Environmental Earth Sciences (2018) 77:249
https://doi.org/10.1007/s12665-018-7434-2

ORIGINAL ARTICLE

Susceptibility mapping of gully erosion using GIS-based statistical bivariate models: a case study from Ali Al-Gharbi District, Maysan Governorate, southern Iraq

Alaa M. Al-Abadi[1] · **Ali K. Al-Ali**[1]

Received: 9 October 2017 / Accepted: 19 March 2018 / Published online: 23 March 2018

Abstract
This work aims to evaluate the predictive capability of three bivariate statistical models, namely information value, frequency ratio, and evidential belief functions, in gully erosion susceptibility mapping in northeastern Maysan Governorate (Ali Al-Gharbi District) in southern Iraq. The gully inventory map, consisting of 21 gullies of different sizes, was prepared based on the interpretation of remotely sensed data supported by field survey. The gully inventory data (polygon format) were randomly partitioned into two sets: 14 gullies for build and training the bivariate model, and the remaining 7 gullies for validating purposes. Twelve gully influential factors were selected based on data availability and the literature review. The selected factors were related to lithology, geomorphology, soil, land cover, and topography (primary and secondary) settings. Analysis of factor importance using information gain ratio proved that out of 12 gully influential factors, eight were of more importance in developing gullies (the average merit was greater than zero). The most important factors and the training gully inventory map were used to generate three gully erosion susceptibility maps based on the three bivariate models used. For validation, the area under the operating characteristics curves for both success and prediction rates was used. The results indicated that the highest prediction rate of 82.9% was achieved using the information value technique. All the bivariate models had prediction rates greater than 80%, and thus they were regarded as very good estimators. The final conclusion was that the bivariate models offer advanced techniques for mapping gully erosion susceptibility.

Keywords Gully erosion · Bivariate statistical models · Information gain ratio · Maysan Governorate · Iraq

Introduction

Gully erosion is an erosive process that significantly contributes to shaping the earth's surface and plays a major role in the degradation of land and soil loss (Billi and Dramis 2003). Gully erosion is regarded as a potentially destructive process that poses a threat to life and property (Ionita et al. 2015) and is widely accepted now as an indicator of desertification and land degradation worldwide (Poesen et al. 2003). A gully is a relatively deep eroded channel formed by concentrated surface flow in narrow flow paths resulting from removing soil and parent material and may grow into gullies deeper than 13 cm over short-time periods (USDA-SCS 1966; Luffman et al. 2015). Gully forming is a complex natural process controlled by factors mainly related to topography, soil, land use/land cover (LULC), and climate. The increasing interest in the study of gully erosion is derived from the need to increase our understanding of its impact and thus attempt to control its adverse effects. To control and prevent the adverse effects of gully erosion, the spatial extent of the gullies and the factors that control their formation should be evaluated (Le Roux and Sumner 2012).

In this context, various methods have been reported for modeling gully erosion, and they can be classified into two major categories: quantitative and qualitative. The quantitative methods describe the process of detachment, transportation, and deposition of the eroded soil in mathematical equations that can be solved analytically or numerically to quantify soil erosion rate. The chemicals, runoff, and erosion from agricultural management system (CREAMS) (Knisel 1980), ephemeral gully erosion (EGEM) (Capra et al. 2005; Merkel et al. 1988), and the linear erosion package of the

✉ Alaa M. Al-Abadi
alaaatiaa@gmail.com

[1] Department of Geology, College of Science, University of Basra, Basra, Iraq

Water Erosion Prediction Project (WEPP) (Flanagan and Nearing 1995) are good examples of these models. The main limitations of these models are in two points: They are incapable of predicting spatial distribution of gullies which is an important tool in erosion control planning (Conoscenti et al. 2013), and they require input parameters with high accuracy that are often unavailable or difficult to assess (Akgün and Türk 2011). With the advent of geographic information systems and tremendous developments in acquiring valuable remotely sensed data, empirical models (a special case of quantitative-based models) have become more sophisticated in their approach to quantifying soil erosion rate and studying the hazards associated with it. The empirical models estimate erosion rate by combining a prefixed set of physical parameters, such as rainfall erosivity and soil erodibility, based on certain standard coefficients or procedures. The Universal Soil Loss Equation (USLE) (Wischmeier and Smith 1965) and its derivatives are examples of these models.

The second category of methods, namely the qualitative, is mainly used to evaluate the soil erosion risk or assess the erosion susceptibility at different scales (Conoscenti et al. 2014). Susceptibility is a term that is used to identify the probability of spatial occurrence of a phenomenon by defining the relationship between a set of physical factors (predictor variables) and the distribution of the events in the past (response variables) (Lucà et al. 2011). Therefore, the calculated probability values can be used to generate a map of gully erosion susceptibility that demonstrates the spatial proneness of an area to this process. In this context, the statistical bivariate and multivariate models are considered to be the most common models to assess gully erosion susceptibility (Meyer and Martínez-Casasnovas 1999; Akgün and Türk 2011; Conforti et al. 2011; Lucà et al. 2011; Conoscenti et al. 2013, 2014; Rahmati et al. 2016a, b; Conoscenti et al. 2017). In bivariate statistical models, such as statistical index (SI), frequency ratio (FR), and weights of evidence (Wof), each factor that is affecting gully forming is analyzed individually in a straightforward and efficient way (Suzen and Doyuran 2004). Conversely, in multivariate statistical models, such as logistic regression (LG), the relationship between influencing gully-forming factors is analyzed together to study the occurrence of gullies. The response variable in LG is binary (i.e., presence or absence of gullies), and the resulting values of the model are probability values (Lucà et al. 2011) that are usually easy to interpret. In addition, the predictor variables can be numeric, categorical, ordinal, nominal, or a combination of these and it is unnecessary for them to be normally distributed.

In light of successful application of bivariate models—besides their simplicity and ease of implementation in a GIS environment—three bivariate statistical models, namely information value (InVal), frequency ratio (FR), and evidential belief functions (EBF), are used here for assessing proneness to gully erosion in Ali Al-Gharbi District, Maysan Governorate, Republic of Iraq. The northeastern part of the considered area (close to the Iraq–Iran border) is typically prone to the gully creation process that affects different lithological and soil types. The prediction performance of the bivariate models used was assessed and compared to identify the best model using the receiver operating characteristics curve (ROC) technique for both success (training) and prediction (testing) data. Specifically, the data processing and modeling were carried out using various software packages such as ArcGIS 10.2, Weka 3.8, and IDRISI Selva 17.

Description of the study area

The study area is a part of Ali Al-Gharbi District that is located in northern Maysan Governorate (Fig. 1) and covers an area of 1576 km^2. The considered area is bounded on the northeast by the Iraq–Iran border and on the southwest by the Tigris River (outside of the considered area). The surface elevation ranges from 2 to 144 m. The climate of the study area is generally arid to semiarid and is basically characterized by two distinct seasons, with hot and dry summers and cold and wet winters. The minimum and maximum monthly average temperatures for the period 1995–2015 were 12.75 and 38.5 °C, respectively. The average monthly recorded temperature is 26 °C. The monthly average minimum and maximum rainfall totals are 0 and 62 mm. The rainy period extends from October till May. Rainfall in the remaining months of May to October is rather low. Generally, the eastern parts (hills) of the considered area receive greater rainfall than the western parts (plains). The average mean of annual rainfall is 212 mm/year (Al-Abadi et al. 2016).

From the geological point of view, the exposed rocks in the study area belong to the Quaternary sediments (Pleistocene and Holocene Epochs) (Buday and Jassim 1987) (Fig. 2). The Quaternary sediments are unconsolidated and usually finer-grained than the underlying Mukdadiya and Bai Hassan Formations (Bellen et al. 1959). The major lithological units of the Quaternary deposits are alluvial fan, flood plain, depression fill, and eolian (Al-Abadi 2012) (Table 1). Tectonically speaking, Iraq is divided into three distinct areas: stable shelf, unstable shelf, and Zagros suture (Jassim and Goff 2006). Most of the study area is located within the eastern part of the Mesopotamian zone within the stable shelf and bounded on the northeast by the high folded zone (Himmren Hills). Regarding the geomorphology, the study area is a flat and featureless surface bounded by the foothill zone in the northeast (Iraq–Iran border). The trend of these hills is southeastern and parallel to the Zagros Mountains. Generally, hills are not continuous ridges, but a series of ridges

