

Prediction of gully erosion susceptibility mapping using novel ensemble machine learning algorithms

Alireza Arabameri, Subodh Chandra Pal, Romulus Costache, Asish Saha, Fatemeh Rezaie, Amir Seyed Danesh, Biswajeet Pradhan, Saro Lee & Nhat-Duc Hoang

To cite this article: Alireza Arabameri, Subodh Chandra Pal, Romulus Costache, Asish Saha, Fatemeh Rezaie, Amir Seyed Danesh, Biswajeet Pradhan, Saro Lee & Nhat-Duc Hoang (2021) Prediction of gully erosion susceptibility mapping using novel ensemble machine learning algorithms, *Geomatics, Natural Hazards and Risk*, 12:1, 469-498, DOI: [10.1080/19475705.2021.1880977](https://doi.org/10.1080/19475705.2021.1880977)

To link to this article: <https://doi.org/10.1080/19475705.2021.1880977>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 10 Feb 2021.



[Submit your article to this journal](#)



Article views: 2510



[View related articles](#)








[View Crossmark data](#)



Citing articles: 11 [View citing articles](#)

Prediction of gully erosion susceptibility mapping using novel ensemble machine learning algorithms

Alireza Arabameri^a , Subodh Chandra Pal^b, Romulus Costache^{c,d} ,
Asish Saha^b, Fatemeh Rezaie^{e,f} , Amir Seyed Danesh^g, Biswajeet Pradhan^{h,i,j,k} ,
Saro Lee^{e,f}  and Nhat-Duc Hoang^l 

^aDepartment of Geomorphology, Tarbiat Modares University, Tehran, Iran; ^bDepartment of Geography, The University of Burdwan, Bardhaman, West Bengal, India; ^cResearch Institute of the University of Bucharest, Bucharest, Romania; ^dNational Institute of Hydrology and Water Management, Bucharest, Romania; ^eGeoscience Platform Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), Daejeon, Republic of Korea; ^fDepartment of Geophysical Exploration, Korea University of Science and Technology, Daejeon, Republic of Korea; ^gFaculty of Technology and Engineering, East of Guilan, University of Guilan, Rudsar-Vajargah, Iran; ^hFaculty of Engineering and Information Technology, Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney, Ultimo, NSW, Australia; ⁱDepartment of Energy and Mineral Resources Engineering, Sejong University, Seoul, Republic of Korea; ^jDepartment of Meteorology, Center of Excellence for Climate Change Research, King Abdulaziz University, Jeddah, Saudi Arabia; ^kEarth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia; ^lFaculty of Civil Engineering, Institute of Research and Development, Duy Tan University, Danang, Vietnam

ABSTRACT

Spatial modelling of gully erosion at regional level is very relevant for local authorities to establish successful counter-measures and to change land-use planning. This work is exploring and researching the potential of a genetic algorithm-extreme gradient boosting (GE-XGBoost) hybrid computer education solution for spatial mapping of the susceptibility of gully erosion. The new machine learning approach is to combine the extreme gradient boosting machine (XGBoost) and the genetic algorithm (GA). The GA meta-heuristic is being used to improve the efficiency of the XGBoost classification approach. A GIS database has been developed that contains recorded instances of gully erosion incidents and 18 conditioning variables. These parameters are used as predictive variables used to assess the condition of non-erosion or erosion in a given region within the Kohpayeh-Sagzi River Watershed research area in Iran. Exploratory results indicate that the proposed GE-XGBoost model is superior to the other benchmark solution with the desired predictive precision (89.56%). Therefore, the newly built model may be a promising method for large-scale mapping of gully erosion susceptibility.

ARTICLE HISTORY

Received 21 December 2020
Accepted 20 January 2021

KEYWORDS

Gully erosion; spatial modelling; ensemble learning; extreme gradient boosting machine; GIS; genetic algorithm

CONTACT Saro Lee  leesaro@kigam.re.kr

This article has been republished with minor changes. These changes do not impact the academic content of the article.

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Effective and sustainable utilization of natural resources are indeed necessary for proper understanding of environmental processes (Fistikoglu and Harmancioglu 2002; Zhang et al. 2019; Feng et al. 2020; Tian et al. 2020). The sustainability of our environment and societies greatly depends on the optimal uses of various kinds of natural resources (Cheng et al. 2016; Keesstra et al. 2016; Lu, Tian, et al. 2019; Lu, Liao, et al. 2019; He et al. 2020; Yang et al. 2020; Zhang et al. 2020; Chen, Chen, Tsangaratos, et al. 2020). Among the several types of natural resources (He, Chen, et al. 2018; He, Shen, et al. 2018), soil is the most vital and one of the key issues for the sustainability as it is support basic human needs and livelihood on the earth surface. In the era of twenty-first century, the sustainability of soil not only depends on management strategies adopted by land planners and farmers but also impacted through marketing, subsidies and several rules taken in political decisions (Keesstra et al. 2016). In the most effective manner, soil is primarily a land related phenomenon and has significant links to the several functions of United Nations sustainable development goals (SDGs). The United Nations has proposed 17 SDGs that can challenge national governments as well as stakeholders, and SDGs can be categorized into three groups i.e. environment, social and economic accordingly (Keesstra et al. 2016; Visser et al. 2019). The proper functioning of soil is greatly emphasis on ecosystem services and links to the achievement of SDGs. In the effort to achieve several SDGs, the most essential is soil-water system with special emphasis on land degradation neutrally and land restoration (Keesstra et al. 2018). Thus, avoiding and reducing the phenomenon of land degradation is one of the key challenges for the sustainable development of environment and economic activities. Therefore, it has been always required comprehensive planning and management of productivity as well as protection against erosion. Soil erosion is a critical environmental hazard which annually destroys a large amount of fertile soils worldwide (Wang and Shao 2013). Erosion refers to a set of physical and chemical interactions, in which parts of the earth's surface are affected by processes such as weathering, destruction, dissolution, transport and sedimentation and lose their relative stability and are transmitted from the original site to other areas. In other words, erosion is a natural process in the geographic cycle that extracted from disposal of soil particles by water or wind and transferring them to other areas, while some human activities increases the natural process of the erosion rate (Gitas et al. 2009). Generally, there are different forms of water-related erosion (Mekonnen et al. 2017); among them, the gully erosion is widely recognized as the most destructive form. Soil erosion has negative impact on environment as it is reducing the soil fertility and damages agricultural land, downstream areas, destroying network communication and ultimately produced badland topography (Rodrigo-Comino et al. 2018). Water induced soil erosion is also control soil organic carbon in an ecosystem with exchange and distribution of carbon, that is ultimately effect on food production system (Berhe et al. 2012; Novara et al. 2016). The complexity of the gully erosion initiation are facing by many researchers in various in-site and off-site environmental and engineering problems such as land degradation, flooding, falling of groundwater level and fill of reservoirs (Dlapa et al. 2012; Zema et al. 2012; Gessesse et al. 2015; Zhao et al. 2019). Moreover, the progress of gully erosion varies

in different environmental conditions including geological formations, topography, soil characteristics and climatic conditions (Poesen et al. 1998). Notably, this type of soil erosion is also governed by human-caused factors including land use alteration as well as exploitation of natural resource such as water, soil and plants (Ayele et al. 2015; Slimane et al. 2016).

By the reason of several complicated interaction among the triggering factors and the particular conditions of local environment, the formation and expansion of the gullies cannot be limited to the simple relationship between the slope and the area of the watershed located in the gully head-cut and physiological characteristics (Gayen et al. 2019; Pal and Chakraborty 2019). The initiation of ephemeral gullies and its further development is largely dependent on several factors such as slope gradient, drainage area, topographic indices, precipitation, surface runoff, land use characteristics and soil types (Amare et al. 2019). Consequently, in different regions, rate and type of affection of parameters on the formation of the gully may vary (Ayele et al. 2015). As pointed out by previous studies (Pimentel 2006; Bouaziz et al. 2011; Shit et al. 2015), an average amount of 30–40 Ton/Ha/Y is removed due to erosion in Asia, Africa and South America. Particularly in Iran, this amount can be up to 50% with soil losses estimated to be 15 Ton/Ha/Y.

Thus, in order to assessment of gully erosion and associated risk, various models have been proposed. These models can be classified into the categories of experimental and mechanical, dynamic and static and quantitative and qualitative models (Gitas et al. 2009). These models have taken into account a wide range of factors affecting soil erosion such as natural and human factors. The main criteria for selecting these models include goal, available data, time and cost (Helming et al. 2005). In recent decades, the analysis of soil erosion hazards by the above models and their integration with satellite data and advanced data modelling tools are employed to identify the physical hazards of soil erosion at regional scales. These models can provide information about erosion-prone locations and their trends, as well as erosion analysis scenarios. In addition, GIS techniques, with its major advantages of greater capacity to process and analyse big data size in digital layers in vector format and raster, can be used to determine the type and value of factors affecting the prediction of prone-erosion through different models, and finally, the results will be presented in the form of a digital layer (Droogers and Kite 2002).

