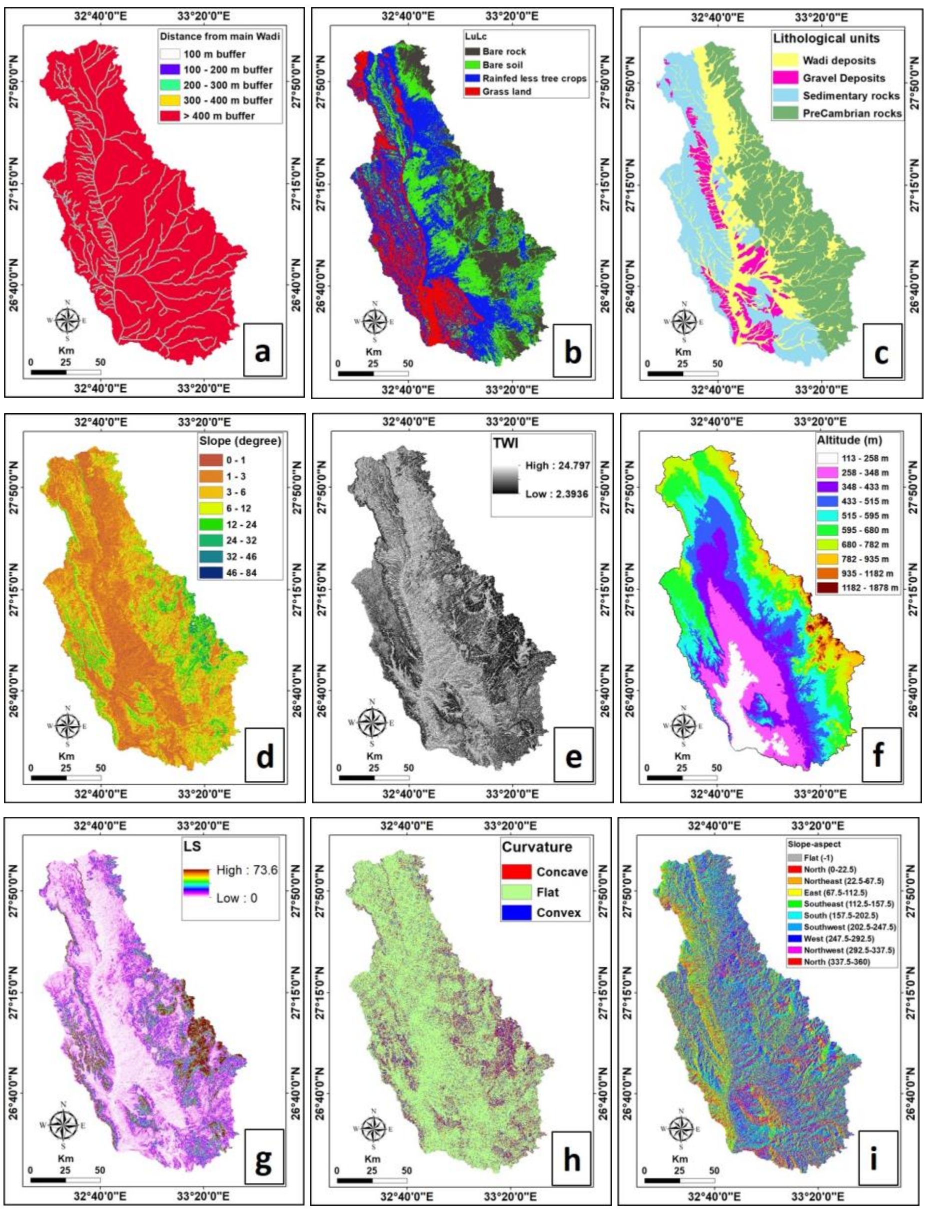

Fig. 5 List of all the flood-influencing data layers. a) Distance from main Wadis, b) LULC c) Lithology unit, d) Slope-angle, e) TWI, f) Altitude, g) LS, h) Curvature, and i) Slope-aspect

3.3. Stage III: Application and validation of machine learning techniques

3.3.1. Application of BRT

The BRT, which has been proposed by Friedman (2001), is a combination of statistical and machine learning methods. It ~~The BRT~~ is aiming to enhance the performance of a single model by fitting and combining many models together (Schapire 2003; Park and Kim 2019). Elith et al. (2008) indicated that the BRT model does not require~~d~~ data transformation or elimination of outliers, and can fit complex nonlinear relationships and automatically address interaction effects between variables. In the BRT model, two algorithms, a regression tree and a boosting algorithm ~~namely boosting and regression,~~ are used where their strengths are combined to enhance the model accuracy and decrease the model variance (Aertsen et al. 2010; Rahmati et al. 2018). Boosting technique, a powerful learning method, is improving model accuracy due to iteratively fitting new trees to the residual errors (RE) of the existing tree assemblage (Cao et al. 2010; Döpke et al. 2017; Pourghasemi and Rahmati 2018). For example, by using the dataset D, the boosting algorithm enhances the regression tree model, F(x) by adding an estimator, h(x) to derive a new BRT model, $F_{new}(x)$ as shown in Equation (2). This is an iteration process, where the number of iterations (M) plays a crucial role in the performance of the final BRT model. To construct the loss function~~,~~ equation (3) is used.

$$F_{new}(x) = F(x) + \gamma h(x) \qquad (2)$$

where γ ∈ (0, 1) is the learning rate which is applied to control the problem of over-fitting.

$$L = \frac{1}{2}[y - F(x)]^2 \qquad (3)$$

At each iteration, a new tree add to the original model must confirm the reduction of the loss function. The BRT training phase will be completed when the pre-defined number of iterations is achieved.

3.3.2. Application of FDA

The FDA method, was firstly proposed by Ramsay and Dalzell (1991), is suitable for the observation data consisting of a series of real functions. FDA is efficient in solving the problem that some key data points may be omitted or deleted. In addition, with the data described as function forms, some dynamic information hidden in data sets can be analyzed by derivation and dimension reduction. Battista et al., (2016) and Wagner-Muns et al., (2018) indicated that the main point of FDA is to consider all data of an observation object containing functional properties as an integral instead of a sequence of observed values. FDA has been widely applied in the problem of classification (Cho et al., 2016; Seifi Majdar and Ghassemian, 2017; Chen et al., 2019). The basic analysis objects of FDA are a sequence of observations expressed as functions. FDA can be applied with machine learning methods in classification problems. The basic steps to apply FDA include: 1) selecting training and testing data sets and executing functional data representation; 2) extracting function data features using functional principal component analysis (FPCA); 3) categorizing data features via machine learning methods; and 4) verifying the validation of the classification model by testing data sets. In the current study, the FDA method was utilized to develop the flood susceptibility assessment model based on existing methodologies and theories according to species distribution modeling (SDM) package in R (Naimi and Araújo, 2016).

3.3.3. Application of GLM

Generalized linear model (GLM) is an extension of linear regression models in which the special and temporal variables could be quantified and incorporated (McCullagh and Nelder 1989; Dobson 2001; Guisan et al. 2002). The GLM is a very popular statistical model due to its capability to carry out non-linear relationships and various statistical distributions characterizing spatial data types (Hjort et al. 2007; Marmion et al. 2008). The relationship between the expectation of the response variable and the linear combination of explanatory variables can be established using the link function of GLM (Venables and Dichmont 2004; Ahmedou et al. 2016; Kéry and Royle 2016; Soch et al. 2017). The expectations and variances of the response variables can be calculated by equations (4, 5):

$$\mu_i = E[Y_i] = g^{-1}\left(\sum_j X_{ij}\beta_j + \varepsilon_i\right) \quad (4)$$

$$var[Y_i] = \frac{\phi V(\mu_i)}{\omega_i} \quad (5)$$

where Y_i is the vector of response variables, X_{ij} is the matrix of explanatory variables, β_j is the vector of pending parameters, ε_i is the interference terms, g(x) is the corresponding link function, V(x) is the variance function, ϕ is the dispersion parameter of V(x), and ω_i is the weight of the i-th observed value.

