Multi-hazards modeling using machine learning algorithms in Southwestern Saudi Arabia

Ahmed M. Youssef
Saudi Geological Survey
Ali M. Mahdi
South Valley University
Mohamed M. Al-Katheri
Saudi Geological Survey
Soheila Pouyan
Shiraz University
Hamid Reza Pourghasemi (hamidreza.pourghasemi@yahoo.com)
Shiraz University https://orcid.org/0000-0003-2328-2998

Research Article

Keywords: Multi-hazards, Susceptibility, Future development, Sustainability, KSA

Posted Date: May 25th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1554302/v1

Abstract

The current study aimed at producing a multi-hazard risk map for Hasher-Fayfa Basin. The Basin is part of Jazan region in the southwestern Saudi Arabia and is distinguished by mountainous terrain. Recently, this area has experienced many extreme natural processes that become natural hazard events when it intersects with human activities (urban areas and infrastructures). This work is mapping the probabilities of three main hazards; landslides, floods, and gully erosion using machine learning algorithms named boosted regression tree (BRT), a generalized linear model (GLM), Flexible discriminant analysis (FDA), random forest (RF), and multivariate discriminant analysis (MDA). Several factors obtained from various sources, including topographical, geological, meteorological, hydrological, and human activities were incorporated to produce the final multi-hazard risk model. Area under the curve (AUC) was applied to identify the best predictive model for each natural hazard type. AUC values between 80 and 90% indicated that the model is very good and above 90% indicated that the model is excellent in predictive capability. Based on the accuracy evaluation, it was found that the FDAFDA model is the most accurate for predicting landslides with AUC values of (92.7%, excellent performance), RF model is the most accurate for predicting floods and erosion with AUC values of (97.2%-excellent and 83.3%-very good performance, respectively). Finally, multi-hazard risk map was prepared by coupling of the mentioned three hazards. Results showed that 33.5% of the total area is safe (no-hazard), whereas 66.5% is characterized by at least a single hazard and 2–32 – 3 hazard combination. Machine learning approaches are useful tools as baselines for management and mitigation processes, based on multi-hazard modeling. It could help planners and decision-makers to manage future human activities and expansions.

1. Introduction

Managing disasters is challenging due to various kinds of hazards (e.g., natural and anthropogenic hazards) (De Silva and Kawasaki, 2018). Natural processes will become natural hazards, when they intersect and have a negative effect on people life (Gill and Malamud, 2017; Ward et al., 2020). Subsequently, natural hazards are the physical events that pose devastating impacts on communities, damaging critical infrastructure, affecting economic viability, and claiming lives (IPCC, 2007). Worldwide, an exponential increase in the human population leads to lateral expansion in development (urbanization, agricultural activities, infrastructures, and life lines), subsequently increasing the exposure and interaction with different types of natural hazards (Weinkle et al., 2018). Globally, several areas are vulnerable to events involving several types of hazards that are related to the same area and at the same time (Leonard et al., 2014). Natural hazards have impacted people and natural environments in many undeveloped countries more severely than developed countries, causing substantial loss of lives and enormous economic plunge (Iglesias et al., 2021). In a simple definition, a hazard refers to the probability of potential damage in certain areas (location or where?) and a specific period (time or when?) due to an event with specific magnitude (how large?) (Shi, 2019).

Arid and semi-arid areas around the world are impacted by different natural hazards events, ranging from insignificant influence events to extremely events (IPCC, 2012; Tabari 2020; Rutgersson et al., 2021). These natural hazards frequently occur, threatening everything (modern cities, ruler areas, infrastructures, and life lines) (De Silva and Kawasaki, 2018). These disasters can cause great economic damage, disruption of transportation systems, injury, and claiming lives. Mountainous provinces are highly likely among the most disaster-susceptible areas due to their various characteristics such as lithology, tectonics, climate, and hydrology (Baig et al., 2021). Risk related to natural hazards is high in middle East countries due to lack of disaster preparedness and public awareness, and inadequate funding support (AlQahtanya and Abubakarb, 2020). Saudi Arabia is one of these countries that experienced many natural hazards, yearly (Alyami et al., 2021). These events cost lives and bring the economy to a static state. Moreover, not much has been done to map different prone areas related to these natural hazards.

Mountainous regions that are inhabited by people and crossed by infrastructures are prone to not only one type of hazard, but they are frequently the scene of multiple disasters that interact together, such as earthquakes, avalanches, landslides, floods, mudslides, ground subsidence, soil erosion, and wildfires (Gill and Malamud, 2014; Shah et al., 2018). Duncan et al. (2016) defined multi-hazard as all potential and cascading hazards, in any particular area at a certain period. Most of the time, these hazards can cause severe damage to the various human activities located and intercept with the hazard prone areas by causing fatalities and injuries, damaging their urban and industrial areas, destroying infrastructures (roads, tunnels, and railways), and disrupt lifelines (power lines, water systems, and gas mains) (Bell and Glade, 2004).

Due to the complexity of natural hazards, many studies deal with individual hazard types as an independent approach (Wastl et al., 2011). Also, problems related to these problems (e.g., floods, landslides, and gully erosions) have been evacuated individually by applying various machine learning algorithms-MLAs (Sarkar and Mishra, 2018; Ghorbanzadeh et al., 2019; Hosseiny et al., 2020; Zhou et al., 2021). However, Earth system science approach indicates that a significant interaction between different types of hazard components due to the interaction between the component systems (e.g., the lithosphere, atmosphere, hydrosphere, and biosphere). Accordingly, a holistic approach to understand all different hazards in a certain area is a must and can prevent cascading hazards that can be formed by the interaction of different hazard types. A multi-hazard risk evaluation could be significant to control hazard interactions (Komendantova et al., 2014). To prevent these natural hazards and protect people and their properties and the country economy from long-term plunging, effective predictive models must develop (Bathrellos et al., 2017). These models should base on profound understanding of the most influential and triggering factors that have a significant contribution on these hazards (van Westen and Greiving, 2017).

Multi-hazard modeling becomes an essential tool in land use development at both regional and national scales (Saunders and Kilvington, 2016). These multi-hazard modeling approaches gain good attention recently worldwide based on their ability to consider different types of hazards that could impact the area (Schmidt et al., 2011; Skilodimou et al., 2019; Lombardo et al., 2020).

Different approaches were used to conduct multi-hazards modeling e.g., by using two decision-making tools including, sequential Monte Carlo method and decision-making tool (Komendantova et al., 2014); using a deterministic equation based (theoretical) and empirical understanding (Bout et al., 2018); and by using multi-criteria analysis and GIS (Skilodimou et al., 2019). Recently, MLAs have been widely used to predict various hazards based on random forest (RF), and support vector machine-SVM (Nachappa et al., 2020), boosted regression tree-BRT-, and generalized additive model-GAM (Ye et al., 2020).

In this study, no multi-hazard assessment has been done in Saudi Arabia before and the application of machine learning techniques in multi-hazard assessment will be a novel study in the area. In this study, developing a multi-hazard risk map using machine learning techniques for the most impacted natural hazards (landslides, floods, and gully erosion) in this area are crucial for effective landuse management. The current study evaluates machine learning models (MDA, GLM, FDA, BRT, and RF) as effective and accurate models to produce multi-hazard risk map for Hasher-Fayfa Basin that might be used by authorities, developers, and decision-makers.

2. Study Area

The Jazan province has an area of 11,671 km^2, is situated at the southwestern corner of Saudi Arabia, bordering the Red Sea to the west and Asir region to the north (Fig. 1a). The Jazan region includes the study area that is bounded by latitudes 17°10' to 17°30' N and longitudes 42°58' and 43°13' E. Hasher-Fayfa Basin covers 513 Km^2. Its elevation ranges from 220 m to 2,340 m above sea level (e.g. Fayfa Mountain up to 2.3 km high). This makes it the one of the highest areas in KSA. The population of the Jazan region is about 1.67 million persons in 2020 (https://knoema.com/atlas/Saudi-Arabia/Jazan/Total-Population). The area under study inhabited by many people who live in the mountainous areas as in (Fayfa Mountain, Fig. 1(b-c)). Since about six decades, Saudi Arabia was discovered the oil fields, many urban centers and infrastructures were established. The Jazan area has been expanded drastically. The area has all the elements that could put it under a high risk of floods, landslides, and erosion that could impact the economic development of the region. These hazards will pose a high risk to people and their property (buildings, highways, roads, infrastructure, and lifelines). The basin is characterized by a rugged landscape with the iconic features of the Hasher and Fayfa mountains. These mountains have steep reliefs (up to 88°). Rainfall is the main trigger factor for landslide, flood, and erosion in the area under consideration. The region receives up to 500 mm/year of precipitation, with a high rainfall value of 1,400 mm in 1979 (Hasanean and Almazroui 2015). The high rainfall in the region is mainly due to the tropical air masses affecting the Jazan and Asir regions (the southwestern provinces of KSA) and the high altitudes that produce orographic rainfall (Şen et al., 2017).

Additionally, the area is distinguished by the presence of intense tectonic deformation (folds, faults, shear zones, fractures, and joints). The Red Sea Mountains of Saudi Arabia consists of extrusive and intrusive igneous rocks, and metamorphic rocks. These rocks were formed by the collection of island arcs and closure of the interleaving oceanic arcs 700 – 1,000 Ma (Greenwood, 1982; Stoeser and Camp, 1985). These mountains were raised up due to the spreading of the Red Sea about thirty million years ago.

These Precambrian rocks are unconformably overlain by sedimentary sequences of Paleozoic and Cenozoic age (Agar, 1987).

3. Methodology

The current work involved five steps (Fig. 2): (1) data collection of the various events based on intensive field work, analysis of various civil defense reports, technical reports, and questionnaires with the local population (for the period between 2000 and 2020); (2) determine the various events based on a through literature review; (3) model different hazard types using various machine learning techniques (; (4) model validation and selection of the most appropriate model for each hazard type; and (5) generation of multi-hazard risk map (MHRM) by integrating the high-AUC model for each hazard.

3.1. The inventory of hazard data (multi-hazards inventory)

Creating an inventory of susceptible landforms is an essential and crucial stage in hazard susceptibility mapping. The inventory of landslides, gully erosion, and flooding in the watershed of the study area was compiled based on field investigations with the help of GPS instrument (732 sites were investigated through the field work between 2015 and 2018) (Fig. 3). National and regional data from various sources, including the Civil Defense, Jazan Region Authority, Ministry of Transportation, Saudi Geological Survey, review of technical reports and scientific publications and the private sector of the Hasher and Fayfa region were collected and reviewed. As some of the vulnerable areas are located in mountainous areas and may have been overlooked during the field survey; so, high-resolution Google Earth images were also used to detect landslides, gully erosion features, and flooded areas. Based on this information, 100 landslide locations, 60 sites for gully erosion, and 70 sites for flood locations were used to establish the multi-hazard inventory map (Fig. 1a). In this study, we used a hazard and non-hazard locations and machine learning techniques. Accordingly, equal numbers of hazards and non-hazardous locations were randomly selected (Fang et al., 2021). A random partitioning method was used to select training and validation sites for each hazard type in the inventory map. In the current work, training datasets (70% of each hazard) were used for model generation and the remaining 30% of hazard locations were used for model validation (Zhao et al., 2018; Wang et al., 2020).