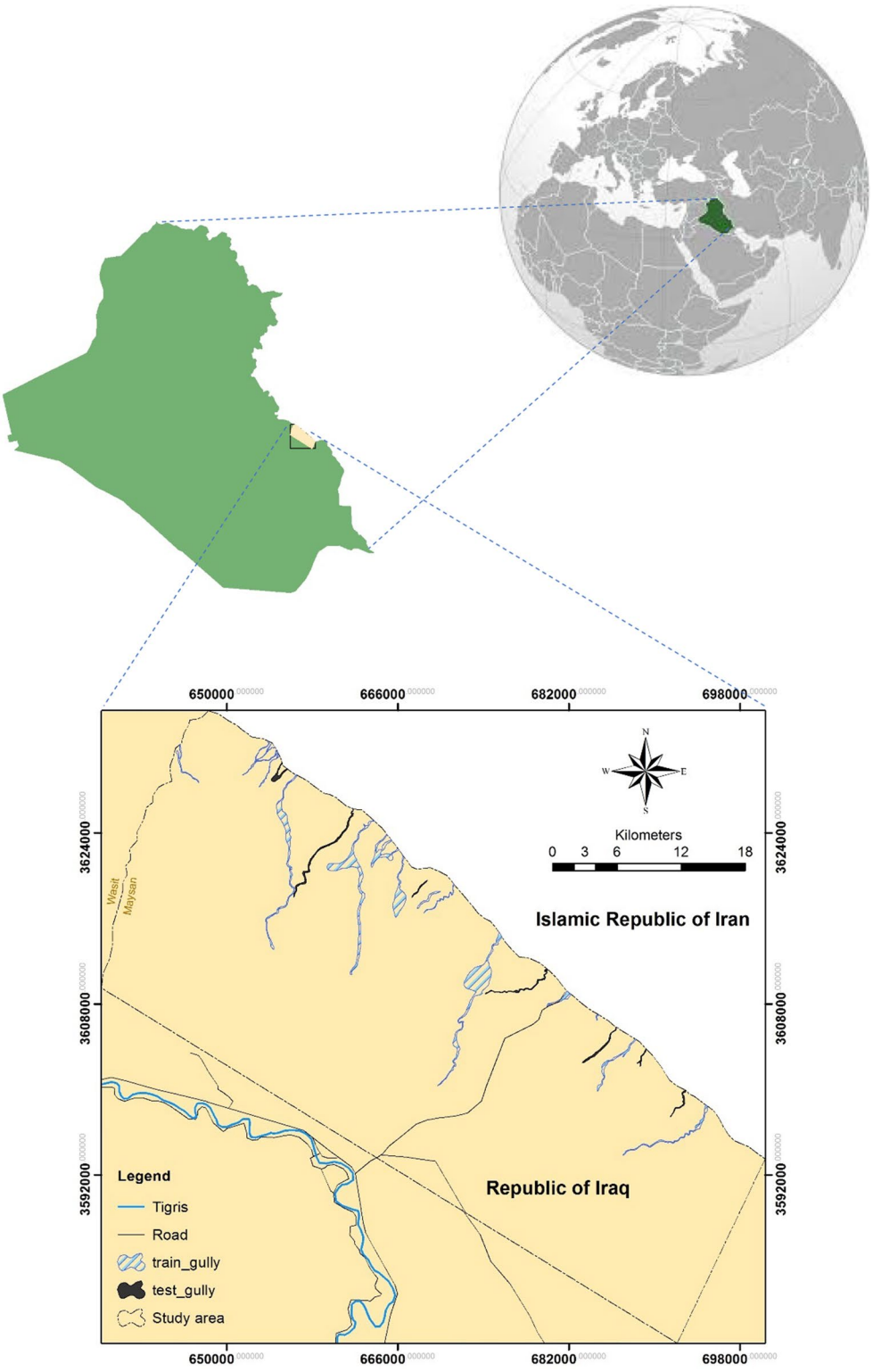


Fig. 1 Geographical location of the study area

with their axes oriented in a northwest–southeast direction (Al-Abadi et al. 2016). The upper parts of the hills are generally exposed to weathering and the erosion process, and thus a thick soil layer of fine texture has formed over them. The dominant landforms within the study areas are sand dunes, glacis, alluvial fan, floodplains, and depressions (Fig. 3). More detail on these landforms can be found in Al-Abadi (2012). Two types of soil found in the study area are mollisols and aridisols. In the examination area, the principle erosive process that influences the landscape

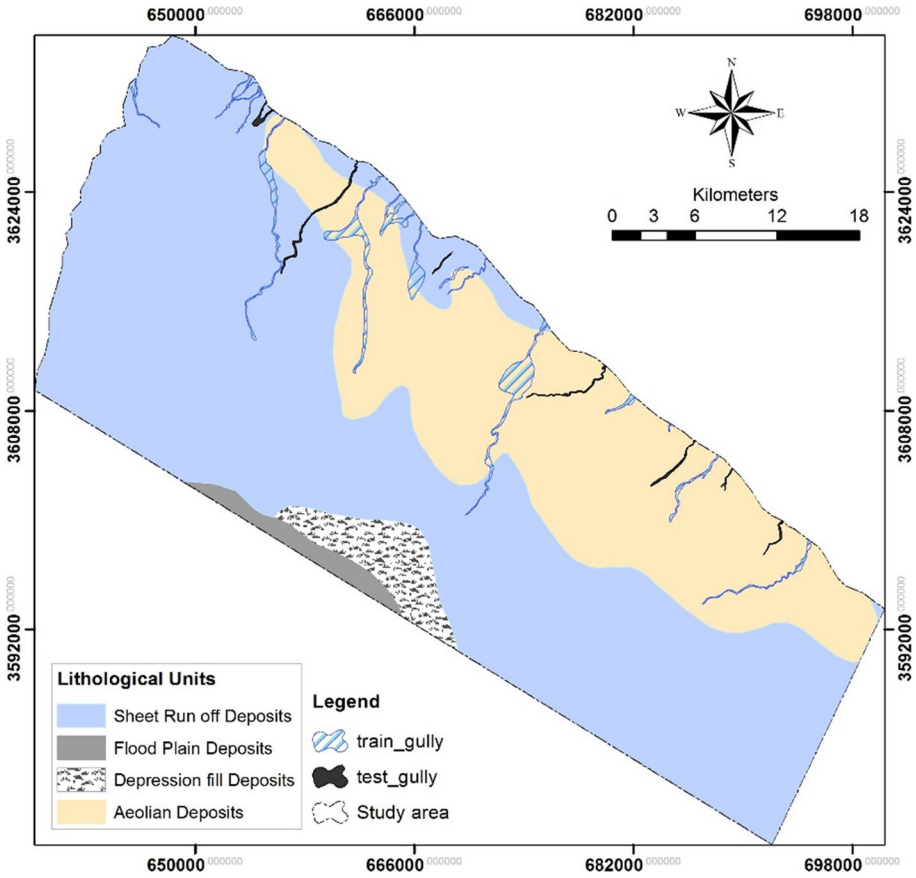


Fig. 2 Exposed lithological units in the study area

Table 1 Lithological units in the study area. (Summarized after Al-Abadi 2012)

Lithological units	Description	Age	Occupied area (km^2)	Are (%)
Alluvial	These deposits mainly comprise gravel, sand, and silty sand with a maximum thickness of 15 m. Poorly sorted cabbles and boulder occur in apical part passing into finer-grained better sorted layered fluvial deposits. Gypcrete also developed on the exposed surface of the fan	Pleistocene–Holocene	1018	65.0
Flood plain	Comprise layers of silt clay and clay (10–20 cm) and sometimes up to 1 m thick	Holocene	22	1.40
Depression fill	Comprise reddish brown fine sand, silt, and clayey silt	Holocene	48	3.10
Eolian	Comprise silty reddish brown and calcareous sand	Holocene	488	31.0

is identified with overland flow, which can take the form of a flash flood coming from Iranian territory during heavy rainfall events.

Methodology

The methodology adapted in this study (Fig. 4) consisted of four steps: (1) preparing a gully erosion inventory map, (2) selecting and preparing gully erosion influential factors, (3) applying bivariate models and creating gully erosion susceptibility maps, and (4) validating gully erosion susceptibility maps using the receiver operating characteristic (ROC) curve and identifying the best one according to the area under the ROC curve (AUC).

Gully landform inventory map

The preparation of a gully inventory is the first step in developing a map of gully erosion susceptibility. With the aid of interpretation of remotely sensed data, supported by an extensive field survey that was carried out in 2016, 21 gully areas were identified and mapped (Fig. 1). The identified gully areas encompass 26.74 km^2 (1.7%) of the total study area. To build bivariate models, these 21 gullies were randomly partitioned into two sets: 14 gullies for training and

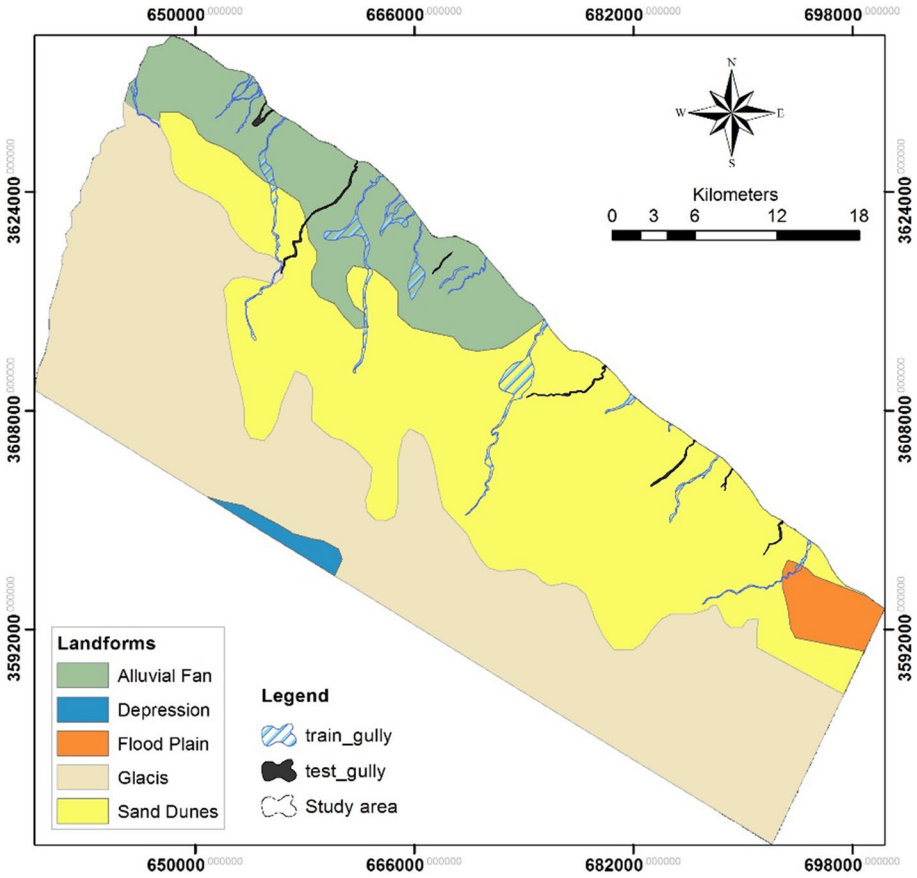


Fig. 3 Major landforms distributed in the study area

the remaining 7 gullies for testing and validating the bivariate models (Fig. 1). In the considered area, gullies have linear plan forms and their length ranges from a few meters to thousands of meters with an average depth of 1.75 m. In general, the cross sections of these gullies are U shaped although sometimes V-shaped ones are found, too.

Selecting and preparing gully erosion influential factors

The second step in developing a map of soil erosion susceptibility of an area was to select the gully erosion influential factors. Basically, the proneness of gully erosion is a function of two properties: the erodibility of the soil material and the erosivity of overland flow (Conoscenti et al. 2008). These properties are essentially related to the lithology, topography, climate, soil, and land use/land cover (LULC) of the considered area. For this work, 12 factors were selected based on the literature review (Svoray and Markovitch 2009; Conforti et al. 2011; Akgün and Türk 2011; Lucà et al. 2011; Conoscenti et al. 2013, 2014; Shit et al. 2015), availability of data, and knowledge of the study area. The selected factors were first these topographic primary factors: altitude, slope, plan curvature, profile curvature, and aspect. Secondary topographic factors were length–slope (LS), stream power index (SPI), and topographic wetness index (TWI). Additional factors were lithology, geomorphology, soil texture, and LULC. The selected factors are well known and comprehensively discussed in the literature for studying gully erosion susceptibility. All the topographic-related factors were derived from Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) with a 1 arc spatial resolution (approximately 30 × 30 m). Two tiles of SRTM were downloaded from the official site of the US Geological Survey (USGS) website (https://earthexplorer.usgs.gov/), merged, sinks filled, and reprojected to Universal Transverse Mercator (UTM)-projected coordinate system with (38 N WGS 1984) datum. Altitude was directly derived from SRTM DEM and classified into four classes (McDonald et al. 1990): < 9 m (plains), 9–30 m (rises), 30–90 m (hills), and 90–300 m (hills) (Fig. 5a). The other primary topographic factors—i.e., slope, plan curvature, profile curvature, and aspect—were also derived from SRTM DEM using the Spatial Analyst extension in ArcGIS 10.2 software. In this context, slope was classified into five classes according to de Winnaar et al. (2007): < 2% (flat), 2–8% (undulating), 8–15% (rolling), 15–30% (hilly), and > 30% (mountainous) (Fig. 5b). These classes are distributed unevenly through the study area, with the major part of the study area (98%) belonging to the flat and undulating classes. The plan

