Advanced machine learning algorithms for flood susceptibility modeling - comparison of their performance: Safaga-Ras Gharib area, Red Sea, Egypt

Ahmed M. Youssef

Saudi Geological Survey

Hamid Reza Pourghasemi (✉ hr.pourghasemi@shirazu.ac.ir)

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

Bosy A. El-Haddad

Sohag University

Research Article

Keywords: Flood susceptibility, Machine learning algorithms, Variables importance, Egypt

Posted Date: February 2nd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-893301/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Floods are among the most devastating environmental hazards that directly and indirectly affect people's lives and activities. In many countries, sustainable environmental management requires identifying inundated-prone areas to avoid potential hazards. In this study, the performance and capabilities of seven machine learning algorithms (MLAs) were tested, evaluated, and compared for flood susceptibility mapping. These MLAs include support vector machine (SVM), random forest (RF), multivariate adaptive regression spline (MARS), boosted regression tree (BRT), functional data analysis (FDA), general linear model (GLM), and multivariate discriminant analysis (MDA). In this study, machine learning algorithms (using R open source software) and spatial datasets (extracted using GIS) for flood susceptibility modeling were conducted between the cities of Safaga and Ras Gharib, Red Sea, Egypt. Initially, 420 actual inundated areas were collected from the study area to create a flood inventory map. Different sources were utilized, including remote sensing data (high and medium resolution), geologic and topographic maps, previous documents, and field investigations. The inventory data were divided randomly into two training groups: 70% and validation group: 30%. Ten flood-related factors were generated for flood susceptibility modeling: altitude, slope aspect, lithology, land use/land cover (LULC), slope length (LS), topographic wetness index (TWI), slope angle, profile curvature, plan curvature, stream power index (SPI), and hydrolithology units. Flood-related factors were tested using a multicollinearity test, variance inflation factors (VIF), and tolerance (TOL), and their importance was evaluated using a partial least squares (PLS) technique. In addition, model performances and comparisons were calculated using the area under the curve (AUC-ROC) approach. Our findings showed that the AUC values for these seven MLTs were divided into two groups: high fitness models (AUC > 80%), including RF, MARS, and GLM, and moderate performance models (AUC between 70% and 80%), including BRT, FDA, and SVM. The results of this study and the flood susceptibility maps could be useful for mitigating environmental problems, future development activities in the area, and flood control areas.

Introduction

In recent times, many urban areas, infrastructure and lifelines (bridges, railways, highways, power, and gas lines), and agricultural lands in all countries have been affected by flood hazards, which are considered the most common catastrophic natural hazards (Taylor et al. 2011; Dawod et al. 2012; Du et al. 2013; Vojtek and Vojteková 2019; Mishra and Sinha 2020; Sarkar and Mondal 2020). The rapid increase in population is forcing people to settle the low-lying areas that intersect or are close to the wadis/rivers. These areas are becoming more vulnerable to flooding due to current and future predicted climate changes and extreme meteorological events (Kjeldsen 2010; Karmaoui et al. 2014; Bubeck and Thielen 2018; Alexander et al. 2019; Ali et al. 2019; Kumar et al. 2019; Huang et al. 2019; Xu et al. 2019). Floods are often the most dangerous natural hazards (more dangerous than landslides, earthquakes, and volcanoes), causing enormous loss of life and injury, massive economic disruption, and contamination of areas with disease (Bolt et al. 2013; Ceola et al. 2014; Dandapat and Panda 2017). According to the statistical data from the United Nations Office for Disaster Risk Reduction (UNISDR), between 1995 and 2015, approximately 150,061 flood events occurred globally, and about 157,000 fatalities due to floods, accounting for about 11% of global disaster victims (Wahlstrom and Guha-Sapir 2015). In 2019, EM-DAT recorded 396 natural disasters

with 11,755 fatalities, 95 million people affected, and \$103 billion in economic losses around the world. The impact was not evenly distributed, as Asia was the hardest hit. Floods were the deadliest disaster type, accounting for 43.5% of fatalities (CRED 2020). Several studies have reported that floods affect approximately 200 million people annually and cause economic losses of US\$95 billion worldwide (Ceola et al. 2014; Mabuku et al. 2018).

Most flash floods occur with severe intensity, in a short period of time, and with high flow velocity occurring suddenly and with little or no time to react, posing a great danger to human and property (Sene 2013). Precipitation in arid areas is limited by cloud size (Laity 2008). In the last two decades, arid regions have faced many extreme rainfall events that caused immense devastation and loss of life, including Morocco 2008 2014; Algeria 2007, 2008 and 2013; and Saudi Arabia 2009, 2001, 2015, 2017 and 2018) (Kenyon 2007; Yamani et al. 2016; Echogdali et al. 2018; Abu-Abdallah et al. 2020). Egypt has been invariably affected by increasingly frequent flood events in recent decades. Accordingly, flood disasters have greatly increased, mainly due to increasing urbanization (residential areas and buildings) and infrastructure (highways, railways, and roads) in flood-prone areas (e.g., floods in 1994, 2010, 2014 and 2016) (Moawad 2013; Moawad et al. 2016; Youssef and Hegab 2019; El-Haddad et al. 2021). Therefore, it is crucial to manage floods and minimize or avoid their risk, which requires flood forecasting programs and inundation modeling. Bubeck et al. (2012) pointed out that flood forecasting minimizes flood-related hazards (fatalities and associated economic losses). The core concept of the flood control strategy is the ability to delineate flood-prone areas (Sarhadi et al. 2012).

Accordingly, flood vulnerability mapping is an extremely important step for predicting flood probability and mitigating and controlling future floods (Kourgialas and Karatzas 2011). Recently, various techniques and models have been applied to delineate flood-prone areas using remote sensing data and GIS (Wanders et al. 2014; Mandal and Chakrabarty 2016; Luu et al. 2018; Mahmoud and Gan 2018; Siahkamari et al. 2018; Dano et al. 2019; Kanani-Sadati et al. 2019; Liu et al. 2019; Mahmood and Rahman 2019; Wang et al. 2019; Sahana et al. 2020). These studies are mainly based on flood-related factors that have a significant impact on flood hazard assessment (e.g., lithology, slope, aspect, curvature, elevation, distance from streams, drain type, slope length (LS), topographic wetness index (TWI), and land use/land cover pattern).

Several authors have applied various techniques and approaches to assess the flood susceptibility of a region, such as (1) heuristic (multi-criteria analysis), (2) inundation methods, (3) statistical techniques, and (4) machine learning techniques. Each approach has its own advantages and disadvantages. For example, heuristic models are highly subjective and largely depend on human perception, judgment, and experience to determine the weighting of each flood-related factor (Dandapat and Panda 2017; Souissi et al. 2019; Vojtek and Vojteková 2019; Youssef and Hegab 2019; Nachappa et al. 2020).

The concept of flood inundation depends mainly on the hydrological characteristics of a given watershed to estimate the peak discharge during a given return period. The high-resolution digital elevation model of urban regions is used to apply the inundation model and produce inundation maps with water depth and flood velocity using hydrological models such as the Soil Water Assessment Tool (SWAT) and Hydraulic

Engineering Center-River Analysis System (HEC-RAS) (Oeurng et al. 2011; Getahun and Gebre 2015; Pal and Pani 2016; Prasad and Pani 2017).

Bivariate and multivariate models have recently been used to overcome human judgment and enhance the accuracy of flood vulnerability by using various computational methods (e.g., Weights-of-Evidence, Frequency Ratio (FR), Information Value (IV), Shannon entropy (SE), statistical index (SI), weighting factor, logistic regression (LR), fuzzy logic (FL), and neuro-fuzzy logic) (Liao and Carin 2009; Mukerji et al. 2009; Kourgialas and Karatzas 2011; Kia et al. 2012; Xu, 2013; Feng et al. 2015; Samantal et al. 2018 a, b; Park et al. 2019; Paul et al. 2019; Sahana and Patel 2019; Ali et al. 2020; Sahana et al. 2020).

The most common machine learning techniques (MLTs) are artificial neural networks (ANNs), general linear models (GLMs), adaptive neuro-fuzzy interface systems (ANFIS), decision trees (DT), random forest (RF), support vector regression (SVR), boosted regression tree (BRT), general linear models (GLMs), and classification and regression trees (CART), which can identify and predict flood-prone areas (Feng et al. 2015; Albers et al. 2016; Gizaw and Gan 2016; Muñoz et al. 2018; Zhao et al. 2018; Park et al. 2019; Dodangeh et al. 2020; Nhu et al. 2020).

Flood susceptibility maps (FSMs) could play an extremely important role in establishing early warning systems, contingency plans, reducing and preventing future inundations, and implementing flood management policies and regulations (Mandal and Chakrabarty 2016; Tehrany et al. 2019). In this study, seven advanced MLTs were applied to generate flood susceptibility models using remote sensing and GIS approaches. These models include BRT, FDA, GLM, MARS, MDA, RF, and SVM. They were used for several characteristics, the flood modeling analysis using MLTs is novel to be applied in Egypt, is suitable for small and medium scale uses, has an objective statistical basis, can quantitatively analyze the contribution of factors to flood development, and relies mainly on RS data rather than detailed field work. The results of our investigation provide important scientific contributions. Moreover, these flood susceptibility maps may be suitable for disaster management analysis to identify and outline flood-prone areas so that decision makers and land use planners can select favorable locations for future urban development.

Study Area

The study area extends from Safaga to Ras Gharib, with an area of approximately 10,537 km². It is situated between latitudes 26°40'00" " and 28°20'00" N and longitudes 32°50'00" and 34°0'00" E (Fig. 1). The watershed is characterized by various physiographic features, including mountains, hills, main wadis, and streams. The elevation ranges from 0 m (Red Sea coast in the east) to 2,068 m (mountainous areas in the west) above the mean sea level. The slope angle varied between 0° and 72° (with an average of 8.2° and a standard deviation of 10.4°). Approximately 16.1% of the total area has a slope greater than 30°, 16.4% of the area has a slope between 15° and 30°, 61.7% of the area has a slope between 5° and 15°, and 5.8% of the area has a slope of less than 5°. Precipitation is typically infrequent and occurs in the form of heavy thunderstorms from November to April. Unfortunately, there are no records of precipitation, as there are few precipitation stations in the area.