3.2. Hazard-predictive factors (HPFs)

It indicated that hazard-predictive factors (HPFs) are the first step of the modeling process (Rusk et al., 2022). Accordingly, intensive review of previous studies was done to identify the most effective factors for each hazard type (Roy and Saha, 2019; Razavi-Termeh et al., 2020; Wang et al., 2020). Raster database layers of different topography, climatology, hydrology, vegetation and landuse/landcover, geology, and anthropogenic were selected and prepared using ArcGIS 10.8 of 5 × 5 m resolution (Table 1, Fig. 4). These factors include 16 HPFs; elevation (Fig. 4a), slope angle (Fig. 4b), topographic wetness index (TWI – Eq. 1) (Beven and Kirkby, 1979) (Fig. 4c), plan curvature (Fig. 4d), aspect (Fig. 4e), topographic roughness index (TRI - Eq. 2) (Nellemann and Reynolds, 1997) (Fig. 4f), profile curvature (Fig. 4g), and valley depth (VD) (Fig. 4h) were extracted from DEM 5 m resolution that were generated

using 1:10,000 topographic map; lithology (Fig. 4n) and distance to fault (Fig. 4o) were extracted from a geologic map of 1:250,000 scale obtained from Saudi Geological Survey; rainfall distribution map was generated from 20 rainfall stations (acquired from the MEWA, covering a period from 1965 to 2020) (Fig. 4i), distance to wadis (Fig. 4j) and stream density ((Fig. 4k) were generated from topographic map 1:10,000 scale and converted to distance to wadis and stream density using Euclidian distance tool and density tools in ArcGIS 10.8; distance to road (Fig. 4p) was extracted from high resolution images (Google Earth, then converted to distance to road using Euclidian distance tool in ArcGIS 10.8, landuse/land cover (LULC) (Fig. 4m) and normalized differential vegetation index (NDVI) (Fig. 4l) were generated from sentenil-2 image (10 m resolution) acquired in May 2020, LULC map was generated using maximum likelihood classification method, and finally NDVI was extracted using NDVI equation (Eq. 3). Shalaby and Tateishi (2007) pointed out that the maximum likelihood method is a stable and accurate model for image classification because it can apply multiple statistical features in its work.

Table 1
The most influencing indicators for multi-hazard susceptibility mapping.

Factor Type	Hazard-indicator factors and abbreviation	Data type	Factor range /Units	Hazard type		
				Landslide (LS)	Flood (FL)	Gully erosion (GE)
Topography	Elevation (EL)	Grid	220-2,340 m	✓	✓	✓
	Slope (S)	Grid	0–88 degree	✓	✓	✓
	Topographic wetness index (TWI)	Grid	-1.78–24.74	✓	✓	✓
	Plan curvature (PC)	Grid	-271.5–1452.2	✓	✓	✓
	Aspect (A)	Grid	9 directions	✓	✓	✓
	Terrain ruggedness index (TRI)	Grid	0–174			✓
	Profile curvature (PrC)	Grid	-1627.8–917.3			✓
	Valley depth (VD)	Grid	0–261 m			✓
Climatology	Rainfall (R)	Point	383–495 mm		✓	✓
Hydrology	Distance to wadis (DtW)	Polyline	0–2,900 m	✓	✓	
	Stream density (SD)	Polyline	0.95–23.7 km/km^2		✓	✓
Vegetation and landuse/land cover	Normalized difference vegetation index (NDVI)	Grid	-0.639–0.675	✓	✓	✓
	Land use/Landcover (LULC)	Grid	5 classes	✓	✓	✓
Geology	Lithology (Lth)	Polygon	6 groups	✓	✓	✓
	Distance to fault (DtF)	Polyline	0–3,153 m	✓		
Anthropogenic	Distance to road (DtR)	Polyline	0–5,780 m	✓	✓	

All, layers were converted to have 5 m * 5 m pixel size using ArcGIS 10.8. Table (1) shows that different factors used to prepare hazard susceptibility maps; 11 factors used for landslides, 12 factors were used for floods, and 13 factors were used for gully erosion.

$$TWI = lin\frac{\mathrm{A}}{\tan B}$$

1

Where, A is the cumulative basin area (m^2), and β is the slope (in degrees) at a point.

$$\mathrm{TRI} = \mathrm{Y}\left[\sum (x_{ij} - x_{00})^2\right]^{1/2} \quad (2)$$

Where, xii is the height of each pixel adjacent to the pixel (0, 0). Areas with a slope of 0 have a TRI value of zero, while the roughened areas with steep elevations have positive values of TRI.

$$NDVI = \frac{(NIR - R)}{(NIR + R)}$$

3

Where, NIR is the near infrared band (wave length = 832.8 nm) and R is the red band (wave length = 664.6 nm) of the spectral bands.

3.3. Machine learning algorithms

In the current work, five machine learning algorithms (MLAs) were used to generate landslide, flood, and gully erosion susceptibility maps that affect the study area. Various literatures were conducted using various MLAs to overcome the problems related to natural hazards by predicting their susceptibility, such as landslides (Park and Kim, 2019; He et al., 2021), floods (Janizadeh et al., 2019; El-Haddad et al., 2021), and gully erosion (Ghorbanzadeh et al., 2020; Amare et al., 2021; Yang et al., 2021). A detailed description of the MLAs that used in this study is given in the following parts:

3.3.1. Random forest

Random Forest (RF) is an ensemble classification technique (Breiman, 2001) and considered as one of the learning techniques that involves multiple steps, starting with training datasets, followed by bootstrapping, then ensemble of trees, and finally aggregation (classification phase) (Hawryło et al., 2018; Sarker et al., 2019). Bootstrapping refers to the parallel training of each decision tree with other training subsets using different available features (Herrera et al., 2019). In this phase, each individual decision tree is unique and is used to minimize the overall variance of the RF model. In the last stage, the RF model summarizes the decisions of the individual trees; therefore, the RF model shows good generalization. The RF algorithm tends to outperform most other classification approaches in terms of accuracy, with no problems of overfitting (Pedregosa et al., 2011). The RF algorithm is robust to training

sample selection and noise in datasets and does not require feature scaling (fits both categorical and continuous values) (Kim et al., 2018). RF can overcome outliers in predictors, automatically handle omitted data and increase the diversity of classification trees.

3.2. Multivariate discriminant analysis

Multivariate discriminant analysis (MDA) is a trained classification algorithm as an extension of linear discriminant analysis (LDA). It can evaluate the multivariate nonlinear relationships between different classes within each group (Lombardo et al. 2006). To calculate the distance to the nearest cluster, the normal distribution (the variability and correlation between variables is uniform) of each class is used (Lombardo et al., 2006). It can be derived using Eq. (4) (Hair et al., 1998).

$$D_v = D_{w1}V_1 + D_{w1}V_1 + \cdots + D_{wn}V_n$$

4

D_v and D_{w1} (i = 1,2,3, ..., n) represent the discriminant value and weights, respectively, and V_1 (i = 1,2,3,..., n) are independent indicators. The MDA model was run using “mda” package in R software (Hastie et al., 2017).

3.3.3. Generalized linear model

The generalized linear model (GLM) is a linear regression model that can quantify and incorporate specific and temporal factors (Goetz et al., 2015). The GLM can scale the accuracy and quality of the outcomes by using multiple regression to establish a unique link between the dependent and independent factors (Scott et al., 1991). By identifying the best regression model, multiple events can be predicted (Payne, 2015). It is an adequate model for numeric factors because of its regression characteristic; however, insignificant and correlated factors can reduce its performance, it also, deal with non-normal distribution data (Kalantar et al., 2020). The GLM linkage function is used to establish the relationship between the dependent variable and the independent variables (Soch et al., 2017). In this study, the occurrence probability of hazard event Y can be represented by Eq. (5). By the logistic transformation, the link function $g(y_i)$ is represented by Eq. 6:

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

5

$$g\left(y_i\right) = c_0 + \sum c_ix_i + \epsilon_i$$

6

where P is the event “Y” probability; c_0 is the intercept; $c_1 \ldots c_i \ldots$ are logistic regression coefficients, $X_1 \ldots X_i$ are the independent variables; and the error residual expressed by ε_i. In this study, R software was used to construct the GLM model (R Core Team, 2019). A simple Gaussian family is determined as the link function for normally distributed response data. The independent indicators were to be included in the model individually, using a smoothing spline based on two degrees of freedom to avoid overfitting (Aertsen et al., 2009).

3.3.4. Flexible discriminant analysis

Ramsay and Dalzell (1991) proposed the Flexible discriminant analysis (FDA) as a nonlinear statistical classification method. The crucial concept of FDA is to treat an observed object with functional characteristics as an integral, regardless of the order of the observed values (Wagner-Muns et al., 2018). It can distinguish unsupervised work where each class is divided into subclasses with a unique value (Zou et al., 2019). FDA can provide a framework by coupling various models and methods (regression algorithms, discriminant analysis, and classification techniques) (Hastie and Tibshirani, 1996). In this study, the FDA technique was used to create a hazard susceptibility map using the species distribution package in R software (Naimi and Araújo, 2016).

3.3.5 Boosted regression trees

BRT is a combination of regression and boosting techniques. It was proposed by Friedman (2001). The boosting algorithm is used to improve model accuracy, where the residual errors of the existing tree will be used to adjust the new trees (Ye et al., 2009). Shin et al. (2012) pointed out that the BRT model has several advantages because it can be improved by fitting and combining multiple methods and increasing the number of trees, it uses different types of data without conversion or outlier removal, it works when some data are missing, and it can use sophisticated nonlinear relationships to fit the data. The BRT model is defined by Eq. (7) (Friedman 2001).

$$F(X; \left[B_m a_m\right]^{m_o} = \sum_{m=0}^{m} B_m h(X; a_m)$$

7

Where is $h(X; m)$ a simple classification function with variable “a” and factors “x.” Factor “m” is representing the step of the algorithm, and is B_m the weighting coefficient at step m. A new tree will be added to the original model at each iteration step, which leads to reducing of the loss function. The BRT training stage will be ended as soon as the predefined iteration numbers achieved. The BRT model was run using the package "brt" in R software (Ridgeway and Southworth 2013).

3.4. HFIs multicollinearity and importance, and model validation

Multicollinearity method is used to evaluate variables effectiveness in different model construction (Yoo and Cho 2019). In this method, two indices were used, variance inflation factors (VIF) and tolerance (TOL)

as shown in equations (8 and 9):

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

8

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

9

R_j^2 Represents the regression coefficient of the explanatory factor J on all other descriptive factors. Studies have shown that a TOL < 0.10 and a VIF > 5 account for multicollinearity issues (Rahman et al., 2019).