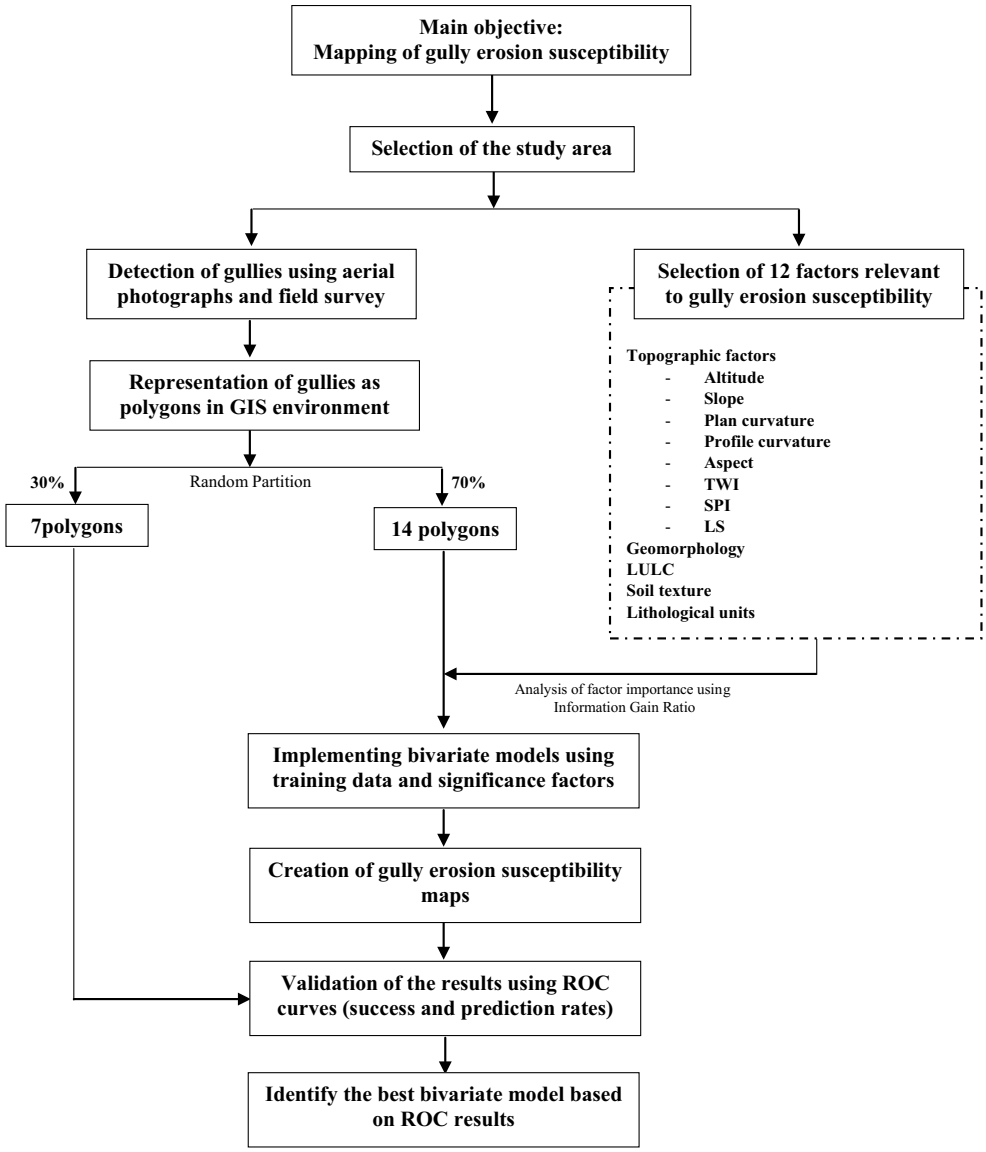


Fig. 4 Flowchart showing the methodology followed in this study

curvature was classified into three classes: < 0 (concave), 0 (flat), and > 0 (convex) (Fig. 5c), while profile curvature was also classified into three classes: < − 0.001, − 0.001 to 0.001, and > 0.001 (Fig. 5d). In the same way, aspect was also created from SRTM DEM and classified into nine classes: flat, north, northeast, east, southeast, south, southwest, west, and northwest (Fig. 5e). The secondary topographic factors—i.e., TWI, SPI, and LS—were created from SRTM DEM using the Raster Calculator and ArcHydro tools in ArcMap 10.2 software. The following equations were applied to create these factors (Moore et al. 1991; Conforti et al. 2011)

$$\mathrm{TWI} = \ln\left(\frac{\mathrm{As}}{\tan\theta}\right), \tag{1}$$

$$\mathrm{SPI} = \mathrm{As} \times \tan\theta, \tag{2}$$

$$\mathrm{LS} = (\mathrm{fa} \times \text{cell size}/22.13)^{0.4} \times (\sin\theta/0.0896)^{1.3}, \tag{3}$$

where As is the specific catchment area (m), θ is the slope gradient in degrees, fa is the flow accumulation layer derived using the ArcHydro tool, and cell size is the resolution of used DEM (here 30 m). All these factors were classified into five categories (Fig. 5f–h).

The lithological and geomorphological maps of the study area were derived from the hard copies of these maps that were provided by the Iraqi Geological Survey with 1: 1,000,000 scale (Figs. 2 and 3). To prepare soil texture map, a total of 25 soil samples were collected and analyzed using Sizer Master Instrument in the Department of Geology, College of Science, University of Basra. The soil samples were assigned appropriate texture name using the triangle of US

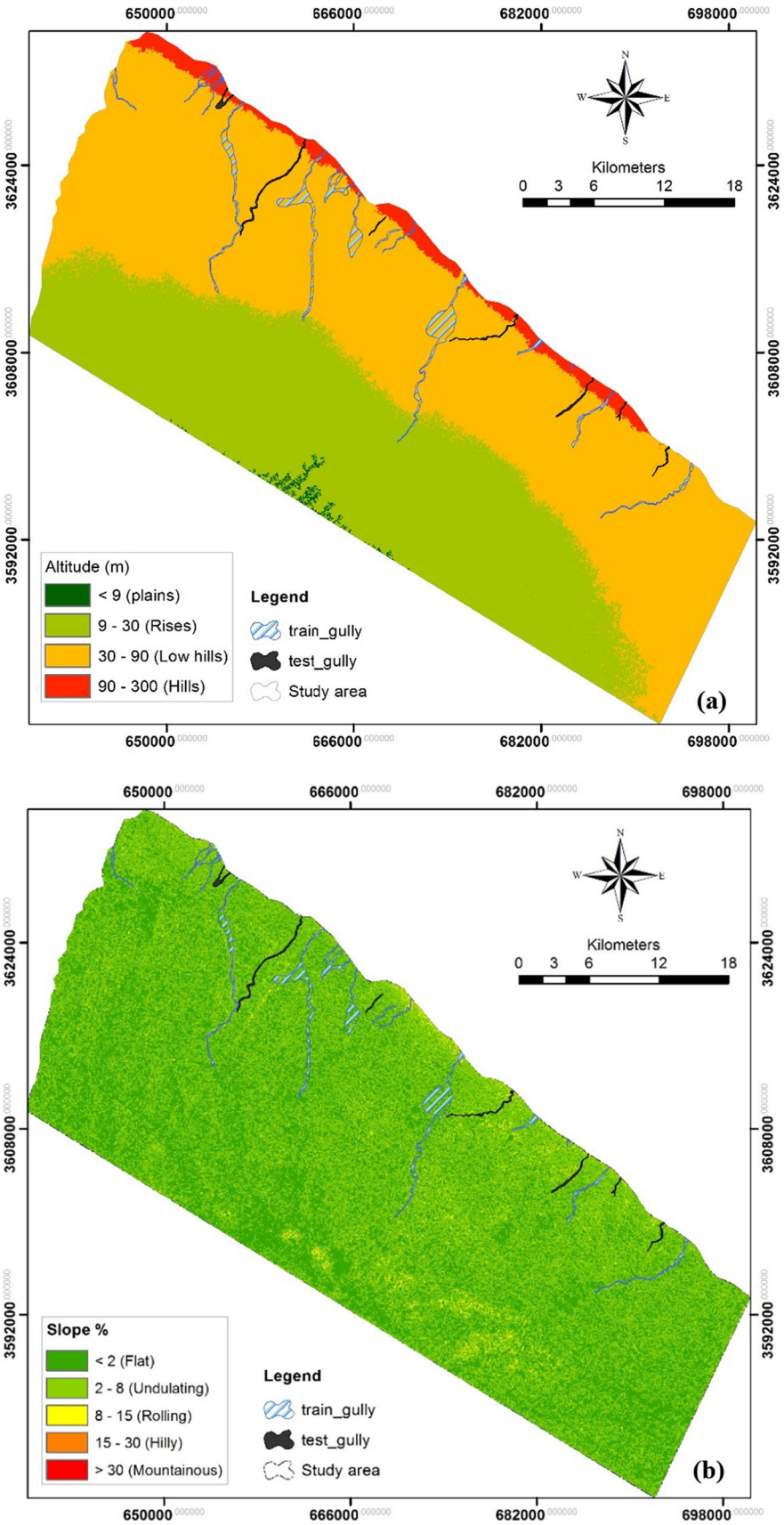


Fig. 5 Gully influential factors **a** altitude (m), **b** slope (%), **c** plan curvature, **d** profile curvature (100/m), **e** aspect, **f** TWI, **g** SPI, **h** LS, **i** soil texture, **j** LULC

Fig. 5 (continued)