In recent years, the greater advancement in computer technology (Zhu 2020; Cao, Zhao, et al. 2020; Cao, Wang, et al. 2020; Lv and Qiao 2020; Chen and Li 2020) along with gradually development of GIS knowledge (Zhu, Wang, Chen, et al. 2019; Zhu, Wang, Wang, et al. 2019; Zhu et al. 2020) and several platforms (Han et al. 2019; Yan et al. 2020), a wide variety of models have been proposed for solving scientific problems such as gully erosion susceptibility mapping (GESM). Therefore, based on the methodological approaches, the methods used for natural hazards susceptibility assessment has been classified into quantitative, qualitative and most recently machine learning approaches by several research group of people. Consequently, the aforementioned methods have also been widely used in assessment of gully erosion. These methods can be briefly reviewed as follows:

- i. Qualitative methods include techniques such as analytical hierarchy process (AHP) employed in (Arabameri, Rezaei, et al. 2018) and TOPSIS which is utilized in (Arabameri, Pradhan, Rezaei, et al. 2019).
- ii. Quantitative models includes bivariate and multivariate statistical models: information value (Arabameri, Pradhan, and Rezaei 2019a), frequency ratio (Rahmati, Haghizadeh, et al. 2016; Meliho et al. 2018), probability models (Rahmati et al. 2017), weights-of-evidence (Gayen et al. 2019), logistic regression (Arabameri, Pradhan, et al. 2018) and linear discriminant analysis (Arabameri and Pourghasemi 2019).
- iii. Artificial intelligence models employs advanced data-driven methods including artificial neural network (Sarkar and Mishra 2018), multivariate adaptive regression spline (Conoscenti et al. 2018). decision tree ensembles (Arabameri, Yamani, Pradhan, et al. 2019; Hosseinalizadeh et al. 2019); (Tien Bui, Shirzadi, et al. 2019) and flexible discriminant analysis (Gayen et al. 2019).

It can be seen from the literature that the predictive capabilities of artificial intelligence models are often better than those of conventional qualitative/quantitative methods (Arabameri and Pourghasemi 2019; Gayen et al. 2019). This fact is explained by the superiority in nonlinear and multivariate data modelling of the former approaches. Although various artificial intelligence models have been put forward for gully erosion, other alternative models should be investigated. It is due to the complex nature of this hazard which involves many influencing factors and the unique environmental conditions of each study area. Evidences of this fact are well demonstrated in previous studies on susceptibility mapping of gully erosion (Arabameri, Pradhan, et al. 2018; Conoscenti et al. 2018; Sarkar and Mishra 2018; Tien Bui, Shirzadi, et al. 2019) as well as other phenomena (Siahkamari et al. 2018; Jaafari et al. 2019; Kavzoglu et al. 2019; Tien Bui, Hoang, et al. 2019).

In this research study, the Kohpayeh-Sagzi watershed is chosen for assessment of GESM. The Kohpayeh-Sagzi watershed locates in Esfahan Province (central part of Iran). In this region, human pressure on land and soil resulting from population growth along with climatic condition is arid and semi-arid types, over-grazing and human induced deforestation and mismanagement of land use have led to serious soil erosion and related problems. The present research work is carried out by using 18 gully conditioning factors, all of these variables have been selected based on local geo-environmental conditions i.e. topography, hydrology, climatological, soil and environmental related conditions. The modelling of gully erosion susceptibility (GES) was carried out by using a novel ensemble of extreme gradient boosting machine (XGBoost) and evolutionary genetic optimization machine learning models and its comparison with three standalone models such as support SVM, RF and LR. Extensive literature studies have been shown that there are till now no research work which is based on the ensemble approach of genetic algorithm-extreme gradient boosting (GE-XGBoost), thus GESM using novel ensemble of GE-XGBoost is the novelty of this research study. XGBoost is an efficient algorithm that has the capacity to find out missing values and enhanced the prediction performance result. On the other hand, the major advantage of genetic algorithm (GA) is the multi objective

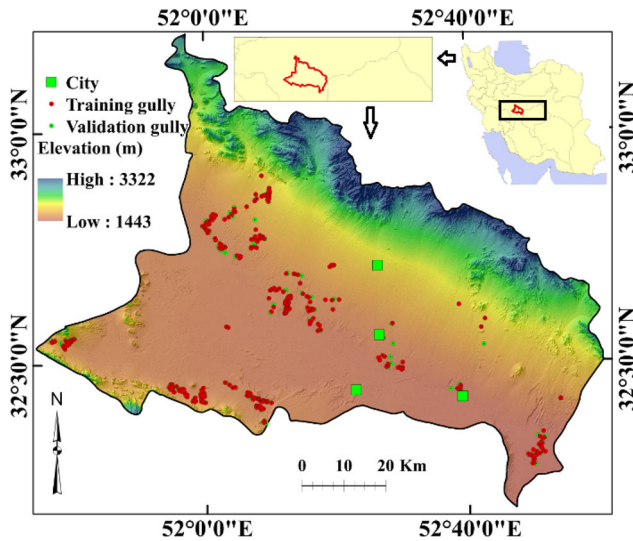


Figure 1. Location map of the study area.

optimization and search from a population of points and not from single points. The stand alone machine learning models used in this study has been previously also widely used by several researchers in different fields and the models give optimal prediction performance result due to their unique advantages. The present research article is prepared into the following subsequent sections: In the first section, we already described about the basic information of gully erosion and associated methods used in GESM. The second section represent about the details of the study area and data used for further progress of our research work. Subsequently, next section explained in details about the back ground of research methods used in this study. The machine learning models used for GESM is presented in the fourth section. The subsequent section deals with investigational result and discussion. The concluding section of the article summarizes the research findings with several closing remarks.

2. Study area and data

2.1. Description of the study area

Kohpayeh-Sagzi watershed is situated between longitudes $51^{\circ} 34' 02''$ to $53^{\circ} 00' 56''$ E and latitudes $32^{\circ} 10' 59''$ to $33^{\circ} 12' 50''$ N. The study area is within the Esfahan Province (the central part of Iran; Figure 1). The maximum and minimum elevations of Kohpayeh-Sagzi watershed are 1443 and 3322 m a.s.l., respectively. The area is 1736 m above sea level in average (a.s.l.). In that study area, the mean annual precipitation and temperature is 192.8 mm and 12.9°C . The majority of this area has a flat topography whereas north, north-west and north-east parts have covered by undulating rugged mountain.

The highest slope in the study area is 69.56° with a mean of 3.4° . The depth of soil is varied from 6.8 to 165 m. Poor rangeland and agriculture-dry farming are the major land uses practices in the study area. Pebble fan shaped debris from alluvial

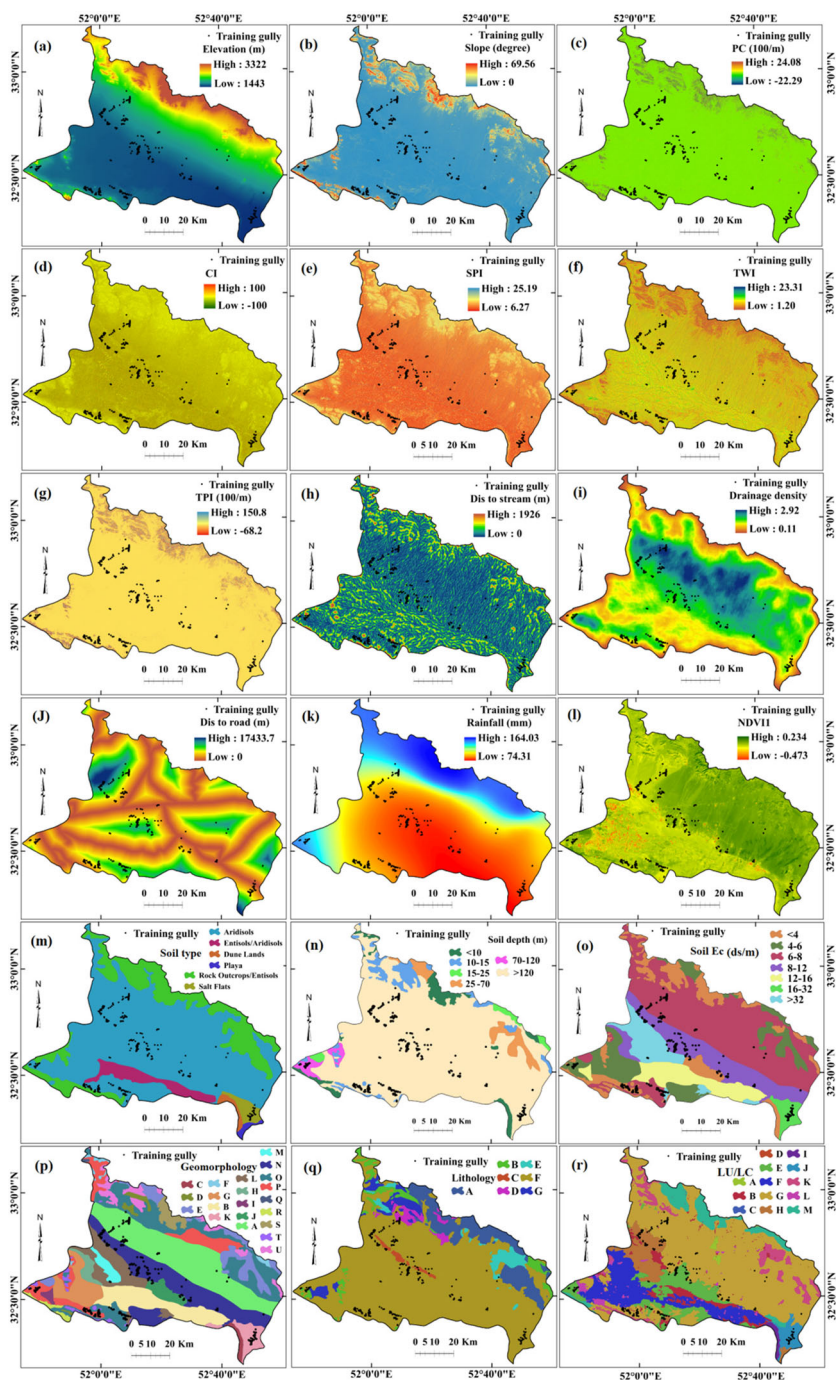


Figure 2. Eighteen variables used in this study (a) elevation, (b) slope, (c) PC, (d) CI, (e) SPI, (f) TWI, (g) TPI, (h) distance to stream, (i) drainage density, (j) distance to road, (k) rainfall, (l) NDVI, (m) soil type, (n) soil depth, (o) soil EC, (p) geomorphology, (q) lithology and (r) LULC.

deposits of upper maternal materials with total slope of 3 to 8%, coarse limestone fan shaped debris along with argillaceous limestone, limestone shale, limestone marl and old alluvial deposits with slope 5% are among the most important geomorphologic units. Quaternary lithotypes including pediment fan of low level and deposition in valley terrace, salt flat, sand dunes and sand sheet and clay flat are widely observed. In addition, aridisols is the dominant soil type in the region.