The study area is characterized by various rock units including bedrock complex in the west (40.8% of the study area), sedimentary rocks, and alluvial soils (wadi deposits), which occupy approximately 40.8%, 15.1%, and 44.1% of the study area, respectively. The area under study has been largely developed. Future planning and development in the study area will be affected by flood hazards. In 2014 and 2016, the study area experienced numerous flood events caused by heavy and short-duration rainfall that caused devastating destruction (Fig. 2). The area is characterized by numerous main streams that cut through the area, making it a flood-prone area (e.g., Wadi Abu Naakhra, Bali, Aish, Milaha, Abu Had, and Gharib). The eastern part of the study area, the low-lying areas, are particularly vulnerable to flooding from the western part. The area, which is constantly subject to flood damage, undergoes cascading changes over time. This presents a constraint for the spatial flood assessment. If the site information is erroneous, it can cause significant problems in the spatial analysis. However, drainage structures and water supply systems may have an impact on flood vulnerability assessments. The change in land use in the eastern portion of the study area from desert to residential and infrastructure, and the lack of action plans or inadequate engineering solutions to prevent flood events.

Data and Methodology

This study demonstrates the use of different datasets that can be used in flood susceptibility mapping. Several critical steps of the methodology were followed in this study to ensure the reliability of the yield models. These critical steps are shown in Fig. 3 and are explained in the following sections.

Data used

Table 1 describes the various datasets that were collected and extracted for this study. Field surveys were conducted to collect various features and evidence related to the consequences of flood events that affected the study area. Questionnaires with residents of the area (local people and Badwins) and historical documents (from the Civil Defense Agency and the Department for Transport) were collected and used to understand previous flood events. Photographs were taken and maintained documenting various flood events that affected different parts of the study area. Remote sensing data were acquired for the study area, including Landsat 8, OLI sensor (Operational Land Imager) (acquired in 2019, 30-m spatial resolution) from the Earth Explorer website (<https://earthexplorer.usgs.gov>). The image mosaic (30-m resolution) was created by overlaying the bands (1–7) and then fused with the panchromatic band (15-m resolution) to generate the final image mosaic (15-m resolution). Additional high-resolution images were obtained using an astro digital 2.5-m resolution and Professional Google Earth. Remote sensing imagery was used to create land use/land cover, flood inventory, lithology, and hydrolithology unit layers. In addition, a 30-m resolution DEM was obtained from ALOS World 3D-30m. DEM was used to generate various datasets (for example, elevation, slope aspect, lithology, LULC, LS, TWI, slope angle, plan and profile curvatures, SPI, and hydrolithology units. Finally, a 1:100,000-scale geologic map was prepared and digitized to delineate different lithological units and hydrolithological units. The data of this study were stored in GIS in a digital database with a uniform projection (UTM zone 36 and WGS84 datum).

Table 1
Data utilized and applied in the current work.

Dataset No.	Dataset Source	Dataset year & Characteristics	Data Style	Resolution & scale	Generated Layers
1	Remote sensing data Earth explorer website (https://earthexplorer.usgs.gov)	Landsat-8 (OLI-11 bands) 2014, 2016, 2019 Astro digital (2014 & 2016) Google Earth (various years)	Grid Grid Grid	30, 15m 2.5m < 1m	- LULC layer, - Flooded areas after 2014 & 2016 events. - Inundating areas after flood events in 2016 - Verify the flood locations after the events 2014 & 2016. - Verify and update hydrolithology units.
2	Geologic map Topographic map	Quadrangle 1985 Sheets 1975	Polygon Lines	1: 100,000	- Lithology units - Soil-drain - Verifying wadis and streams.
3	Digital Elevation Model (ALOS World 3D-30m)	DEM	Grid	30m	- Altitude, Slope-aspect, Slope-angle, TWI, LS, SPI, Plan- and profile-curvature
4	Field investigation; Field questionnaires, Historical data, & Photographs	Information on the flooded and destructed areas by 2014 & 2016 flood events	Points/ Polygon	Field trips	- Inundated and damaged areas in 2014 & 2016 events. - Verify lithology and hydrolithology units maps

Flood inventory map

Based on historical data and previous flood events, flooded areas were extracted to construct an inventory map. The inventory map is an extremely crucial element in flood susceptibility modeling (Sarkar and Mondal 2020). Several authors have pointed out that areas that have been exposed to past flood events under the same conditions are most likely to be vulnerable to current flood events (Fotovatikhah et al. 2018). To prepare susceptibility maps, it is necessary to determine the relationship between the inventory map (existing problems) and various factors that are relevant to susceptibility (Petley 2008). Different types of data (e.g., historical records, field visits, and satellite imagery interpretation) were used to generate an inundated inventory layer (Fig. 2b). Previously flooded areas (in the form of points) were extracted by comparing the study area before and after the flood events (2014 and 2016) using visual inspection of 1) high-resolution imagery (Google Earth and astrodigital imagery) and 2) medium-resolution imagery (Landsat-8 OLI). Flooded site data were examined and identified during field investigations following the 2014 and 2016 flood events (Fig. 2). Additional inundation data in the form of coordinated locations were collected from the Civil Defense Agency and past news over the past three decades. To isolate the exact flooded areas using medium-and high-resolution remote sensing images, Landsat-8 (2014) imagery with a spatial resolution of 15 m and Astro digital (2016) imagery with a spatial resolution of 2.5 m were used in two time periods. Cloud-free images were acquired before and after the flood events in 2014 and 2016. Visual inspection of the true color images (bands 1, 2, and 3 in RGB) using ArcGIS 10.8 software was used to extract the flooded areas (Fig. 4). The inundated areas identified using satellite imagery were verified using field investigations and civil defense data. Finally, a point feature layer (420 flooded locations) of the inundated locations was created to produce the flood inventory layer (Fig. 1b). The data points were randomly partitioned using R statistical software to divide the data into training and validation datasets (Naimi and Araújo 2016). In the current study, a general trend from previous literature was applied in which the inventory dataset was divided into 70% (including 295 flood locations) for training and 30% (125 flood locations) validation datasets (Wang et al. 2019) (Fig. 1b).

Flood-related factors (FRFs)

The determination of key flood-related factors (FRFs) is essential for flood susceptibility modeling (Sanyal and Lu 2004), and they vary according to catchment characteristics (Waqas et al. 2021). Rainfall is considered the most influential factor in the occurrence of floods. Lawal et al. (2012) pointed out that there are several other flood-related factors that contribute significantly to flood hazards. Runoff along the catchment depends on the characteristics of the catchment (e.g., catchment area, topography, and LULC types) (Hölting and Coldewey 2019). In the current study, eleven flood-related factors (FRFs) were selected as thematic layers based on the sound information from different types of literature (the most commonly used factors in flood vulnerability assessment literature), data availability related to the current study area and field investigation (Al-Juaidi et al. 2018; Kanani-Sadata et al. 2019; Liu et al. 2019; Paul et al. 2019; Wang et al. 2019; Vojtek and Vojteková 2019). These FRFs include altitude, slope aspect, lithology, land use/land cover (LULC), slope length (LS), topographic wetness index (TWI), slope angle, profile curvature, plan curvature, stream power index (SPI), and hydrolithology units (Fig. 5). They were generated and stored in spatial database themes with a grid cell size of 30 × 30 m in an ArcGIS environment for data processing. A digital elevation model (DEM) of the study area with a spatial resolution of 30-m was obtained from ALOS World 3D-30m), from which eight layers were generated. Of these, five factors, slope aspect, slope angle,

altitude, plan curvature, and profile curvature were extracted using ArcGIS 10.8 software. The other three themes, including TWI, LS, and SPI, were generated using the SAGA software. Other factors such as lithology, land use/land cover (LULC), and hydrolithology unit maps were extracted using remote sensing images (Landsat 8 - OLI and Google Earth), geological maps, and field surveys. Different types of FRFs were used in the present study, such as nominal (lithology, slope aspect, land use/land cover, and hydrolithology unit layers) and ordinal (altitude, TWI, slope angle, LS, profile curvature, plan curvature, and SPI).

Altitude

According to several authors, altitude is influenced by various factors (e.g., lithologic unit, wind action, precipitation, and erosion) (Waqas et al. 2021). The occurrence of flooding is likely to be influenced by elevation, which is considered an influential factor in flooding. Low-elevation regions (flat areas) are more susceptible to flooding than higher-elevation areas because water flows from high altitudes to lower-elevation areas (Kia et al. 2012; Cao et al. 2016). The altitude layer was extracted from the DEM using ArcGIS and ranged from 0 m to 2,173 m (Fig. 5a).

Slope-aspect

The slope aspect is the direction of the maximum inclination of the Earth's surface. It affects the direction of runoff, which maintains the soil moisture (Chu et al. 2020). The slope aspect may indirectly affect flooding, as the inclined shaded regions are characterized by relatively high soil moisture, indicating high runoff (Islam et al. 2021). The slope-aspect theme was created from the DEM map of the ArcGIS platform. The slope aspect map was divided into nine categories (Fig. 5b).

Lithology

Because of the varying permeability of rocks and sediments in a watershed, lithological units play a crucial role in hydrological processes (variations in the quantity and rate of water flow and sediment production) (Ward and Robinson 2000). The drainage density depends on the type of material used. Çelik et al. (2012) and Srivastava et al. (2014) indicated that a low drainage density is associated with highly resistant rock or highly permeable subsoil material. Stefanidis and Stathis (2013) concluded that flood hazard zones are influenced by geological units, especially torrential formations characterized by erodibility and permeability. In the current study, lithological units were generated from lithological maps (1:100,000-scale). Four main geological units were identified: (1) wadi deposits (alluvium), (2) sandstone, (3) limestone, (4) evaporates, (5) shales, and (6) basement rocks (Fig. 5c).