Additionally, the importance of independent indicators is essential. It measures their contribution in each hazard type modeling. Several techniques were used among them the learning vector quantization (LVQ), and random forest algorithm (RF) (Zhang and Xie, 2012; Sun et al., 2020). In this study, HIFs importance and contribution to landslide, floods, and gully erosion occurrence were assessed using RF.

Finally, to evaluate the accuracy of the results, the receiver operating characteristics (ROC) and the area under the curves (AUC) can also be applied to evaluate the MLMs that used to provide valuable information (Bradley 1997). Without validated maps, the models and their results would have no scientific significance or if the AUC is less than 0.5, the model is considered a random model (Marzban, 2004). Evaluation of the accuracy of the different machine learning models for each hazard type is a profound step to create the final hazard model (Amare et al., 2021). In this study, the performance of all hazard models was determined using the validation dataset of each hazard type. This will provide quantitative measures of the model accuracy. The predictive performance was done using the AUC. AUC value ranges from 0 to 1 (0 to 100%). A range of 1 (100%) represents high performance, while a range of ≤ 0.5 (50%) indicates a poor classification. Evaluation of the accuracies of these models based on (Guzzetti et al., 2005). They divided the AUC values into three groups: between 0.75 and 0.8 for an acceptable model, between 0.8 and 0.9for a good susceptibility model, and above 0.9 for an excellent model.

4. Results And Analysis

4.1. Multicollinearity test

The multicollinearity analysis of the hazard-predictor factors used in this study for landslide susceptibility (11 factors), flood susceptibility (12 factors), and gully erosion susceptibility (13 factors) is shown in Table (2). Results showed that both tolerance (ToL) and (VIF) values of landslide-predictor factors used in

this study to run the different machine learning models were appropriate, ToL values were more than 0.1 (the lower ToL value is 0.20) and the VIF values were less than 5 (the large VIF value is 4.9) for slope factor. The tolerance (ToL) and (VIF) values of flood-predictor factors were appropriate, ToL values were more than 0.1 (the lower ToL value is 0.257) and the VIF values were less than 5 (the large VIF value is 3.898) for elevation factor. For the gully erosion predictor factors, all factors are adequate except the TRI that provide VIF value more than 5. For the rest of these factors the ToL values were more than 0.1 (the lower ToL value is 0.271) and the VIF values were less than 5 (the large VIF value is 3.697) for the slope facto. Accordingly, the results showed that there was no multicollinearity among the selected FCFs, which means that they made a significant contribution to the model construction in this study.

4.2. The importance of hazard-predictors

The analysis of landslide hazard-predictor importance is shown in Figure (5a). Our findings indicated that the top predictors that controlled landslide susceptibility are the slope angle (S), followed by land use/land cover (LULC), elevation (EL), the normalized differential vegetation index (NDVI), followed by aspect (As), and distance to wadis (DtW), distance to road (DtR), topographic witness index (TWI) and distance to fault (DtF). Less importance predictor is lithology (Lth). We believed that this is due to the higher slope angle for the mountainous areas and the gravity forces play a great role than material types (Çellek, 2020). However, negative importance factor comes to be plan curvature (PC).

Gully erosion hazard-predictor importance is shown in Figure (5b). Results showed that the top predictors-controlled gully erosion susceptibility is TWI followed by VD, EL, NDVI, SD, PrC, S, and R. The less important predictors are aspect (AS), TRI, and PC, followed by LULCs, and Lth.

The analysis of flood hazard-predictor importance is shown in Figure (5c). Results showed that the top predictors that controlled flood susceptibility belonged to the hydrologic group, which was DtW, followed by equal importance of slope angle (S), elevation (EL), and stream density (SD), then followed by both TWI and rainfall (R), followed by normalized differential vegetation index (NDVI) and plan curvature (PC). Less importance predictors are LULCs, aspect (AS), and lithology (Lth). However, negative importance factor comes to be distance to road (DtR).

4.3. Different susceptibility maps

Maps of landslides, flood, and gully erosion susceptibility were generated based on both hazard-predictive factors (independent factors) and actual hazard locations (dependent variable) by applying various types of machine learning models (MLMs). Each susceptibility map was divided into five classes very low, low, moderate, high, and very high based on Natural Breaks Classifier-Jenks (Jenks and Caspall, 1971) (Fig. 6). Landslide susceptibility maps for the Hasher-Fayfa Basin were created on the basis of RF, FDA, MDA, GLM, and BRT models (Fig. 6 (a-e)); gully erosion susceptibility maps for the Hasher-Fayfa Basin were created according to the RF, FDA, MDA, GLM, and BRT models (Fig. 6 (f-j)); and flood susceptibility maps were created using RF, FDA, and MDA models (Fig. 6 (k-m)).

Detailed analysis was conducted for each model. Landslide hazard models indicated that high and very high susceptible regions based on RF and BRT covered more areas (21.2 and 22.5%, respectively) than the other three models (FDA = 17.8%, MDA = 18.6%, and GLM = 17.6%) (Fig. 7a). Gully erosion models indicated that high and very high susceptible classes in MDA and BRT, covering more areas (37.5 and 34.2%, respectively) than the other three models (RF = 32.8%, FDA = 31.4%, and GLM = 32.6%) (Fig. 7b). Flood models showed that high and very high susceptible classes in the MDA and FDA, covering more areas (15.0 and 14.1%, respectively) than RF model (RF = 6.5%) (Fig. 7c).

4.4. Models' validation

Figure 8 (a) and Table (3) show the AUC curves and accuracy values for the landslide's five ML models. Results show that landslides susceptibility models give high performance where AUC > 0.9 (90%), the AUC values for ML models BRT, RF, GLM, MDA, and FDA are 0.909 (90.9%), 0.918 (91.8%), 0.923 (92.3%), 0.923 (92.3%), and 0.927 (92.7%), respectively. Results show that the most accurate ML model that has a high predictive accuracy for landslide susceptibility is FDA model with AUC value of 92.7%. Figure 8(b) and Table (3) show the AUC curves and values for the five ML models that were applied to create the gully erosion susceptibility models. Results show that gully erosion susceptibility models give AUC values range between 0.739 and 0.833 (medium to high performance). The ML models MDA, GLM, FDA, BRT and RF give AUC values of 0.739 (73.9%), 0.776 (77.6%), 0.792 (79.2%), 0.827 (82.7%), and 0.833 (83.3%), respectively. Results show that the most accurate ML model that has high predictive accuracy for gully erosion susceptibility is RF model with AUC value of 83.3%. Figure 8(c) and Table (3) show the AUC curves and values for the three ML models that were used to generate the flood susceptibility maps. Results show that flood susceptibility models give AUC values range between 0.932 and 0.972 (high performance). The ML models MDA, FDA, and RF give AUC values of 0.932 (93.2%), 0.945 (94.5%), and 0.972 (97.2%), respectively. Results show that the most accurate ML algorithm that has high predictive performance for flood susceptibility is RF model with AUC value of 97.2%.

4.5. Multi-hazard risk map

To produce multi-hazard risk map for these three hazard types, three steps were applied: 1) hazard susceptibility map to each hazard type (LS, GE, and FL) was created according to the relationship between the independent factors (hazard-predictors) and the dependent factors (hazard locations of landslides, gully erosions, and floods) using various types of machine learning techniques (Fig. 6), 2) the ROC-AUC method was used to determine the highest performance models for each hazard type and then integrated to map multiple hazards (Table 2). The most accurate models in this study are FDA model for landslides where AUC = 92.7%, RF for gully erosion with AUC = 83.3%, and RF for floods with AUC = 97.2%. These three susceptibility maps of the three natural hazards (landslides, gully erosion, and floods) were combined in ArcGIS 10.8 to produce an integrated multi-hazard risk map.

Table 2
Multicollinearity results of hazard-predictor factors.

Hazard-causative factors	LS		FL		GE	
	Tolerance	VIF	Tolerance	VIF	Tolerance	VIF
Elevation (EL)	0.290	3.448	0.257	3.898	0.476	2.101
Slope (S)	0.200	4.900	0.316	3.168	0.271	3.697
Topographic wetness index (TWI)	0.307	3.258	0.304	3.290	0.326	3.067
Plan curvature (PC)	0.869	1.151	0.926	1.080	0.446	2.243
Aspect (A)	0.677	1.477	0.897	1.115	0.843	1.186
Terrain ruggedness index (TRI)	-	-	-	-	0.020	49.302
Profile curvature (PrC)	-	-	-	-	0.523	1.911
Valley depth (VD)	-	-	-	-	0.502	1.991
Rainfall (R)	-	-	0.719	1.391	0.716	1.396
Distance to wadis (DtW)	0.461	2.168	0.398	2.511	-	-
Stream density (SD)	-	-	0.723	1.384	0.613	1.632
Normalized difference vegetation index (NDVI)	0.382	2.620	0.460	2.176	0.699	1.430
Land use/Landcover (LULC)	0.660	1.514	0.741	1.350	0.703	1.422
Lithology (Lth)	0.718	1.393	0.586	1.705	0.672	1.488
Distance to fault (DtF)	0.842	1.188	-	-	-	-
Distance to road (DtR)	0.767	1.304	0.705	1.418	-	-

Table 3
AUC values for different machine learning models used in modeling landslide, gully erosion, and flood.

Model Type	Test Result Variable(s)	Area (AUC)	Standard Errors
Landslides	BRT	0.909	0.054
	RF	0.918	0.052
	GLM	0.923	0.051
	MDA	0.923	0.050
	FDA	0.927	0.049
Gully Erosion	MDA	0.739	0.064
	GLM	0.776	0.059
	FDA	0.792	0.057
	BRT	0.827	0.053
	RF	0.833	0.051
Floods	MDA	0.932	0.049
	FDA	0.945	0.035
	RF	0.972	0.000

Results showed that the multi-hazard risk map included eight susceptibility classes (Figs. 9 and 10). These classes include; no-hazard or safe areas (No-H covers 33.6% of the study area). However, about 66.4% of the total area are subjected to various types of hazards as follows: landslides (LS = 22.4%), gully erosion (GE = 28.9%), floods (FL = 1.6%), landslides-erosion (LS-GE = 6.5%), floods-landslides (FL-LS = 0.3%), floods-erosion (FL-GE = 6.5%), and flood-landslides-erosion (FL-LS-GE = 0.3%).

5. Discussion

Understanding hazard management could be enhanced using different types of modeling approaches (Hill and Minsker, 2010). These models could provide planners and policymakers with useful, efficient, and informative results. Across the world, countless literature has individually studied different types of natural hazards (e.g., Amare et al., 2021; Msabi and Makonyo, 2021). All these studies provide important and useful results. However, most natural hazards do not occur individually due to cascading effects. One hazard may lead to another; therefore, multiple hazards and their linkages, interactions and cascading impacts can provide much understanding of their processes and optimal ideas for averting and minimizing disaster losses and for effective land management (Godschall et al., 2020). These maps can provide valuable information that is critical for planning and managing existing and future human

activities. Dealing with multi-hazards has shown that the interaction of numerous hazard types can cause a higher risk than the risk of a single hazard type (Liu et al., 2021).