In the current study, suppose Y is the response variable, which represents where flood inundation has happened in a raster, and X_i is the i-th flood conditioning factor. So, the occurrence probability of event Y can be expressed as equation (6). By logistic transformation, the link function $g(y_i)$ is shown in equation (7).

$$P = \frac{\exp(c_0 + c_1 X_2 + c_2 X_2 + \cdots + c_i X_i)}{1 + \exp(c_0 + c_1 X_2 + c_2 X_2 + \cdots + c_i X_i)} \qquad (6)$$

$$g(y_i) = c_0 + \sum c_i x_i + \varepsilon_i \qquad (7)$$

where P is the occurrence probability of event Y, and c_0; c_1;...; c_i are logistic regression coefficients, ε_i is the residual errors.

In the current research, R statistical package was used to build the GLM model. A simple Gaussian family was identified to be the link function for the normally distributed response data. Aertsen et al. (2009) indicated that independent variables should enter the model individually using a smoothing spline with only 2 degrees of freedom in a polynomial fit of degree 2 to avoid over fitting.

3.3.2. Application of MDA

The MDA is considered ~~to be~~being a linear discriminate analysis (LDA). In LDA, a collection is assumed to be a portion of the nearest cluster. The distance is generally calculated by the normal distribution of the variables, and in each category, it is assumed that the variability and correlation among the variables are equal (Lombardo et al. 2006). In MDA, multiple normal distributions are used within each category. According to Hair et al. (1998), the MDA can derive the linear combinations using equation (~~2~~7).

$$Y = W1X1 + W2X2 + WnXn \qquad (\cancel{2}7)$$

where Y is a discriminant score, Wi (i = 1,2,3, …, n) are discriminant weights, and Xi (i=1,2,3,…, n) are independent variables.

3.3.3. ~~Multicolinearity~~Multicollinearity of flood e~~E~~ffective f~~F~~actors

Before models run, a multicollinearity analysis of the independent variables needs to be conducted. Multicolinearity is a statistical approach in which numbers of independent variables in a multiple regression model are strongly correlated, the variables with significant collinearity are eliminated

Commented [BP1]: Reference plz

(Chen et al. 2017; Pourghasemi et al. 2017; Saha 2017). Commonly -used indicators are two exponents, Variance Inflation Factors (VIF) and Tolerance (TOL), which applied for considering multicollinearity of variables. They can be calculated using equations (8, 9):

$$TOL = 1 - R_J^2 \qquad (8)$$

$$VIF = \left[\frac{1}{T}\right] \qquad (9)$$

where, R_J^2 is the coefficient of determination of a regression of explanatory J on all the other explanatory.

Various literatures indicated that a TOL of less than 0.10 and VIF of more than 5 indicate multi-~~colinearity~~collinearity problems (Hosmer and Lemeshow 1989; Menard 2001).

3.3.4. Factors importance

In the recent ~~days~~years, research on the ~~stability of~~ factor impact measurements based on machine learning algorithm (random forest) has received a great deal of ~~high~~ attention (Wang et al. 2016). Factor impact measurement in a random forest can be calculated based on two representative methods. These methods are divided into two categories: Mean Decrease Impurity (MDI) and Mean Decrease Accuracy (MDA), which proposed by Breiman (2001). The Mean Decrease Impurity (MDI) index measures the classification impact of variables by totaling the amount of decrease in impurity as the classification is performed. The sum of the impurity reductions in all the trees is calculated as the importance of the variable. For impurity reduction, classification trees use Gini coefficient index or information gain, and regression trees use the mean value of variables. The variable importance (VI) for MDI method is calculated using equations (10) (Strobl et al. 2008), it adds up the decrease of Gini index of each of the variables from 1 to n_{tree}, which means the number of trees, and gets the average of all. The advantage of MDI method is being easy to compute, but it has the disadvantage that it can be biased only for categorical variables that contain multidimensional attributes.

$$VI(x_j) = \frac{1}{n_{tree}}\left[1 - \sum_{k=1}^{n_{tree}} Gini(j)^k\right] \qquad (10)$$

The Mean Decrease Accuracy measures the classification impact of variables by the sum of the amount of decrease in accuracy depending on the presence or absence of specific variables. MDA method calculates variable importance by permutation. The method uses OOB (Out-Of-Bag) to divide its sample data. The OOB is one of the subsampling techniques to calculate prediction error of each of the training samples using bootstrap aggregation. MDA calculates variable importance using equation (11) (Strobl et al. 2008). OOB estimates more accurate prediction value by computing OOB accuracy before and after the permutation of variable x_j and compute the difference. Since $t \in \{1,2,3,...,n\text{tree}\}$, the variable importance of x_j in tree t is the averaged value of the difference between predicted class before permuting x_j , which is $y_i = f(x_i)$, and after permuting variable x_j, which is $y_i = f\left(x_i^j\right)$, in certain observation i.

$$VI\left(x_j\right) = \frac{1}{n_{tree}} \sum_{t=1}^{n_{tree}} \frac{\sum_{i \epsilon OOB} I\left(y_i = f(x_i)\right) - \sum_{i \epsilon OOB} I\left(y_i = f\left(x_i^j\right)\right)}{|OOB|} \qquad (11)$$

~~3.3.3.~~3.3.5. Model validation

Remondo et al. (2003) mentioned that validation approach could be used as a guidance in data collection and field practice for susceptibility mapping, Chung and Fabbri (2003) used sensitivity analysis for individual factors and combinations of factors to test the validation of various map-producing methods, Tien Bui et al. (2012) indicated that the accuracy and success rate used to validate the flood susceptibility models. The receiver operating curve (ROC) is the most crucial method applied for verification of the susceptibility models (e.g. landslides and flood), in which the prediction accuracy and quality of the constructed models are examined using the area under the curve (AUC) (e.g., Lee and Pradhan 2007; Chauhan et al. 2010; Akgun et al. 2012; Mohammady et al. 2012; Tien Bui et al. 2012; Pourghasemi et al. 2012; Ozdemir and Altural 2013; Jaafari et al. 2014; Youssef et al. 2016; Youssef and Hegab 2019). A suitable flood model should have~~s~~ an AUC value ranges from 0.5 to 1, and the quality of the model is increased by increasing the AUC value. The model considered ~~to be~~being random, if the AUC value below 0.5. The susceptibility models might produce the highest accuracy and reliability when the AUC value is equal or close to 1.0 which showing the capability of the model to predict disaster occurrence without any bias (Pradhan et al. 2010; Tien Bui et al. 2012).

4.1. Multi- multicollinearity test

The results of the multicollinearity analysis among nine flood-influencing factors used in this study are presented in Table 3. This analysis indicated that the tolerance and VIF of all flood-influencing factors used in this study were > 0.1 (0.539) and < 10 (1.857), respectively. As a result, there is no multicollinearity among the independent flood-influencing factors, which enables them to participate in model establishing~~.~~ ~~As a result, there is no multicollinearity among the independent flood-influencing factors used~~ in the current study.