2.2. Data used

As shown in previous studies, the interaction of several geo-environmental factors like topographical, hydrological, characteristics and properties of soil, present LULC and human interaction causes gulling in prone area (Nyssen et al. 2002; Poesen et al. 2003; Nyssen et al. 2006; El Maaoui et al. 2012). Based on literature review of related works (Conoscenti et al. 2018; Arabameri, Yamani, Pradhan, et al. 2019); (Tien Bui, Shirzadi, et al. 2019); physiographic characteristic in study area, multi collinearity test, availability of data, and scale of research, 18 conditioning factors are selected as influencing factors of gully erosion in this research area. These factors include elevation, slope, PC, CI, SPI, TWI, TPI, distance to stream, drainage density, distance to road, rainfall, NDVI, soil type, soil depth, soil EC, geomorphology, lithology and LULC (Figure 2(a-r)).

These conditioning factors were extracted from various resources including:

- i. A 12.5 m spatial resolution of ALOS PALSAR DEM is collected from Alaska Satellite System and used for extraction and calculation of topographical and hydrological factors consisting of elevation, slope, PC, CI, TWI, SPI, drainage density, TPI and distance to stream. Equation 1 is used for calculation of TPI (Guisan et al. 1999; De Reu et al. 2013):

$$TPI = \frac{E_{pixel}}{E_{surrounding}} \quad (1)$$

where, E_{pixel} represent the respective cell elevation and $E_{surrounding}$ represents the neighbor pixels mean elevation.

- ii. The Geological Society of Iran (GSI) produced geological map of Iran with the scale of 1:100,000 (<http://www.gsi.ir/>) and here we used this map to generate the lithology map.
- iii. The topographical map with the scale of 1:50,000 were collected from the secondary sources (www.ngo-org.ir). The road network map was digitized from Google Earth satellite images.
- iv. The NDVI map was prepared by using LANDSAT-8 OLI satellite image, which was collected from USGS earth explorer website (<https://earthexplorer.usgs.gov/>) with spatial resolution of 30 meters.
- v. The mean annual rainfall map was prepared by using 30 years (1986 to 2016) rainfall statistics of several synoptic stations.
- vi. Land-use types have a have a main role in runoff and infiltration (Zhang et al. 2020). The LULC map for present study area was collected from the secondary

Table 1. LULC units found in the Kohpayeh-Sagzi watershed.

Classes	Description
A	Abkhan
B	Agriculture
C	Airport
D	Orchard
E	Bare land
F	Agri-dryfarming
G	Poor-range
H	Salt land
I	Sand dune
J	Rock
K	Urban
L	Woodland

source (<https://www.scwmri.ac.ir>). Various LULC units of present study area are shown in Table 1.

- vii. The following (<http://esfahan.areeo.ac.ir/>) website was used for collecting various soil characteristics maps and prepared soil type, soil depth and soil EC maps for this research study area.
- viii. The geomorphology layer is mapped using Landsat 8 OLI satellite images for 2016, the RGB853 false color, lines of geomorphic surfaces and Google Earth images. To distinguish geomorphic units from geomorphologic knowledge, land area processes, slope map obtained from ALOS-PALSAR DEM and geological map were used. The boundaries of the units were determined based on field surveys, and with the help of GPS. All steps of geomorphology mapping were performed in ArcGIS10.5 environment. The detail descriptions of geomorphology units for this study area are presented in Table 2. The details about the data sources are presented in Table 3.

Additionally, here we used ENVIv4.8, ArcGISV10.5, Google Earth pro 7.8 and Arc HydroV10.4 software to prepare several conditioning factors which are used for GESM. Alongside, several statistical and validation analysis was carried out by using the SPSSv24 and Microsoft Excel 2016 packages. Beside this, enormous readers can follow the previous works of (Arabameri, Rezaei, et al. 2018; Tien Bui, Shirzadi, et al. 2019) for more details of the gully erosion affecting factors.

3. Background of the methods used

3.1. Extreme gradient boosting machine

The XGBoost was introduced by Chen and Guestrin (2016) is selected to be used in this study because it represents the state-of-the-art within the machine learning community. This algorithm relies on classification trees (Breiman et al. 1984) and the gradient boosting framework (Friedman 2001). XGBoost is a popular machine learning system of scalable used for boosting the performance of classification trees (Li 2010). A classification tree typically establishes a set of rules to categorize each gully erosion instance as a function of a set of predisposing factors in a graph structure. The graph is built as a single tree whose leaves are assigned with a score which expresses how

Table 2. Geomorphological units in the Kohpayeh-Sagzi watershed.

Units	Description
A	Pebble fan shaped debris from alluvial deposits of upper maternal materials with total slope of 3 to 8 %
B	The final part of the alluvial plain of the Zayandehrud river consists of alluvial deposits with a slope of 0 to 1%
C	The stabilized sand dunes with a slope of 2 to 15 %
D	Low elevation hills along with limestone, marl, shale and sandstone with moderate to high rock outcrops with a steep slope of over 35%
E	The relatively high hills include limestone and limestone shale with marl and sandstone bedding with high rocky outcrops and averaging slope of 18%
F	High hills include of grey shale, limestone, marl and limestone shale with a steep slope of 15%
G	Alluvial plain with very deep soils and heavy to very heavy texture with slope of 0-1%
H	Alluvial plain consisting of alluvial deposits with limestone and marl with slope of 1 to 4%
I	Hill slope plain consisting of alluvial deposits with limestone and sandstone with slope of 1 to 5%
J	Flood plain composed of alluvial and flood deposits, along with limestone, salt and sandstone and clay and salt slabs with slope of 0-3%
K	The flood plain consisting of alluvial and flood deposits along with limestone, salt, limestone marl and sandstone with clay and salt slabs with slope of 0 to 3 %
L	The flood plain consists of alluvial and flood deposits, along with gypsum, salt, limestone, limestone marl, sandstone and clay
M	The low lands consist of very fine alluvial deposits, with clay and salt slabs with slope of 0 to 0.5%
N	old Plateau consists of limestone, salt, grey and red marls, argillaceous limestone, sandstone, conglomerate and limestone marl
O	High and very high mountains along with limestone, marl and limestone shale, with a steep slope of 33%
P	Very high mountains with limestone and conglomerate stones along with rocky outcrops with a steep slope of 43 %
Q	Low-elevation mountains of limestone, shale and limestone marl, along with limestone sandstones with slope of 26%
R	Relatively high to low mountains consisting of metamorphic and pyrogenic rocks along with limestone, marl and shale with slope of 26%
S	High mountains of igneous rocks and volcanic lavas along with metamorphic rocks with slope of 45%
T	Coarse limestone fan shaped debris along with argillaceous limestone, limestone shale, limestone marl and old alluvial deposits with slope 5%
U	Coarse limestone fan shaped debris along with argillaceous limestone, limestone shale, conglomerate, sandstone, calcareous marl with a slope of 2 to 8 %

likely is a gully to fall in a given factor class, be it a categorical (e.g. lithotypes) or ordinal (e.g. reclassified slope steepness).

Notably, in the framework of XGBoost, the loss function used to train the ensemble model is combined with a procedure known as regularization which penalizes the complexity of the trees. This regularization method can potentially improve the performance of gully erosion model since over fitting phenomenon is alleviated. Over fitting occurs when a new dataset of gully erosion used for prediction may not be explained as well as the one stored in the training set. Thus, regularization contributes to the limitation of over fitting and to the improvement of the gully erosion prediction model flexibility.

XGBoost combines the results of different tree models by taking their weighted averages (Gusain et al. 2018). An iterative process is employed by using weak prediction models to learn the overall prediction model at every step by correcting the misclassification in the preceding iteration. The XGBoost model is constructed by optimizing the following objective function:

$$OF(\theta) = \sum_{i=1}^n l(y_i, \bar{y}_i) a + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

Table 3. Details about the data sources used in the research study.

Conditioning factors	Data source	Time (year)	Spatial resolution/scale
Elevation, slope, plan curvature, convergence index, drainage density (DD), stream power index (SPI), topographic wetness index (TWI), topographic position index, distance to stream, drainage density	ALOS PALSAR DEM (Alaska Satellite Facility)	2012	12.5 m
Rainfall	Iran Meteorological Organization (IMO) (http://www.weather.ir/)	1986 to 2016	–
Distance to road	Google earth image	2019	1:40,000
Geomorphology, Lithology	Geological Survey of Iran (GSI) (http://www.gsi.ir/), Landsat OLI 8 satellite image	2016	1:10,00,000
NDVI, Land use	Landsat OLI 8 satellite image (USGS) (https://earthexplorer.usgs.gov/), www.ngo-org.ir , Google earth image	2019	30 m
Soil type, soil depth, soil EC	http://esfahan.areeo.ac.ir/	2019	1:10,00,000

where, $\sum_{i=1}^n l(y_i, \bar{y}_i)$ represents the loss function of root mean square error used for fitting the model on the training data. $\sum_{k=1}^K \Omega(f_k)$ refers to the regularization term which helps to avoid overfitting. K denotes to the number of individual trees. f_k is a tree within the ensemble. y_i and \bar{y}_i are an actual and predicted class output, respectively.

The training process of XGBoost aims at minimizing of the aforementioned objective function. To do so, a weak learner is added to the ensemble iteratively. Notably, while the level of model complexities grows as the learning of trees progresses, the regularization term prevents the over fitting problems by regulating the number of leaf nodes in the tree (Chang et al. 2018; Gusain et al. 2018). The detail about the model construction phases of XGBoost was find out in the previous works of (Chen and Guestrin 2016; Gusain et al. 2018).