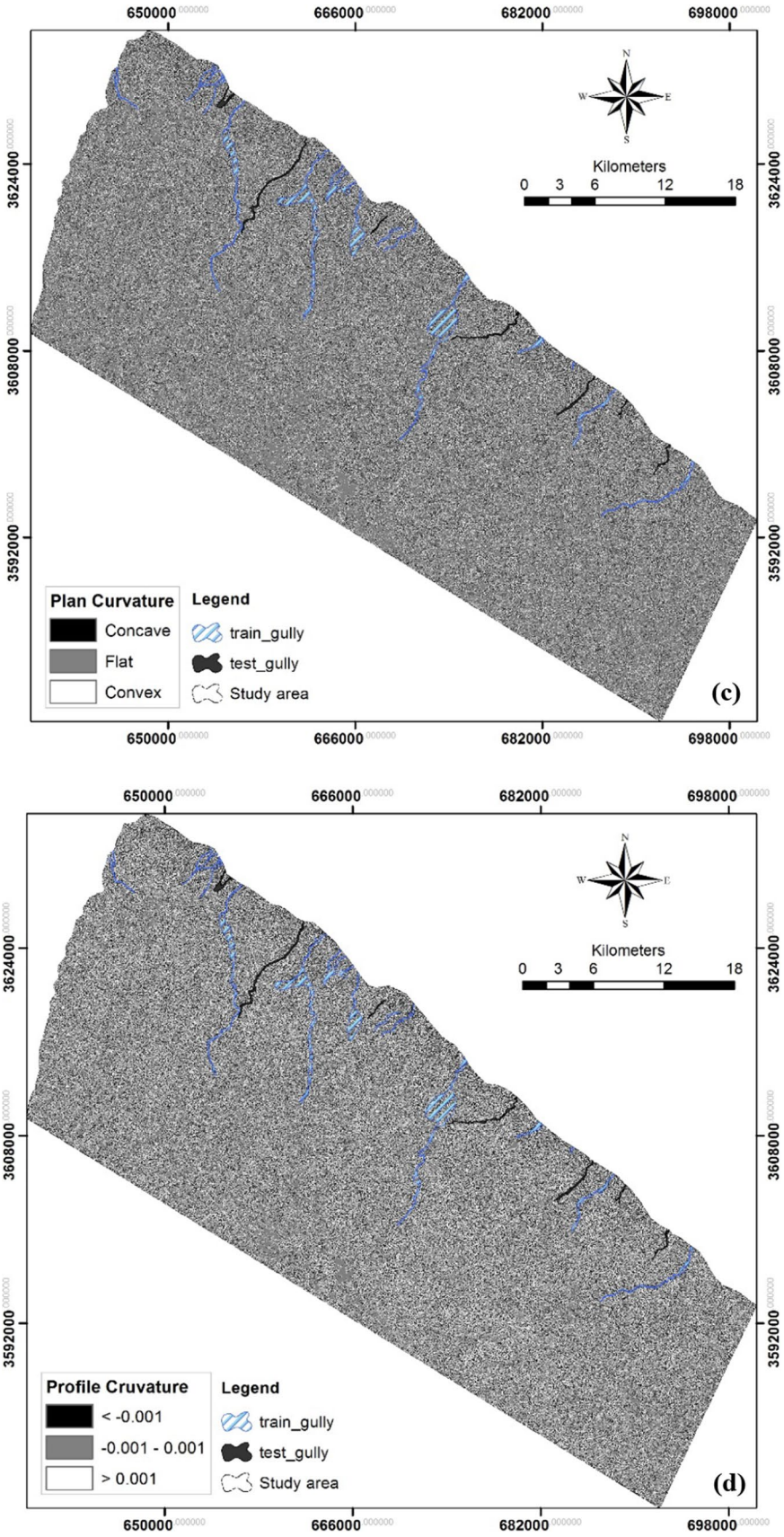


Fig. 5 (continued)

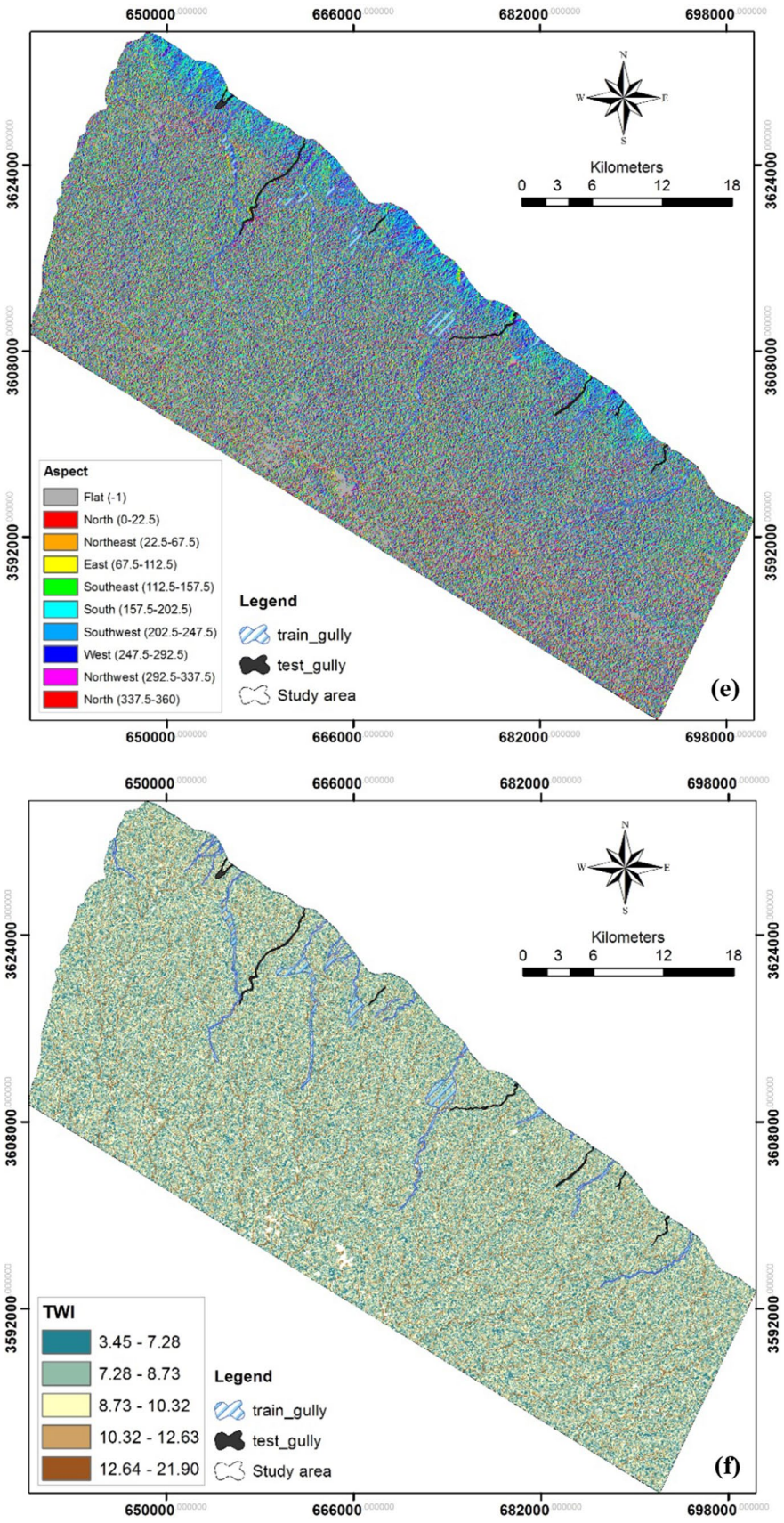
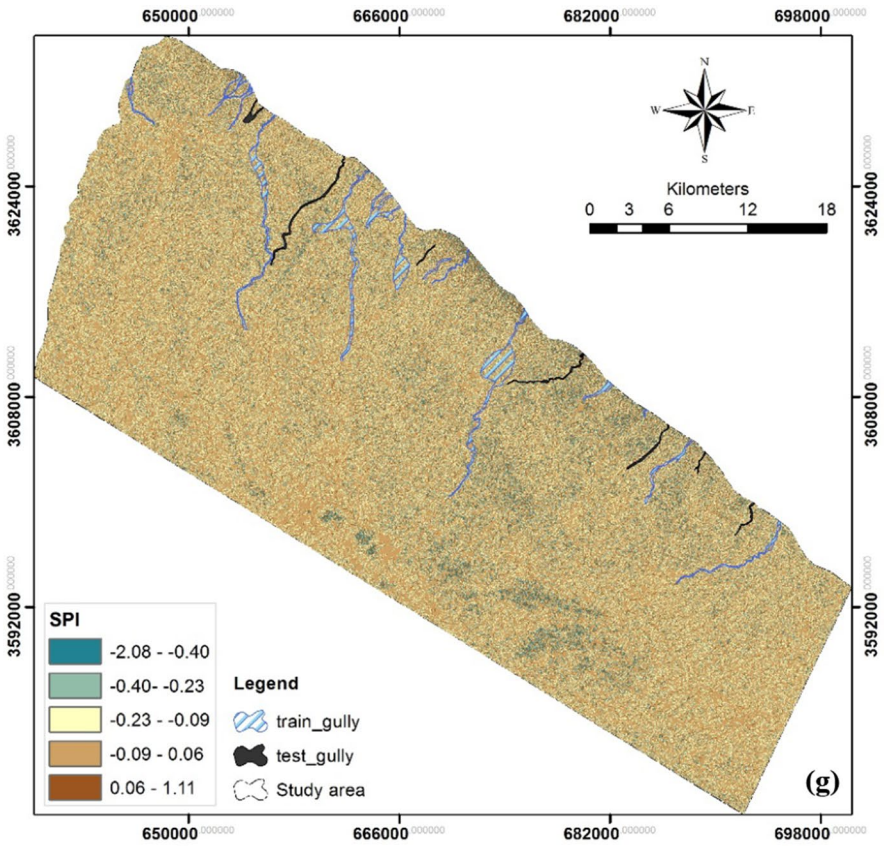


Fig. 5 (continued)