Flood-influencing Factors	**Unstandardized Coefficients**		**Standardized Coefficients**			**Collinearity Statistics**	
	B	**Std. Error**	**Beta**	**T**	**Sig.**	**Tol**	**VIF**
Slope Length	5.924E-6	.000	.008	.221	.825	.971	1.029
Slope-angle	-.013	.004	-.137	-3.190	.002	.639	1.564
Distance from main Wadi	-8.075E-5	.000	-.292	-7.439	.000	.764	1.308
LULC	.125	.022	.254	5.741	.000	.599	1.669
Lithology	.087	.019	.211	4.528	.000	.539	1.857
Altitude	-5.819E-5	.000	-.021	-.456	.649	.547	1.827
Curvature	.075	.045	.065	1.662	.097	.779	1.284
Slope-aAspect	-.006	.008	-.027	-.764	.445	.958	1.044
TWI	6.157E-6	.000	.008	.229	.819	.972	1.029

4.2. Variables importance

In the current study, an attempt was carried out to evaluate the importance of effective flood-influencing factors using a random forest data-mining technique. The results, is shown in Fig. 6, depicted that the river distance, LULC, and lithology factors are the most important, followed by slope, TWI, altitude, and LS which are moderately important flood-influencing factors, and then curvature and aspect are less important. However, according to the mean decrease gini, it was found that river distance factor is the most important, followed by altitude, lithology, LULC, slope,

Formatted: Font: Bold

TWI, and LS which are moderately important flood-influencing factors, and then curvature and aspect are less important. The results indicated that river distance is extremely important in the occurrence of floods.

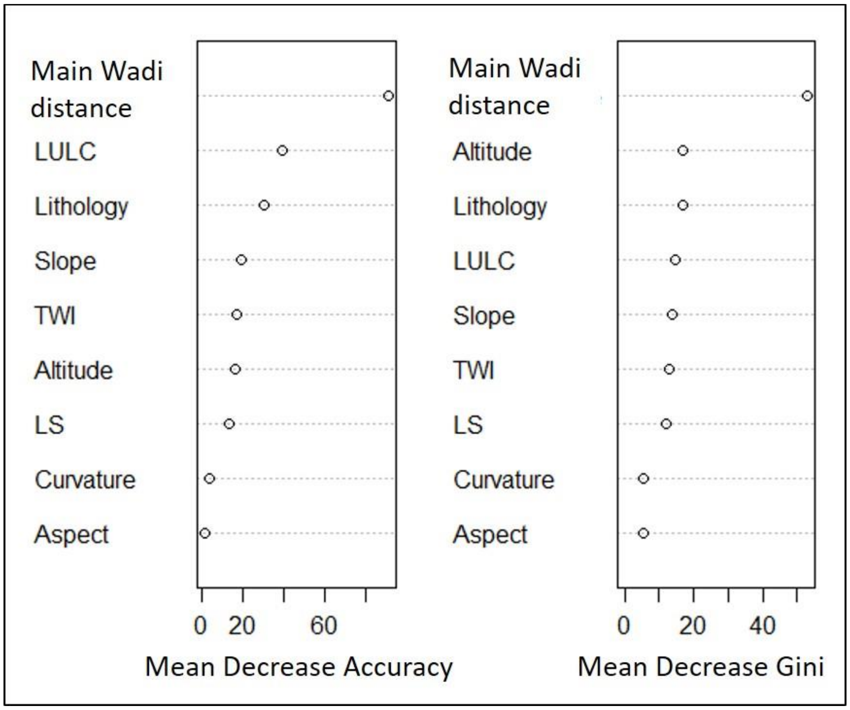

Figure 6. The importance of flood-influencing factors using a random forest model

4.3. Flood susceptibility maps

Using the training dataset, the MDA, GLM, FDA, and BRT models were established to obtain the flood susceptibility index (FSI) for the study area (Fig. 7 (a-d)). Subsequently, the LSI pixels of the study area were classified ~~applied~~ into different zones of susceptibility to produce the flood susceptibility maps using the ArcGIS 10.5 software. The most common methods used in natural hazard susceptibility index classification are natural break, equal interval, and quantile (Ayalew and Yamagishi 2005). In the current work, the flood susceptibility maps were finally divided into five classes based on the natural break method scheme (Nicu, 2018) (Fig. ~~7~~ 8 (a–d)). Finally, results revealed that very low, low, moderate, high, and very high flood susceptibility map (FSM) classes derived using the MDA model cover 19.5, 21.5, 20.0, 19.6, and 19.4% of the total area, respectively (Fig. ~~7a~~8a); 19.5, 21.9, 19.9, 19.4, and 19.3% of the total area covered by very low, low, moderate, high, and very high respectively on the FSM map obtained from the GLM method (Fig. ~~7b~~8b); 19.4, 20.9, 20.3, 19.7, and 19.7 % of the total area are related to very low, low, moderate, high, and very high FSM zones, respectively, using the FDA model (Fig. ~~7c~~8c). According to the BRT model, 16.2, 23.9, 20.2, 20.5, and 19.2% of the study areas were classified as very~~-~~ low, low, moderate, high, and very~~-~~ high susceptibility respectively (Fig. ~~7d~~8d). The real inundating flood zones were extracted from the sentinel images (10 m resolution) after the flood event in 2016 in order to test the performance of the used models (Fig. 5). The comparison shows good matches between the areas were inundated in 2016, along wadi Qena basin, and the results of the susceptibility models. Finally, it can be noticed that the high flood susceptible zones in all produced models are mainly located along the main course of wadi Qena and its tributaries. In addition, these models indicated that a large portion of the study area were classified as very low, low, and moderate susceptible zones (61%, 61.3%, 61.5%, and 60.3% for MDA, GLM, FDA, and BRT models respectively).