3.2. GA for global optimization

GA (Goldberg 1989) is a stochastic approach and widely used by several research group of people. The algorithm for this model is based on the idea of organisms' natural selection and genetic processes. In this model for better identification of individuals, several phenomena of reproduction, crossover and mutation in nature are simulating and compete with each other among the group of solutions candidates (Kilinc and Caicedo 2019). Individuals with better fitness combine their genetic features to generate new candidate solutions. Here, the whole process is repeated in several times until the criterion of stopping is satisfied. The operation of the GA is demonstrated in Table 4. More details about GA approach has been found in recent established works (Amjad et al. 2018; Mirjalili et al. 2020). In this study, the GA is

Table 4. The genetic algorithm.

Define population size (<i>Pop Size</i>), number of searched variable (<i>D</i>)
Generating initial population <i>Pop</i>
Compute the fitness value of each member
While stopping criterion is not satisfied
Selecting parent members based on fitness values
Create new members based on crossover and mutation process
End While

applied to optimize the hyper-parameters of the aforementioned XGBoost model used for gully erosion prediction.

3.3. Benchmark models

3.3.1. Random forest (RF)

The RF, proposed in (Breiman 2001), is a type of decision tree ensemble. This method is used to an overall classifier by combining various individual trees. Its structure can be boiled down to few important steps. From a gully dataset made of gully erosion/no-gully, erosion instances associated to a set of conditioning factors. RF starts by sampling with replacement one subset of the data upon which a classification tree will be applied.

Notably, the initial number of predictors of individual trees is kept constant and the tree is allowed to grow to its largest extent. Each leaf of a tree is made of a predictor's binary split which best fits the data. This operation ensures that each tree in the forest can learn or differentiate substantially from the others. This fact helps to reduce the tendency of decision trees to over fit the data. The aggregation among all trees is then computed via a majority vote which is achieved by combining the learning outcomes of all the weak tree based learners (Deng et al. 2019). Successful implementation of the RF model has been demonstrated in previous works on spatial mapping of natural hazards (Chen et al. 2018; Arabameri, Yamani, Pradhan, et al. 2019; Dang et al. 2019).

3.3.2. Support vector machine (SVM)

Vapnik (1998) proposed the algorithm of support vector machine, is a classification model which has its root from the statistical learning theory. The main task of SVM algorithm at hand is to develop a decision surface that separates the data instances into two classes: No-gully and gully erosion. The generalized form of a SVM model in decision surface is to set of training data points $\{x_k, y_k\}_{k=1}^N$ along with the input data of $x_k \in R^n$ and a group of class labels represent by $y_k \in \{-1, +1\}$. Notably, nonlinear mapping function and kernel techniques is used to deal with nonlinear classification problems, through increases the data structure within the SVM model (Hamel 2011). Accordingly, a hyper-plane is constructed in the high dimensional space and used for data classification. Several research paper related to this work indicates that SVM is very much appropriate; thus, here we have selected SVM is a benchmark method (Tehrany et al. 2015; Ngoc Thach et al. 2018; Abedini et al. 2019).

3.3.3. Logistic regression (LR)

The LR is a very popular statistical method that employs a logistic function to model a binary dependent variable. This machine learning method establishes a division boundary that classified the data structures into two classes of binary values i.e. 0 and 1, which indicates negative and positive class, respectively (Agresti 2003). The advantages of the LR are straightforward learning phase and transparent model structure. To identify the model structure of a LR model, the loss function of log likelihood and fast iterative methods such as the stochastic gradient descent can be used (Gormley et al. 2016). Although being simple, this classifier has been shown to achieve acceptable prediction accuracy in various recent studies (Lombardo and Mai 2018; Abedini et al. 2019; Hoang et al. 2019).

4. Proposed EG-XGBoost ensemble model for predicting gully erosion potential

4.1. Gully erosion database

The understanding of spatial distribution and forms of gullies is most important prerequisite for the preparation of GESM, and it is show by the gully erosion inventory map (GEIM; Conoscenti et al. 2014). To prepare the GEIM various resources were used in this research work:

- i. Gully erosion points were identified by the published map from the (<http://esfahan.areeo.ac.ir/>) website.
- ii. Interpretation and verified them from Google Earth images satellite image.
- iii. To identify exact location of gullies, extensive field surveys were carried out by using a GPS (Global Positioning System).

A total of 529 gully locations were recognized throughout this study area. Thereafter, this total gully head cut points were randomly classified into two groups of 70/30 ratio for training (370 gullies) and validating (159 gullies) purposes. Also, an equal amount of gully absence locations were created by random point generation tool in ArcGIS platform to support the calibration and validation procedures (Kornejady et al. 2017). Figure 3 demonstrates the field photograph in this field of research.

As mentioned earlier, 18 variables namely elevation, slope, PC, CI, SPI, TWI, TPI, distance to stream, drainage density, distance to road, rainfall, NDVI, soil type, soil depth, soil EC, geomorphology, LULC and lithology are employed as erosion influencing factors. These factors are integrated into a GIS database which is illustrated in Figure 4.

4.2. Multicollinearity (MC) diagnosis and feature selection

MC is defined as the linear relationship among the two or more variables (Alin 2010). Basically, in a regression model, MC analysis has been carried out when two or more variables are correlated among each other (Chowdhuri et al. 2020; Chen,



Figure 3. Ground photograph of gullies in the study area of Kohpayeh-Sagzi watershed.

Chen, Tsangaratos, et al. 2020; Chen and Li 2020). The result of traditional statistical method in feature selection procedure has ambiguity and affected by extreme values because of the presence of large dataset (Pal et al. 2020; Chen, Chen, et al. 2021; Chen, Li, et al. 2021). Thus, in GESM, analysing the collinearity among the gully related conditioning factors is very important because the presence of collinearity decreases the model's predictive performance (Arabameri, Rezaei, et al. 2018). Variance inflation factor (VIF) and tolerance (TOL) are commonly used statistical techniques for collinearity analysis among the independent variables by many researchers (Band, Janizadeh, Chandra Pal, Saha, Chakraborty, Shokri, et al. 2020; Roy et al. 2020; Chen, Zhao, et al. 2020) and are employed in the present study. The threshold value of TOL is < 0.1 and VIF is > 10 can indicates collinearity present among the variables (Oh et al. 2017). Therefore, first we have extracting the values of gully and non-gully points from several conditioning factors used in this study and finally VIF and TOL values were calculated by using SPSS 16 package. The following equations were used to calculate TOL and VIF (Arabameri, Karimi-Sangchini, et al. 2020; Band, Janizadeh, Chandra Pal, Saha, Chakraborty, Melesse, et al. 2020).

$$\text{TOL} = 1 - R_j^2 \quad (3)$$

$$\text{VIF} = \frac{1}{\text{TOL}} \quad (4)$$

where, R_j^2 indicate the regression of coefficient.

4.3. The GA optimized XGBoost model

The hyper-parameters to be set appropriately in the construction phase of the XGBoost model. These hyper-parameters include:

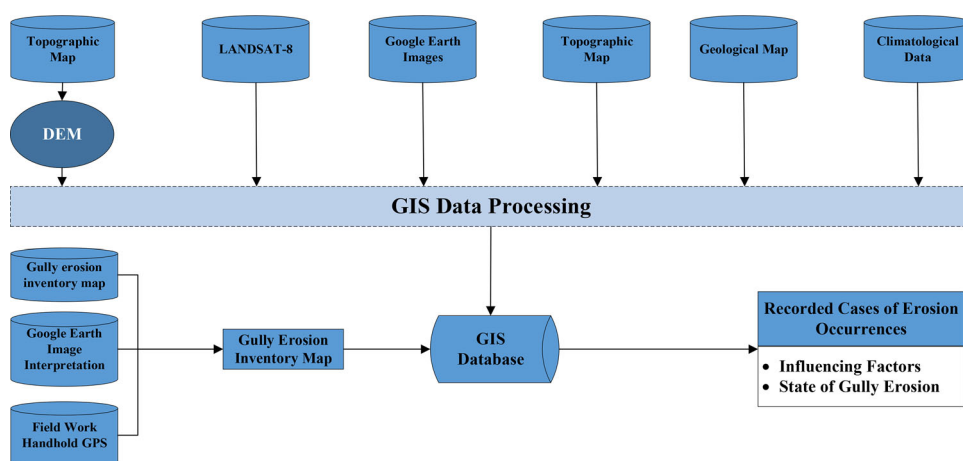


Figure 4. The established GIS database.

- i. Maximum tree depth (MTD): this parameter strongly affects the model accuracy as well as complexity. Generally, trees with shallow depths may to under fit the data; meanwhile, large values of MTD can lead to over fitted models.
- ii. Learning rate (η): This parameter used to prevent the over fitting problems through the step size shrinkage phase.
- iii. Minimum split loss (γ): This parameter governs the data partition on a leaf node of the tree.
- iv. Regularization coefficient (α): The parameter α affects the model regularization. Therefore, it helps to alleviate over fitting.
- v. The subsample ratio (r): The parameter r affects the number of data columns when each decision tree is constructed.
- vi. The number of decision trees (TN): This is the maximum number of individual decision trees in the ensemble.

Thus, the proper selection of the above hyper-parameters is very much necessary for the analysis of training and prediction outcome of the XGBoost model. In addition, due to the presence of infinite number of candidate solutions, it is a challenging task to identifying a suitable set of the hyper-parameters in XGBoost model. Because an optimization problem has been formulated during the hyper-parameter selection task (Sayed et al. 2018; Xiong et al. 2018; Deng et al. 2019; Lu, Tian, et al. 2019; Lu, Liao, et al. 2019), this study employs the GA metaheuristic to optimize the model construction phase of the XGBoost.

The general concept of the GA optimized XGBoost is presented in Figure 5(a). At the first generation ($g=0$), all of the model hyper-parameters are initialized randomly. The GA gradually evaluates the fitness of each population member to identify robust candidates during the searching of optimization within the model. After the optimization process terminates, the optimized XGBoost model used for gully erosion prediction can be constructed as illustrated in Figure 5(b).