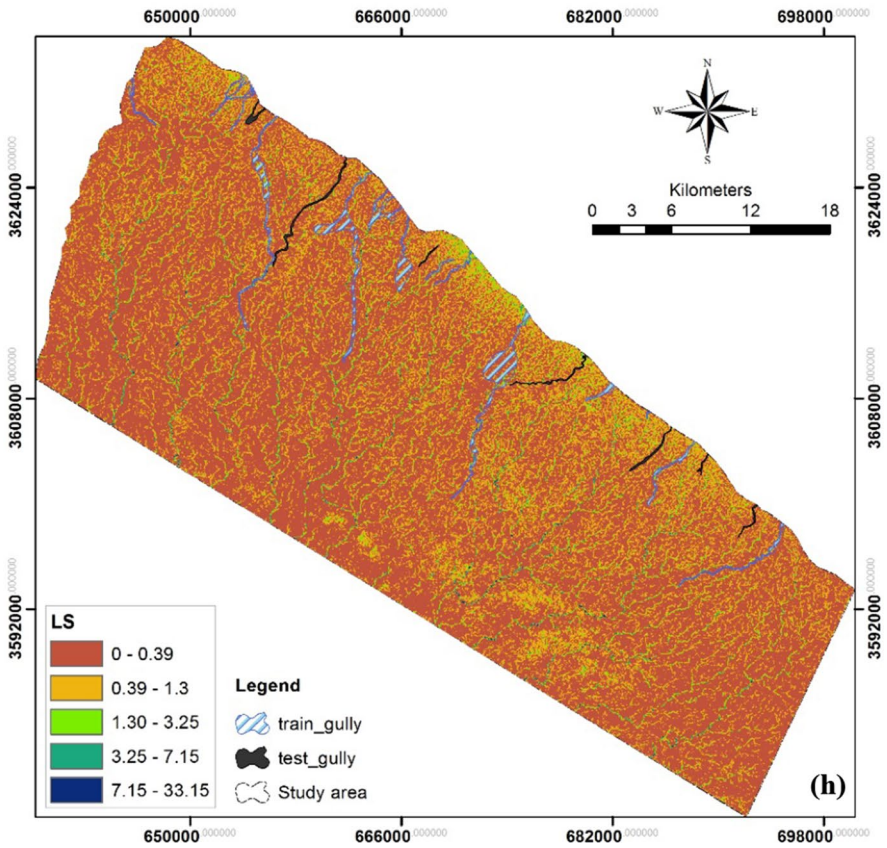
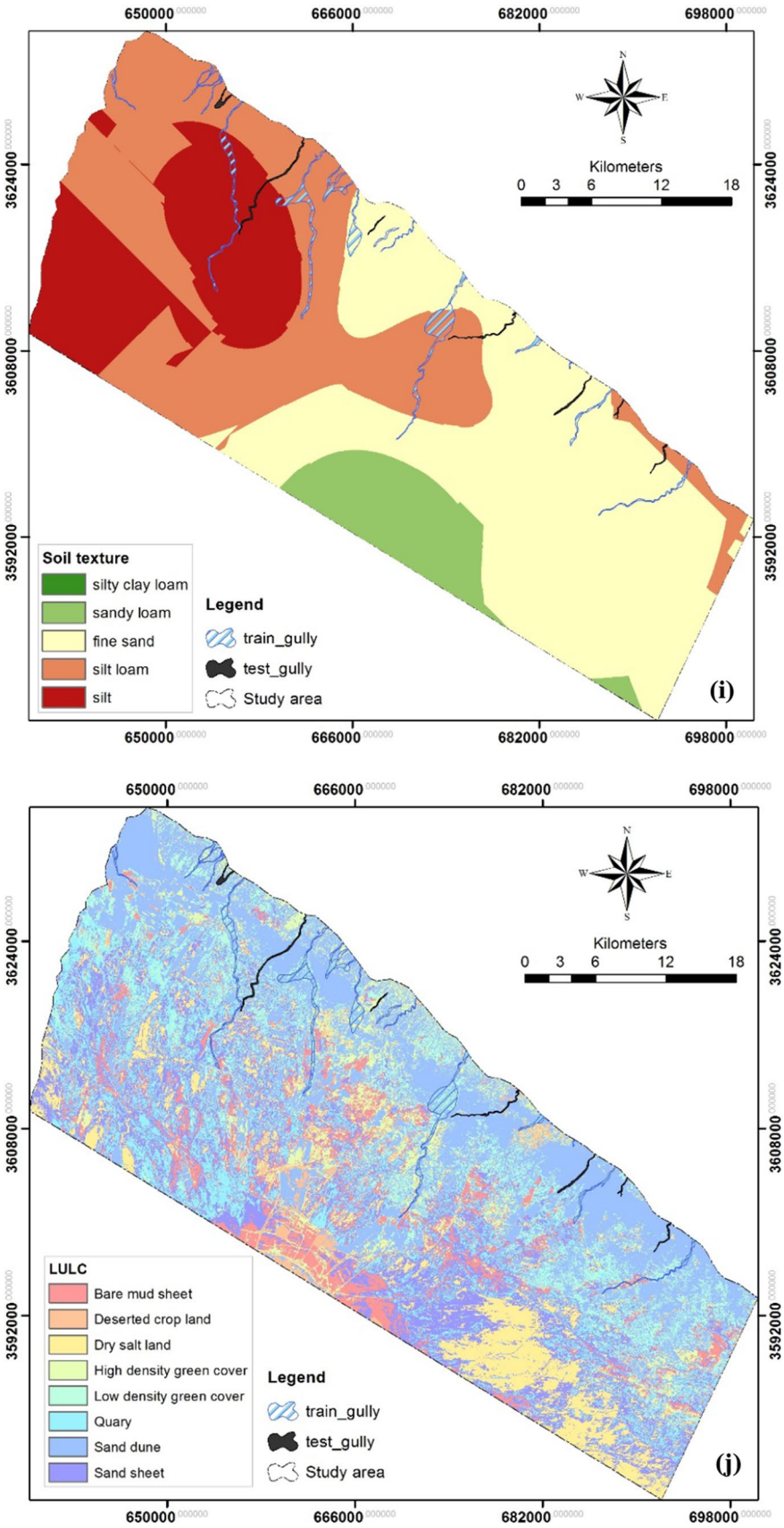


Fig. 5 (continued)



Department of Agriculture (USDA), Fig. 5i. Finally, the LULC map was derived from the interpretation of Landsat 8 imagery and field checks. The maximum likelihood algorithm as a supervised classification system was used after creation of a signature file from ground truth. The following LULC classes were predicted for the study area according to the Anderson et al. (1976) system: bare mud sheet, deserted crop land, dry salt land, high-density green cover, low-density green cover, quarry, sand dune, and sand sheet (Fig. 5j). About 50% of the study area was covered with sand dunes and sand sheet, with the other classes distributed unevenly through the study area.

All influential factors were prepared as raster factors with 30 × 30 m spatial resolution to use in modeling gully erosion susceptibility in the considered area. It is accepted in the literature that the initial set of influential factors has different predictive abilities to create a realistic gully erosion susceptibility map. Hence, incorporating all influential factors into the analysis may reduce the predictive ability of the developed bivariate models. Therefore, it is preferable to carry out a feature selection process to evaluate and exclude those factors that may have adverse effects on the accuracy of the developed models. In this study, the information gain ratio (IGR) was used to screen the factors used and select the most influential in the gully erosion process in the considered area.

Modeling techniques

Information value (InVal)

In this simple bivariate statistical analysis, each individual influential gully factor was compared to the gully distribution inventory map (training data set), and the weighted value for each class *i* in a factor was determined on the basis of gully density for each individual class. Mathematically, InVal method is written as (Yalcin 2008):

$$W_i = \ln \frac{DensClass}{DensMap} = \ln \frac{NpixS_i / NpixN_i}{\sum NpixS_i / \sum NpixN_i}, \quad (4)$$

where W_i is the weighting value of the class *i* of an influential gully factor; *DensClass* is the density of the gullies in the class *i*; *DensMap* is the density of gullies in the total area being studied; $NpixS_i$ is the number of training gully pixels in the class *i*; $NpixN_i$ is the number of pixels of the class *i*; $\sum NpixS_i$ is the total number of gully pixels in the study area; and $\sum NpixN_i$ is the total number of pixels in the study area. A positive value of W_i means that there is a strong relationship between the presence of gully and the factor class (Shit et al. 2015), while the negative value of W_i demonstrates a weak relationship and, thus, the factor class has no effect on the development of the gully landform.

Finally, the W_i for all classes are summed in the GIS environment (overlay and index method) to produce the proneness value to delineate the gully eroded area.

Frequency ratio (FR)

FR is the ratio of probability of an occurrence to that of a nonoccurrence for a given attribute and is defined as (Bonham-Carter 1994):

$$\mathrm{FR} = \frac{NpixS_i / NpixN_i}{\sum NpixS_i / \sum NpixN_i}. \quad (5)$$

The FR ratio defines the degree of correlation between gully location and the class of influential factor that contains the gully. So, the FR > 1 indicates a positive correlation and, thus, a high probability of gully occurrence, whereas FR < 1 indicates a negative correlation and, thus, a low probability of gully occurrence. To produce the gully erosion susceptibility index using this technique, the following equation was used:

$$\mathrm{GES} = \sum_{i=1}^{n} \mathrm{FR}_i, \quad (6)$$

where GES is gully erosion susceptibility and *n* is the total number of influential factors involved in the analysis.

Evidential belief function (EBF)

The EBF is a bivariate statistical model based on the Dempster–Shafer theory of evidence (Dempster 1968) and is regarded as a generalization of the Bayesian theory of subjective probability (Tien Bui et al. 2012). The EBF model has an ability to combine beliefs from multiple sources of evidence and relative flexibility to deal with uncertainty (Thiam 2005). EBFs consist of four mass functions: belief (*Bel*), disbelief (*Dis*), uncertainty (*Unc*), and plausibility (*Pls*). All these functions have the range [0, 1]. In this bivariate technique, generalized Bayesian lower and upper probabilities represent *Bel* and *Pls* functions, respectively (Park 2011), while Unc represents the difference between *Pls* and *Bel*. On the other hand, the *Dis* function is equal to 1 − *Pls*. Hence, the sum of *Bel*, *Unc*, and *Dis* mass functions is equal to 1 (Carranza et al. 2005). In a GIS platform, the Bel function is calculated as (Al-Abadi 2015):

$$Bel_{c_{ij}} = \frac{W_{cijF}}{\sum_{j=1}^{m} W_{cijF}}, \quad (7)$$

where C_{ij} is a factor class ($j = 1, 2, \ldots, m$) and

$$W_{CijF} = \frac{NpixS_i / NpixN_i}{\left(\sum NpixN_i - NpixS_i\right) / \left(\sum NpixN_i - NpixN_i\right)}. \quad (8)$$

The *Dis* function is computed as:

$$Dis_{c_{ij}} = \frac{W_{cij\bar{F}}}{\sum_{j=1}^{m} W_{cij\bar{F}}}, \quad (9)$$

where

$$W_{Cij\bar{F}} = \frac{\left(NpixN_i - NpixS_i\right) / NpixN_i}{\sum NpixN_i - \sum NpixS_i - \left(NpixN_i + NpixS_i\right) / \left(\sum NpixN_i - NpixN_i\right)}. \quad (10)$$

The *Unc* and *Pls* functions are determined via:

$$Unc = 1 - Dis - Bel, \quad (11)$$

$$Pls = 1 - Dis. \quad (12)$$

Once the four mass functions were computed, Dempster's rule of combination was used to obtain the integrated mass functions (Dempster 1968). Dempster's rules for combining two factor maps A and B are as follows (Carranza et al. 2005):

$$Bel_X = \frac{Bel_A Bel_B + Bel_A Unc_B + Bel_B Unc_A}{\beta}, \quad (13)$$

$$Dis_X = \frac{Dis_A Dis_B + Dis_A Unc_B + Dis_B Unc_A}{\beta}, \quad (14)$$

$$Unc_X = \frac{Unc_A Unc_B}{\beta}, \quad (15)$$

$$Pls_X = Unc_X + Bel_X. \quad (16)$$

The denominator β is called the normalization factor and is computed as:

$$\beta = 1 - Bel_A Dis_B - Dis_A Bel_B. \quad (17)$$

Information gain ratio (IGR)

IGR is the ratio of information gain to the intrinsic information. It is used to reduce a bias toward multivalued attributes by taking the number and size of branches into account when choosing an attribute. The IGR was proposed by Quinlan (1993) to overcome disadvantages associated with gain ratio technique. Assume there is a training data set defined as *S* that contains *n* input samples. If $n(F_i, S)$ is the number of instants in the training data set S belonging to the class F_i (gully, non-gully), the information needed to classify *S* is computed as:

$$Info(S) = -\sum_{i=1}^{2} \frac{n(F_i, S)}{|S|} \log_2 \frac{n(F_i, S)}{|S|}. \quad (18)$$

The amount of information to split *S* into (S_1, S_2, …, S_m) regarding the gully influential factor A is computed as:

$$Info(S, A) = -\sum_{j=1}^{m} \frac{S_j}{|S|} Info(S). \quad (19)$$

The IGR for a specific factor A is calculated as:

$$\mathrm{IGR}(S, A) = \frac{Info(S) - Info(S, A)}{SplitInfo(S, A)}, \quad (20)$$

where *SplitInfo* is the potential information produced by partitioning *S* into *m* subset and is computed as:

$$SplitInfo(S, A) = -\sum_{j=1}^{m} \frac{|S_j|}{|S|} \log_2 \frac{|S_j|}{|S|}. \quad (21)$$

Results and discussion

Feature selection using IGR

The analysis of feature selection using IGR was implanted in Weka software 3.9, and the result is demonstrated in Table 2. The average merit here was used to represent the average information gain ratio and its standard deviation using the tenfold cross-validation technique (Tien Bui et al. 2016). From Table 2, the highest average merit (0.221) was obtained for geomorphology, followed by altitude (0.218). This was followed by land cover (0.055), lithology (0.054), soil texture (0.045), LS (0.024), SPI (0.011), and TWI (0.009). In contrast, the four remaining influential factors (slope, profile curvature, aspect, and plan curvature) had 0 average merit, and thus, the inclusion of these factors in the analysis might have decreased the accuracy of the developed model. Therefore, these factors were removed and the rest of the analysis was carried out using only 8 factors.