Land use/land cover (LULC)

Land use/land cover type (LULC) plays a critical role in runoff velocity, interception, infiltration, and evaporative transport (Benito et al. 2010; Yalcin et al. 2011). Various LULC features can affect the infiltration and surface flow generation in a catchment (Rahmati et al. 2015). Tehrany et al. (2019) indicated that forested areas can infiltrate more water into the subsurface than other LU/LC types. Many studies have shown that LULC types have a significant impact on distinguishing flood-vulnerable areas (Karlsson et al. 2017; Komolafe et al. 2018). The LULC layer was generated based on 2018 Landsat-8 satellite imagery (OLI) and classified into five categories using supervised classification in ENVI 5.4 software: wetlands, bare rock, bare soil, built-up area, and sandy soil with trees (Fig. 5d).

Slope length (LS)

The slope length (LS) is one of the influential factors determining soil erosion, where soil erosion accelerates with increasing slope length owing to the effects of higher accumulation of surface runoff (Bera 2017). LS shows the combined impacts of gradient length and steepness and affects particle transport "soil loss" and upland (mountainous) hydrological processes (Park et al. 2019). In this study, LS was calculated from the DEM layer according to the slope gradient and specific basin area using SAGA software based on the universal soil loss equation (USLE) (Eq. 1) (Moore and Burch, 1986):

$$LS = \left(\frac{As}{22.13} \right)^{0.4} \left(\frac{\sin \beta}{0.0896} \right)^{1.3}$$

1

As (m²) is the specific area of the catchment, and β is the slope angle in degrees. In this study, the slope length (LS) ranged from 0 to 59.1 (Fig. 5e).

Topographic wetness index (TWI)

The TWI reflects the variation in the quantity of water gathered in a basin (wetness values) and is the relationship between the specific basin area and the gradient (Beven, 2011; Gokceoglu et al., 2005). TWI can be strongly correlated with locations within a catchment that have a high potential for flooding (Chen and Yu 2011; Manfreda et al. 2011; Abdel Hamid et al. 2020). Tehrani et al. (2019) pointed out that flat areas can absorb more water than steep terrain. Accordingly, areas near drainage networks and flat lands (flood-prone areas) have higher TWI values than those in areas with slopes (Meles et al. 2020; Zhang et al. 2020). The TWI index value was extracted based on Eq. (2) (Beven and Kirkby 1979):

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

2

where A is the cumulative basin area (m²), and β is the slope angle (in degrees) at a point. In this work, the TWI was created using SAGA-GIS software ranging from 1.5 to 22.8 (Fig. 5f).

Slope angle

The slope angle is a crucial physiographic element for flood behavior and occurrence (Mukerji et al. 2009; Meraj et al. 2015). High-gradient areas have less time for percolation, which leads to the acceleration of runoff velocity, resulting in the accumulation of immense runoff in the lower lying areas (around the river or in flat areas) and are more vulnerable to flooding (Steva et al. 2020). The slope-angle layer was generated from a DEM map using ArcGIS. The slope angle ranged from 0° to 72° (Fig. 5g).

Plan and profile curvatures

The curvature represents the slope shape and the terrain morphology. It is one of the key terrain elements used in several geomorphometric works (Rau et al. 2019; Torcivia and López 2020). Curvature is a major flood-controlling factor in flood vulnerability mapping (Ahmadlou et al. 2019). Cao et al. (2016) reported that

the curvature has a significant impact on surface flow and infiltration. Shahabi et al. (2020) stated that areas with zero curvature values are more prone to flooding than areas with positive or negative curvatures. The curvature can be represented by the plan and profile curvatures. The plan curvature is directly correlated with the convergence and dispersion of surface runoff (Nasiri Aghdam et al. 2016). At the same time, Xiao et al. (2019) indicated that profile curvature impacts material deposition on the slope (by controlling the deposition increasing or decreasing of these materials). In the following study, plan and profile curvatures were generated from the DEM layer using ArcGIS software. The values of the plane and profile curvatures ranged from - 0.0249 to 0.0233 and - 0.0193 to 0.0208, respectively (Fig. 4 (h) (i) and i).

Stream power index (SPI)

The SPI is a crucial hydrological factor that plays a vital role in assessing the spatial variation of flood-vulnerable areas (Deepak et al. 2020). SPI directly correlates with the erosive power of the catchment, soil water content status in a basin, and discharge relative to a specific area within the watershed (the power of flood water to flow downward) (Cao et al. 2016). High SPI values indicate high flood power, and lower values indicate that the terrain in the watershed has the potential for impound flow (Turoglu and Dolke 2011). The SPI of the catchment was calculated using Eq. (3) (Moore and Wilson 1992; Wu et al. 2020a).

$$SPI = A_s \tan \beta$$

3

A_s is the specific basin area, and β is the local slope angle (in degrees). The SPI values in the current study range from 0 to 3.24 (Fig. 5j).

Hydrolithology units

Some soil types have a decisive influence on rainfall runoff mechanisms. The higher the infiltration rate, the less likely is the occurrence of flooding (Fluegel 1995; Phillips et al. 2019; Xie et al. 2019). In this study, a hydrolithology unit map was created by integrating Landsat 8 satellite imagery (OLI), Google Earth imagery, geological data, and field investigations. According to the national soil classes and soil taxonomy, the hydrolithology unit map of the study area was classified into three categories: well-drained, semi-drained, and impervious (Fig. 5k).

Theoretical background of methods used

Problems related to natural hazards, such as floods, landslides, and ground subsidence, have been identified and solved using various machine learning techniques (MLTs) (Park et al. 2014; Shi et al. 2016; Zhou et al. 2018; Ghorbanzadeh et al. 2019; Kavzoglu et al. 2019; Sevgen et al. 2019; Eini et al. 2020). Despite the continued advantages of MLTs as a powerful method, human expertise still plays an essential role in hazard assessment (Marjanović et al. 2011). In the current study, seven MLTs were utilized to evaluate their effectiveness in flood susceptibility mapping. These include SVM, RF, MARS, BRT, FDA, GLM, and MDA, which are discussed in detail in the following sections.

SVM model

The key elements of the SVM model are the utilization of classification and regression, which relate to the learning control approach (Vapnik 1998, 2013; Christianini and Shawe-Taylor 2000). SVM is a supervised learning method that deals with binary classification models (Amiri et al. 2019). The results provided minimal clustering errors and determined the optimal response (Vapnik 1998). This provides a key advantage in effectively identifying and analyzing factors (Micheletti et al. 2014). SVM has been used to create flood-prone areas (Yang and Cervone 2019). Many authors have provided detailed studies on SVM techniques (e.g., Yao et al. 2008).

RF model

Random forest (RF) is an ensemble learning approach based on regression trees, where many classification trees are aggregated to quantify a classification (Calle and Urrea 2010; Micheletti et al. 2014; Thanh Noi and Kappas 2018; Hawryło et al. 2018). The RF model is a robust ML model owing to several advantages, including a large number of trees in the analysis, insensitivity to noise, unbiased estimation of generalization error, acceptance of most types of data, and determination of significant variables (Breiman 2001; Rodrigues and De la Riva 2014; Kim et al. 2018). RF can overcome outliers in predictors, automatically deal with omitted data, and increase diversity among classification trees (Breiman and Cutler 2004). The model RF was run in R software version R 3.5.3, using the package "randomForest" (Briman and Cutler 2015).

MDA model

MDA, a supervised classification algorithm, is a linear discriminant analysis (LDA) in which a cluster is suggested as part of the closest group (Fraley and Raftery 2002). The normal distribution of variables is utilized to calculate the distance to the nearest collection, assuming that the variability and correlation between variables is uniform (Lombardo et al. 2006). MDA applies multiple normal distributions in every class. MDA can be derived from linear combinations using Eq. (4) (Hair et al. 1998).

$$Y = W_1X_1 + W_2X_2 + W_nX_n$$

4

Y represents the discriminant value, W_i ($i = 1, 2, 3, \dots, n$) are discriminant weights, and X_i ($i = 1, 2, 3, \dots, n$) are independent variables. The MDA analysis was run in R software version R 3.5.3, using the package "mda" (Hastie et al. 2017).

MARS model

MARS is a powerful regression algorithm owing to its flexibility in predicting events (Adnan et al. 2019). MARS considers both linear and nonlinear relationships between independent and dependent factors and reflects these functions as coefficients used to calculate the effects of these factors separately (Gu and Wahba 1991; Busto Serrano et al. 2020). MARS has been used in various uses to evaluate relationships between different disciplines (e.g., geophysics, climatology, ecology, and geomorphology) (Deichmann et al. 2002; Hjort and Luoto 2013; Abdulelah et al. 2019). It also allows the determination of the relative importance of the independent variables in the predictions (Adnan et al. 2019). It is also used to split the data sets into multiple splines on an equivalent interval basis; each spline can be subdivided into

subclasses by generating knots (Friedman 1991). The predictor MARS can be determined using Eq. 5, according to Hastie et al. (2001):

$$f(x) = \beta_0 \sum_{j=1}^p \sum_{b=1}^B [\beta_{jb}(+) \text{Max}(0, x_j - H_{bj}) + \beta_{jb}(-) \text{Max}(0, H_{bj} - x_j)] \quad (5)$$

where x , $f(x)$, p , and B are the input, output, predictor variable, and basis function, respectively. $\text{Max}(0, x-H)$ and $\text{Max}(0, H-x)$ are BF and do not need to exist if their coefficients are 0. The H values are referred to as the nodes.

Three steps in applying the MARS algorithm: (1) applying a stepwise forward algorithm to select spline basis functions, (2) deleting BFs until the "best" set is found by applying a stepwise backward algorithm, and (3) providing the final MARS approximation with some degree of continuity by performing a smoothing method. Generalized cross-validation (GCV) criteria were applied to delete the BF in order of least contribution using Eq. 6 (Craven and Wahba 1979).

$$\text{GCV} = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - f(x_i)]^2}{[1 - \frac{C(B)}{N}]^2} \quad (6)$$

N stands for the number of data points, and $C(B)$ stands for a complexity penalty escalated with the number of BFs in the technique and is determined by Eq. 7:

$$C(B) = (B + 1) + dB$$

7

Here, d represents the penalty for every BF incorporated in the technique. This can also be considered a smoothing variable. The MARS technique was run in the software R 3.5.3, using the package "MARS" (Deichmann et al. 2002).