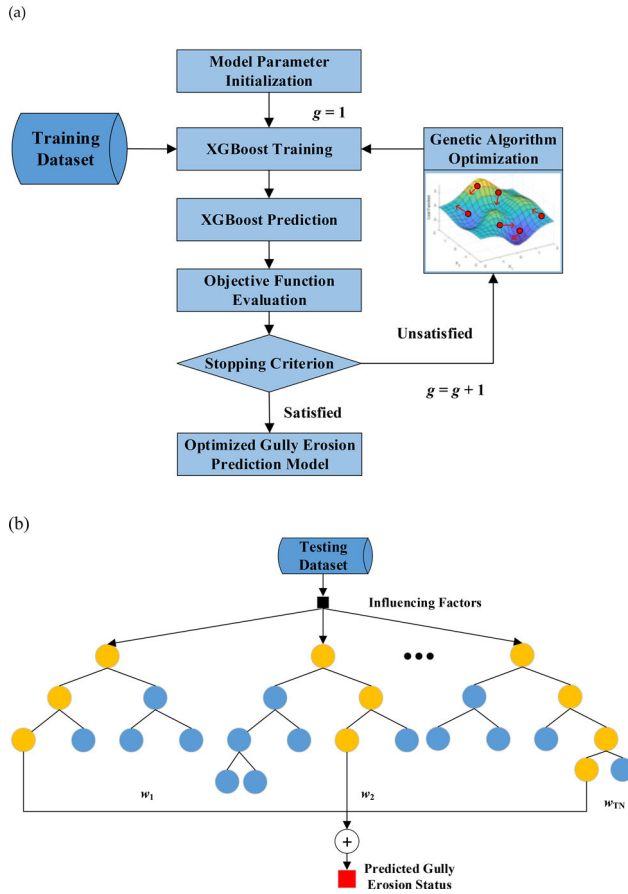


Figure 5. The proposed EG-XGBoost for predicting gully erosion potential: (a) model construction phase and (b) model prediction phase.

Moreover, an objective function is necessary to described for exhibit the suitability of each solution. The objective function (OF) of the EG-XGBoost algorithm is described as follows:

$$OF = \frac{1}{N} \sum_{n=1}^N (Y_A - Y_P)^2 \tag{5}$$

where, Y_A indicates actual output, Y_P indicates predicted output and N refers to the number of training data samples.

4.4. Validation and accuracy assessment

The validation and accuracy assessment in modelling is crucial for its proper evaluation in management studies (Chao et al. 2018; Wang et al. 2020). Without validation, the evaluation of machine learning models output result has less significance in reality. In the current study, several statistical indexes namely classification accuracy

(CAR), sensitivity, specificity, receiver operating characteristics (ROC) curve, Kappa coefficient and many were used to validation and accuracy assessment of the model's result. The classification accuracy (CAR) is computed as follows:

$$CAR = \frac{R_C}{R_A} \times 100(\%) \quad (6)$$

where, R_C represents the sample numbers being classified correctly and R_A denote the total number of samples.

Beside CAR, another validation performance tools i.e. the ROC curve (Moayedi et al. 2019; Nguyen et al. 2019) and Cohen's kappa coefficient (McHugh 2012) are also employed to express the overall performance of a classifier. Furthermore, the specificity, sensitivity and ROC-area under curve (AUC) are also widely employed for model assessment and the respective equations are as follows:

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (8)$$

$$AUC = \frac{(\sum TP + \sum TN)}{(P + N)} \quad (9)$$

where, TP refer to the true positive; TN denotes the true negative; FP represents the false positive and FN is the false negative (van Erkel and Pattynama 1998).

5. Experimental results

The results of the multicollinearity analysis on the relevance of gully erosion conditioning factors are reported in Table 5. As stated previously, collinearity among the predictors are investigated by VIF and TOL. It is worth reminding that TOL values is < 0.1 and VIF is > 10 indicate collinearity problems among the variables (Oh et al. 2017). As can be seen from the analysis outcomes, TOL and VIF values of the employed variables do not exceed such thresholds.

Moreover, the relevancy of conditioning factors is presented by using the Information Gain Ratio. Based on the results, highest Information Gain Ratio occupied by soil depth, followed by soil type TWI, lithology and NDVI. Influencing factors of geomorphology, elevation, plan curvature, slope, TPI and drainage density feature moderate values of Information Gain Ratio. Distance to stream and SPI has comparatively low relevancy. Notably, convergence index has a null value of Information Gain Ratio. Thus, this variable is excluded from the construction phase of the gully erosion prediction model.

Additionally, the performance of the newly constructed EG-XGBoost is benchmarked with those of the RF, SVM and LR. These benchmark models are constructed in Python. The grid search method (Hoang and Bui 2018) is used to identify the

Table 5. Multicollinearity diagnosis using VIF, TOL and Information Gain Ratio.

No.	Influencing factor	VIF	TOL	Information gain ratio
1	Soil depth	2.513	0.398	0.1779
2	Soil type	1.644	0.608	0.1683
3	TWI	1.362	0.734	0.1653
4	Lithology	1.739	0.575	0.1549
5	NDVI	1.271	0.787	0.1405
6	Geomorphology	1.395	0.717	0.1188
7	Elevation	5.419	0.185	0.1103
8	Plan curvature	1.994	0.502	0.1095
9	Slope	2.671	0.374	0.1094
10	TPI	2.339	0.428	0.1089
11	Drainage density	1.775	0.563	0.1074
12	Rainfall	3.237	0.309	0.0968
13	Landuse/Landcover	1.567	0.638	0.0859
14	Distance to road	1.098	0.911	0.0792
15	Soil EC	1.57	0.637	0.0649
16	Distance to stream	1.265	0.791	0.0287
17	SPI	1.484	0.674	0.0173
18	Convergence index	1.443	0.693	0

Table 6. Goodness-of-fit of the proposed GE-XGBoost, RF, SVM and LR model using the training dataset.

Statistical index	GE-XGBoost	RF	SVM	LR
True positive	346	329	301	275
True negative	364	362	342	308
False positive	25	42	70	96
False negative	7	9	29	63
Positive predictive value (%)	93.26	88.68	81.13	74.12
Negative predictive value (%)	98.11	97.57	92.18	83.02
Sensitivity (%)	98.02	97.34	91.21	81.36
Specificity (%)	93.57	89.60	83.01	76.24
Classification accuracy (%)	95.69	93.13	86.66	78.57
MSE	0.061	0.126	0.099	0.150
Kappa	0.914	0.736	0.733	0.571

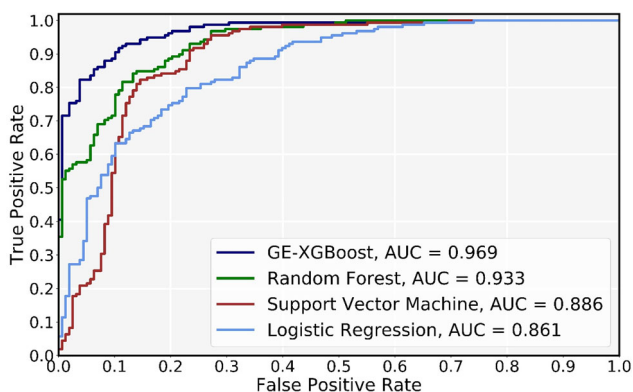
hyper-parameters within the SVM model. In addition, the tuning parameters of RF and LR are set based on the authors’ experience and trial-and-error processes.

The model performances are presented in [Tables 6 and 7](#) for training and testing phases, respectively. From the output result, it was noticed that the newly developed GE-XGBoost model has attained the most preferred accuracy for training (CAR = 95.69%) and testing (CAR = 89.56%) data predictions, followed by RF (testing CAR = 87.03%), SVM (testing CAR = 83.54%) and LR (testing CAR = 77.22%). Moreover, the ROCs of the prediction models are illustrated in [Figure 6](#). Based on the overall results of training and testing dataset, it is stated that GE-XGBoost is the best fit model for gully erosion prediction analysis.

In the next step of the current research, the proposed GE-XGBoost and other benchmark model are utilized to calculate the GES within this study area. The calculated susceptibility is then transformed to a grid format and opened in ArcGIS 10.4 package. By using the aforementioned MLA models, the GES maps ([Figure 7](#)) are obtained and each map classified into five categories for better understanding the variation. These five groups of classes are found from very high to very low. It is noted that the thresholds for categorizing these measured indices into the five classes

Table 7. Prediction performances of the proposed GE-XGBoost, RF, SVM and LR model using the testing dataset.

Statistical index	GE-XGBoost	RF	SVM	LR
True positive	136	130	120	114
True negative	147	145	144	130
False positive	22	28	38	44
False negative	11	13	14	28
Positive predictive value (%)	86.08	82.28	75.95	72.15
Negative predictive value (%)	93.04	91.77	91.14	82.28
Sensitivity (%)	92.52	90.91	89.55	80.28
Specificity (%)	86.98	83.82	79.12	74.71
Classification accuracy (%)	89.56	87.03	83.54	77.22
MSE	0.095	0.135	0.123	0.16
Kappa	0.791	0.684	0.671	0.54
AUC	0.969	0.933	0.886	0.870

**Figure 6.** ROCs of the prediction models.

mentioned above were reached by using the natural break classification system of Jenk mentioned in Bednarik et al. (2010).

6. Discussion

Among the several types of water induced soil erosion, gully erosion is the most destructive form, in terms of its negative impact on humankind throughout the world (Arabameri, Asadi Nalivan, et al. 2020). The water induced gully erosion is responsible for several natural causes but recently, human interventions on the environment and associated changes of the equilibrium of ecosystem also influences the formation and development of gullies (Chaplot et al. 2005; Band, Janizadeh, Chandra Pal, Saha, Chakraborty, Shokri, et al. 2020). This type of erosion is responsible for destroying the environment in two ways: first, it causes extensive land degradation through surface and sub-surface soil erosion and associated result i.e. sedimentation into the downstream area, degradation of agricultural land and adverse effect on human's life; second, the gradual reduction of groundwater recharge as the rate of surface runoff increases (Kou et al. 2016). Extensive literature study has shown that gullies exist in most semiarid and arid landscapes around the world, with substantial morphological activities and dynamics (Arabameri, Yamani, Pradhan, et al. 2019; Gayen et al. 2019).