Applying bivariate models and generation of gully erosion susceptibility maps

The reclassified influential gully erosion factor maps were firstly interested with the training polygons (14 polygons) using the Tabulated Intersection command of ArcGIS 10.2 software to produce a table that contained the $NpixS_i$ for each class. The $NpixN_i$ was directly obtained from the attribute table of each reclassified factor. In this study, the total number of training gully pixels $\left(\sum NpixS_i\right)$ equaled 25,867, while

Table 2 GRI for the influential gully factors

Rank	Influential factors	Average merit	SD
1	Geomorphology	0.221	±0.003
2	Altitude	0.218	±0.002
3	Land cover	0.055	±0.002
4	Lithology	0.054	±0.002
5	Soil texture	0.045	±0.002
6	LS	0.024	±0.002
7	SPI	0.011	±0.001
8	TWI	0.009	±0.001
9	Slope	0	0
10	Profile curvature	0	0
11	Aspect	0	0
12	Plan curvature	0	0

the number of pixels in the total study area $\left(\sum NpixS_i\right)$ was 1,753,516. All four of these values were used to calculate weights in InVal, probability ratio in FR, and mass functions of EBF (Table 3), using Eqs. 4, 5, 7–12, respectively.

The spatial relationship between each influential gully factor and gully distribution by InVal (Table 3) indicated that alluvial fan and sand dune landforms had positive W_i and subsequently high probability to develop gullies. The W_i for flood plain and glacis were negative, indicating a lower probability of gully occurrence and that, therefore, these classes besides depression ($W_i = 0$) were less susceptible to the gully process. In the case of altitude, it can be seen that W_i was positive for the range 30–300 m (low hill–hill), indicating a higher probability of gully occurrence for this range. With respect to land cover, sand dune and low-density green cover lands have the highest W_i values (0.66 and 0.86, respectively), indicating that these classes were more susceptible to gully process. The other land cover classes had negative values, and as a result they were less prone to gully erosion. For the lithology, eolian deposits had the highest value and positive W_i (0.77), indicating that these lithological units were more prone to gully. All other lithological units had negative *Wi* and were thus less prone to the gully erosion process. For the soil texture, it can be seen that silty loam texture ($W_i = 0.60$) had a high proneness for gully occurrence. LS's proneness to gully erosion was higher in areas where the LS was greater than 0.40, and the susceptibility to erosion increases with the increase in LS values. For the SPI and TWI, similar results were obtained. The areas with the highest values of SPI and TWI had higher positive W_i, and thus these areas were more prone to gully erosion.

To map gully erosion susceptibility (GES) in the study area using the InVal technique, the calculated W_i was summed up by means of an overlay procedure to get the final scores for the susceptibility map. The GES values were classified into five classes based on the quantile classification scheme: very low, low, moderate, high, and very high, Fig. 6a. Selection of this classification scheme is based on the studies of Youssef et al. (2015) and Rahmati et al. (2016b), which indicated that the quantile classification scheme is a good classifier in susceptibility mapping. The very low–low classes cover an area of 492 km^2 (31%), while the moderate class encompasses an area of 356 km^2 (23%). The area occupied by high–very classes is 727 km^2 (46%).

The result of applying the FR model is also shown in Table 3. From this table, the alluvial fan landform had a higher FR value (3.57), followed by sand dune (1.33), implying the highest prone-to-gully susceptibility, whereas other geomorphological land forms had a low FR, indicating less vulnerability to gully erosion. In the case of altitude, the 30–90 and 90–300 m classes had FR values of 1.65 and 1.69, respectively, indicating that gullies were more likely to occur at these elevations. For land cover factor, the most vulnerable classes to gully erosion with high FR were low-density green cover (FR = 2.37) and sand dune (FR = 1.93), followed by high-density green cover (1.41). Analysis of lithology showed that the eolian deposits had the highest FR value (2.16), followed by sheet runoff deposits (0.516), implying that these classes were more prone to gully. Regarding soil texture, the highest FR values were observed for silty loam, silt, and fine sand (1.82, 0.63, and 0.63, respectively), indicating that these classes were more prone to gully erosion. The other classes had FR values equal to "0", and thus they were regarded as insignificant in developing gully landforms. For the LS factor, the highest FR values (1.94, 1.62, 1.20, and 1.19) were recorded for the classes (3.25–7.14, 1.30–3.25, 0.39–1.30, and 7.15–33.15), respectively, indicating that they were highly susceptible to gully erosion. Referring to SPI, FR values greater than 1 were recognized for the last range (0.06–1.11), indicating a significant relationship between gully occurrence and SPI for this range. The other classes had FR values greater than "0", and thus these classes played a minor role in developing gully erosion phenomena in the study area. And finally, regarding TWI, the most significant class that affects gully occurrence was (12.63–21.90) and (10.32–12.63) with FR values of 1.42 and 1.24, respectively, indicating a higher probability of gully development compared to other classes.

The map of GES using the FR model (Fig. 6b) was obtained by applying Eq. 6 and using the *Weighted Sum* module in ArcGIS 10.2 software. The obtained GES values varied from 3.23 to 16.66 and were classified into five classes using the quantile classification scheme: very low, low, moderate, high, and very high. The areas encompassed by these classes were distributed as: 683 km^2 (43%) for very low–low classes, 356 km^2 (22%) for moderate class, and 588 km^2 (73%) for high–very high classes.

Table 3 Calculation of FR, Wi, and mass functions for the influential flood factors

Factor	Class	$NpixN_i$	$NpixN_i$ %	$NpixS_i$	$NpixS_i$ %	FR	W_i	Mass functions of EBF			
								Bel	*Dis*	*Unc*	*Pls*
Geomorphology	Sand dunes	690,334	0.394	13,548	0.524	1.330	0.285	0.232	0.149	0.619	0.851
	Alluvial fan	221,963	0.127	11,700	0.452	3.573	1.273	0.644	0.119	0.238	0.881
	Flood plain	32,368	0.018	334	0.013	0.700	−0.357	0.121	0.191	0.688	0.809
	Glacis	796,016	0.454	286	0.011	0.024	−3.715	0.004	0.349	0.647	0.651
	Depression	12,835	0.007	0	0	0	0	0.000	0.192	0.808	0.808
Altitude (m)	< 9 (plains)	6523	0.004	0	0	0	0	0.000	0.265	0.735	0.735
	9–30 (rises)	697,229	0.398	307	0.012	0.030	−3.512	0.009	0.438	0.554	0.562
	30–90 (low hills)	994,765	0.567	24,188	0.935	1.648	0.500	0.489	0.039	0.472	0.961
	90–300 (hills)	54,999	0.031	1373	0.053	1.692	0.526	0.502	0.258	0.240	0.742
Land cover	Bare mud sheet	161,046	0.092	535	0.021	0.225	−1.491	0.032	0.136	0.832	0.864
	Deserted crop land	23,627	0.013	151	0.006	0.434	−0.836	0.063	0.127	0.811	0.873
	Dry salt land	150,198	0.086	11	0.000	0.005	−5.305	0.001	0.138	0.862	0.862
	High-density green cover	132,673	0.076	2764	0.107	1.412	−0.345	0.207	0.121	0.672	0.879
	Low-density green cover	189,462	0.108	6612	0.256	2.366	0.861	0.352	0.105	0.544	0.895
	Quarry	227,506	0.130	518	0.020	0.154	−1.868	0.022	0.142	0.836	0.858
	Sand dune	483,964	0.276	13,786	0.533	1.931	0.658	0.285	0.081	0.634	0.919
	Sand sheet	384,990	0.220	1490	0.058	0.262	−1.338	0.038	0.152	0.810	0.848
Lithology	Eolian deposits	542,839	0.310	17,255	0.667	2.155	0.768	0.811	0.108	0.081	0.892
	Flood plain deposits	54,312	0.031	0	0	0	0	0.000	0.233	0.767	0.767
	Depression fill deposits	24,744	0.014	0	0	0	0	0.000	0.229	0.771	0.771
	Sheet runoff deposits	1,131,571	0.645	8612	0.333	0.516	−0.662	0.189	0.430	0.381	0.570
Soil texture	Silty clay loam	50	0.000	0	0	0	0	0.000	0.200	0.800	0.800
	Sandy loam	138,649	0.079	0	0	0	0	0.000	0.217	0.783	0.783
	Fine sand	745,306	0.425	6913	0.267	0.629	−0.464	0.187	0.255	0.557	0.745
	Silty loam	557,029	0.318	14,911	0.576	1.815	0.596	0.550	0.123	0.326	0.877
	Silt	312,482	0.178	4043	0.156	0.877	−0.131	0.262	0.205	0.533	0.795
LS	0–0.38	1,268,422	0.723	16,717	0.646	0.893	−0.113	0.129	0.247	0.623	0.753
	0.39–1.3	400,866	0.229	7090	0.274	1.199	0.181	0.174	0.181	0.645	0.819
	1.30–3.25	70,461	0.040	1685	0.065	1.621	0.483	0.237	0.188	0.575	0.812
	3.25–7.14	12,068	0.007	346	0.013	1.942	0.664	0.286	0.191	0.523	0.809
	7.15–33.15	1649	0.001	29	0.001	1.192	0.176	0.173	0.193	0.634	0.807
SPI	−2.07 to (−0.41)	49,976	0.02850127	529	0.020	0.718	−0.332	0.148	0.202	0.651	0.798
	−0.41 to (−0.23)	198,950	0.113461	2524	0.098	0.860	−0.151	0.177	0.204	0.619	0.796
	−0.23 to (−0.09)	375,604	0.21420661	5348	0.207	0.965	−0.035	0.199	0.202	0.599	0.798
	−0.09 to 0.06	942,214	0.53734375	13,892	0.537	0.999	−0.001	0.206	0.200	0.593	0.800
	0.06 to 1.11	186,722	0.10648738	3573	0.138	1.297	0.260	0.269	0.193	0.538	0.807
TWI	3.44–7.28	345,864	0.197	4792	0.185	0.939	−0.063	0.171	0.202	0.626	0.798
	7.28–8.73	604,910	0.345	7904	0.306	0.886	−0.121	0.161	0.211	0.627	0.789
	8.73–10.32	470,417	0.268	6856	0.265	0.988	−0.012	0.180	0.200	0.620	0.800
	10.32–12.63	241,517	0.138	4419	0.171	1.240	0.215	0.227	0.191	0.581	0.809
	12.63–21.90	90,758	0.052	1895	0.073	1.415	0.347	0.260	0.195	0.545	0.805