GLM model

The generalized linear model (GLM) is a linear regression model that can quantify and incorporate specific and temporal variables (McCullagh and Nelder 1989; Guisan et al. 2002; Ozdemir and Altural 2013). The use of GLM can increase the accuracy and quality of the results because it uses multiple regression to develop a clear relationship between the dependent and independent variables (Scott et al. 1991). Moreover, it can predict numerous events as it can identify the best regression model (Federici et al. 2007; Payne 2015). Several authors have applied GLM to different spatial models (Bolker et al. 2009; Dumbser et al. 2020). The relationship between the response variable and explanatory variables can be constructed using the GLM link function (Ahmedou et al. 2016; Kéry and Royle 2016; Soch et al. 2017). The predictions and variances of the response factors were estimated using Equations (8) and (9):

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

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

Y_i denotes the vector of response parameters, X_{ij} is the matrix of explanatory parameters, β_j is the vector of floating variables, 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 this work, it is assumed that Y is the response parameter representing the flooded area in a grid, and X_i is the i -th flood-related parameter. Thus, the occurrence probability of a flooding event Y is represented by Eq. (10): By logistic transformation, the link function $g(y_i)$ is represented by Eq. (11).

$$P = \frac{\exp(c_0 + c_1X_1 + c_2X_2 + \dots + c_iX_i)}{1 + \exp(c_0 + c_1X_1 + c_2X_2 + \dots + c_iX_i)}$$

$$g(y_i) = c_0 + \sum c_iX_i + \varepsilon_i$$

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

In the present study, R software was used to construct the GLM model. A simple Gaussian family is determined as a link function for normally distributed response data. The independent factors were included in the model separately, using a smoothing spline with only two degrees of freedom in a polynomial of degree 2 to avoid overfitting (Aertsen et al. 2009)

FDA model

Ramsay and Dalzell (1991) proposed the FDA model as a statistical method to analyze the effect factors. The crucial concept of FDA is to treat an observed object with functional properties as an integral, regardless of the order of the observed values (Battista et al. 2016; Wagner-Muns et al. 2018). It can discriminate unsupervised work, where each class is divided into subcategories with a unique value (Chamroukhi et al. 2012; Zou et al. 2019). The FDA is a nonparametric method that is widely used in problem classification (Lu 2007; Seifi Majdar and Ghassemian 2017). Ray et al. (2019) summarized that the FDA model is a combination of regression models that perform an unseen operation for each category in the modeling analysis when applying complex class models. The basic tasks in applying the FDA model include 1) implementing a functional data representation by selecting training and testing datasets, 2) using functional principal component analysis (FPCA) to extract functional data features, 3) using machine learning methods to classify data features, and 4) testing datasets to verify the validity of the classification method.

In this study, the FDA model was used to generate a flood vulnerability map using the species distribution modeling (SDM) package in R software (Naimi and Araújo 2016).

BRT model

Friedman (2001) proposed BRT, which uses an integration of statistical and machine-learning techniques. The advantages of the BRT model are: 1) the ability to improve the model performance by fitting and combining several models, 2) no data transformation or outlier removal is required, 3) sophisticated nonlinear relationships can be fitted, and 4) interaction influences between variables are automatically accounted for (Schapire 2003; Elith et al. 2008; Park and Kim 2019). The combined strength of the regression tree and boosting algorithms can improve model accuracy and minimize variance (Aertsen et al. 2010). Model accuracy is improved by boosting, a powerful learning method that iteratively fits new trees to the residual errors of the existing tree composition (Doepke et al. 2017). The BRT model was run in R software version R 3.5.3, using the package "brt" (Ridgeway and Southworth 2013).

The FRFs effectiveness and contribution

Multicollinearity analysis is a technique used to determine the effectiveness of independent variables in a model (Dormann et al. 2012). It is a statistical method in which independent parameters in a model are highly correlated using multiple regression techniques, and the parameters with high collinearity are deleted (Saha 2017). The multicollinearity technique uses two indicators, namely variance inflation factors (VIF) and tolerance (TOL) (Eqs. 12 and 13):

$$TOL = 1 - R_J^2$$

12

$$VIF = \left[\frac{1}{T} \right]$$

13

R_J^2 represents the regression coefficient of explanatory factor J for the rest of the descriptive factors. Previous studies have shown that a $TOL < 0.10$ and a $VIF > 5$, account for problems of multicollinearity (Hosmer and Lemeshow 1989; Menard 2001).

Evaluating the importance of independent factors is crucial for flood susceptibility analysis. It can be applied to determine the contribution of various flood-related factors and accurately determine their role in modeling production. Several methods have been applied to evaluate relationships between related factors and events. These techniques have been used and received much attention, including random forest (RF) and partial least squares (PLS) (Wang et al. 2016; Huang et al. 2018). PLS was used in this study. PLS is a strong multivariate regression technique that enables a broad spectrum of processes to be performed (Martens and Martens 2000). It has many advantages, such as: it allows a quick understanding of the essential sequence of variations in the data; it is suitable for the analysis of noisy data, collinear, and even incomplete parameters, and it helps to detect errors in the input data (Wold et al. 2001). PLS was used for

multivariate calibration of a dependent parameter against many independent parameters. Accordingly, it was suitable for the selected critical factors in the analysis. Details of PLS functions and applications have been explained in various studies (e.g. Hastie et al. 2001; Abdi 2010; Lowry and Gaskin 2014). In the present study, the contribution and importance of all FRFs to flooding occurrence were evaluated using partial least squares (PLS).

Modelling prediction and performance

Evaluating the predictive and performance accuracy of the susceptibility models used was critical. The cross-validation approach using receiver operating characteristic (ROC) and area under the curve (AUC) has been applied quantitatively and graphically by various authors (Akgun et al. 2012; Ozdemir and Altural 2013; Youssef and Hegab 2019). The cross-validation approach offers many advantages, including quantitative evaluation of model prediction, determination of a better prediction approach, ability to compare the predictive capabilities of different models, ability to distinguish the less and most vulnerable areas, identification of the influencing factors and their contribution to prediction, evaluation of the effectiveness of the input parameters to the models, and improvement of the quality of model prediction. The ROC method is a statistical indicator of model performance based on the rates of true and false positives (sensitivity and 1-specificity) (Chung and Fabbri 2003; Mathew et al. 2009). The acceptable susceptibility model must have an AUC value between 0.5 and 1. The effectiveness, accuracy, and reliability of the model were improved by a higher AUC value (equal to or close to 1.0). An AUC value of less than 0.5 is considered a random model (Marzban, 2004). Sajedi-Hosseini et al. (2018) stated that the overall performance of the model can be identified by categorizing the AUC values as follows: incompetent model (AUC from 0.5 to 0.6), model with poor performance (AUC from 0.6 to 0.7), model with moderate performance (AUC between 0.7 and 0.8), and model with high fitness and performance (AUC 0.8).

Results And Discussions

Multicollinearity test and variable importance

Multicollinearity analysis of the 11 flood-related factors utilized in this study is shown in Table 2. The tolerance ToL and VIF values indicated that flood-related variables selected in the current work were more than 0.1 (ToL = 0.468) and less than 5 (VIF = 2.135), respectively. Consequently, there is no multicollinearity among the selected FRFs so that they can contribute significantly to the model construction in this study.

Furthermore, a partial least squares (PLS) method was applied to evaluate the significance of influential flood-related factors. Figure 6 presents the PLS results, which show that slope angle, altitude, LS, and TWI are the most important factors, followed by LULC, SPI, slope aspect, hydrolithology units, lithology, and plan curvature, which are moderately important flood-related factors. Thus, profile curvature was less critical.

Table 2
Multicollinearity results of flood-related factors.

FRFs	Collinearity Statistics	
	Tol	VIF
Hydrolithology	.752	1.330
Slope angle	.492	2.033
Profile curvature	.745	1.342
Plan curvature	.798	1.253
LULC	.581	1.721
Lithology	.780	1.282
Altitude	.468	2.135
TWI	.864	1.157
Slope aspect	.936	1.069
SPI	.925	1.081
LS	.901	1.091

Flood susceptibility models (FSMs)

Due to exponential population growth, future sustainable development requires an accurate understanding of the spread of natural hazards in each area. Predictions made through the application of modeling and simulation techniques are critical to natural resources and sustainable development studies. This provides agencies, decision makers, planners, and engineers with practical, reliable, and accurate information about an area's level of vulnerability to natural hazards. Seven MLT models (SVM, RF, MDA, MARS, GLM, FDA, and BRT) were successfully used for comparative analysis. Using the training dataset, these MLT models were used to generate the flood susceptibility models using ArcGIS 10.8 software for the study area (Fig. 7 (a-g)). The natural break classifier (Jenks) was used in this study (Nicu 2018). These FSMs were then categorized into four groups: low, moderate, high, and very high susceptibility zones. The percentage of relative areas in each group was calculated for each model (Fig. 8). The results showed that the areas of low, moderate, high, and very high classes correspond to the flood susceptibility map (FSM): 38.0%, 17.3%, 10.8%, and 33.9% of the total area for SVM; 36.2%, 19.5%, 27.4%, and 16.9% of the total area for RF; 28.3%, 22.1%, 26.8%, and 22.8% of the total area for MDA; 40.9%, 21.2%, 22.2%, and 15.7% of the total area for MARS, 35.1%, 17.3%, 26.6%, and 21.1% of the total area for GLM; 27.3%, 22.3%, 26.9%, and 23.6% of the total area for FDA; and 35.1%, 4.7%, 10.4%, and 49.8% of the total area for BRT. The actual floodplains were extracted from the high-resolution imagery (Astro Digital and Google Earth imagery with spatial resolution of 2.5 m and 1 m, respectively) subsequent to the 2016 flood event to verify the accuracy and performance of the models

used. The comparison indicates that there is good agreement, and the flood-prone areas in these seven models mainly occupy the main wadis and low-lying areas. However, low-susceptibility zones are located in highland regions.