In comparison, semi-arid climate and precipitation regimes promote soil erosion through low vegetation cover and recurring heavy rainfall events. Therefore, the adverse effect of climatic conditions is very much significant for large scale land degradation in the form of gully formation and development. Keeping in view the adverse damages of our environment due to gully erosion, several research works has been conducted to find out the responsible factors of gully erosion and prepared the map of susceptible areas. Hence, it is indeed necessary to manage the adverse effect of gully induced soil erosion through sustainable measurement techniques.

The present study area of Kohpayeh-Sagzi watershed, Iran having several problems related to gully erosion; therefore, it is the main obstacle for sustainable land uses practices and associated land management strategies. Thus, it is very much necessary to use suitable models for optimal prediction of GES and apply several suitable measurement techniques for soil and water management. In this regard, here we used machine learning algorithm (MLA) of XGBoost, GA, RF, SVM and LR for modelling GES.

Given the shortcomings and limitations of each of the stand alone models (Li et al. 2019; Feng et al. 2020; Qian et al. 2020; Qu et al. 2020; Shi et al. 2020), scientists have proposed and developed integrated methods to overcome their disadvantages and increase their efficiency (Cao, Dong, et al. 2020; Peng et al. 2020; Liu et al. 2021). In this research, A novel ensemble of GE-XGBoost was proposed for better prediction analysis of gully erosion than the stand-alone model. In this study, 18 conditioning factors was used which is responsible for gully erosion based on several geo-environmental conditions. To accurately represent the output result here, we used multicollinearity test of VIF and TOL techniques and information gain ratio (IGR) to proper understand the relationship among the variables. The result of VIF, TOL and IGR indicates that there is no multi-collinearity problem among the variables as all the variable's value is within the threshold value of multicollinearity. In VIF, among the 18 variables, rainfall is the highest value of 3.237 and distance to road is the low value of 1.265. In the case of TOL, distance to road and rainfall consist of the highest and lowest value of 0.911 and 0.309, respectively. On the other side, the result of IGR is shown that soil depth has high priority with the value of 0.1779 and followed by soil type (0.1683), TWI (0.1653) and lithology (0.1549) and zero value has been found in convergence index factor. Thus, in the current study area soil play an important role for the occurrences and further development of gullies, as the soil texture determine the rate of infiltration, associated surface runoff and soil resistance (Deng et al. 2015). The hydrological process of a terrain is quantify through TWI and determine the erosive power of water and greatly influences on gully erosion (Rahmati, Pourghasemi, et al. 2016). Whereas, the characteristics of lithology determine the water percolation rate and significantly influences on gully initiation and extension (Band, Janizadeh, Chandra Pal, Saha, Chakraborty, Shokri, et al. 2020). The same result i.e. soil characteristics, hydrological and lithological properties has been highlighted by others established research work. Alongside, elevation (0.1103), slope (0.1094), drainage density (0.1074), rainfall (0.0968) and LULC (0.0859) also play significant role for development of gullies and associated land degradation problems.

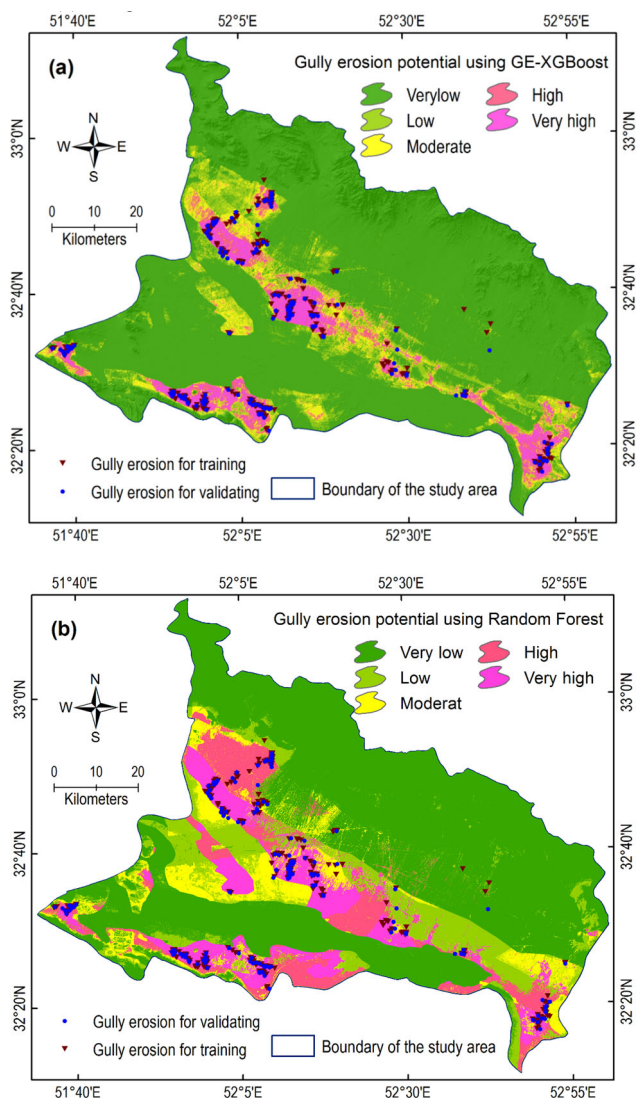


Figure 7. Gully erosion susceptibility maps: (a) using the GE-XGBoost, (b) using the RF, (c) using SVM and (d) using LR.

Several statistical indices have been used for prediction performance of the model's output. In this study, we have carried out goodness-of-fit and prediction performance of several models used for gully erosion modelling. The goodness-of-fit among the four models of GE-XGBoost, RF, SVM and LR, statistical index of true positive, true negative, positive predictive value, negative predictive value, sensitivity, specificity, classification accuracy, MSE and kappa with their highest value of 346, 364, 93.26, 98.11, 98.02, 93.57, 95.69, 0.061 and 0.914, respectively, is shows GE-XGBoost is the best fit model, in the training phase. In the case of false positive and false negative with their highest value of 96 and 63 indicates LR is the goodness-of-fit model. The result of predictive performance in all the models have shown high accuracy, but the

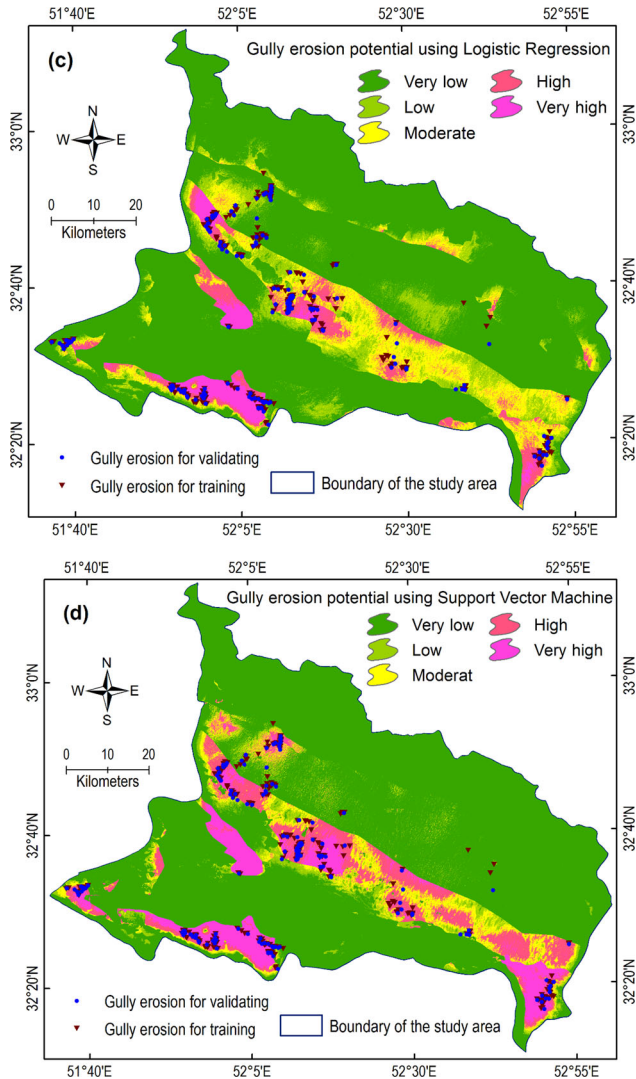


Figure 7. (Continued).

ensemble of GE-XGBoost is the top most optimal with the AUC value is 0.969 and followed by RF (AUC = 0.933), SVM (AUC = 0.886) and LR (AUC = 0.870) in the testing phase. The other statistical indices used in this study for validation and accuracy assessment has also shows that GE-XGBoost is the best prediction model for GES flowed by RF, SVM and LR. However, false positive and true positive has shown that LR is the best model with their value of 44 and 28, respectively, followed by SVM, RF and GE-XGBoost. Finally, it is concluded that assessment of gully erosion in proper way always faced uncertainties due to lack of knowledge regarding physical condition, complex function of the model and spatial inconsistency (Arabameri, Pradhan, and Rezaei 2019b). It is also very much difficult to proper validation of the model in an accurate way. It has also been stated that the framework of the model, the input data structure are accurate and proper validation techniques is used than

the uncertainty of the prediction result will be reduced (Rojas et al. 2010). Thus, the novel ensemble of GE-XGBoost with their proper validation performance resolved the problems related to gully erosion in various regions of the world.