The results of applying EBF are also summarized in Table 3 in terms of four mass functions (*Bel*, *Dis*, *Unc*, and *Pls*). A comparatively high value of *Bel* function implies a higher probability of gully proneness and vice versa. Analysis of *Bel* values in Table 3 indicated the gully erosion was more likely to occur in alluvial fan (0.65), elevation range (90–300) (0.50) and (30–90 m) (0.49), low-density green land cover class (0.86) and sand dune (0.66), eolian deposits (0.77), silty loam soil texture class (0.55), LS class (3.25–7.14) (0.30), SPI class (0.06–1.11) (0.27), and TWI class (12.63–21.90) (0.26). The other classes had low values

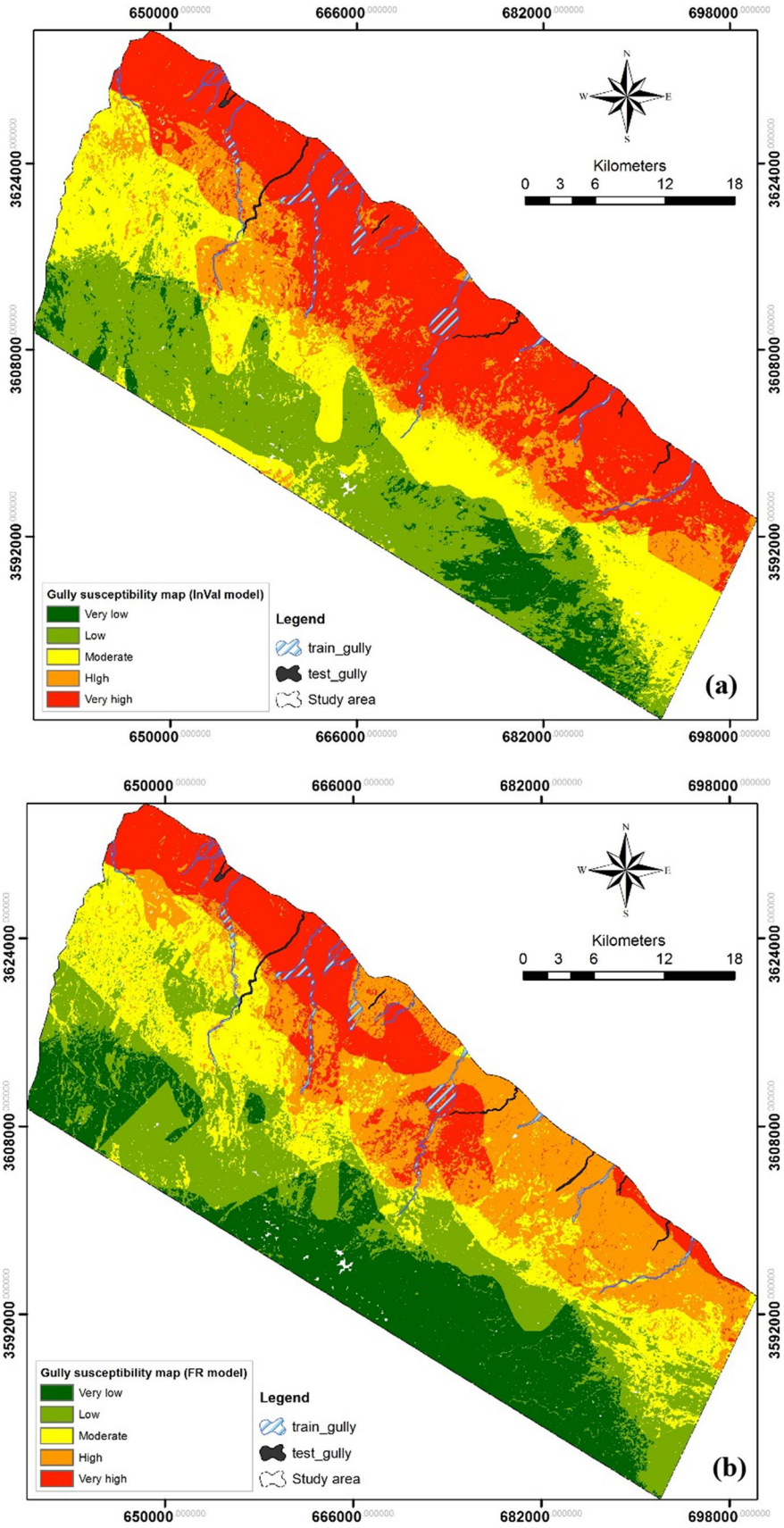
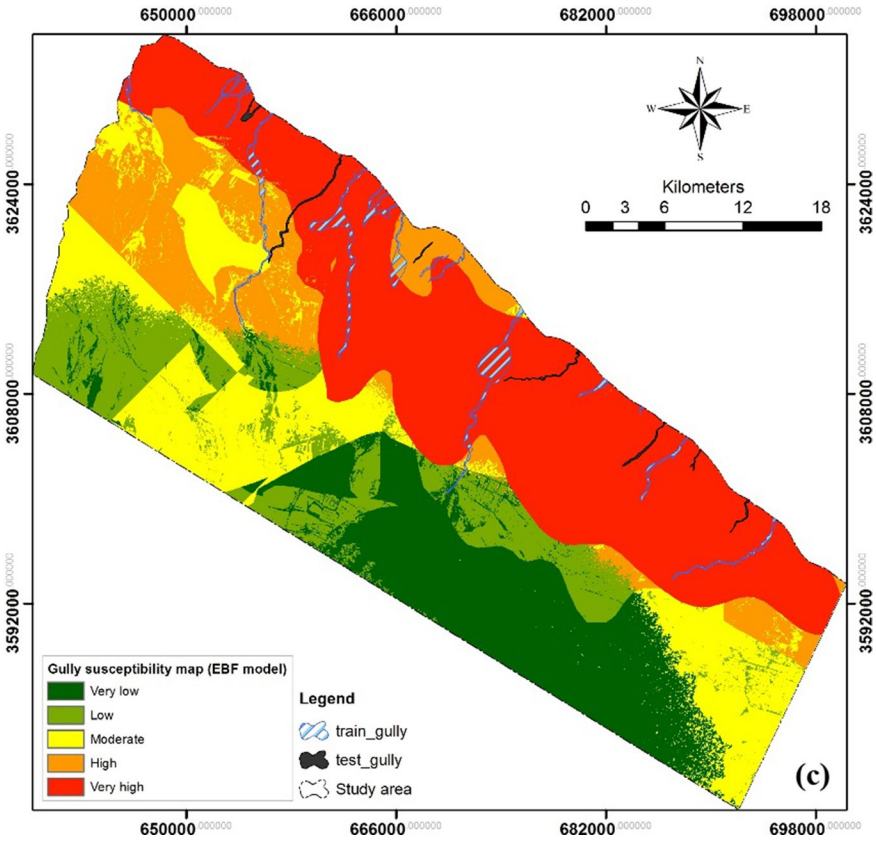


Fig. 6 Maps of gully erosion susceptibility **a** InVal model, **b** FR model, **c** EBF model

Fig. 6 (continued)



and were considered to have less impact on gully erosion phenomena.

To map gully erosion susceptibility using the EBF model, the *lookup function* module of the *spatial analysis* extension of ArcGIS 10.2 was used to generate mass functions raster layers to be used in combination with the process of Dempster's rules (Eqs. 13–17). The two mass functions of geomorphology and altitude factors were first combined. The resulting combined layer was combined with the land cover mass functions layer to generate another combined layer and so on. In total, seven combining processes were carried out to get the final integrated mass functions maps. The integrated *Bel* map (Fig. 6c) was used as an index to reveal the condition of gully erosion susceptibility. The pixel value of *Bel* was classified into five classes similar to the way of InVal and the FR model: very low, low, moderate, high, and very high. These classes were distributed over the study area as: 489 km^2 (31%) very low–low, 298 km^2 (19%) moderate, and 788 km^2 (50%) high–very high.

Validation of gully erosion susceptibility map using ROC curve

The results of the model validation using the ROC technique (implemented in EDRIS Selva 17.0) in terms of success (training) and predicting (testing) rates are presented in Fig. 7. The best bivariate model in the training stage was FR (AUC = 0.855), followed by InVal (AUC = 0.849). The worst model in this stage was EBF with AUC equal to 0.834. In contrast, the best model in the testing stage was InVal (AUC = 0.829), followed by FR (AUC = 0.824) and EBF (AUC = 0.815). From these results, it can be concluded that the InVal model was the best bivariate model for describing gully erosion susceptibility in the study area, while the worst model was EBF.

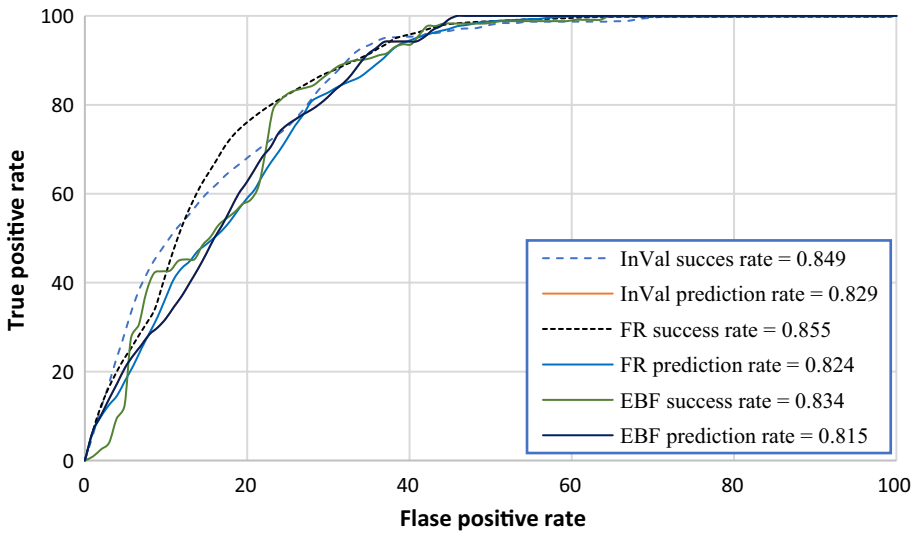


Fig. 7 Validation results

Conclusions

Mapping of gully erosion susceptibility is an essential step in controlling and mitigating the effects of soil erosion in an area. To develop gully erosion susceptibility maps, different modeling techniques have been adopted so far. Due to their simplicity and ease of implementation under the GIS platform, three bivariate techniques, namely information value, frequency ratio, and evidential belief functions, were used here to map gully erosion susceptibility at Ali Al-Gharbi District in northern Maysan Governorate, southern Iraq. To apply bivariate models, two things should be prepared: a gully area inventory map and gully influential factors. Through interpretation of remotely sensed data and field survey, 21 areas affected by gully erosion were mapped. The area affected by gully erosion was partitioned into two sets: 14 gully areas (polygon format in GIS) were used for training the bivariate models, and the remaining areas (7) were used for testing and validating the developed models. Twelve gully erosion influential factors were chosen based on the literature review and availability of data. These factors were altitude, slope, profile curvature, plan curvature, aspect, TWI, SPI, LS, lithology, geomorphology, soil texture, and land cover. Screening of influential factors using the information gain ratio proved that eight factors (altitude, TWI, SPI, LS, lithology, geomorphology, soil texture, and land cover) were important in developing gully landforms in the study area, while the other four factors (slope, profile curvature, plan curvature, and aspect) were insignificant in the gullying process. Using the most important factors and training gullies inventory map, three bivariate models were developed and used to generate gully erosion susceptibility maps. The developed maps were validated using the relative operating characteristics curve, and the results showed that the InVal model was the best in terms of prediction rate. The worst model was the evidential belief function as it compared with the other models, InVal and FR. For the best performance model, InVal, the gully erosion susceptibility index values were classified into five classes using the quantile classification scheme: very low, low, moderate, high, and very high. The very low–low classes covered an area of 492 km^2 (31%), the moderate class encompassed an area of 356 km^2 (23%), while the high–very high classes extended over an area of 727 km^2 (46%). About 50% of the study area is exposed to gully erosion; therefore, creation of mitigation and erosion control plans is highly necessary in the study area.