FSMs Validation

The predictive performance of the MLTs producing flood vulnerability maps was evaluated using the ROC–AUC method. This method plots the sensitivity (percentage of currently inundated pixels correctly predicted by the model) against the 1-specificity (percentage of predicted inundated pixels in the entire area) (Fig. 7). The extracted FSMs were validated using the prediction rate method. The validation datasets (30% of the total flood locations), which were not previously used in building the models, were run to test how well the model predicted flooding (see Fig. (9)). According to the AUC classification, Sajedi-Hosseini et al. (2018). As shown in Table 3, the AUC results of the current study indicated that the differences in model performance among MLTs were relatively moderate to high. Models with high performance were RF (AUC = 81.3%), GLM (AUC = 80.2%), and MARS (AUC = 80.1%). This was followed by the models with moderate performance, including BRT (AUC = 77.7%), MDA (AUC = 76.8%), FDA (AUC = 76.3%), and SVM (AUC = 73.3%) (Fig. 9).

Table 3
AUC results of MLTs models in the current study.

Models	Area under curve (AUC)	Std. Error ^a	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
SVM	0.733	0.034	0.667	0.800
RF	0.813	0.029	0.757	0.869
MDA	0.768	0.032	0.705	0.830
MARS	0.801	0.030	0.742	0.860
GLM	0.802	0.030	0.744	0.860
FDA	0.763	0.032	0.700	0.826
BRT	0.777	0.032	0.715	0.840

Discussion

Severe flood events, which are becoming more frequent in different areas as part of the effects of climate change, require significant efforts through the analysis of flood hazards, vulnerabilities, and risks. Therefore, flood management is the most important requirement for averting and reducing flood hazards. One such technique is flood vulnerability modeling, which is crucial for protecting people and developing viable and effective mitigation and management strategies worldwide (Sahana et al. 2020; Wang et al. 2020). The MLTs used in the present work, such as SVM, RF, MARS, BRT, FDA, GLM, and MDA, to generate flood-susceptibility maps provide remarkable results. One of these models is BRT, which is used in this study and provides a prediction rate of 77.7%, which is considered a reclosable value, and is used by other authors,

suggesting that BRT is one of the best approaches for accurately determining flood-vulnerable areas (Darabi 2020). Selecting 11 flood-related factors that were evaluated to determine the most flood-prone areas according to seven MLTs, the resulting map shows the statistical relationships between flood inventory data (real flooded areas) and flood-related factors. These eleven thematic layers were extracted and generated from various sources, including remote sensing data (high to medium resolution from 1 m to 30 m), digital elevation models (30 m resolution), geological and topographic maps, and various field surveys. The inventory map (actual flooded areas) was created through various methods such as field visits (survey of flooded areas with GPS), historical documents (civil defense and local population), and high-resolution satellite imagery (1 m, 2.5 m, and 15 m resolution). Potentially flood-prone areas were mapped using seven MLTs in the current study. The results showed a significant correlation between model results. The validation of the resulting models was based on validation datasets not used in the training phase and provided relevant results ranging from 73.3 to 81.3% (values above 70%). The Red Sea region (especially the section between Safaga-Ras Gharib), which includes many urban areas (Ras-Gharib, Hurghada, and Safaga), various resorts, and attractive tourist sites, which are crossed by many lifelines (roads, power lines, railways, and highways), is crossed by many wadis that affect the area. For example, one of these events struck the area on October 18, 2016, in the proximity of Ras-Gharib City. This resulted in extensive flooding problems, 22 fatalities, many injuries, destruction of approximately 5000 houses, damage to many vehicles, erosion effects, and destruction of lifelines (main roads and streets) (Youssef and Hegab 2019).

The current work is new and innovative and, as expected, provides relevant mapping results, especially for preliminary planning management. The comparison of flood susceptibility maps produced by different machine learning models in the current study with flood hazard maps produced using LULC and DEM as the main parameters showed some differences, possibly because of the different types of data used in each model (Asare-Kyei et al. 2015; Mousavi et al. 2019). Flood hazards generated using LULC and DEM data are based on the extraction of runoff coefficients from land use/land cover (detection from remote sensing images), slope gradient layer and hydrological soil types, and precipitation intensity. Peak flows can be integrated with elevation data to generate a flood hazard map with a high performance rate of 87.83% (Mousavi et al. 2019). Rahmati et al. (2019) pointed out that when using SVM, BRT, and GAM for modelling multi-hazard delling, BRT has the highest validation and shows high accuracy for flood hazards (AUC = 94.2%). In the current study, MLTs were used to identify flood-susceptible areas. The model prediction shows a minimum value of prediction performance of 73.3% for the SVM method and a maximum value of 81.3% for the RF approach. Even so, the prediction values are not high enough, and the machine learning techniques to produce appropriate flood-susceptibility maps are still high. Our results for the prediction accuracy of RF (81.3%) and BRT (77.7%) are in agreement with the findings of Lee et al. (2017). They indicated that the RF approach provides a higher performance than the boosted tree technique. In the current work, the prediction performance based on RF was 79.18%, whereas it was 77.26% for BRT. Various new and robust machine learning models have been applied by many researchers to achieve highly precise and accurate flood vulnerability mapping (Al-Abadi 2018; Paul et al. 2019; Wang et al. 2019; Ali et al. 2020; Nachappa et al. 2020; Wang et al. 2020; El-Haddad et al. 2021; Islam et al. 2021). However, the prediction accuracy of these models is still difficult because of the different characteristics of the studied areas. The main advantages of MLTs are enormous, including their relative ease of use, which for reducing workload

and time, and a variety of applications, it can handle different types of data. Their predictive accuracy usually outperforms some conventional methods, and they help us find ways to modernize technology. However, there are some limitations and side effects, such as the possibility of a high error rate due to the large amount of data, the inconsistency of data due to the large amount of data for training and testing, and selection of an algorithm is still a manual task, so the process is very time consuming. This helps us find various innovative ways to reduce these problems. Finally, these MLTs can automatically detect the relationships between flood-related factors and overcome uncertainties when reliable inventory maps (actual flooded areas) are acquired.

This work, using various MLTs, helped to understand the extent of flood hazard in the study area, which is crucial because many important facilities, critical highways, and tourist sites are located in this section of the area along the Red Sea and the Gulf of Suez. In addition, the MLTs provided deep insights into the importance of flood risk management. As millions of people visit these areas each year, they represent future income for the country, in addition to the residents of these areas. Therefore, a safe environment for this area is essential. This study has shown that many coastal areas are in worrisome areas. For this reason, we recommend that agencies, managers, and developers in the area study areas of high flood vulnerability and carefully consider the results of future development models to select the areas with the lowest risk. Areas with high or very high flood vulnerability should be studied in detail to determine flood protection measures.

Conclusions

In the context of climate change, floods are considered the most damaging natural disasters, causing socio-economic disruptions, loss of life, and property damage. Accordingly, low-lying areas surrounded by mountainous regions are at risk of flooding. These areas are prone to flooding. Effective and reliable techniques are required to delineate flood-prone areas. In this study, the spatial distribution of flood-prone areas in the coastal area of the Red Sea (between Safaga-Ras Gharib) in Egypt was investigated using seven machine learning models (SVM, RF, MARS, BRT, FDA, GLM, and MDA). The identification of flood-prone areas is crucial for most planners, the private sector, and decision makers. The analysis was mainly based on the identification of 420 flood points (295 points were used as training and 125 points for validation). Eleven flood-related factors were used, including elevation, slope, lithology, LULC, LS, TWI, slope, profile curvature, plan curvature, SPI, and soil runoff. Multicollinearity diagnostic tests (VIF and TOL) were used to test the suitability of flood-related factors. The partial least squares (PLS) approach was applied to identify the significance of flood-related factors. The ROC curve was constructed to check the flood-vulnerability models based on the validation datasets. Accordingly, the ability of seven MLTs to map most flood-prone areas is presented in this paper. The evaluation of the reliability and predictive performance of the FSMs produced by the SVM, RF, MARS, BRT, FDA, GLM, and MDA models showed that RF performed the best, followed by GLM and MARS, which produced more than 80%, indicating significantly better results. The BRT, MDA, and FDA algorithms provided moderately significant results. Finally, the SVM provides less significant results than the other models. The average results of all MLTs showed that 34.4% of the study area had low flood susceptibility, 17.8% had moderate flood susceptibility, 21.6% had high flood susceptibility, and 26.3% had very high flood susceptibility (extremely flood prone). Our results show that

MLTs provide prediction values greater than 0.7% (70%), indicating that the models are adequate for flood susceptibility mapping in the area under consideration. Furthermore, the results indicate that these techniques have moderate to high performance in analyzing flood susceptibility, with such small differences. Therefore, these results can be viewed with greater confidence and applied in future studies to investigate the flood hazard distribution and provide helpful knowledge for decision makers to be proactive in flood management, hazard mitigation measures, and land use regulations. Recently, developers, planners, local governments, and other agencies have acquired flood susceptibility modeling as an important step in identifying flood-prone areas that need to be studied in more detail to prevent future flooding.

Declarations

Acknowledgment

This work was supported by the College of Agriculture, Shiraz University (Grant No. 99GRC1M271143).

Ethical Approval

Not applicable.

Consent to Participate

Not applicable.

Consent to Publish

Not applicable.

Authors Contributions

AMY, HRP, and BAEI-H designed the experiments, ran models, analyzed the results, and wrote and reviewed the manuscript.

Availability of data and materials

Data will send based on request.

Conflicts of Interest: “The authors declare no conflicts of interest.”

Funding: This study was supported by the College of Agriculture, Shiraz University (Grant No. 99GRC1M271143).