The worldwide soil erosion phenomena and associated loss of soil fertility has been gradually decreasing the sustainability of environment and national economy (Lal 2003). Therefore, measurement of soil sustainability is an important goal among the farmers and decision makers (Novara et al. 2021). Several kinds of soil erosion control strategies should be implemented in various scale and regions to optimal solution for this type of devastating problems. Various literature studies have been shown that cover crops have positive impact in the control of soil erosion and has been considered for soil erosion management in several regions across the world (López-Vicente et al. 2020; Rodrigo-Comino et al. 2020; Novara et al. 2021). In general, cover crops protected the soil surface from direct falling of raindrop and reduced the runoff velocity and associated erosional activities. Alongside, vegetation cover is also improved the soil quality and play a key role to reduced soil losses through plant cover and root system (Van Hall et al. 2017). Hence, in this study area also vegetation coverage, construction of small check dams and proper land use planning should be implemented to mitigate and control the soil erosion problem.

7. Conclusion

The present research work has been carried out by using a new methodology of GIS technology, remote sensing and MLA for spatial mapping of gully erosion hazard. GIS technology and remote sensing are employed to detect erosion locations and construct a database containing erosion influencing factors. The GIS database established in the present study takes into account the historical cases of gully erosion events and 18 conditioning factors. Elevation, slope, PC, CI, SPI, TWI, TPI, distance to stream, drainage density, distance to road, rainfall, NDVI, soil type, soil depth, soil EC, geomorphology, lithology and LULC are utilized as predicting variables used to determine the status of non-erosion or erosion for a certain area within the study region.

A novel machine learning ensemble method that combined the XGBoost and GA, named as GE-XGBoost, is developed in the study. The division of gully from non-gully was carried by construction of decision boundary through XGBoost. In addition, GA is used to optimize the performance of the XGBoost method. Experimental outcomes demonstrate that the hybrid method can deliver predictive performance which is superior to those of RF, SVM and LR. Therefore, the proposed GE-XGBoost can be an effective model to assist local authorities in establishing effective countermeasures against gully erosion and proper planning of land-use.

The novel ensemble model of GE-XGBoost which is used in this study area can also be extended in future for predicting gully erosion in several parts of the world and spatial modelling of other natural hazards such as landslide or flash flood. Hence, every research study has some limitations, thus this study also not included the hydrological modelling for gully system analysis and we need emphasize the same for future studies. Moreover, the proposed method's performance used in this work

can be improving enhancing with other advanced metaheuristic optimization algorithms for an appropriate potential research direction.

Data availability statement

The data that support the findings of this study are available on request from the authors.

Disclosure statement

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

Funding

This research was supported by the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) and Project of Environmental Business Big Data Platform and Center Construction funded by the Ministry of Science and ICT.

ORCID

Alireza Arabameri  <http://orcid.org/0000-0002-1142-1666>

Romulus Costache  <http://orcid.org/0000-0002-6876-8572>

Fatemeh Rezaie  <http://orcid.org/0000-0003-1771-6753>

Biswajeet Pradhan  <http://orcid.org/0000-0001-9863-2054>

Saro Lee  <http://orcid.org/0000-0003-0409-8263>

Nhat-Duc Hoang  <http://orcid.org/0000-0001-9450-4637>

References

- Abedini M, Ghasemian B, Shirzadi A, Shahabi H, Chapi K, Pham BT, Ahmad BB, Bui DT. 2019. A novel hybrid approach of Bayesian Logistic Regression and its ensembles for landslide susceptibility assessment. *Geocarto Int.* 34(13):1427–1457.
- Agresti A. 2003. *Categorical data analysis*. Wiley, Hoboken, New Jersey, United States.
- Alin A. 2010. Multicollinearity. *WIREs Comp Stat.* 2(3):370–374.
- Amare S, Keesstra S, van der Ploeg M, Langendoen E, Steenhuis T, Tilahun S. 2019. Causes and controlling factors of valley bottom gullies. *Land.* 8(9):141.
- Amjad MK, Butt SI, Kousar R, Ahmad R, Agha MH, Faping Z, Anjum N, Asgher U. 2018. Recent research trends in genetic algorithm based flexible job shop scheduling problems. *Math Prob Eng [Internet]*. 2018:1–32.
- Arabameri A, Asadi Nalivan O, Chandra Pal S, Chakraborty R, Saha A, Lee S, Pradhan B, Tien Bui D. 2020. Novel machine learning approaches for modelling the gully erosion susceptibility. *Remote Sens.* 12(17):2833.
- Arabameri A, Karimi-Sangchini E, Pal SC, Saha A, Chowdhuri I, Lee S, Tien Bui D. 2020. Novel credal decision tree-based ensemble approaches for predicting the landslide susceptibility. *Remote Sens.* 12(20):3389.
- Arabameri A, Pourghasemi HR. 2019. 13 - Spatial modeling of gully erosion using linear and quadratic discriminant analyses in GIS and R. In: Pourghasemi HR, Gokceoglu C, editors. *Spatial modeling in GIS and R for earth and environmental sciences [Internet]*. Elsevier; p.

- 299–321. [accessed 2020 Oct 13]. <http://www.sciencedirect.com/science/article/pii/B9780128152263000132>. Elsevier, Amsterdam, Netherlands
- Arabameri A, Pradhan B, Rezaei K, Conoscenti C. 2019. Gully erosion susceptibility mapping using GIS-based multi-criteria decision analysis techniques. *CATENA*. 180:282–297.
- Arabameri A, Pradhan B, Rezaei K, Yamani M, Pourghasemi HR, Lombardo L. 2018. Spatial modelling of gully erosion using evidential belief function, logistic regression, and a new ensemble of evidential belief function–logistic regression algorithm. *Land Degrad Dev*. 29(11):4035–4049.
- Arabameri A, Pradhan B, Rezaei K. 2019a. Spatial prediction of gully erosion using ALOS PALSAR data and ensemble bivariate and data mining models. *Geosci J*. 23(4):669–686.
- Arabameri A, Pradhan B, Rezaei K. 2019b. Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS. *J Environ Manage*. 232:928–942.
- Arabameri A, Rezaei K, Pourghasemi HR, Lee S, Yamani M. 2018. GIS-based gully erosion susceptibility mapping: a comparison among three data-driven models and AHP knowledge-based technique. *Environ Earth Sci*. 77(17):628.
- Arabameri A, Yamani M, Pradhan B, Melesse A, Shirani K, Tien Bui D. 2019. Novel ensembles of COPRAS multi-criteria decision-making with logistic regression, boosted regression tree, and random forest for spatial prediction of gully erosion susceptibility. *Sci Total Environ*. 688:903–916.
- Ayele GK, Gessess AA, Addisie MB, Tilahun SA, Tenessa DB, Langendoen EJ, Steenhuis TS, Nicholson CF. 2015. The economic cost of upland and gully erosion on subsistence agriculture for a watershed in the Ethiopian highlands [Internet]. *Afr J Agric Resour Econ*. 10(4): 265–278.
- Band SS, Janizadeh S, Chandra Pal S, Saha A, Chakraborty R, Melesse AM, Mosavi A. 2020. Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. *Remote Sens*. 12(21):3568.
- Band SS, Janizadeh S, Chandra Pal S, Saha A, Chakraborty R, Shokri M, Mosavi A. 2020. Novel ensemble approach of deep learning neural network (DLNN) model and particle swarm optimization (PSO) algorithm for prediction of gully erosion susceptibility. *Sensors*. 20(19):5609.
- Bednarik M, Magulová B, Matys M, Marschalko M. 2010. Landslide susceptibility assessment of the Kralovany–Liptovský Mikuláš railway case study. *Phys Chem Earth Parts A/B/C*. 35(3–5):162–171.
- Berhe AA, Harden JW, Torn MS, Kleber M, Burton SD, Harte J. 2012. Persistence of soil organic matter in eroding versus depositional landform positions. *J Geophys Res*. 117(G2): 1–16.
- Bouaziz M, Leidig M, Gloaguen R. 2011. Optimal parameter selection for qualitative regional erosion risk monitoring: a remote sensing study of SE Ethiopia. *Geosci Front*. 2(2):237–245.
- Breiman L, Friedman J, Stone CJ, Olshen RA. 1984. Classification and regression trees. Taylor & Francis, Milton Park, Abingdon-on-Thames, Oxfordshire United Kingdom.
- Breiman L. 2001. Random forests. *Mach Learn*. 45(1):5–32.
- Cao B, Dong W, Lv Z, Gu Y, Singh S, Kumar P. 2020. Hybrid microgrid many-objective sizing optimization with fuzzy decision. *IEEE Trans Fuzzy Syst*. 28(11):2702–2710.
- Cao B, Wang X, Zhang W, Song H, Lv Z. 2020. A many-objective optimization model of industrial internet of things based on private blockchain. *IEEE Netw*. 34(5):78–83.
- Cao B, Zhao J, Lv Z, Gu Y, Yang P, Halgamuge SK. 2020. Multiobjective evolution of fuzzy rough neural network via distributed parallelism for stock prediction. *IEEE Trans Fuzzy Syst*. 28(5):939–952.
- Chang Y-C, Chang K-H, Wu G-J. 2018. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Appl Soft Comput*. 73:914–920.
- Chao L, Zhang K, Li Z, Zhu Y, Wang J, Yu Z. 2018. Geographically weighted regression based methods for merging satellite and gauge precipitation. *J Hydrol*. 558:275–289.