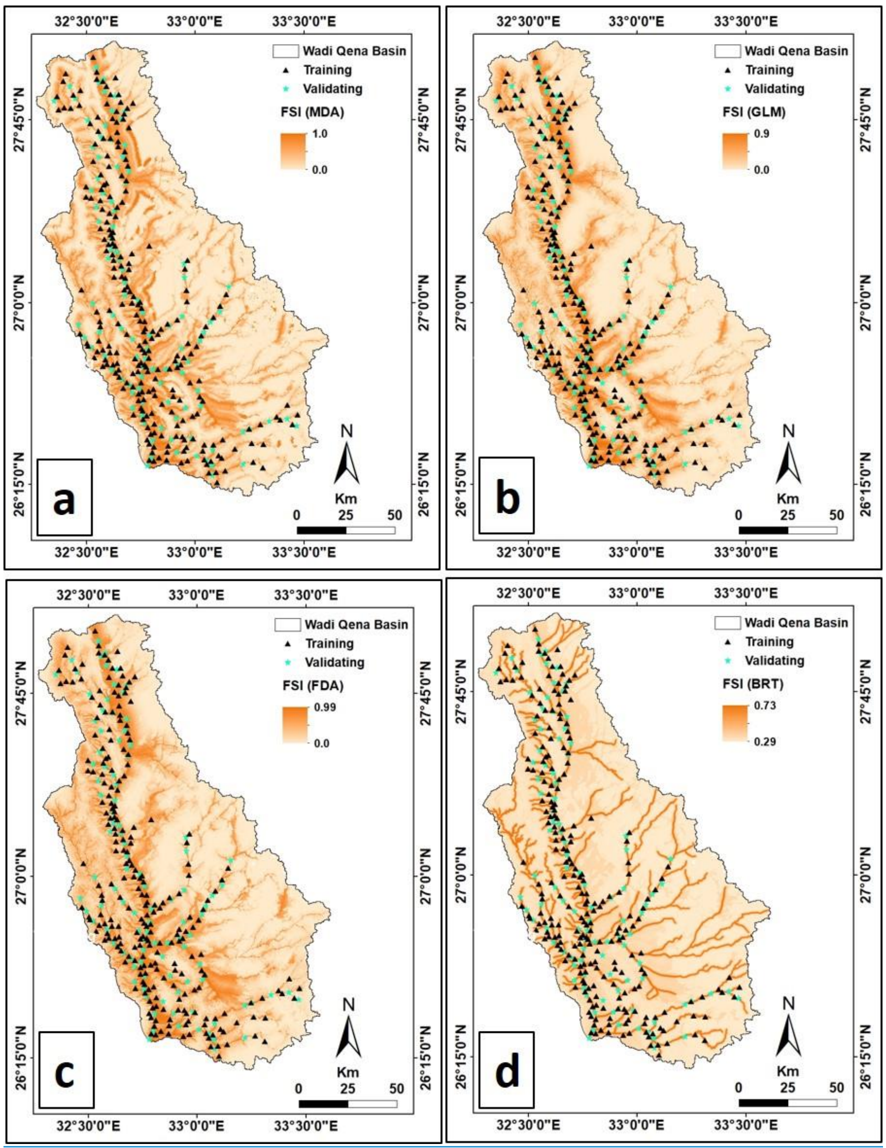

Fig. 7 Flood susceptibility index maps derived from: (a) MDA, (b) GLM, (c) FDA, and (d) BRT.

Fig. ~~7~~ 8 ~~Generated f~~Flood susceptibility maps using a) MDA, b) GLM, c) FDA, and d) BRT

To evaluate the reliability of the obtained susceptibility maps, an accuracy assessment was performed using the AUC method. Many authors emphasize the importance of validation method for susceptibility maps. In the current study, the (ROC) curve was used to identify true- and false-positive rates (plot the sensitivity of the model (the percentage of existing flood pixels correctly predicted by the model) against 1-specificity (the percentage of predicted flood pixels over the total study area). The derived flood susceptibility index maps (Fig. 7) have been validated through both success rate method (using the training flood locations that were used in establishing the flood models) and prediction rate method (using validating flood locations which examine how well the model predicts the flood). The success and prediction rate curves were used to understand the effectiveness of each model and their validation as shown in Fig. ~~8~~ 9 (a, b). In the success rate curves, the AUC values for the MDA, GLM, FDA, and BRT models are 0.919, 0.918, 0.917, and 0.921, respectively (Fig. ~~8a~~9a). In addition, the prediction rate curve showed that the AUC values for the MDA, GLM, FDA, and BRT models are 0.968, 0.967, 0.966, and 0.974, respectively (Fig. ~~8b~~9b). It can be concluded that all these models give the success and prediction rate curve values above 0.9, showing that models for flood susceptibility mapping in the study area are reasonable. These represent reasonable models for flood susceptibility mapping in the study area. In addition, the results show that these models show an excellent accuracy in flood susceptibility analysis with so tiny differences.

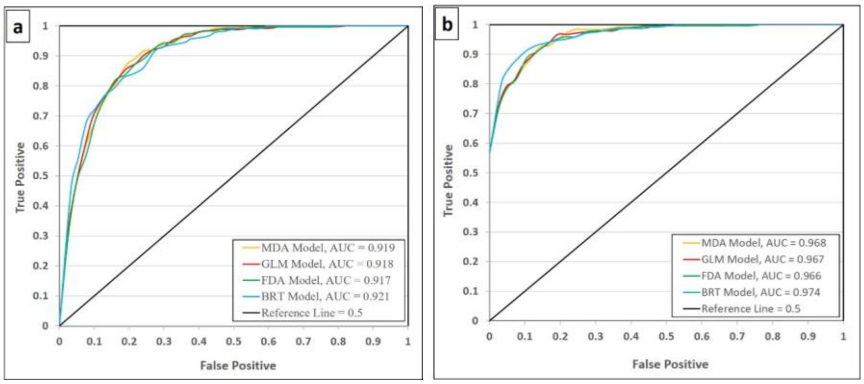

Fig. ~~8~~ 9 Success rate (a) and prediction rate (b) curves for models derived from the MDA, GLM, FDA, and BRT.

Flood hazard, vulnerability and risk should be analysed effectively, specifically for major events that come more frequently as a part of the climate change impact. According to (Tehrany et al., 2014b), categorization of the outputs from several methods into maps of flood susceptibility analysis is a crucial step. The models that were employed such as MDA, GLM, FDA and the BRT in this research for flood-susceptibility mapping out of which all outputs provide unique results based on natural break classification technique with different significance. Results are close to each other and a little difference can be found in map based on BRT than other three methods. The finding of BRT in the current study confirm previous results, which indicated that BRT is one of the most accurate model for identifying flood-vulnerable areas (Rahmati and Pourghasemi 2017). Selection of 9 variables that contribute to flooding as contributing factors helped in calculating susceptible areas according to four models, which demonstrate the relationships between inventory data of flooded-area with the applied flood-influencing factors. These nine thematic maps were extracted from different sources, such as remote sensing images (30m resolution), digital elevation models (30m resolution), and geologic map. Wadi map was verified using field investigation and topographic map 1: 50000 resolution. In addition to that inventory map was prepared based on field visits, historical records, and high-resolution image analysis (slicing technique). Therefore, set up of spatial datasets that justify the relevant factors help to execute and map the areas of flood occurrences and the indicated correlation between four methods. Validation was successfully conducted with accuracy more than 90 % using the flooding data that was employed for training.

Wadi Qena basin is suffering a great loss because of unforeseen weather conditions and floods. In this work, our resulting map was based on four different methods which were analysed, compared, and helped to understand the usefulness of several models and applications. The very high and high areas in all those maps were distributed adjacent to the border areas of the Wadi Qena basin (~~built-up area in~~ Fig. ~~7~~8). However, flood frequency and intensity have been increased in the twenty-first century in this basin. Accordingly, future planning and development of this wadi area will be under the flood hazard. This wadi area during the last decade was impacted by different flood events. This wadi is characterised by low drainage density, which is the primary reason for the high susceptibility in the basin. The eastern part of the basin is less susceptible to flooding because of its high elevation; however, the neighbouring regions from the northern to southern along the basin showed a high susceptibility. The area continuously experiencing damages inflicted by floods undergoes a series of changes over time. It imposes a limitation on a spatial flood analysis. If the location information is incorrect, this could lead to substantial spatial analysis problems. However, the drainage facilities, water-supply system can create an effect on flood susceptibility assessment.

Our findings are innovative and provide good mapping results as expected. According to the study by Al-Abadi (2018), AdaBoost model with significant results outperformed Random Forest and other models as per the validation dataset. According to the results, the RF and AdaBoost models achieved 94% accuracy and outperformed the RTF model, which is 92%. However, Lee et al. (2017) described that classification accuracy can be achieved better in random forest than boosted-tree model. The accuracy for regression and classification model based on RF was 78.78% and 79.18%, while 77.55% and 77.26% in the case of BRT. Khosravi et al. (2018) presented in the assessment of a flash flood susceptibility mapping at the Haraz Watershed in Iran, showed that the Alternating Decision Trees (ADT) and BRT model had the highest predictive accuracy than other models.