In the current work, we investigated three hazards in a Hasher-Fayfa Basin (mountainous catchment) in Saudi Arabia. To evaluate disaster events in this area, multi-hazard modeling was performed. To achieve that five machine learning models (RF, FDA, MDA, GLM and BRT) for landslides and gully erosion and three models (RF, FDA and MDA) for floods. Based on (Guzzetti et al., 2005) classification, our study shows that the three best models had predictive ability of above 0.8 (good and excellent performances). The FDA model for landslides has an AUC = 92.7% (excellent performance), RF for gully erosion has an AUC = 83.3% (good performance), and RF for flooding has an AUC = 97.2% (excellent performance).

The MHM map was produced by coupling the results of the FDA (for landslides) and RF (for gully erosion and flooding) approaches. Our work is in agreement with that of Nachappa et al. (2020) that used different machine learning approaches in multi-hazard evaluation. Also, our results in choosing RF for multi-hazard map that provides excellent accuracy for landslides AUC = 0.93 and for floods AUC = 0.97, are in agreement with the Nachappa et al. (2020) in which they applied multi-hazard evaluation use of random forest (RF), which provides a suitable result for both flood (AUC = 0.87) and landslide (AUC = 0.90), and SVM, which provides AUC = 0.87 for flood and AUC = 0.89 for landslides. They constructed the multi-hazard map using RF and SVM and produced an accurate model for planners and managers. The results show that no hazard areas cover approximately 33.5% of the study area. However, 66.5% of the total area is probably affected by one or more hazards.

Our findings indicating the application of the optimal MLTs to predict multiple problems provides crucial informative data about their interactions. These relationships are highly correlated with the scale of the indicators used in the analysis and the specific types of hazards in the area. The current study bridged the gap between the different hazards by fully identifying the cascading interactions between these different types of natural hazards.

6. Conclusions

The western and southern parts of Saudi Arabia face various natural hazards. Many mountainous areas are inhabited by people and these areas are vulnerable to several (compound) natural hazards. The delineation of high-risk areas is the critical stage and the most difficult task for all developers, and decision-makers. In the current work, we applied various machine learning techniques (MLTs) to map individual hazards affecting the Hasher - Fayfa basin (landslides, gully erosion, and flooding). The final map of this study is a multi-hazard risk map created by linking the three predominant natural hazards in the area (landslides, gully erosion, and flooding).

The multi-hazard risk map is divided into 8 zones, the no hazard zone (covers 20% of the study area). Three single hazard zones namely landslide hazard zone (3%), which is confined to high altitude and steeply sloping areas, gully erosion hazard zone (2%) which is concentrated along gullies and mainly characterized by soil and weathered materials that can be easily eroded by rainfall, and flood hazard

zone (4%) which is confined to main wadis and mainly downstream parts as water collects from various tributaries. Three zones include two types of interacting hazards, such as landslide-flood hazard zone (5%), landslide-erosion hazard zone (9%), and flood- erosion hazard zone (10%). Finally, there is a multiple hazard zone affected by the three hazard types landslide - erosion - flood, which covers 10% of the study area.

The machine learning methods achieved acceptable accuracy in predicting the different hazard types; so, the multi-hazard map was produced with a high confidence level. Due to population growth and expansion of infrastructure and urban areas to mountainous and floodplains, and to ensure sustainable development, multi-hazard assessments are essential. Our ability to produce multi-hazard map will help planners and decision-makers make concrete management decisions for existing and future developments. This will make communities more resilient and able to act proactively to minimize future damage not only caused by a single hazard type, but also that may result from cascades of hazards or combined hazards.

Declarations

Authors Contributions

Ahmed M. Youssef, Ali M. Mahdi, Mohamed M. Al-Katheri, Soheila Pouyan, Hamid Reza Pourghasemi designed the experiments, ran models, analyzed the results, and wrote and reviewed the manuscript.

Declaration of competing interest

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

References

1. Agar R (1987) The Najd fault system revisited: a two-way strike-slip orogeny in the Saudi Arabian shield. J Struct Geol 9:41–48
2. AlQahtanya AM, Abubakarb IR (2020) Public perception and attitudes to disaster risks in a coastal metropolis of Saudi Arabia. Int J Disaster Risk Reduct 44:101422. Doi: 10.1016/j.ijdrr.2019.101422
3. Alyami A, Dulong CL, Younis MZ, Mansoor S (2021) Disaster Preparedness in the Kingdom of Saudi Arabia: Exploring and Evaluating the Policy, Legislative Organisational Arrangements Particularly During the Hajj Period. Eur J Environ Public Health 5(1):em0053. https://doi.org/10.29333/ejeph/8424
4. Amare S, Langendoen E, Keesstra S, Ploeg Mvd, Gelagay H, Lemma H, Zee SEATMvd (2021) Susceptibility to Gully Erosion: Applying Random Forest (RF) and Frequency Ratio (FR) Approaches to a Small Catchment in Ethiopia. Water 13:216. https://doi.org/10.3390/w13020216

5. Baig SU, Rehman MU, Janjua NN (2021) District-level disaster risk and vulnerability in the Northern mountains of Pakistan, Geomatics. Nat Hazards Risk 12(1):2002–2022. DOI: 10.1080/19475705.2021.1944331
6. Bathrellos GD, Skilodimou HD, Chousianitis K, Youssef AM, Pradhan B (2017) Suitability estimation for urban development using multi-hazard assessment map. Sci Total Environ 575:119–134
7. Bell R, Glade T (2004) Multi-hazard analysis in natural risk assessments. WIT Trans Ecol Environ 77:1–10
8. Bout B, Lombardo L, van Westen CJ, Jetten VG (2018) Integration of two-phase solid fluidequations in a catchment model for flashfloods, debris flows and shallow slope failures. Environ Model Softw 105:1–16. DOI. 10.1016/j.envsoft.2018.03.017
9. Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn 30:1145–1159
10. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
11. Çellek S (2020) Effect of the Slope Angle and Its Classification on Landslide. Nat. Hazards Earth Syst. Sci. Discuss. https://doi.org/10.5194/nhess-2020-87, 2020
12. De Silva MMGT, Kawasaki A (2018) Socioeconomic Vulnerability to Disaster Risk: A Case Study of Flood and Drought Impact in a Rural Sri Lankan Community. Ecol Econ 152:131–140
13. Duncan M, Edwards S, Kilburn C, Twigg J, Crowley K (2016) An interrelated hazards approach to anticipating evolving risk GFDRR (Ed.), The Making of a Riskier Future: How Our Decisions Are Shaping Future Disaster Risk, Global Facility for Disaster Reduction and Recovery, Washington, USA, pp. 114–121
14. El-Haddad BA, Youssef AM, Pourghasemi HR, Pradhan 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 (2021). https://doi.org/10.1007/s11069-020-04296-y
15. Fang Z, Wang Y, Duan G, Peng L (2021) Landslide Susceptibility Mapping Using Rotation Forest Ensemble Technique with Different Decision Trees in the Three Gorges Reservoir Area, China. Remote Sens 13(2):238. https://doi.org/10.3390/rs13020238
16. Friedman JH (2001) Greedy function approximation: A gradient boosting machine. Ann Stat 29:1189–1232
17. 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(2):196. https://doi.org/10.3390/rs11020196
18. Ghorbanzadeh O, Shahabi H, Mirchooli F, Kamran KV, Lim S, Aryal J, Jarihani B, Blaschke T (2020) Gully erosion susceptibility mapping (GESM) using machine learning methods optimized by the multi–collinearity analysis and K-fold cross-validation. Geomatics Nat Hazards Risk 11(1):1653–1678. DOI: 10.1080/19475705.2020.1810138

19. Gill JC, Malamud BD (2014) Reviewing and visualizing the interactions of natural hazards. Rev Geophys 52(4):680–722. https://doi.org/10.1002/2013RG000445
20. Gill JC, Malamud BD (2017) Anthropogenic processes, natural hazards, and interactions in a multi-hazard framework. Earth Sci Rev 166:246–269. https://doi.org/10.1016/j.earscirev.2017.01.002
21. Godschall S, Smith V, Hubler J, Kremer P (2020) A Decision Process for Optimizing Multi-Hazard Shelter Location Using Global Data. Sustainability 2020, 12, 6252. https://doi.org/10.3390/su12156252
22. Goetz JN, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci 81:1–11
23. Greenwood WR (1982) Late Proterozoic Island-Arc Complexes and Tectonic Belts in the Southern Part of the Arabian Shield. Ministry of Petroleum and Mineral Resources, Riyadh, Saudi Arabia
24. Guzzetti F, Reichenbach P, Cardinali M, Galli M, Ardizzone F (2005) Probablistic Landslide Hazard Assessment at the Basin Scale. Geophys J Roy Astron Soc 72:272–299. https://doi.org/10.1016/j.geomorph.2005.06.002
25. Hasanean H, Almazroui M (2015) Rainfall: Features and Variations over Saudi Arabia. Rev Clim 3(3):578–626. https://doi.org/10.3390/cli3030578
26. Hastie T, Tibshiran R, Leisch F, Hornik K, Ripley BD (2017) Mixture and flexible discriminant analysis. (2017). https://cran.r-project.org/web/packages/mda/mda.pdf
27. Hastie T, Tibshirani R (1996) Discriminant Analysis by Gaussian Mixtures. J. R. Stat. Soc. Ser. B 1996, 58, 155–176
28. He Q, Jiang Z, Wang M, Liu K (2021) Landslide and wildfire susceptibility assessment in southeast asia using ensemble machine learning methods. Remote Sens 13:1572
29. Herrera VM, Khoshgoftaar TM, Villanustre F, Furht B (2019) Random forest implementation and optimization for Big Data analytics on LexisNexis's high performance computing cluster platform. J Big Data 6:68. https://doi.org/10.1186/s40537-019-0232-1
30. Hill DJ, Minsker BS (2010) Anomaly detection in streaming environmental sensor data: A data-driven modeling approach. Environ Model Softw 25:1014–1022
31. Hosseiny H, Nazari F, Smith V, Nataraj C (2020) A Framework for Modeling Flood Depth Using a Hybrid of Hydraulics and Machine Learning. Sci Rep 10, 8222 (2020). https://doi.org/10.1038/s41598-020-65232-5
32. Iglesias V, Braswell AE, Rossi MW, Joseph MB, McShane C, Cattau M, Koontz MJ, McGlinchy J, Nagy RC, Balch J, Leyk S, Travis WR (2021) Risky Development: Increasing Exposure to Natural Hazards in the United States. Earth's Future, https://doi.org/10.1029/2020EF001795
33. IPCC (2007) Climate change 2007. Synthesis report: Contribution of Working Groups I, II, and III to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press. Retrieved from http://www.ipcc.ch/pdf/assessment-report/ar4/wg2/ar4-wg2-ts.pdf