References

Akgün A, Türk N (2011) Mapping erosion susceptibility by a multivariate statistical method: a case study from the Ayvalık region, NW Turkey. Comput Geosci 37(9):1515–1524. https://doi.org/10.1016/j.cageo.2010.09.006

Al-Abadi A (2012) Hydrological and hydrogeological analysis of northeaster Missan governorate, south of Iraq using geographic information system. Doctoral Thesis, Baghdad University

Al-Abadi AM (2015) The application of Dempster–Shafer theory of evidence for assessing groundwater vulnerability at Galal Badra basin, Wasit governorate, east of Iraq. Appl Water Sci 7(4):1725–1740. https://doi.org/10.1007/s13201-015-0342-7

Al-Abadi AM, Al-Temmeme AA, Al-Ghanimy MA (2016) A GIS-based combining of frequency ratio and index of entropy approaches for mapping groundwater availability zones at Badra–Al Al-Gharbi–Teeb areas, Iraq. Sustain Water Resour Manag 2(3):265–283. https://doi.org/10.1007/s40899-016-0056-5

Anderson JR, Hardy EE, Roach JT, Witmer RE (1976) A land use and land cover classification for use with remote sensor data. U.S. Geological Survey Professional Paper 964, Government Printing Office, Washington, U.S.

Bellen RC, Dunnington HV, Wetzel R, Morton D (1959) Lexique Stratigraphique International, Asie. Iraq. Intern Geol Conger Comm Stratigr, 3, Fasc, 10a

Billi P, Dramis F (2003) Geomorphological investigation on gully erosion in the Rift valley and the northern highlands of Ethiopia. Catena 50(2–4):353–368. https://doi.org/10.1016/S0341-8162(02)00131-5

Bonham-Carter GF (1994) Geographic information system for geoscientists: modelling with GIS. Pergamon/Elsevier Science Ltd., New York

Buday T, Jassim SZ (1987) The regional geology of Iraq. Vol. 2, tectonism, magmatism and metamorphism. S. E. Geol. Surv. and Mineral Invest., Baghdad, p 352

Capra A, Mazzara LM, Scicolone B (2005) Application of the EGEM model to predict ephemeral gully erosion in Sicily, Italy. Catena 59(2):133–146. https://doi.org/10.1016/j.catena.2004.07.001

Carranza EJM, Woldai T, Chikambwe EM (2005) Application of data-driven evidential belief functions to prospectivity mapping for aquamarine-bearing pegmatites, Lundazi district, Zambia. Nat Resour Res 14(1):47–63. https://doi.org/10.1007/s11053-005-4678-9

Conforti M, Pietro A, Gaetano R, Fabio S (2011) Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). Nat Hazards 56(3):881–898. https://doi.org/10.1007/s11069-010-9598-2

Conoscenti C, Di Maggio C, Rotigliano E (2008) Soil erosion susceptibility assessment and validation using a geostatistical multivariate approach: a test in Southern Sicily. Nat Hazards 46:287–305. https://doi.org/10.1007/s11069-007-9188-0

Conoscenti C, Agnesi V, Angileri S, Cappadonia C, Rotigliano E, Märker M (2013) A GIS-based approach for gully erosion susceptibility modelling: a test in Sicily, Italy. Environ Earth Sci 70(3):1179–1195. https://doi.org/10.1007/s12665-012-2205-y

Conoscenti C, Angileri A, Cappadonia C, Rotigliano E, Agnesi V, Märker M (2014) Gully erosion susceptibility assessment by means of GIS-based logistic regression: a case of Sicily (Italy). Geomorphology 204:399–411. https://doi.org/10.1016/j.geomorph.2013.08.021

Conoscenti C, Agnesi V, Cama M, Caraballo-Arias NA, Rotigliano E (2017) Assessment of gully erosion susceptibility using multivariate adaptive regression splines and accounting for terrain connectivity. Land Degrad Dev. https://doi.org/10.1002/ldr.2772

de Winnaar G, Jewitt GPW, Horan M (2007) A GIS-based approach for identifying potential runoff harvesting sites in the Thukela River basin, South Africa. Phys Chem Earth 32:1058–1067. https://doi.org/10.1016/j.pce.2007.07.009

Dempster AP (1968) A generalization of Bayesian inference. J Royal Stat Soc 30(2):205–247

Flanagan DC, Nearing MA (1995) USDA-water erosion prediction project: hillslope pro- file and watershed model documentation. NSERL Report #10.USDA-ARS National Soil Erosion Research Laboratory, West Lafayette, Indiana

Ionita I, Fullen MA, Zgłobicki W, Poesen J (2015) Gully erosion as a natural and human-induced hazard. Nat Hazards 79(Suppl 1):1. https://doi.org/10.1007/s11069-015-1935-z

Jassim SZ, Goff JC (2006) Geology of Iraq. Dolin, Prague and Moravian Museum, Brno

Knisel WG (1980) CREAMS: a field scale model for chemicals, runoff, and erosion from agricultural management systems. US department of agriculture, conservation report 26

Le Roux JJ, Sumner PD (2012) Factors controlling gully development: comparing continuous and discontinuous gullies. Land Degrad Dev 23(5):440–449. https://doi.org/10.1002/ldr.1083

Lucà F, Conforti M, Robustelli G (2011) Comparison of GIS-based gullying susceptibility mapping using bivariate and multivariate statistics: Northern Calabria, South Italy. Geomorphology 134(3–4):297–308. https://doi.org/10.1016/j.geomorph.2011.07.006

Luffman IE, Arpita N, Tim S (2015) Gully morphology, hillslope erosion, and precipitation characteristics in the Appalachian Valley and Ridge province, southeastern USA. Catena 133:221–232. https://doi.org/10.1016/j.catena.2015.05.015

McDonald RC, Isbell RF, Speight JG, Walker J, Hopkins MS (1990) Australian soil and land survey—field handbook, 2nd edn. Inkata Press, Melbourne

Merkel WH, Woodward DE, Clarke CD (1988) Ephemeral gully erosion model (EGEM). In: Agricultural, forest, and Rangeland hydrology. American Society of Agricultural Engineers Publication 07-88, pp 315–323

Meyer A, Martínez-Casasnovas JA (1999) Prediction of existing gully erosion in vineyard parcels of the NE Spain: a logistic modelling approach. Soil Tillage Res 50(3–4):319–331. https://doi.org/10.1016/S0167-1987(99)00020-3

Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modeling—a review of hydrological, geomorphological, and biological applications. Hydrol Process 5(1):3–30. https://doi.org/10.1002/hyp.3360050103

Park NW (2011) Application of Dempster–Shafer theory of evidence to GIS-based landslide susceptibility analysis. Environ Earth Sci 62(2):367–376. https://doi.org/10.1007/s12665-010-0531-5

Poesen J, Nachtergaele J, Verstraeten G, Valentin C (2003) Gully erosion and environmental change: importance and research needs. Catena 50(2–4):91–133. https://doi.org/10.1016/S0341-8162(02)00143-1

Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann, San Mateo

Rahmati O, Ali H, Pourghasemi HR, Noormohamadi F (2016a) Gully erosion susceptibility mapping: the role of GIS-based bivariate statistical models and their comparison. Nat Hazards 82(2):1231–1258. https://doi.org/10.1007/s11069-016-2239-7

Rahmati O, Tahmasebipour N, Pourghasemi HR, Feizizadeh B (2016b) Evaluating the influence of geo-environmental factors on gully erosion in a semi-arid region of Iran: an integrated framework. Sci Total Environ 579:913–927. https://doi.org/10.1016/j.scitotenv.2016.10.176

Shit PK, Paira R, Bhunia GS, Maiti R (2015) Modeling of potential gully erosion hazard using geo-spatial technology at Garbheta block, West Bengal in India. Model Earth Syst Environ 1:2. https://doi.org/10.1007/s40808-015-0001-x

Suzen ML, Doyuran V (2004) A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. Environ Geol 45(5):665–679. https://doi.org/10.1007/s00254-003-0917-8

Svoray T, Markovitch H (2009) Catchment scale analysis of the effect of topography, tillage direction and unpaved roads on ephemeral gully incision. Earth Surf Process Landforms 34(14):1970–1984. https://doi.org/10.1002/esp.1873

Thiam AK (2005) An evidential reasoning approach to land degradation evaluation: Dempster–Shafer theory of evidence. Trans GIS 9:507–520

Tien Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. Catena 96:28–40. https://doi.org/10.1016/j.catena.2012.04.001

Tien Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaut I (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13(2):361–378. https://doi.org/10.1007/s10346-015-0557-6

USDA-SCS (1966) Procedure for determining rates of land damage, land depreciation, and volume of sediment produced by gully erosion. Technical Release No. 32. US GPO 1990-261-419:20727/SCS. US Government Printing Office, Washington, DC

Wischmeier WH, Smith DD (1965) Predicting rainfall-erosion losses from cropland east of the rocky mountains, guide for selection of practices for soil and water conservation. Agriculture Handbook No. 282, Agricultural Research Service, U. S. Dept. of Agric, Washington DC, p. 47

Yalcin A (2008) GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. Catena 72(1):1–12. https://doi.org/10.1016/j.catena.2007.01.003

Youssef AM, Pourghasemi HR, El-Haddad BA, Dhahry BK (2015) Landslide susceptibility maps using different probabilistic and bivariate statistical models and comparison of their performance at Wadi Itwad Basin. Bull Eng Geol Environ. https://doi.org/10.1007/s10064-015-0734-9