However, according to our results, BRT model achieved the highest accuracy concerning the mapping of flood susceptible areas, followed by the MDA, GLM and FDA models. Our results is in agreement with Rahmati et al. (2019) study. They indicated that the highest validation methods in the application of support vector machine (SVM), boosted regression tree (BRT), and generalized additive model (GAM) for multi hazard mapping is the BRT which demonstrated the best performance for flood hazards (AUC = 94.2%). To produce an outcome, the major concern is the computational time, and there is a requirement of considerable time to produce an appropriate form of spatial data. The transformation of data into maps using the above three methods is a time-consuming process involving the usage of several third-party software. This work provides sophisticated numerical results of flood-susceptibility-map that can be applied for vulnerability and risk assessment in the future.

Although all four models successfully identified flood susceptibility areas in the Wadi Qena basin, however, susceptibility maps obtained from MDA, GLM and FDA and BRT could reflect the spatial heterogeneity of the build-up areas and describe more details of expected susceptible areas. In general, BRT model provided slightly better than the other methods. Nevertheless, to determine the best classifier in this study is difficult because all the employed models performed similarly. However, the success rate curves and the prediction rate curve showed that BRT achieved 0.921 and 0.974 with the highest prediction ability based on the used statistical measures. Thus, at the end it is confirmed that BRT classifier can be consider as a base classifier which exhibit the best performance in flood susceptibility mapping in Wadi Qena basin. Therefore, the local government agencies and decision makers could adopt the produced map to implement suitable plans to mitigate future flood damages.

~~6.~~5. Conclusions

Regarding the current and future climate changes, floods have been represented to be the most devastating natural hazards causing loss of lives and properties damages worldwide. Accordingly, effective methods are required to delineate the most vulnerable areas for floods. Flood susceptibility models represent a crucial approach to map and delineate the flood vulnerable areas. These flood susceptibility models can be achieved using advanced statistical approaches that could be integrated in R and GIS environment. The current work aiming at investigating and applying four data mining models named MDA, GLM, FDA, and BRT, which considered ~~to be~~being novel techniques to perform the flood susceptibility mapping in the Wadi Qena Basin, Egypt. Nine flood-influencing variables (slope-angle, slope-aspect, altitude, distance from main wadis, lithology, curvature, land use, slope length, and topographic wetness index) were constructed and utilized with the aid of a flood inventory data (training and validating data) to build the FSMs. The success rate and prediction rate curves were applied to evaluate the stability and predictability performances of the four flood susceptibility maps produced from the proposed models. The area under the curve (AUC) was calculated based on the training and the validating datasets. The AUC values of the success rates are 91.9%, 91.8%, 91.7%, and 92.1%, and of the prediction rates are 96.8%, 96.7%, 96.6%, and 97.4%, respectively for the MDA, GLM, FDA, and BRT models. Findings from this current work was verified using flood inundated areas, which extracted from the sentinel images after flood event in 2016. Results indicated that the applied models are adequately representing the quantitative relationships between flood occurrences and multiple spatial data variables (flood-influencing variables). Many countries (decision-makings, planners, and private sectors) have been adapting flood susceptibility modeling as a preliminarily essential step in overall flood management program to identify the flood-vulnerable areas that could prevent excessive urbanization extension in susceptible flood-prone areas and/or minimize the potential damages and losses caused by existing and future floods.

References

Abdelkarim A, Gaber AFD, Youssef AM, Pradhan B (2019) Flood Hazard Assessment of the Urban Area of Tabuk City, Kingdom of Saudi Arabia by Integrating Spatial-Based Hydrologic and Hydrodynamic Modeling. Sensors, 19, 1024.

Aertsen W, Kint V, Van Orshoven J, Ozkan K, Muys B (2009) Performance of modelling techniques for the prediction of forest site index: a case study for pine and cedar in the Taurus mountains, Turkey. XIII World Forestry Congress, Buenos Aires, pp 1–12

Aertsen W, Kint V, Van Orshoven J, Özkan K, Muys B (2010) Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. Ecol. Model., 221, 1119–1130.

Ahmedou A, Marion JM, Pumo, B. (2016) Generalized linear model with functional predictors and their derivatives. Journal of Multivariate Analysis, 146(Supplement C), 313-324. https://doi.org/10.1016/j.jmva.2015.10.009.

random forest modeling approach to flood analysis in a regulated reservoir system. Can. Water Resour. J., 41, 250–260

Al-Abadi AM (2018) Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study. *Arabian Journal of Geosciences*, *11*(9), 218.

Al-Juaidi AM, Nassar AM, Al-Juaidi OEM (2018) Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. Arabian Journal of Geosciences (2018) 11: 765. https://doi.org/10.1007/s12517-018-4095-0

Ashour MM (2002) Flashfloods in Egypt (a case study of Drunka village – Upper Egypt). Bull Soc Geog Egypte. 75:101–114.

Battista TD, Fortuna F, Maturo F (2016) BioFTF: An R package for biodiversity assessment with the functional data analysis approach. Ecological Indicators, 73, 726-732.

Bera A (2017) Estimation of soil loss by USLE model using GIS and Remote Sensing techniques: A case study of Muhuri River Basin, Tripura, India. Eurasian Journal of Soil Science, 6(3), 206-215. DOI: 10.18393/ejss.288350

Botzen W, Aerts J, Van den Bergh J (2013) Individual preferences for reducing flood risk to near zero through elevation. Mitigation and Adaptation Strategies for Global, 18(2), 229–244 DOI 10.1007/s11027-012-9359-5.

Breiman L (2001) Random forests. Machine Learning, 45(1), 5-32.

Cao DS, Xu QS, Liang YZ, Zhang LX, Li HD (2010) The boosting: A new idea of building models. Chemom. Intell. Lab., 100, 1–11

Chauhan S, Sharma M, Arora MK (2010) Flood susceptibility zonation of the Chamoli region, Garhwal Himalayas, using logistic regression model. Floods 7:411–423. doi: 10.1007/s10346-010-0202-3

Chen YR, Yeh CH, Yu B (2011) Integrated application of the analytic hierarchy process and the geographic information system for flood risk assessment and flood plain management in Taiwan. Nat Hazards 59(3):1261–1276

Chen W, Pourghasemi HR, Panahi M, Kornejady A, Wang J, Xie X, Cao S (2017) Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. Geomorphology, 297, 69–85. doi: 10.1016/j.geomorph.2017.09.007.

Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A (2019) An Ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Elsevier Sci. Total Environ., 651, 2087–2096.

Dano UL, Balogun AL, Matori AN, Wan Yusouf K, Abubakar IR, Said Mohamed MA, Aina YA, Pradhan B (2019) Said Mohamed, M.A.; Aina, Y.A.; Pradhan, B. Flood Susceptibility Mapping Using GIS-Based Analytic Network Process: A Case Study of Perlis, Malaysia. Water, 11, 615. https://doi.org/10.3390/w11030615

Dandapat K, Panda GK (2017) Flood vulnerability analysis and risk assessment using analytical hierarchy process model. Earth Syst. Environ. 3:1627–1646. DOI 10.1007/s40808-017-0388-7

Dawod GM, Mirza MN, Al-Ghamdi KA (2012) GIS-based estimation of flood hazard impacts on road network in Makkah city, Saudi Arabia. Environ Earth Sci 67:2205-–2215. doi: 10.1007/s12665-012-1660-9

Döpke J, Fritsche U, Pierdzioch C (2017) Predicting recessions with boosted regression trees. Int. J. Forecast 33: 745-759.