34. IPCC (2012) Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change. The Intergovernmental Panel on Climate Change
35. Janizadeh S, Avand M, Jaafari A, Phong TV, Bayat M, Ahmadisharaf E, Prakash I, Pham BT, Lee S (2019) Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran. Sustainability 2019, 11, 5426. https://doi.org/10.3390/su11195426
36. Jenks GF, Caspall FC (1971) Error on choroplethic maps: Definition, measurement, reduction Ann. Assoc Am Geogr 61(2):217–244
37. Kalantar B, Ueda N, Saeidi V, Ahmadi K, Halin AA, Shabani F (2020) Landslide Susceptibility Mapping: Machine and Ensemble Learning Based on Remote Sensing Big Data. Remote Sens. 2020, 12, 1737. https://doi.org/10.3390/rs12111737
38. Komendantova N, Mrzyglocki R, Mignan A, Khazai B, Wenzel F, Patt A, Fleming K K (2014) Multi-hazard and multi-risk decision-support tools as a part of participatory risk governance: feedback from civil protection stakeholders. Int J Disaster Risk Reduct 8:50–67. DOI: 10.1016/j.ijdrr.2013.12.006
39. Leonard M, Westra S, Phatak A, Lambert M, van den Hurk B, McInnes K, Risbey J, Schuster S, Jakob D, Stafford-Smith M (2014) A compound event framework for understanding extreme impacts Wiley Interdiscip. Rev Clim Chang 5:113–128. DOI:10.1002/wcc.252
40. Liu B, Han X, Qin L, Xu W, Fan J (2021) Multi-hazard risk mapping for coupling of natural and technological hazards. Geomatics Nat Hazards Risk 12(1):2544–2560. DOI: 10.1080/19475705.2021.1969451
41. 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
42. Lombardo L, Tanyas H, Nicu IC (2020) Spatial modeling of multi-hazard threat to cultural heritage sites. Eng Geol 277:105776
43. Msabi MM, Makonyo M (2021) Flood susceptibility mapping using GIS and multi-criteria decision analysis: A case of Dodoma region, central Tanzania, Remote Sensing Applications: Society and Environment, 21: 100445. https://doi.org/10.1016/j.rsase.2020.100445
44. Nachappa TG, Ghorbanzadeh O, Gholamnia K, Blaschke T (2020) Multi-Hazard Exposure Mapping Using Machine Learning for the State of Salzburg. Austria Remote Sens 12:2757. https://doi.org/10.3390/rs12172757
45. Park S, Kim J (2019) Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance.Appl Sci9:942Return to ref 2019 in article
46. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011) Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–30

47. R Core Team (2019) R: A Language and Environment for Statistical Computing. Foundation for Statistical Computing, Vienna: R [Available at https://www.R-project.org/]
48. Rahman M, Ningsheng C, Islam MM, Dewan A, Iqbal J, Washakh RMA, Shufeng T (2019) Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis. Earth Syst Environ 3, 585–601 (2019). https://doi.org/10.1007/s41748-019-00123-y
49. Razavi-Termeh SV, Sadeghi-Niaraki A, Choi S-M (2020) Gully erosion susceptibility mapping using artificial intelligence and statistical models, Geomatics. Nat Hazards Risk 11(1):821–844. DOI: 10.1080/19475705.2020.1753824
50. Ridgeway G, Southworth MH (2013) RUnit S. Package 'gbm.' Viitattu; 10:40
51. Roy J, Saha S (2019) Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. Geoenviron Disasters 6:11. https://doi.org/10.1186/s40677-019-0126-8
52. Rusk J, Maharjan A, Tiwari P, Chen T-HK, Shneiderman S, Turin M, Seto KC (2022) Multi-hazard susceptibility and exposure assessment of the Hindu Kush Himalaya. Sci Total Environ 804:150039. https://doi.org/10.1016/j.scitotenv.2021.150039
53. Rutgersson A, Kjellström E, Haapala J, Stendel M, Danilovich I, Drews M, Jylhä K, Kujala P, Guo Larsén X, Halsnæs K, Lehtonen I, Luomaranta A, Nilsson E, Olsson T, Särkkä J, Tuomi L, Wasmund N (2021) Natural Hazards and Extreme Events in the Baltic Sea region, Earth Syst. Dynam Discuss. https://doi.org/10.5194/esd-2021-13
54. Sarkar T, Mishra M (2018) Soil erosion susceptibility mapping with the application of logistic regression and artificial neural network. J Geovis Spat Anal 2(1):8
55. Sarker IH, Watters P, Kayes ASM (2019) Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. J Big Data. 2019;6(1):1–28
56. Saunders WSA, Kilvington M (2016) Innovative land use planning for natural hazard risk reduction: A consequence-driven approach from New Zealand. Int J Disaster Risk Reduct 18:244–255. https://doi.org/10.1016/j.ijdrr.2016.07.002
57. Schmidt J, Matcham I, Reese S, King A, Bell R, Henderson R, Smart G, Cousins J, Smith W, Heron D (2011) Quantitative multi-risk analysis for natural hazards: a framework for multi-risk modelling. Nat. Hazards 58, 1169–1192 (2011)
58. Şen Z, Al-Harithy S, As-Sefry S, Almazroui M (2017) Aridity and Risk Calculations in Saudi Arabian Wadis: Wadi Fatimah Case. Earth Syst Environ 1, 26 (2017). https://doi.org/10.1007/s41748-017-0030-x
59. Shah AA, Khwaja S, Shah BA, Reduan Q, Jawi Z (2018) Living With Earthquake and Flood Hazards in Jammu and Kashmir, NW Himalaya. Front Earth Sci 6:179. https://doi.org/10.3389/feart.2018.00179
60. Shalaby A, Tateishi R (2007) Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. Appl Geogr. 27 28 – 41 https://doi.org/10.1016/j.apgeog.2006.09.004

61. Shin Y, Kim T, Cho H, Kang KI (2012) A formwork method selection model based on boosted decision trees in tall building construction. Autom Constr 23:47–54
62. Skilodimou HD, Bathrellos GD, Chousianitis K, Youssef AM, Pradhan B (2019) Multi-hazard assessment modeling via multi-criteria analysis and GIS: a case study. Environ Earth Sci 78, 47 (2019). https://doi.org/10.1007/s12665-018-8003-4
63. Solberg AHS (1996) Texture fusion and classification based on flexible discriminant analysis. In Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, 25–29 August; 596–600
64. Stoeser DB, Camp VE (1985) Pan-African microplate accretion of the Arabian shield. Geol Soc Amer Bull 96:817–826
65. Sun D, Wen H, Wang D, Xu J (2020) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. Geomorphology 362:107201. https://doi.org/10.1016/j.geomorph.2020.107201
66. Tabari H (2020) Climate change impact on flood and extreme precipitation increases with water availability. Sci Rep 10:13768. https://doi.org/10.1038/s41598-020-70816-2
67. Van Westen CJ, Greiving S (2017) Multi-hazard risk assessment and decision making. Environmental Hazards Methodologies for Risk Assessment and Management. IWA publishing, London, UK, pp 31–94
68. Wang Y, Fang Z, Hong H, Peng L (2020) Flood susceptibility mapping using convolutional neural network frameworks. J Hydrol 582:124482. https://doi.org/10.1016/j.jhydrol.2019.124482
69. Wang Y, Fang Z, Wang M, Peng L, Hong H (2020) Comparative study of landslide susceptibility mapping with different recurrent neural networks. Comput Geosci 138:104445. doi: 10.1016/j.cageo.2020.104445
70. Ward PJ, Blauhut V, Bloemendaal N, Daniell JE, de Ruiter MC, Duncan MJ, Emberson R, Jenkins SF, Kirschbaum D, Kunz M, Mohr S, Muis S, Riddell GA, Schäfer A, Stanley T, Veldkamp TIE, Winsemius HC (2020) Review article: Natural hazard risk assessments at the global scale. Nat Hazards Earth Syst Sci 20:1069–1096. https://doi.org/10.5194/nhess-20-1069-2020
71. Wastl M, Stötter J, Kleindienst H (2011) Avalanche risk assessment for mountain roads: A case study from Iceland. Nat Hazards 56(2):465–480
72. Weinkle J, Landsea C, Collins D, Musulin R, Crompton RP, Klotzbach PJ, Pielke R (2018) Normalized hurricane damage in the continental United States 1900–2017. Nat Sustain 1(12):808–813. https://doi.org/10.1038/s41893-018-0165-2
73. Yang A, Wang C, Pang G, Long Y, Wang L, Cruse RM, Yang Q (2021) Gully Erosion Susceptibility Mapping in Highly Complex Terrain Using Machine Learning Models. ISPRS Int J Geo-Information 10(10):680. https://doi.org/10.3390/ijgi10100680
74. Ye J, Chow J-H, Chen J, Zheng Z (2009) Stochastic gradient boosted distributed decision trees, in Proceedings of the ACM 18th International Conference on Information and Knowledge Management (CIKM '09), 2061–2064, Hong Kong, November 2009

75. Ye T, Liu W, Mu Q, Zong S, Li Y, Shi P (2020) Quantifying livestock vulnerability to snow disasters in the Tibetan Plateau: Comparing different modeling techniques for prediction. Int J Disaster Risk Reduct 48101578. https://doi.org/10.1016/j.ijdrr.2020.101578
76. Yoo C, Cho E (2019) Effect of Multicollinearity on the Bivariate Frequency Analysis of Annual Maximum Rainfall Events. Water 11:905. https://doi.org/10.3390/w11050905
77. Zhang C, Xie Z (2012) Combining object-based texture measures with a neural network for vegetation mapping in the Everglades from hyperspectral imagery. Remote Sens Environ 124:310–320
78. 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. doi: 10.1016/j.scitotenv.2017.10.037
79. Zhou X, Wen H, Zhang Y, Xu J, Zhang W (2021) Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. Geosci Front 12(5):101211. https://doi.org/10.1016/j.gsf.2021.101211