References

1. Abdel Hamid HT, Wenlong W, Qiaomin L (2020) Environmental sensitivity of flash flood hazard using geospatial techniques Global. J Environ Sci Manag 6(1):31–46

2. Abdi H (2010) Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdiscip Rev Comput Stat* 2:97–106
3. Abdulelah Al-Sudani Z, Salih SQ, Sharafati A, Yaseen ZM (2019) Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. *J Hydrol* 573:1–12
4. Abu-Abdullah MM, Youssef AM, Maerz NH, Abu-AlFadail E, Al-Harbi HM, Al-Saadi NS (2020) A Flood Risk Management Program of Wadi Baysh Dam on the Downstream Area: An Integration of Hydrologic and Hydraulic Models, Jizan Region. *KSA Sustainability* 12:1069
5. Adnan RM, Liang Z, Heddarn S, Zounemat-Kermani M, Kisi O, Li B (2019) Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *J. Hydrol.* 124371 (2019).
<https://doi.org/10.1016/J.JHYDROL.2019.124371>
6. Aertsens 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, 1–12
7. Aertsens 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
8. Ahmadi M, Karimi M, Alizadeh S, Shirzadi A, Parvinnejhad D, Shahabi H, Panahi M (2019) Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA) *Geocarto Int.*, 34 (11): 1252–1272
9. 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>
10. Akgun A, Sezer EA, Nefeslioglu HA, Gokceoglu C, Pradhan B (2012) An easy-to-use MATLAB program (MamLand) for the assessment of flood susceptibility using a Mamdani fuzzy algorithm. *Comput Geosci* 38(1):23–34. doi:10.1016/j.cageo.2011.04.012
11. Albers SJ, Déry SJ, Petticrew EL (2016) Flooding in the Nechako River Basin of Canada: A random forest modeling approach to flood analysis in a regulated reservoir system. *Can Water Resour J* 41:250–260
12. Alexander K, Hettiarachchi S, Ou Y, Sharma A (2019) Can integrated green spaces and storage facilities absorb the increased risk of flooding due to climate change in developed urban environments? *J. Hydrol.*, 579 (2019), Article 124201
13. Ali R, Kuriqi A, Abubaker S, Kisi O (2019) Long-Term Trends and Seasonality Detection of the Observed Flow in Yangtze River Using Mann-Kendall and Sen's Innovative Trend Method. *Water* 11(9):1855
14. Ali SA, Parvin F, Pham QB, Wojtek M, Wojteková J, Costache R, Linh NTT, Nguyen HQ, Ahmad A, Ghorbani MA (2020) GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-

- criteria decision-making approach, naïve Bayes tree, bivariate statistics and logistic regression: A case of Topľa basin, Slovakia *Ecol. Indic.*, 117, Article 106620
15. Al-Abadi AM (2018) Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study. *Arab J Geosci* 11(9):218
 16. Al-Juaidi AM, Nassar AM, Al-Juaidi OEM (2018) Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. *Arab J Geosci* 11:765.
<https://doi.org/10.1007/s12517-018-4095-0>
 17. Asare-Kyei D, Forkuor G, Venus V (2015) Modeling flood hazard zones at the sub-district level with the rational model integrated with GIS and remote sensing approaches. *Water* 7:3531–3564
 18. Battista TD, Fortuna F, Maturo F (2016) BioFTF: An R package for biodiversity assessment with the functional data analysis approach. *Ecol Ind* 73:726–732
 19. 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
 20. Benito G, Rico M, Sánchez-Moya Y, Sopeña A, Thorndycraft VR, Barriendos M (2010) The impact of late Holocene climatic variability and land use change on the flood hydrology of the Guadalentín River, southeast Spain *Glob. Planet Chang* 70(1–4):53–63
 21. Beven K, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sci J* 24(1):43–69
 22. Beven KJ (2011) *Rainfall-runoff Modelling: the Primer*. John Wiley & Sons
 23. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MH, White JS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol.* 24(3):127 – 35. doi: 10.1016/j.tree.2008.10.008. PMID: 19185386
 24. Bolt BA, Horn W, MacDonald GA, Scott R (2013) *Geological Hazards: Earthquakes-tsunamis-volcanoes-avalanches-landslides-floods*. Springer Science & Business Media
 25. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
 26. Breiman L, Cutler A (2004) http://www.stat.berkeley.edu/users/Breiman/RandomForests/cc_papers.html
 27. Briman L, Cutler A (2015) Package 'randomForest', 29 (Date/Publication 2015-10-07)
 28. Bubeck P, Botzen W, Aerts J (2012) A review of risk perceptions and other factors that influence flood mitigation behavior. *Risk Anal* 32:1481–1495. doi:10.1111/j.1539-6924.2011.01783.x
 29. Bubeck P, Thielen AH (2018) What helps people recover from floods? Insights from a survey among flood-affected residents in Germany *Reg. Environ Chang* 18(1):287–296
 30. Busto Serrano N, Suárez Sánchez A, Sánchez Lasheras F, Iglesias-Rodríguez FJ, Fidalgo Valverde G (2020) Identification of gender differences in the factors influencing shoulders, neck and upper limb MSD by means of multivariate adaptive regression splines (MARS). *Appl Ergon* 82:102981

31. Calle ML, Urrea V (2010) Letter to the Editor: Stability of random forest importance measures. *Brief Bioinform* 12(1):86–89. doi:10.1093/bib/bbq011
32. Cao C, Xu P, Wang Y, Chen J, Zheng L, Niu C (2016) Flash flood hazard susceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas. *Sustainability* 8(9):948
33. Çelik HE, Coskun G, Cigizoglu HK, Ağralioğlu N, Aydın A, Esin A (2012) The analysis of 2004 flood on kozdere stream in Istanbul. *Nat hazards* 63(2):461–477
34. Ceola S, Laio F, Montanari A (2014) Satellite nighttime lights reveal increasing human exposure to floods worldwide. *Geophys Res Lett* 41(20):7184–7190
35. Chamroukhi F, Glotin H, Rabouy C (2012) Functional mixture discriminant analysis with hidden process regression for curve classification. *ESANN 2012 Proceedings, 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* 281–286
36. Craven P, Wahba G (1979) Smoothing noisy data with spline functions. *Numer Math* 31:377–390
37. CRED (2020) Natural Disasters 2019. Brussels. https://emdat.be/sites/default/files/adsr_2019.pdf
38. Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge
39. Chu H, Wu W, Wang QJ, Nathan R, Wei J (2020) An ANN-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions *Environ. Model. Softw.*, 124, Article 104587
40. Chung C-JF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. *Nat Hazards* 30(3):451–472. <https://doi.org/10.1023/B:NHAZ.00000007172.62651.2b>
41. Dano UL, Balogun AL, Matori AN, Wan Yusouf K, Abubakar IR, Said Mohamed MA, Aina YA, Pradhan B (2019) Flood Susceptibility Mapping Using GIS-Based Analytic Network Process: A Case Study of Perlis, Malaysia. *Water* 11(3):615. <https://doi.org/10.3390/w11030615>
42. 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
43. Darabi H, Haghighi AT, Mohamadi MA, Rashidpour M, Ziegler AD, Hekmatzadeh AA, Kløve B (2020) Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran. *Hydrol Res* 51(1):127–142. <https://doi.org/10.2166/nh.2019.090>
44. 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
45. Deepak S, Rajan G, Jairaj PG (2020) Geospatial approach for assessment of vulnerability to flood in local self governments. *Geoenviron Disasters* 7:35. <https://doi.org/10.1186/s40677-020-00172-w>
46. Deichmann J, Eshghi A, Haughton D, Sayek S, Teebagy N (2002) Application of multiple adaptive regression splines (mars) in direct response modeling. *J Interact Mark* 16:15–27
47. Dodangeh E, Choubin B, Eigdir AN, Nabipour N, Panahi M, Shamshirband S, Mosavi A (2020) Integrated machine learning methods with resampling algorithms for flood susceptibility prediction *Sci. Total Environ.*, 705, Article 135983

48. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2012) Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
49. Döpke J, Fritsche U, Pierdzioch C (2017) Predicting recessions with boosted regression trees. *Int J Forecast* 33:745–759
50. Du J, Fang J, Xu W, Shi P (2013) Analysis of dry/wet conditions using the standardized precipitation index and its potential usefulness for drought/flood monitoring in Hunan Province China. *Stoch Env Res Risk Assess* 27(2):377–387
51. Dumbser M, Fambri F, Gaburro E, Reinartz A (2020) On GLM curl cleaning for a first order reduction of the CCZ4 formulation of the Einstein field equations. *J Comput Phys* 404:109088
52. 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
53. Eini M, Kaboli HS, Rashidian M, Hedayat H (2020) Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts. *Int. J. Disaster Risk Reduct.* 101687
54. Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *J Anim Ecol* 77:802–813
55. El-Haddad BA, Youssef AM, Pourghasemi HR, Pardhan B, El-Shater A, El-Khashab MH (2021) Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin, Egypt. *Nat Hazards* 105:83–114. <https://doi.org/10.1007/s11069-020-04296-y>
56. Federici PR, Puccinelli A, Cantarelli E, Casarosa N, Avanzi GDA, Falaschi F, Giannecchini R, Pochini A, Ribolini A, Bottai M, Salvati N (2007) Multidisciplinary investigations in evaluating landslide susceptibility: an example in the Serchio River valley (Italy). *Quatern Int* 171–172:52–63
57. 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
58. Flügel WA (1995) Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany *Hydrol. Process* 9(3–4):423–436
59. Fotovatikhah F, Herrera M, Shamshirband S, Chau K-W, Ardabili S, Faizollahzadeh, 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
60. Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc* 97:611–631
61. Friedman JH (2001) Greedy Function Approximation: A Gradient Boosting Machine. *Ann Stat* 29:1189–1232