Echogdali FZ, Boutaleb S, Jauregui J, Elmouden A (2018) Cartography of Flooding Hazard in Semi-Arid Climate: The Case of Tata Valley (South-East of Morocco). J Geogr Nat Disast 8: 214. doi: 10.4172/2167-0587.1000214

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J. Anim. Ecol., 77, 802–813.

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J. Anim. Ecol. 77 (4), 802–813.

Feng Q, Liu J, Gong J (2015) Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—A case of Yuyao, China. Water 2015, 7, 1437–1455

Foody GM, Ghoneim EM, Arnell NW (2004) Predicting locations sensitive to flash flooding in an arid environment. J Hydrology. 292:48–58

Fotovatikhah F, Herrera M, Shamshirband S, Chau K-W, Ardabili SFaizollahzadeh, Piran MJ (2018) Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work. Engineering Applications of Computational Fluid Mechanics 12(1): 411–437. DOI 10.1080/19942060.2018.1448896.

Friedman JH (2001) Greedy Function Approximation: A Gradient Boosting Machine. Ann. Stat 29: 1189-1232.

Gokceoglu C, Sonmez H, Nefeslioglu HA, Duman TY, Can T (2005) The 17March 2005 Kuzulu landslide (Sivas, Turkey) and landslide-susceptibility map of its near vicinity. Eng. Geol. 81, 65–83.

Guisan A, Edwards, TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: Setting the scene. Ecological Modelling, 157(2), 89-100. https://doi.org/10.1016/S0304-3800(02)00204-1.

subarctic Finland: a grid-based modelling approach. Permafrost and Periglacial Processes 18: 115–127.

Hosmer DW, Lemeshow S (1989) Applied Regression Analysis; John Wiley and Sons: New York, NY, USA, 1989; ISBN 978-0-470-58247-3.

Irin (2013) Preparing for floods in West Africa. Available from: http://reliefweb.int/report/nigeria/-preparing-floods-west-africa

Kanani-Sadata Y, Arabsheibani R, Karimipour F, Nasseri M (2019) A new approach to flood susceptibility assessment in data-scarce and ungauged regions based on GIS-based hybrid multi criteria decision-making method. Journal of Hydrology, 572, 17-31. https://doi.org/10.1016/j.jhydrol.2019.02.034

Karlsson CS, Kalantari Z, Mörtberg U, Olofsson B, Lyon SW (2017) Natural hazard susceptibility assessment for road planning using spatialmulti-criteria analysis. Environ. Manag. 60 (5), 823–851.

Karmaoui A, Messouli M, Yacoubi Khebiza M, Ifaadassan I (2014) Environmental Vulnerability to Climate Change and Anthropogenic Impacts in Dryland, (Pilot Study: Middle Draa Valley, South Morocco). J Earth Sci Clim Change 2014, S11, pp: 1-12.

Kéry M, Royle JA (2016) Linear models, generalized linear models (GLMs), and random effects models: The components of hierarchical models. In M. Kéry, and JA. Royle (Eds.), Applied hierar-chical modeling in ecology (pp. 79-122). Boston, MA: Academic Press.

Khosravi K, Nohani E, Maroufinia E, Pourghasemi HR (2016a) A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Nat. Hazards 83 (2), 947–987.

Khosravi K, Pham BT, Chapi K, Shirzadi A, Shahabi H, Revhaug I, Prakash I, Tien Bui DT (2018) A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at haraz watershed, Northern Iran. Sci. Total Environ. 627, 744–755

Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, Dou J, Ly HB, Gróf G, Ho HL, Hong H, Chapi K, Prakash I (2019a) A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. Journal of Hydrology, 573, 311-323. https://doi.org/10.1016/j.jhydrol.2019.03.073

Khosravi K, Melesse AM, Shahabi H, Shirzadi A, Chapi K, Hong H (2019b) Flood susceptibility mapping at ningdu catchment, china using bivariate and data mining techniques. In Extreme Hydrology and Climate Variability; Elsevier: Amsterdam, The Netherlands, 2019; pp. 419–434.

Kia MB, Pirasteh S, Pradhan B, Mahmud AR, Sulaiman WNA, Moradi A (2012) An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environ Earth Sci 67:251–264. doi: 10.1007/s12665-011-1504-z

Kisi O, Nia AM, Gosheh MG, Tajabadi MRJ, Ahmadi A (2012) Intermittent streamflow forecasting by using several data driven techniques. Water Resour Manag 26(2):457–474

Kjeldsen TR (2010) Modelling the impact of urbanization on flood frequency relationships in the UK. Hydrol Res 41:391–405. doi:10.2166/nh.2010.056

Komolafe AA, Herath S, Avtar R (2018) Methodology to assess potential flood damages in urban areas under the influence of climate change. Nat. Hazards Rev. 19 (2), 05018001.

Kourgialas NN, Karatzas GP (2011) Flood management and a GIS modelling method to assess flood-hazard areas-a case study. Hydrol Sci J 56:212–225. doi:10.1080/02626667.2011.555836

Laity JE (2008) Deserts and desert environments. Oxford, UK: Wiley-Blackwell; p. 360.

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

Lee S, Pradhan B (2007) Flood hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Floods 4:33–41. doi: 10.1007/s10346-006-0047-y

Lee MJ, Kang JE, Jeon S (2012) Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS. In: Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International. Munich 895–898

Liao X, Carin L (2009) Migratory logistic regression for learning concept drift between two data sets with application to UXO sensing. IEEE Trans Geosci Remote Sens 47:1454–1466

Liu J, Xu Z, Chen F, Chen F, Zhang L (2019) Flood Hazard Mapping and Assessment on the Angkor World Heritage Site, Cambodia. Remote Sens. 2019, 11, 98. https://doi.org/10.3390/rs11010098

Luu C, von Meding J (2018) A Flood Risk Assessment of Quang Nam, Vietnam Using Spatial Multicriteria Decision Analysis. Water, 10, 461. https://doi.org/10.3390/w10040461

Lombardo F, Obach RS, DiCapua FM, Bakken GA, Lu J, Potter DM, Zhang Y (2006) A hybrid mixture discriminant analysis–random forest computational model for the prediction of volume of distribution of drugs in human. J. Med. Chem. 49 (7), 2262–2267.

Mandal SP, Chakrabarty A (2016) Flash flood risk assessment for upper Teesta river basin: using the hydrological modeling system (HEC-HMS) software. Model Earth Syst Environ 2:59

Marmion M, Hjort J, Thuiller W, Luoto M (2008) A comparison of predictive methods in modelling the distribution of periglacial landforms in Finnish Lapland. Earth Surf. Process. Landforms 33: 2241-2254. DOI: 10.1002/esp

Matori A (2012) Detecting flood susceptible areas using GIS-based analytic hierarchy process. International Conference on Future Environment and Energy, Singapore (pp.).

McCullagh P, Nelder JA (1989) Generalized Linear Models. Chapman and Hall: New York.

Menard S (2001) Applied Logistic Regression Analysis, 2nd ed.; Sage Publication: Thousand Oaks, CA, USA, 2001; pp. 1–101. ISBN 0-7619-2208-3.

Meraj G, Romshoo SA, Yousuf AR, Altaf S, Altaf F (2015) Assessing the influence of watershed characteristics on the flood vulnerability of Jhelum basin in Kashmir Himalaya. Nat. Hazards 77 (1), 153–175.