Figures

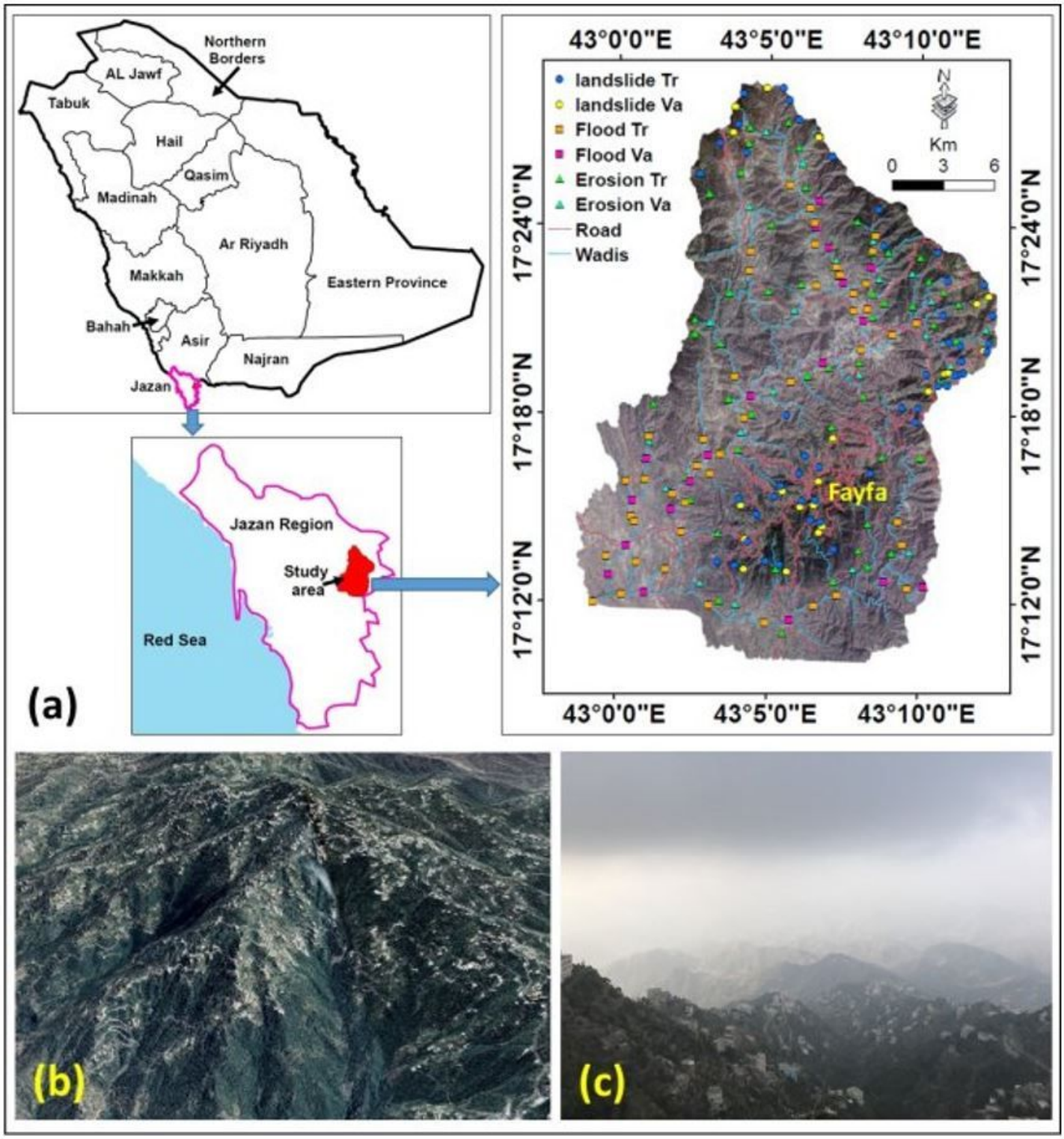

Figure 1

Location of Hasher-Fayfa Basin in Jazan province, KSA, with training and validation locations of landslide, flood, and gully erosion, b, c) 3D perspective from Google Earth and photograph by the author (Ahmed M. Youssef), respectively, showing the densely populated area in Fayfa Mountain.

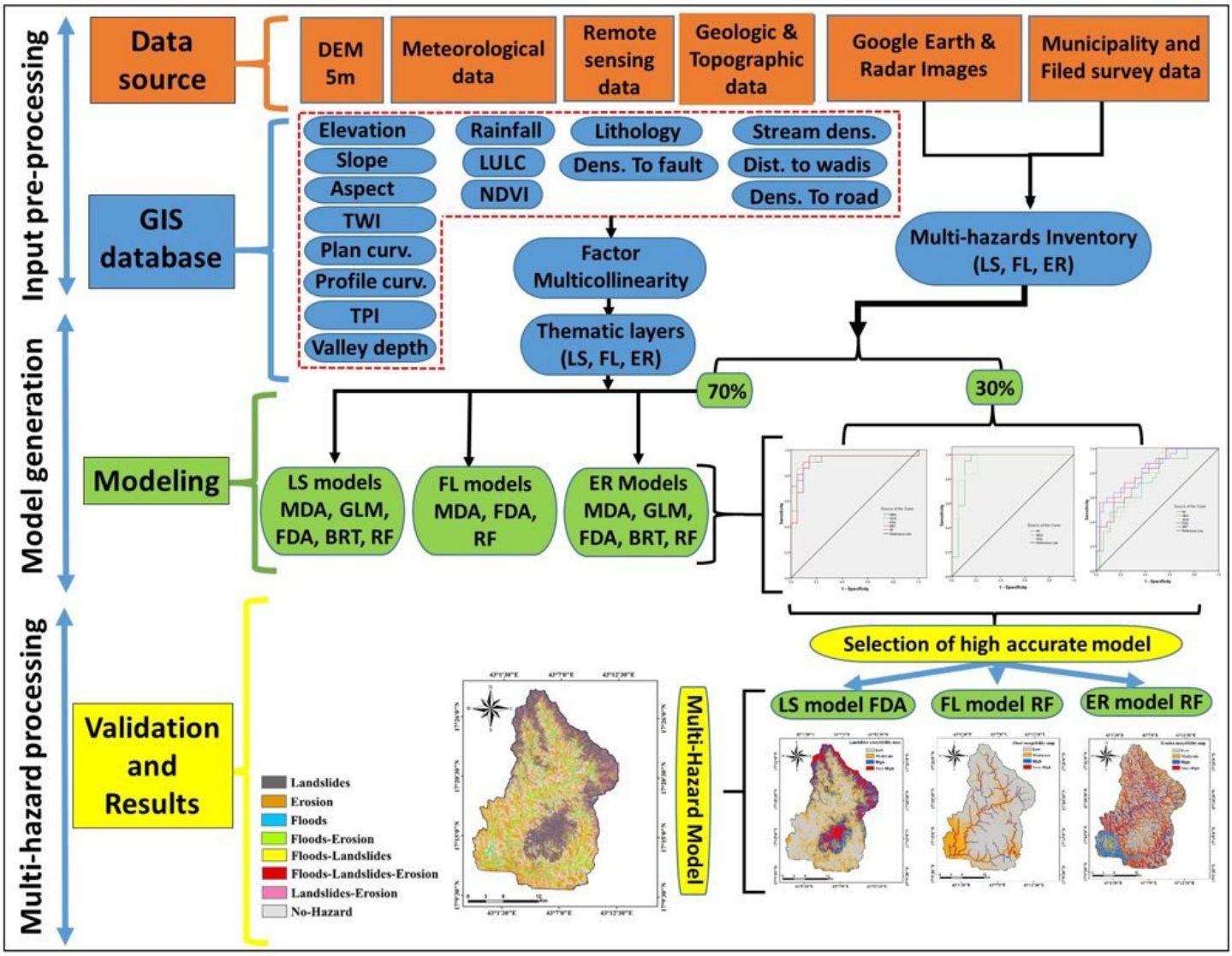

Figure 2

Flowchart showing the sequential steps of this study.

Figure 3

Photographs showing the three hazards collected between 2008 and 2020 in the study area, landslides (a, b, c); gully erosion (d, e, f); and flash floods (g, h). They were taken by Ahmed M. Youssef (first author).

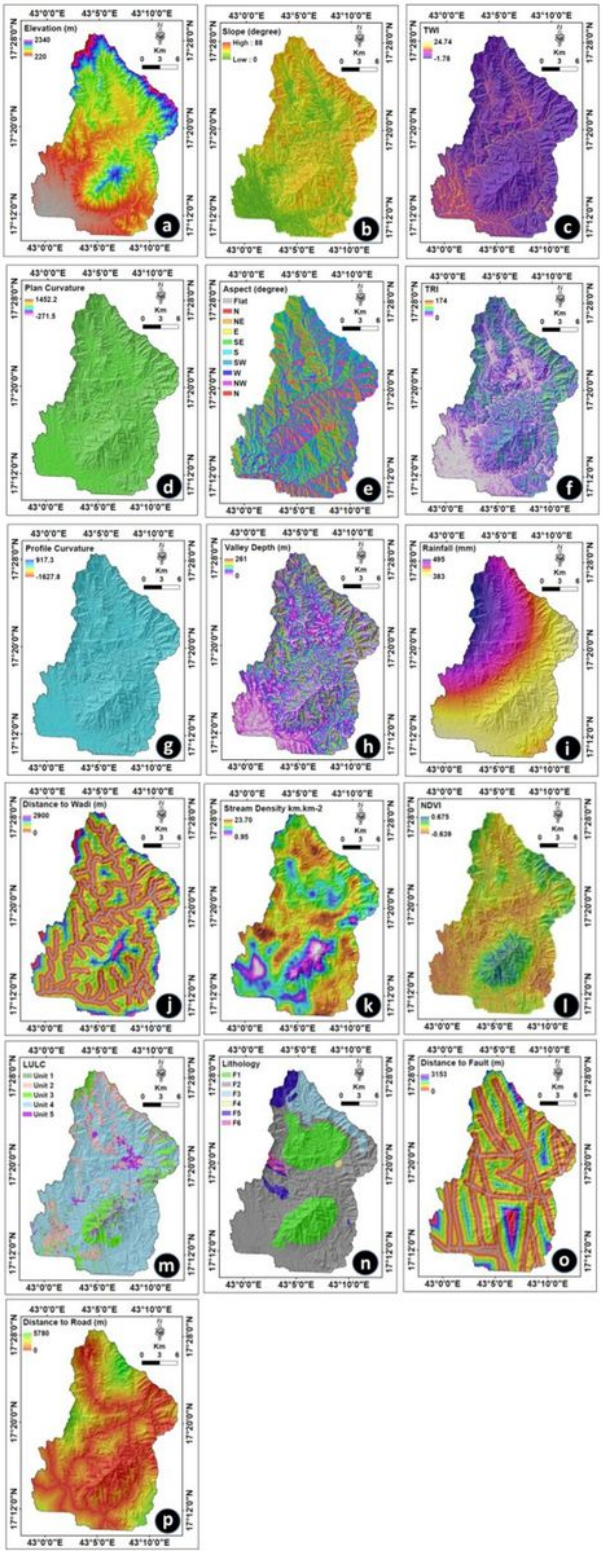

Figure 4

Predictive factors used to construct multi-hazard model: (a) elevation, (b) slope, (c) TWI, (d) plan curvature, (e) aspect, (f) TRI, (g) profile curvature, (h) valley depth, (i) rainfall, (j) distance to wadis, (k) stream density, (l) NDVI, (m) LULC, (n) lithology, (o) distance to fault, and (p) distance to road.