62. Getahun YS, Gebre SL (2015) Flood hazard assessment and mapping of flood inundation area of the Awash River Basin in Ethiopia using GIS and HEC-GeoRAS/HEC-RAS model J. Civil Environ Eng 5(4):1
63. Ghorbanzadeh O, Blaschke T, Gholamnia K, Meena SR, Tiede D, Aryal J (2019) Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sens 11:196
64. Gizaw MS, Gan TY (2016) Regional flood frequency analysis using support vector regression under historical and future climate. J Hydrol 538:387–398
65. Gokceoglu C, Sonmez H, Nefeslioglu HA, Duman TY, Can T (2005) The 17 March 2005 Kuzulu landslide (Sivas, Turkey) and landslide-susceptibility map of its near vicinity. Eng Geol 81:65–83
66. Gu C, Wahba G (1991) Discussion: Multivariate adaptive regression splines. Ann Stat 19:115–123
67. Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: Setting the scene. Ecol Model 157(2):89–100. [https://doi.org/10.1016/S0304-3800\(02\)00204-1](https://doi.org/10.1016/S0304-3800(02)00204-1)
68. Hair J, Anderson R, Tatham RL, Black WC (1998) Multivariate data analysis, 5th edn. Upper Saddle River, Prentice-Hall, NJ
69. Hastie T, Tibshirani R, Friedman J (2001) The Elements of Statistical Learning; Data Mining, Inference, and Prediction. Springer, New York
70. Hastie T, Tibshirani R, Leisch F, Hornik K, Ripley BD (2017) Mixture and flexible discriminant analysis. <https://cran.r-project.org/web/packages/mda/mda.pdf>
71. Hawryło P, Bednarsz B, Wężyk P, Szostak M (2018) Estimating defoliation of Scots pine stands using machine learning methods and vegetation indices of Sentinel-2. Eur. J. Remote Sens. 51
72. Hjort J, Luoto M (2013) Statistical methods for geomorphic distribution modeling. Treatise Geomorphol 2:59–73
73. Hosmer DW, Lemeshow S (1989) Applied Regression Analysis; John Wiley and Sons: New York, NY, USA, ISBN 978-0-470-58247-3
74. Hölting B, Coldewey WG (2019) Hydrogeology. Surface water infiltration. Springer, Berlin, pp 33–37
75. Huang X-d, Wang L, Han P-p, Wang W-c (2018) Spatial and Temporal Patterns in Nonstationary Flood Frequency across a Forest Watershed: Linkage with Rainfall and Land Use Types. Forests 9:339. <https://doi.org/10.3390/f9060339>
76. Huang K, Li X, Liu X, Seto KC (2019) Projecting global urban land expansion and heat island intensification through 2050 Environ. Res. Lett., 14 (11), Article 114037
77. Islam ARMT, Talukdar S, Mahato S, Kundu S, Kutub Eibek KU, Pham QB, Kuriqi A, Linh NTT (2021) Flood susceptibility modelling using advanced ensemble machine learning models. Geosci Front 12(3):101075
78. Kanani-Sadata Y, Arabsheibani R, Karimipour F, Nasserli 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. J Hydrol 572:17–31. <https://doi.org/10.1016/j.jhydrol.2019.02.034>

79. Karlsson CS, Kalantari Z, Mörtberg U, Olofsson B, Lyon SW (2017) Natural hazard susceptibility assessment for road planning using spatial multi-criteria analysis. *Environ Manag* 60(5):823–851
80. 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: 1–12
81. Kavzoglu T, Colkesen I, Sahin EK (2019) Machine learning techniques in landslide susceptibility mapping: A survey and a case study. In: *Landslides: Theory, Practice and Modelling*. Springer, Berlin, pp 283–301
82. Kenyon P (2007) Climate connections: Algeria vs. the Sahara, NPR's climate connections series with National Geographic. Available from: <http://www.npr.org/templates/story/story.php?storyId¼12903558>
83. Kéry M, Royle JA (2016) Linear models, generalized linear models (GLMs), and random effects models: The components of hierarchical models. In: Kéry M, Royle JA (eds) *Applied hierarchical modeling in ecology*. Academic Press, Boston, pp 79–122
84. 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
85. Kim JC, Lee S, Jung HS, Lee S (2018) Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto Int* 33:1000–1015
86. 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
87. 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
88. 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
89. Kumar R, Nayak PC, Patra JP, Mani P (2019) Chapter-10-Impact of climate change on floods. National Institute of Hydrology
90. Laity JE (2008) *Deserts and desert environments*, 360. Wiley-Blackwell, Oxford
91. Lawal DU, Matori AN, Hashim AM, Wan Yusof K, Chandio IA (2012) Detecting food susceptible areas using GIS-based analytic hierarchy process. In: *Proceedings of the 2012 international conference on future environment and energy IPCBEE*, Kuala Lumpur, Malaysia, 22–23 December 2012, vol 28. IACSIT Press, Singapore, 1–5
92. 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
93. 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*. 11: 98. <https://doi.org/10.3390/rs11010098>
94. Lowry PB, Gaskin J (2014) Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: when to choose it and how to use it. *IEEE Trans Prof Commun* 57:123–146

95. Lu ZQJ (2007) Nonparametric functional data analysis: theory and practice. *Technometrics*
<https://doi.org/10.1198/tech.2007.s483>
96. 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>
97. 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
98. Mabuku MP, Senzanje A, Mudhara M, Jewitt GPW, Mulwafu W (2018) Rural households' flood preparedness and social determinants in Mwandia district of Zambia and Eastern Zambezi Region of Namibia. *International Journal of Disaster Risk Reduction* 28:284–297.
<http://doi.org/10.1016/J.IJDRR.2018.03.014>
99. 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>
100. Mahmood S, Rahman AU (2019) Flash flood susceptibility modeling using geo-morphometric and hydrological approaches in Panjkora Basin, Eastern Hindu Kush, Pakistan *Environ. Earth Sci* 78(1):43
101. 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
102. Manfreda S, Di Leo M, Sole A (2011) Detection of flood-prone areas using digital elevation models. *J Hydrologic Eng* 16(10):781–790
103. Marjanović M, Kovačević M, Bajat B, Voženílek V (2011) Landslide susceptibility assessment using SVM machine learning algorithm. *Eng Geol* 123(3):225–234
104. Martens H, Martens M (2000) Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual Prefer* 11:5–16. 27
105. Marzban C (2004) The ROC Curve and the Area under It as Performance Measures. *Weather and Forecasting*, 1106 – 1114. DOI: <https://doi.org/10.1175/825.1>
106. Mathew J, Jha VK, Rawat GS (2009) Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. *Landslides* 6:17–26. <https://doi.org/10.1007/s10346-008-0138-z>
107. McCullagh P, Nelder JA (1989) *Generalized Linear Models*. Chapman and Hall, New York
108. Meles MB, Younger SE, Jackson CR, Du E, Drover D (2020) Wetness index based on landscape position and topography (WILT): Modifying TWI to reflect landscape position *J. Environ. Manag.*, 255, Article 109863
109. Menard S (2001) *Applied Logistic Regression Analysis*, 2nd ed.; Sage Publication: Thousand Oaks, CA, USA, 1–101. ISBN 0-7619-2208-3
110. 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

111. Micheletti N, Foresti L, Robert S, Leuenberger M, Pedrazzini A, Jaboyedoff M, Kanevski M (2014) Machine learning feature selection methods for landslide susceptibility mapping. *Math Geosci* 46: 33–57. doi:10.1007/s11004-013-9511-0
112. Mishra K, Sinha R (2020) Flood risk assessment in the Kosi megafan using multi-criteria decision analysis: A hydro-geomorphic approach. *Geomorphology*, 350, Article 106861
113. 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
114. 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 Risk* 7(1):215–236. DOI:10.1080/19475705.2014.885467
115. Moore ID, Wilson JP (1992) Length-slope factors for the revised universal soil loss equation: simplified method of estimation. *J soil water conservation* 47(5):423–428
116. 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
117. Muñoz P, Orellana-Alvear J, Willems P, Céleri 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
118. Nachappa TG, Piralilou ST, Gholamnia K, Ghorbanzadeh O, Rahmati O, Blaschke T (2020) Flood Susceptibility Mapping with Machine Learning, Multi-Criteria Decision Analysis and Ensemble Using Dempster Shafer Theory *J. Hydrol.*, 590, Article 125275
119. Naimi B, Araújo MB (2016) sdm: A reproducible and extensible R platform for species distribution modelling. *Ecography* 39(4):368–375
120. Nasiri Aghdam I, Varzandeh MHM, Pradhan B (2016) Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz Mountains (Iran) *Environ. Earth Sci* 75(7):1–20
121. Nfrag (2008) Flood risk management in Australia. *Aust J Emerg Manag* 23(4):21–27
122. Nicu IC (2018) Application of analytic hierarchy process, frequency ratio, and statistical index to landslide susceptibility: An approach to endangered cultural heritage. *Environ Earth Sci* 77(3):79. <https://doi.org/10.1007/s12665-018-7261-5>
123. Oeurng C, Sauvage S, Sánchez-Pérez JM (2011) Assessment of hydrology, sediment and particulate organic carbon yield in a large agricultural catchment using the SWAT. model *J Hydrol* 401(3–4):145–153
124. Ozdemir A, Altural T (2013) A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan mountains, SW Turkey. *J Asian Earth Sci* 64:180–197
125. Pal R, Pani P (2016) Seasonality, barrage (Farakka) regulated hydrology and food scenarios of the Ganga River: a study based on MNDWI and simple Gumbel model. *Modeling Earth Syst Environ* 2(2)

126. Park I, Lee J, Saro L (2014) Ensemble of ground subsidence hazard maps using fuzzy logic. *Open Geosci* 6:207–218
127. Park S, Kim J (2019) Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. *Appl Sci* 9:942
128. 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* 11:5659
129. 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 Sensing Earth Syst Sci* 2(2–3):120–146
130. Payne RA (2015) *Guide to Regression, Nonlinear and Generalized Linear Models in Genstat*, 18th edn. VSN International, Hemel Hempstead
131. Petley DN (2008) The global occurrence of fatal landslides in 2007, *Geophysical research abstract*, EGU General Assembly, 10(3)
132. Phillips TH, Baker ME, Lautar K, Yesilonis I, Pavao-Zuckerman MA (2019) The capacity of urban forest patches to infiltrate stormwater is influenced by soil physical properties and soil moisture. *J Environ Manag* 246:11–18
133. 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>
134. Ramsay JO, Dalzell CJ (1991) Some tools for functional data analysis. *Journal of the Royal Statistical Society Series B (Methodological)* 53(3):539–572
135. Rau P, Bourrel L, Labat D, Ruelland D, Frappart F, Lavado W, Felipe O (2019) Assessing multidecadal runoff (1970–2010) using regional hydrological modelling under data and water scarcity conditions in Peruvian Pacific catchments *Hydrol. Process* 33(1):20–35
136. Ray A, Dhir A, Bala PK, Kaur P (2019) Why do people use food delivery apps (FDA)? A uses and gratification theory perspective. *J Retail Consum Serv* 51:221–230
137. Ridgeway G, Southworth MH, RUnit S. Package ‘gbm.’ *Viitattu*. 2013; 10(2013):40
138. Rodrigues M, De la Riva J (2014) An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environ Model Softw* 57:192–201
139. 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
140. Sahana M, Patel PP (2019) A comparison of frequency ratio and fuzzy logic models for flood susceptibility assessment of the lower Kosi River Basin in India *Environ. Earth Sci* 78(10):289
141. Sahana M, Rehman S, Sajjad H, Hong H (2020) Exploring effectiveness of frequency ratio and support vector machine models in storm surge flood susceptibility assessment: A study of Sundarban Biosphere Reserve, India *Catena*, 189, Article 104450