Milewski A, Sultan M, Yan E, Becker R, Abdeldayem A, Soliman F, Abdel Gelil K (2009) A remote sensing solution for estimating runoff and recharge in arid environments. J Hydrology. 373:1–14

Mind'je R, Li L, Amanambu AC, Nahayo L, Nsengiyumva BJ, Gasirabo A, Mindje M (2019) Flood susceptibility modeling and hazard perception in Rwanda. International Journal of Disaster Risk Reduction, 38, 101211. https://doi.org/10.1016/j.ijdrr.2019.101211.

Moawad BM (2012) Predicting and analyzing flash floods of ungauged small-scale drainage basins in the Eastern Desert of Egypt. J Geomatics, Indian Soc Geomatics. 6:23–30

Moawad BM (2013) Analysis of the flash flood occurred on 18 January 2010 in wadi El Arish, Egypt (a case study). Geomatics, Nat Hazards Risk. 4(3):254–274.

Moawad BM, Abdel Aziz AO, Mamtimin B (2016) Flash floods in the Sahara: a case study for the 28 January 2013 flood in Qena, Egypt, Geomatics, Natural Hazards and Risk, 7:1, 215-236, DOI: 10.1080/19475705.2014.885467

Mahmoud SH, Gan TY (2018) Multi-criteria approach to develop flood susceptibility maps in arid regions of Middle East. J. Clean. Prod., 196, 216–229. https://doi.org/10.1016/j.jclepro.2018.06.047

Mosavi A, Ozturk P, Chau KW (2018) Flood Prediction Using Machine Learning Models: Literature Review. Water, 10, 1536.

Mukerji A, Chatterjee C, Raghuwanshi NS (2009) Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. J Hydrol Eng 14:647–652. doi:10.1061/(ASCE)HE.1943-5584.0000040

Muñoz P, Orellana-Alvear J, Willems P, Célleri R (2018) Flash-Flood Forecasting in an Andean Mountain Catchment—Development of a Step-Wise Methodology Based on the Random Forest Algorithm. Water, 10, 1519

Nfrag (2008) Flood risk management in Australia. Aust J Emerg Manag. 23(4):21–27

Naimi B, Araújo MB (2016) sdm: A reproducible and extensible R platform for species distribution modelling. Ecography, 39(4), 368-375.

Nicu IC (2018) Application of analytic hierarchy process, frequency ratio, and statistical index to landslide susceptibility: An approach to endangered cultural heritage. Environmental Earth Sciences, 77(3), 79. https://doi.org/10.1007/s12665-018-7261-5.

Opperman JJ, Galloway GE, Fargione J, Mount JF, Richter BD, Secchi S (2009) Sustainable floodplains through large-scale reconnection to rivers. Science 326 (5959), 1487–1488.

Papadopoulou-Vrynioti K, Bathrellos GD, Skilodimou HD, Kaviris G, Makropoulos K (2013) Karst collapse susceptibility mapping considering peak ground acceleration in a rapidly growing urban area. Eng Geol 158:77–88. doi:10.1016/j.enggeo.2013.02.009

Park S, Kim J (2019) Landslide ~~Susceptibility~~ susceptibility ~~Mapping~~ mapping ~~Based~~ based on ~~Random~~ random ~~Forest~~ forest and ~~Boosted~~ boosted ~~Regression~~ regression ~~Tree~~ tree ~~Models~~models, and a ~~Comparison~~ comparison of ~~Their~~ their ~~Performance~~performance. Appl. Sci., 9, 942.

Park S, Hamm S-Y, Kim J. (2019) Performance evaluation of the GIS-based data-mining techniques decision tree, random forest, and rotation forest for landslide susceptibility modeling. Sustainability 2019, 11, 5659.

Pourghasemi HR, Rahmati A (2018) Rapid GIS-based spatial and regional modelling of landslide susceptibility using machine learning techniques in the R open source software. Catena, 162: 177–192, https://doi.org/10.1016/j.catena.2017.11.022

Pourghasemi HR, Yousefi S, Kornejady A, Cerdà A (2017) Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. Sci. Total Environ. 2017, 609, 764–775.

Pradhan B (2010a) Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. J Spat Hydrol 9(2):1–18

Pradhan B (2010b) Flood susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. J Indian Soc Remote 38:301–320. doi: 10.1007/s12524-010-0020-z

Pradhan B, Buchroithner MF (2010) Comparison and Validation of Landslide Susceptibility Maps Using an Artificial Neural Network Model for Three Test Areas in Malaysia. Environ Eng Geosci 16(2):107–126

Pradhan B, Youssef AM, Varathrajoo R (2010) Approaches for delineating flood hazard areas using different training sites in an advanced neural network model. Geo-spatial Information Science 13:93–102, doi:10.1007/s11806-010-0236-2

Pradhan B, Hagemann U, Tehrany MS, Prechtel N (2014) An easy to use ArcMap based texture analysis program for extraction of flooded areas from TerraSAR-X satellite image. Comput Geosci 63:34–43. doi:10.1016/j.cageo.2013.10.011

Predick KI, Turner MG (2008) Landscape configuration and flood frequency influence invasive shrubs in floodplain forests of the wisconsin river (USA). J. Ecol., 96, 91–102.

Rahmati O, Pourghasemi HR (2017) Identification of Critical Flood Prone Areas in Data-Scarce and Ungauged Regions: A Comparison of Three Data Mining Models. Water Resour. Manag, 31, 1473–1487.

Rahmati O, Pourghasemi HR, Zeinivand H (2016) Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. Geocarto International. https://doi.org/10.1080/10106 049.2015.10415 59

Rahmati O, Naghibi SA, Shahabi H, Tien Bui D, Pradhan B, Azareh A, Melesse AM (2018) Groundwater spring potential modelling: Comprising the capability and robustness of three different modeling approaches. J. Hydrol., 565, 248–261.

Rahmati O, Yousefi S, Kalantari Z, Uuemaa E, Teimurian T, Keesstra S, Pham TD, Tien Bui D (2019) Multi-Hazard Exposure Mapping Using Machine Learning Techniques: A Case Study from Iran. Remote Sens, 11, 1943. https://doi.org/10.3390/rs11161943

Ramsay JO, Dalzell CJ (1991) Some tools for functional data analysis. Journal of the Royal Statistical Society. Series B (Methodological), 53(3), 539-572.

Reid I, Powell DM, Laronne JB, Garcia C (1994) Flash floods in desert rivers: studying the unexpected. Eos, Trans Am Geophysical Union. 75–39:452.

Regmi AD, Yoshida K, Dhital MR, Pradhan B (2013) Weathering and mineralogical variation in gneissic rocks and their effect in Sangrumba Flood, East Nepal. Environ Earth Sci 71(6):2711-2727. doi: 10.1007/s12665-013-2649-8

Regmi AD, Devkota KC, Yoshida K, Pradhan B, Pourghasemi HR, Kumamoto T, Akgun A (2014) Application of frequency ratio, statistical index, and weights-ofevidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. Arab. J. Geosci. 7 (2), 725–742.

Rozos D, Bathrellos GD, Skillodimou HD (2011) Comparison of the implementation of rock engineering system and analytic hierarchy process methods, upon flood susceptibility mapping, using GIS: a case study from the Eastern Achaia County of Peloponnesus, Greece. Environ Earth Sci 63:49–63. doi: 10.1007/s12665-010-0687-z

Saha S (2017) Groundwater potential mapping using analytical hierarchical process: A study on Md. Bazar Block of Birbhum District, West Bengal. Spat. Inf. Res., 25, 615–626.