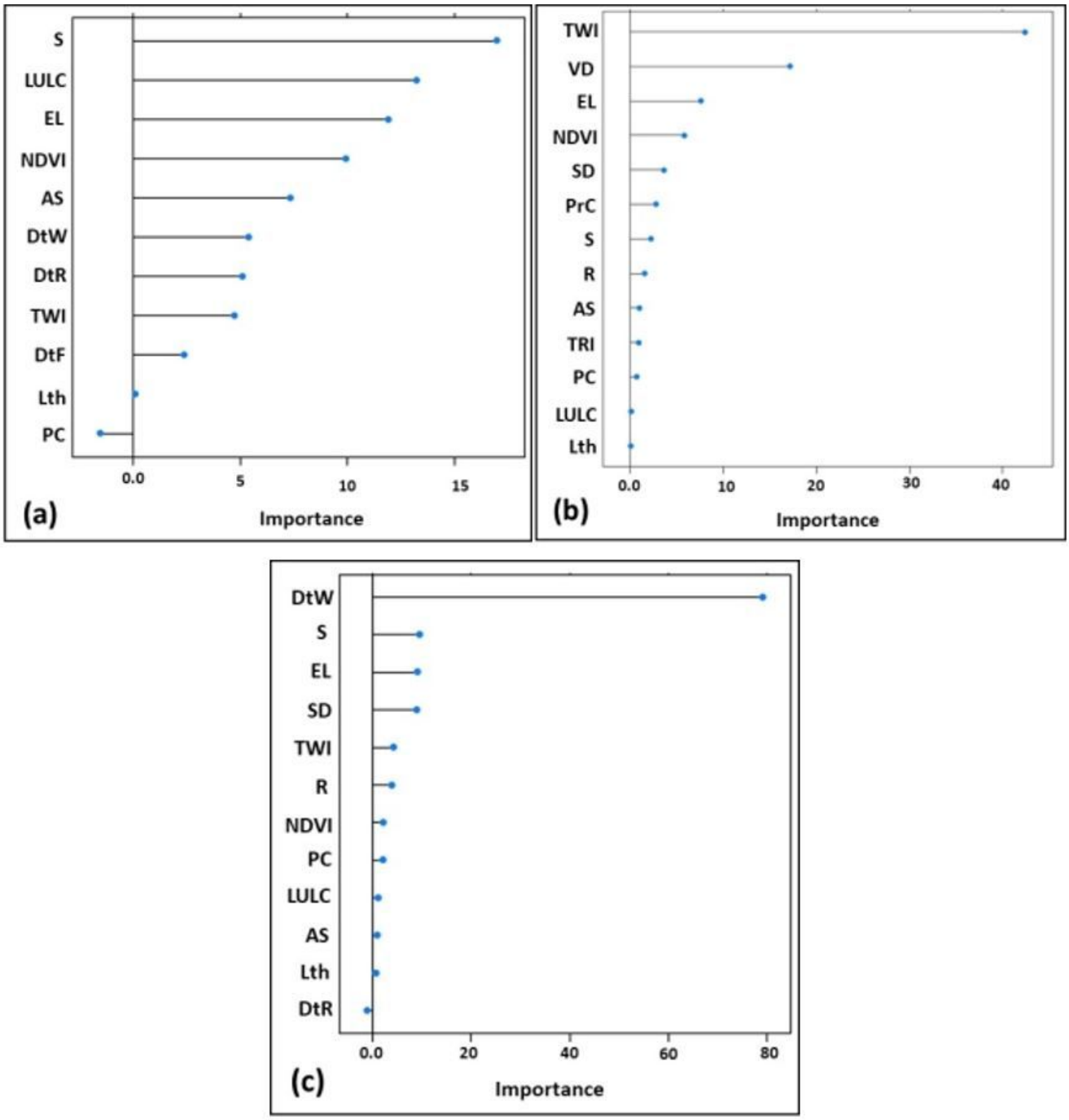

Figure 5

Hazard-predictive factors importance EL; S; TWI; PC, A, TRI, PrC, VD, R, DtW, SD, NDVI, LULC, Lth, DtF, and DtR) for a) landslide; b) gully erosion; and c) flood.

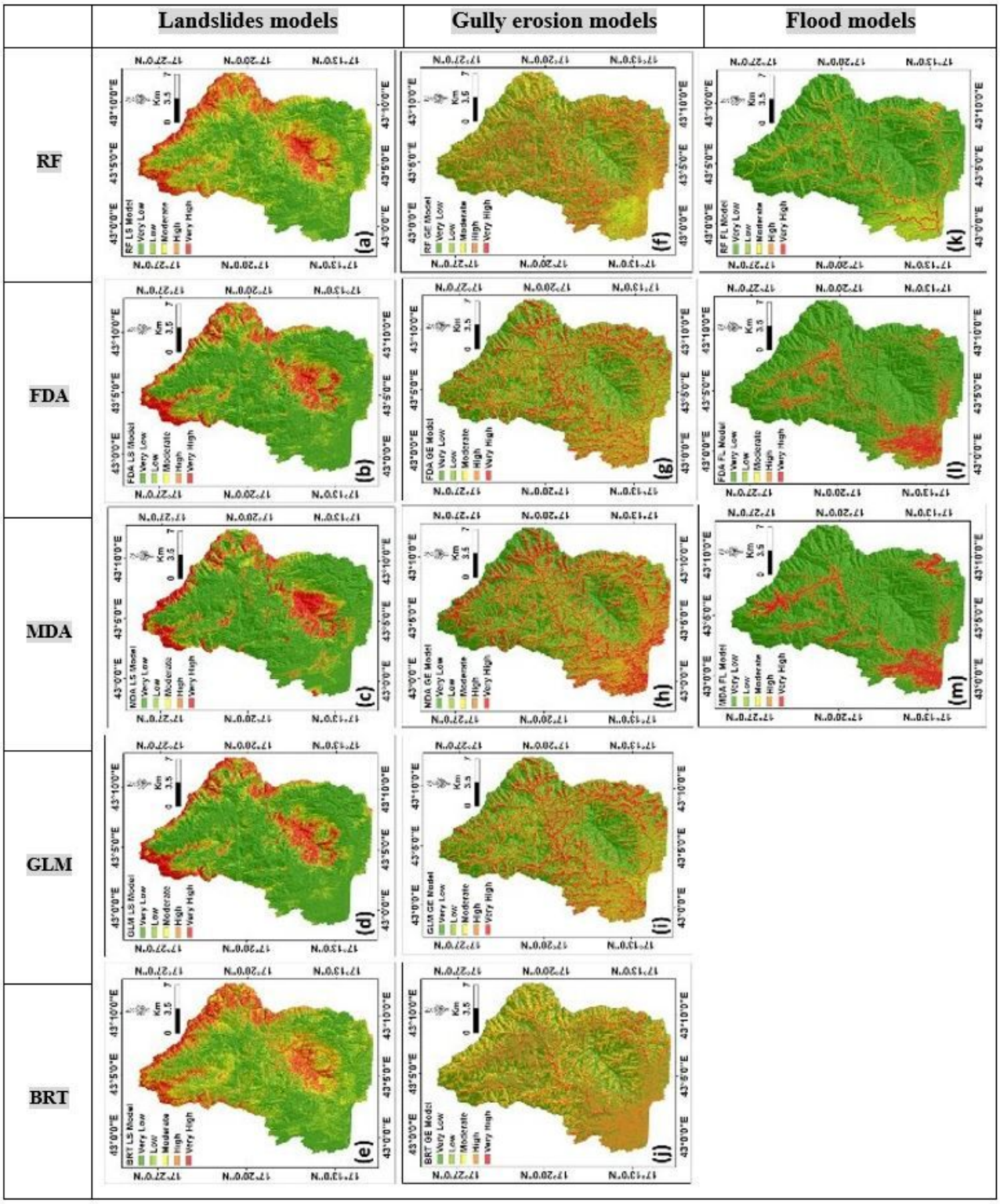

Figure 6

susceptibility maps for landslide, gully erosion, and flood using MLMs.

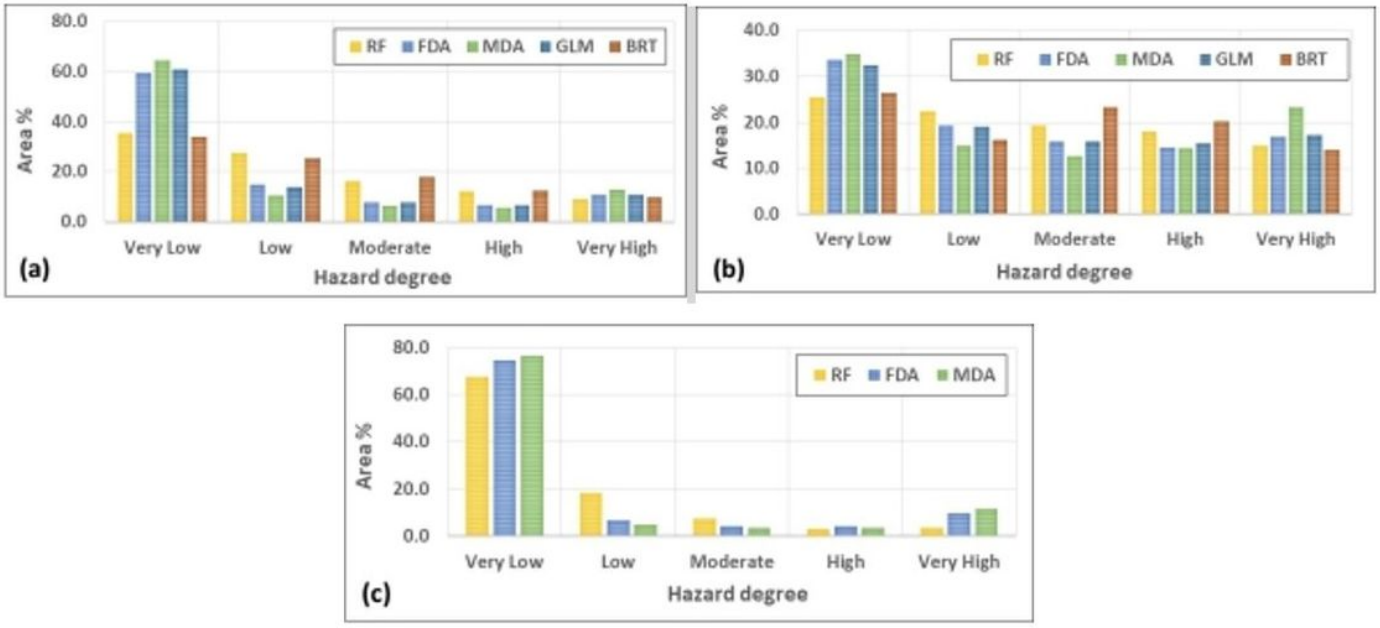

Figure 7

Area percentage of different hazard classes for (a) landslide, (b) gully erosion, and (c) flood.