142. Sajedi-Hosseini F, Malekian A, Choubin B, Rahmati O, Cipullo S, Coulon F, Pradhan B (2018) A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination *Sci. Total Environ* 644:954–962
143. 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
144. 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>
145. Sanyal J, Lu X (2004) Application of remote sensing in flood management with special reference to monsoon Asia: a review. *Nat hazards* 33(2):283–301
146. Sarhadi A, Soltani S, Modarres R (2012) Probabilistic flood inundation mapping of ungauged rivers: linking GIS techniques and frequency analysis. *J. Hydrology* 45
147. Schapire RE (2003) The boosting approach to machine learning: An overview. In: *Nonlinear Estimation and Classification*. Springer, New York, pp 149–171
148. Scott AJ, Hosmer DW, Lemeshow S (1991) Applied logistic regression. *Biometrics* 47:1632
149. Sene K (2013) Flash floods: forecasting and warning. Springer, Dordrecht, p 395
150. 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>
151. Sevgen E, Kocaman S, Nefeslioglu HA, Gokceoglu C (2019) A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ANN and random forest. *Sensors* 19(18):3940. <https://doi.org/10.3390/s19183940>
152. Sarkar D, Mondal P (2020) Flood vulnerability mapping using frequency ratio (FR) model: a case study on Kulik river basin, Indo-Bangladesh Barind region. *Appl Water Sci* 10(1):17
153. Shi Y, Taalab K, Cheng T (2016) Flood prediction using support vector machines (SVM)
154. Siahkamari S, Haghizadeh A, Zeinivand H, Tahmasebipour N, Rahmati O (2018) Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models *Geocarto Int.*, 33 (9): 927–941
155. 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>
156. Souissi D, Zouhri L, Hammami S, Msaddek MH, Zghibi A, Dlala M (2019) GIS-based MCDM–AHP modeling for flood susceptibility mapping of arid areas, southeastern Tunisia *Geocarto Int.*, 1–27
157. Srivastava PK, Han D, Rico-Ramirez MA, Islam T (2014) Sensitivity and uncertainty analysis of mesoscale model downscaled hydro-meteorological variables for discharge prediction. *Hydrol Process* 28(15):4419–4432
158. Stefanidis S, Stathis D (2013) Assessment of flood hazard based on natural and anthropogenic factors using analytic hierarchy process (AHP). *Nat hazards* 68(2):569–585

159. Stevaux JC, de Azevedo Macedo H, Assine ML, Silva A (2020) Changing fluvial styles and backwater flooding along the Upper Paraguay River plains in the Brazilian Pantanal wetland *Geomorphology*, 350, Article 106906
160. 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*
161. 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>
162. Thanh Noi P, Kappas M (2018) Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors* 18:18
163. Torcivia CEG, López NNR (2020) Preliminary Morphometric Analysis: Río Talacasto Basin, Central Precordillera of San Juan, Argentina *Advances in Geomorphology and Quaternary Studies in Argentina*. Springer, Cham, pp 158–168
164. Turoglu H, Dolke (2011) Floods and their likely impacts on ecological environment in the Bolaman river basin (ORDU, TURKEY). *Res J Agric Sci* 43(4):167–173
165. Vapnik V (1998) The support vector method of function estimation. In: *Nonlinear Modeling*. Springer, Boston, pp 55–85
166. Vapnik V (2013) *The nature of statistical learning theory*. Springer science & business media
167. 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>
168. Wahlstrom M, Guha-Sapir D (2015) *The human cost of weather-related disasters 1995–2015* UNISDR, Geneva, Switzerland
169. Wagner-Muns IM, Guardiola IG, Samaranayke VA, Kayani WI (2018) A Functional data analysis approach to traffic volume forecasting. *IEEE Trans Intell Transp Syst* 19(3):878–888
170. Wanders N, Karssenbergh 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
171. Wang H, Yang F, Luo Z (2016) An experimental study of the intrinsic stability of random forest variable importance measures. *BMC Bioinformatics*, 17 (1)
172. Wang Y, Hong H, Chen W, Li S, Pamučar D, Gigović L, Drobnjak S, Tien Bui D, Duan H (2019) A Hybrid GIS Multi-Criteria Decision-Making Method for Flood Susceptibility Mapping at Shangyou, China. *Remote Sens* 11:62. <https://doi.org/10.3390/rs11010062>
173. Wang Y, Fang Z, Hong H, Peng L (2020) Flood susceptibility mapping using convolutional neural network frameworks *J. Hydrol.*, 582, Article 124482
174. 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>
175. Ward RC, Robinson M (2000) *Principles of Hydrology*, 4th edn. McGraw-Hill, Maidenhead

176. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: A basic tool of chemometrics. *Chemom Intell Lab Syst* 58:109–130
177. Wu L, He Y, Ma X (2020) Can soil conservation practices reshape the relationship between sediment yield and slope gradient? *Ecol. Eng.*, 142, Article 105630
178. Xie H, Dong J, Shen Z, Chen L, Lai X, Qiu J, Chen X (2019) Intra-and inter-event characteristics and controlling factors of agricultural nonpoint source pollution under different types of rainfall-runoff events *Catena*, 182, Article 104105
179. Xiao T, Yin K, Yao T, Liu S (2019) Spatial prediction of landslide susceptibility using GIS-based statistical and machine learning models in Wanzhou County, Three Gorges Reservoir, China *Acta Geochim.*, 38: 654–669, 10.1007/s11631-019-00341-1
180. Xu C (2013) Assessment of earthquake-triggered landslide susceptibility based on expert knowledge and information value methods: a case study of the 20 April 2013 Lushan, China Mw6. 6 earthquake *Disaster Advances*, 6 (13): 119–130
181. Xu C, Dai F, Xu X, Lee YH (2012) GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China; *Geomorphology*, doi: 10.1016/j.geomorph.2011.12.040
182. Xu M, Li C (2020) Influencing Factors Analysis of Water Footprint Based on the Extended STIRPAT Model Application of the Water Footprint: Water Stress Analysis and Allocation. Springer, Singapore, pp 105–126
183. Yamani K, Hazzab A, Sekkoum M, Slimane T (2016) Mapping of vulnerability of flooded area in arid region. Case study: area of Ghardaïa-Algeria. *Model Earth Syst Environ* 2:147. <https://doi.org/10.1007/s40808-016-0183-x>
184. 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
185. Yao X, Tham LG, Dai FC (2008) Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* 101:572–582
186. Yang L, Cervone G (2019) Analysis of remote sensing imagery for disaster assessment using deep learning: a case study of flooding event *Soft. Comput.*, 1–16
187. 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*, 229–257. <https://doi.org/10.1016/B978-0-12-815226-3.00010-7>
188. Zhang G, Chen W, Li G, Yang W, Yi S, Luo W (2020) Lake water and glacier mass gains in the northwestern Tibetan Plateau observed from multi-sensor remote sensing data: Implication of an enhanced hydrological cycle *Remote Sens. Environ.*, 237, Article 111554
189. 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
190. Zhou C, Yin K, Cao Y, Ahmed B, Li Y, Catani F, Pourghasemi HR (2018) Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area,

191. Zou M, Sun C, Liang S, Sun Y, Li D, Li L, Fan L, Wu L, Xia W (2019) Fisher discriminant analysis for classification of autism spectrum disorders based on folate-related metabolism markers. J Nutr Biochem 64:25–31

Figures

Figure 1

a) Location map of the study area along the Red Sea and the Gulf of Suez coast; b) Close-up of the study area showing cities, road networks, holiday areas, flood detection points (training and validation points) and comparison sites (1, 2, 3, and 4).

Figure 2

Various photographs taken during different flood events in the study area. Impact of the flood in 2014: a, b show the flooding of the paved road Hurgada-Safaga Sea; c, d show the impact of the flood on Hurgada airport. Impact of flood in 2016: e) flooding of asphalt road Qena-Safaga, f) rock fall along asphalt road due to heavy rainfall, g, and h) show the flooding of different asphalt roads within Hurgada.

Figure 3

A diagram showing data used and modeling steps applied to provide an accurate FSMs.

Figure 4

Close-up medium-resolution Landsat-8 images at sites 1 and 2 (as shown in Fig. 1b) before (a, c) and after (b, d) the 2014 flash flood event; high-resolution astro-digital images at sites 3 and 4 (as shown in Fig. 1b) before (e, g) and after (f, h) the 2016 flash flood event.

Figure 5

FRFs maps applied in the current study: (a) altitude, (b) slope-aspect, (c) lithology, (d) landuse/landcover (LULC), (e) slope length (LS), (f) topographic witness index (TWI), (g) slope-angle, (h) profile curvature, (i) plan curvature, (j) stream power index (SPI), and (k) hydrolithology units.

Figure 6

The importance of Flood-related factors using Partial Least Squares (PLS).

Figure 7

Flood susceptibility models using MLTs models: (a) SVM, (b) RF, (c)MDA, (d) MARS, (e) GLM, (f) FDA, and (g) BRT)

Figure 8

FSM classes' areas % for various MLTs in the current study.

Figure 9

Prediction rate curves for the FSMs generated in this study for different MLTs models.