Saidi ME, Daoudi L, Aresmouk ME, Fniguire F, Boukrim S (2010) The Ourika floods (High Atlas, Morocco), Extreme events in semi-arid mountain context. Comunicações Geológicas 97: 113-128.

Samanta RK, Bhunia GS, Shit PK, Pourghasemi HR (2018a) Flood susceptibility mapping using geospatial frequency ratio technique: a case study of Subarnarekha River basin, India. Model. Earth Syst. Environ. 1–14.

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

Sar N, Chatterjee S, Adhikari MD (2015) Integrated remote sensing and GIS based spatial modelling through analytical hierarchy process (AHP) for water logging hazard, vulnerability and risk assessment in Keleghai river basin, India. Model Earth Syst Environ 1:31. https ://doi.org/10.1007/s4080 8-015-0039-9

Schapire RE (2003) The boosting approach to machine learning: An overview. In Nonlinear Estimation and Classification; Springer: New York, NY, USA, pp. 149–171.

Schumann GP, Vernieuwe H, De Baets B, Verhoest NEC (2014) ROC-based calibration of flood inundation models. Hydrol. Process. 28 (22), 5495–5502.

Sene K (2013) Flash floods: forecasting and warning. Dordrecht, Germany: Springer; p. 395

Seifi Majdar R, Ghassemian H (2017) Spectral-spatial classification of hyperspectral images using functional data analysis. Remote Sensing Letters, 8(5), 488-497. https://doi.org/10.1080/2150704X.2017.1287973.

Sezer EA, Pradhan B, Gokceoglu C (2011) Manifestation of an adaptive neuro-fuzzy model on landslide susceptibility mapping: Klang Valley Malaysia. Expert Syst Appl 38(7):8208–8219.

Shahabi H, Shirzadi A, Ghaderi K, Omidvar E, Al-Ansari N, Clague JJ, Geertsema M, Khosravi K, Amini A, Bahrami S, Rahmati O, Habibi K, Mohammadi A, Nguyen H, Melesse AM, Ahmad BB, Ahmad A (2020) Flood detection and susceptibility mapping using Sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on K-nearest neighbor classifier. Remote Sens. 2020, 12, 266

Soch J, Meyer AP. Haynes JD. Allefeld C (2017) How to improve parameter estimates in GLM-based fMRI data analysis: Cross-validated Bayesian model averaging. NeuroImage, 158 (Supplement C), 186-195. https://doi.org/10.1016/j.neuroimage.2017.06.056.

Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis A (2008) Conditional variable importance for random forests. BMC Bioinformatics, 9(1) 307.

Taylor J, Davies M, Clifton D, Ridley I, Biddulph P (2011) Flood management: prediction of microbial contamination in largescale floods in urban environments. Environ Int 37:1019–1029. doi:10.1016/j.envint.2011.03.015

Tehrany MS, Pradhan B, Jebur MN (2013) Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. J Hydrol 504:69–79. doi:10.1016/j.jhydrol.2014.03.008

Tehrany MS, Pradhan B, Jebur MN (2014a) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J Hydrol 512:332–343. doi:10.1016/j.jhydrol.2014.03.008

Tehrany MS, Kumar L, Shabani F (2019) A novel GIS-based ensemble technique for flood susceptibility mapping using evidential belief function and support vector machine: Brisbane, Australia. PeerJ 7: e7653. http://doi.org/10.7717/peerj.7653

Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena 125:91–101. doi:10.1016/j.catena.2014.10.017

Tehrany MS, Shabani F, Neamah Jebur M, Hong H, Chen W, Xie X (2017) GIS-based spatial prediction of flood prone areas using standalone frequency ratio, logistic regression, weight of evidence and their ensemble techniques. Geomatics Nat. Hazards Risk 8 (2), 1538–1561.

Tien Bui D, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides. 13:361–378.

Tsakiris G (2014) Flood risk assessment: concepts, modelling, applications. Nat Hazards Earth Syst. Sci 14: 1361-1369.

Umar Z, Pradhan B, Ahmad A, Jebur MN, Tehrany MS (2014) Earthquake induced flood susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. Catena 118:124–135. doi:10.1016/j.catena.2014.02.005

Venables, WN, Dichmont, CM (2004) GLMs, GAMs and GLMMs: An overview of theory for applications in fisheries research. Fisheries Research, 70(2), 319-337. https://doi.org/10.1016/j.fishres.2004.08.011.

Wagner-Muns IM, Guardiola IG, Samaranayke, VA, Kayani WI (2018) A Functional data analysis approach to traffic volume forecasting. IEEE Transactions on Intelligent Transportation Systems, 19, (3): 878-888.

Wanders N, Karssenberg D, de Roo A, de Jong SM, Bierkens MFP (2014) The suitability of remotely sensed soil moisture for improving operational flood forecasting. Hydrol Earth Syst Sci 18:2343–2357

Wang H, Yang F, Luo Z (2016) An experimental study of the intrinsic stability of random forest variable importance measures. BMC Bioinformatics, 17 (1).

Ward RC, Robinson M (2000) Principles of Hydrology. 4th edn. McGraw-Hill, Maidenhead.

Warner TT (2004) Desert meteorology. Edinburgh: Cambridge university press; p. 612

Yalcin A, Reis S, Aydinoglu AC, Yomralioglu T (2011) A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. Catena 85 (3), 274–287.

Youssef AM, Hegab MA (2019) Flood-Hazard Assessment Modeling Using Multicriteria Analysis and GIS: A Case Study—Ras Gharib Area, Egypt. Spatial Modeling in GIS and R for Earth and Environmental Sciences. Pp. 229-257. https://doi.org/10.1016/B978-0-12-815226-3.00010-7

Youssef AM (2015) Landslide susceptibility delineation in the Ar-Rayth Area, Jizan, Kingdom of Saudi Arabia, by using analytical hierarchy process, frequency ratio, and logistic regression models. Environ earth sci. doi:10.1007/s12665-014-4008-9

Youssef AM, Pradhan B, Sefry SA (2016) Flash flood Susceptibility mapping in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. Environ. Earth Sci., 75, 12.

Youssef AM, Sefry SA, Pradhan B, Alfadail EA (2009) Analysis on causes of flash flood in Jeddah city (Kingdom of Saudi Arabia) of 2009 and 2011 using multi-sensor remote sensing data and GIS. Geomatics. Nat. Hazards Risk 2016b, 7, 1018–1042.

Youssef AM, Al-kathery M, Pradhan B (2014a) Flood susceptibility mapping at Al-Hasher Area, Jizan (Saudi Arabia) using GIS-based frequency ratio and index of entropy models. Geosci J. doi:10.1007/s12303-014-0032-8

Youssef AM, Pradhan B, Jebur MN, El-Harbi HM (2014b) Flood susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area, Saudi Arabia. Environ Earth Sci. doi:10.1007/s12665-014-3661-3

Zhao G, Pang B, Xu Z, Yue J, Tu T (2018) Mapping flood susceptibility in mountainous areas on a national scale in China. Sci. Total Environ., 615, 1133–1142.

Zou Q, Zhou J, Zhou C, Song L, Guo J (2013) Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. Stoch Env Res Risk Assess 27(2):525–546

Zwenzner H, Voigt S (2009) Improved estimation of flood parameters by combining space based SAR data with very high resolution digital elevation data. Hydrol Earth Syst Sci 13:567–576. doi:10.5194/hess-13-567-2009