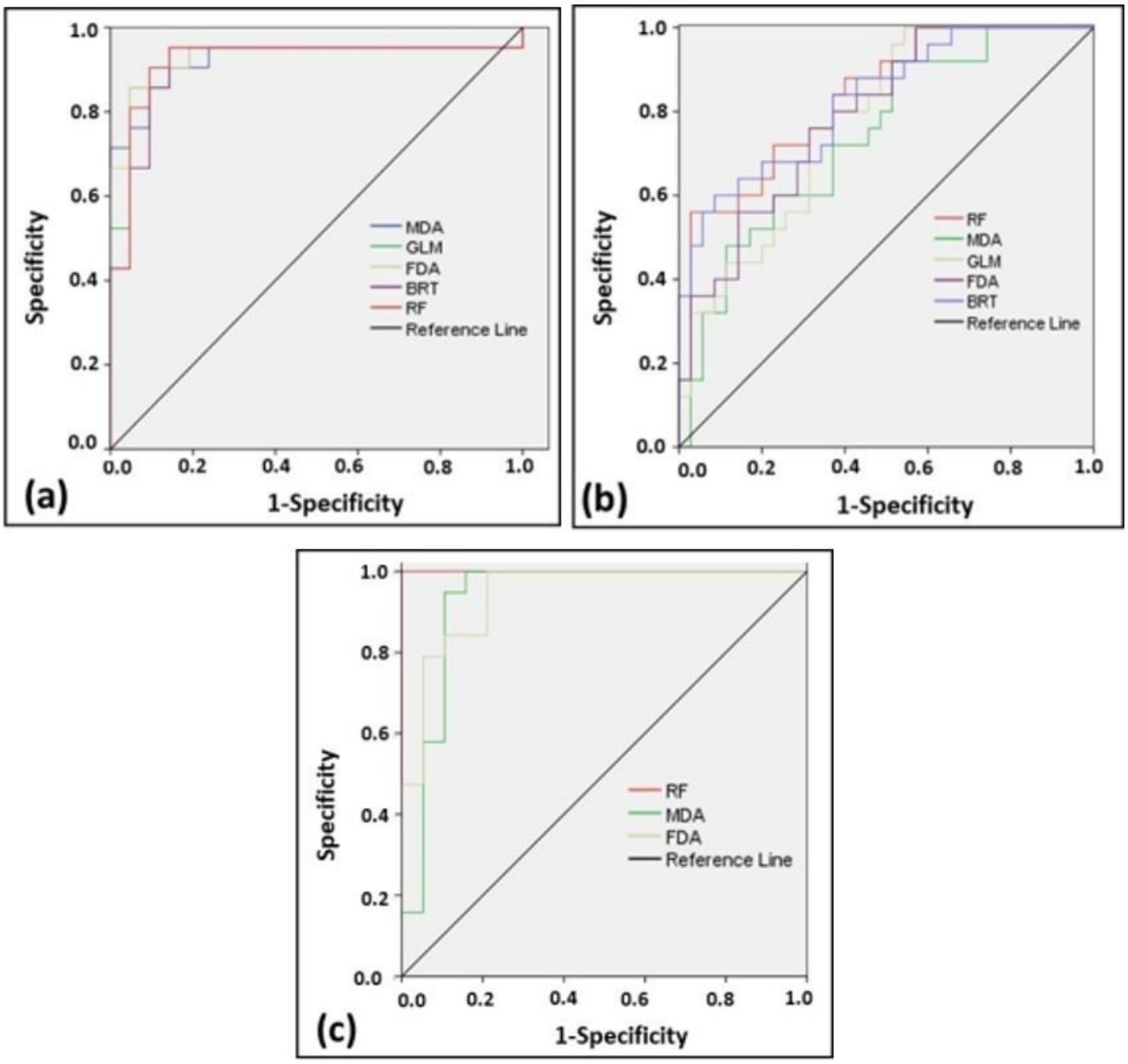

Figure 8

Accuracy assessment of different hazard models for (a) landslides, (b) gully erosion, and (c) floods

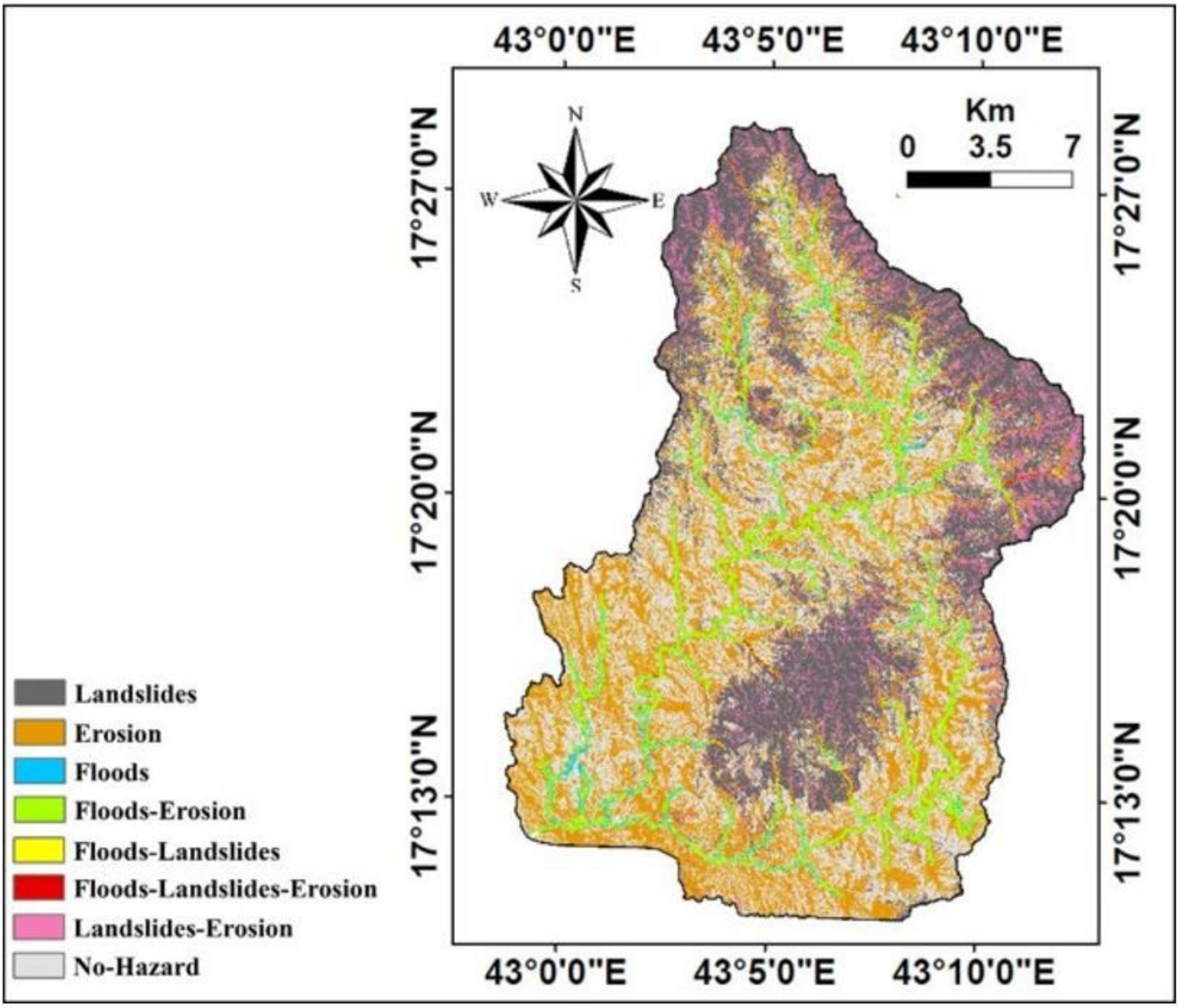

Figure 9

The multi-hazard risk map was constructed by integrating of highly accurate models of FDA for landslide, RF for gully erosion, and RF for floods.

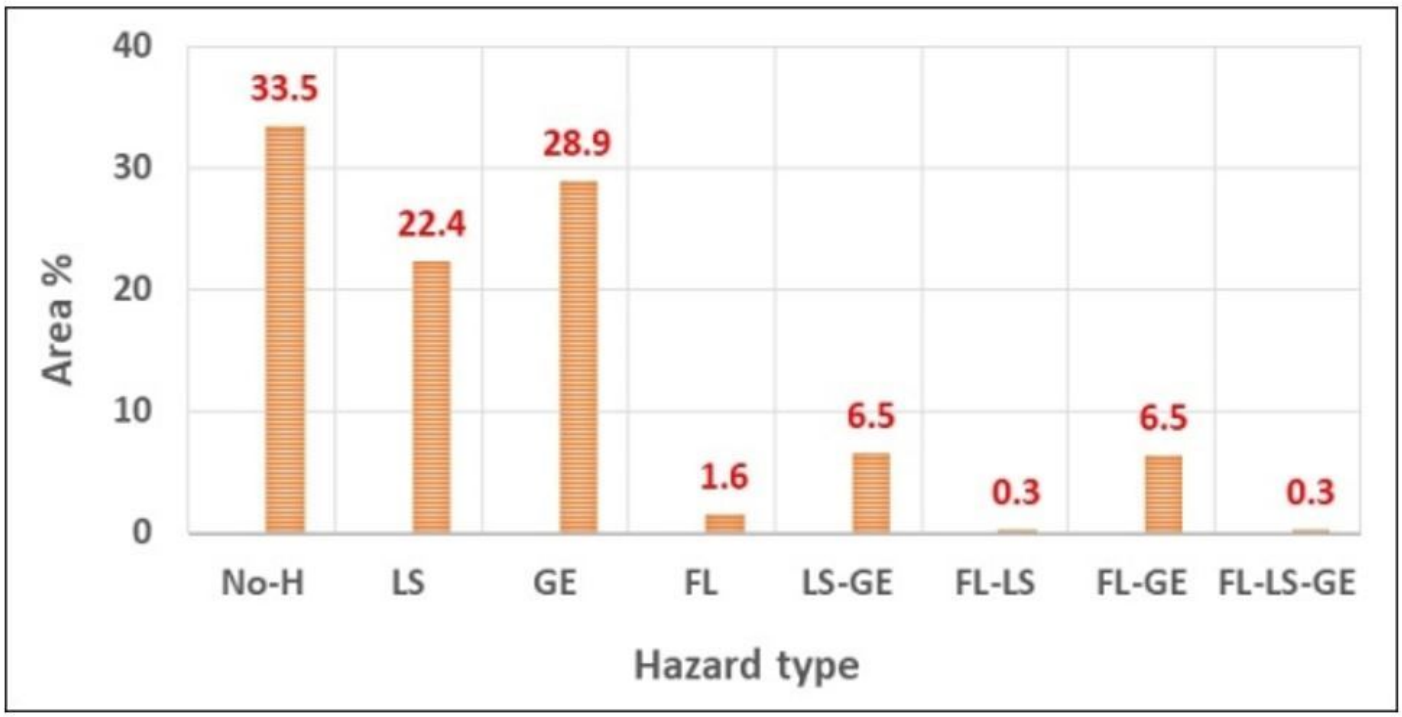

Figure 10

Different types of hazard percentage in the study area. No-H = No hazard, LS = landslide, GE = gully erosion, FL = flood, LS-GE = landslide-erosion, FL-LS = flood-landslide, FL-GE = flood-gully erosion, and FL-LS-GE flood-landslide-gulley erosion.