- Chaplot VAM, Rumpel C, Valentin C. 2005. Water erosion impact on soil and carbon redistributions within uplands of Mekong River. *Global Biogeochem Cycles*. 19(4): 1–13. [accessed 2020 Sep 10] <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GB002493>.
- Chen H, Chen A, Xu L, Xie H, Qiao H, Lin Q, Cai K. 2020. A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources. *Agric Water Manage*. 240:106303.
- Chen T, Guestrin C. 2016. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; p. 785–794; Aug 13–17, San Francisco, USA.
- Chen W, Chen X, Peng J, Panahi M, Lee S. 2021. Landslide susceptibility modeling based on ANFIS with teaching-learning-based optimization and Satin bowerbird optimizer. *Geosci Front*. 12(1):93–107.
- Chen W, Chen Y, Tsangaratos P, Ilia I, Wang X. 2020. Combining evolutionary algorithms and machine learning models in landslide susceptibility assessments. *Remote Sens*. 12(23): 3854.
- Chen W, Li Y. 2020. GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. *CATENA*. 195(104777):104777.
- Chen W, Shahabi H, Zhang S, Khosravi K, Shirzadi A, Chapi K, Pham BT, Zhang T, Zhang L, Chai H, et al. 2018. Landslide susceptibility modeling based on GIS and novel bagging-based kernel logistic regression. *Appl Sci*. 8(12):2540.
- Chen W, Zhao X, Tsangaratos P, Shahabi H, Ilia I, Xue W, Wang X, Ahmad BB. 2020. Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. *J Hydrol*. 583:124602.
- Chen Y, Li J, Lu H, Yan P. 2021. Coupling system dynamics analysis and risk aversion programming for optimizing the mixed noise-driven shale gas-water supply chains. *J Cleaner Prod*. 278:123209.
- Cheng X, He L, Lu H, Chen Y, Ren L. 2016. Optimal water resources management and system benefit for the Marcellus shale-gas reservoir in Pennsylvania and West Virginia. *J Hydrol*. 540:412–422.
- Chowdhuri I, Pal SC, Arabameri A, Saha A, Chakraborty R, Blaschke T, Pradhan B, Band SS. 2020. Implementation of artificial intelligence based ensemble models for gully erosion susceptibility assessment. *Remote Sens*. 12(21):3620.
- Conoscenti C, Agnesi V, Cama M, Caraballo-Arias NA, Rotigliano E. 2018. Assessment of gully erosion susceptibility using multivariate adaptive regression splines and accounting for terrain connectivity. *Land Degrad Dev*. 29(3):724–736.
- Conoscenti C, Angileri S, 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.
- Dang V-H, Dieu TB, Tran X-L, Hoang N-D. 2019. Enhancing the accuracy of rainfall-induced landslide prediction along mountain roads with a GIS-based random forest classifier. *Bull Eng Geol Environ*. 78(4):2835–2849.
- De Reu J, Bourgeois J, Bats M, Zwertvaegher A, Gelorini V, De Smedt P, Chu W, Antrop M, De Maeyer P, Finke P, et al. 2013. Application of the topographic position index to heterogeneous landscapes. *Geomorphology*. 186:39–49.
- Deng L, Zeng G, Fan C, Lu L, Chen X, Chen M, Wu H, He X, He Y. 2015. Response of rhizosphere microbial community structure and diversity to heavy metal co-pollution in arable soil. *Appl Microbiol Biotechnol*. 99(19):8259–8269.
- Deng S, Wang X, Zhu Y, Lv F, Wang J. 2019. Hybrid grey wolf optimization algorithm-based support vector machine for groutability prediction of fractured rock mass. *J Comput Civ Eng*. 33(2):04018065.
- Dlapa P, Chrenková K, Mataix-Solera J, Šimkovic I. 2012. Soil profile improvement as a by-product of gully stabilization measures. *CATENA*. 92:155–161.

- Droogers P, Kite G. 2002. Remotely sensed data used for modelling at different hydrological scales. *Hydrol Process*. 16(8):1543–1556.
- El Maaoui MA, Sfar Felfoul M, Boussema MR, Smane MH. 2012. Sediment yield from irregularly shaped gullies located on the Fortuna lithologic formation in semi-arid area of Tunisia. *CATENA*. 93:97–104.
- Feng W, Lu H, Yao T, Yu Q. 2020. Drought characteristics and its elevation dependence in the Qinghai–Tibet plateau during the last half-century. *Sci Rep*. 10(1):14323.
- Fistikoglu O, Harmancioglu NB. 2002. Integration of GIS with USLE in assessment of soil erosion. *Water Resour Manag*. 16(6):447–467.
- Friedman JH. 2001. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 29(5):1189–1232.
- Gayen A, Pourghasemi HR, Saha S, Keesstra S, Bai S. 2019. Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Sci Total Environ*. 668:124–138.
- Gessesse B, Bewket W, Bräuning A. 2015. Model-based characterization and monitoring of runoff and soil erosion in response to land use/land cover changes in the Modjo Watershed, Ethiopia. *Land Degrad Dev*. 26(7):711–724.
- Gitas IZ, Douros K, Minakou C, Silleos GN, Karydas CG. 2009. Multi-temporal soil erosion risk assessment in N. Chalkidiki using a modified USLE raster model. *EARSeL eProc*. 8(1): 40–52.
- Goldberg DE. 1989. Genetic algorithms in search, optimization, and machine learning. Wesley, p.36. Boston, Massachusetts, United States.
- Gormley MA, Crowe RP, Bentley MA, Levine R. 2016. A national description of violence toward emergency medical services personnel. *Prehosp Emerg Care*. 20(4):439–447.
- Guisan A, Weiss SB, Weiss AD. 1999. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol*. 143(1):107–122.
- Gusain K, Gupta A, Popli B. 2018. Transition-aware human activity recognition using extreme gradient boosted decision trees. In: Choudhary RK, Mandal JK, Bhattacharyya D, editors. *Advanced computing and communication technologies*. Singapore: Springer; p. 41–49.
- Hamel LH. 2011. Knowledge discovery with support vector machines. Wiley, Hoboken, New Jersey, United States.
- Han C, Zhang B, Chen H, Wei Z, Liu Y. 2019. Spatially distributed crop model based on remote sensing. *Agric Water Manage*. 218:165–173.
- He L, Chen Y, Li J. 2018. A three-level framework for balancing the tradeoffs among the energy, water, and air-emission implications within the life-cycle shale gas supply chains. *Resour Conserv Recycl*. 133:206–228.
- He L, Shao F, Ren L. 2020. Sustainability appraisal of desired contaminated groundwater remediation strategies: an information-entropy-based stochastic multi-criteria preference model. *Environ Dev Sustain*. 4:1–21.
- He L, Shen J, Zhang Y. 2018. Ecological vulnerability assessment for ecological conservation and environmental management. *J Environ Manage*. 206:1115–1125.
- Helming K, Auzet A-V, Favis-Mortlock D. 2005. Soil erosion patterns: evolution, spatio-temporal dynamics and connectivity. *Earth Surf Process Landforms*. 30(2):131–132.
- Hoang N-D, Bui DT. 2018. Predicting earthquake-induced soil liquefaction based on a hybridization of kernel Fisher discriminant analysis and a least squares support vector machine: a multi-dataset study. *Bull Eng Geol Environ*. 77(1):191–204.
- Hoang N-D, Nguyen Q-L, Tran X-L. 2019. Automatic detection of concrete spalling using piecewise linear stochastic gradient descent logistic regression and image texture analysis. *Complexity*. 2019:e5910625.
- Hosseinalizadeh M, Kariminejad N, Chen W, Pourghasemi HR, Alinejad M, Mohammadian Behbahani A, Tiefenbacher JP. 2019. Spatial modelling of gully headcuts using UAV data and four best-first decision classifier ensembles (BFTree, Bag-BFTree, RS-BFTree, and RF-BFTree). *Geomorphology*. 329:184–193.

- Jaafari A, Panahi M, Pham BT, Shahabi H, Bui DT, Rezaie F, Lee S. 2019. Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *CATENA*. 175:430–445.
- Kavzoglu T, Colkesen I, Sahin EK. 2019. Machine learning techniques in landslide susceptibility mapping: a survey and a case study. In: Pradhan SP, Vishal V, Singh TN, editors. *Landslides: theory, practice and modelling* [Internet]. Cham: Springer International Publishing; p. 283–301. [accessed 2020 Oct 13]. https://doi.org/10.1007/978-3-319-77377-3_13
- Keesstra S, Mol G, De Leeuw J, Okx J, Molenaar C, De Cleen M, Visser S. 2018. Soil-related sustainable development goals: four concepts to make land degradation neutrality and restoration work. *Land*. 7(4):133.
- Keesstra SD, Bouma J, Wallinga J, Tiftonell P, Smith P, Cerdà A, Montanarella L, Quinton JN, Pachepsky Y, Van Der Putten WH. 2016. The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals. *Soil*. 2(2):111–128.
- Kilinc M, Caicedo JM. 2019. Finding plausible optimal solutions in engineering problems using an adaptive genetic algorithm. *Adv Civ Eng*. 2019:e7475156.
- Kornejady A, Ownegh M, Bahremand A. 2017. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. *CATENA*. 152:144–162.
- Kou M, Jiao J, Yin Q, Wang N, Wang Z, Li Y, Yu W, Wei Y, Yan F, Cao B. 2016. Successional trajectory over 10 years of vegetation restoration of abandoned slope croplands in the Hill-Gully Region of the Loess Plateau. *Land Degrad Dev*. 27(4):919–932.
- Lal R. 2003. Soil erosion and the global carbon budget. *Environ Int*. 29(4):437–450.
- Li P. 2010. Robust logitboost and adaptive base class (ABC) logitboost. *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence, UAI 2010* [Internet]. p. 302–311. [accessed 2020 Oct 13]; <https://www.researchwithrutgers.com/en/publications/robust-logitboost-and-adaptive-base-class-abc-logitboost>. Jul 8–11, Catalina Island, California
- Li T, Xu M, Zhu C, Yang R, Wang Z, Guan Z. 2019. A deep learning approach for multi-frame in-loop filter of HEVC. *IEEE Trans Image Process*. 28(11):5663–5678.
- Liu S, Yu W, Chan FTS, Niu BA. 2021. variable weight-based hybrid approach for multi-attribute group decision making under interval-valued intuitionistic fuzzy sets. *Int J Intell Syst*. 36(2):1015–1052.
- Lombardo L, Mai PM. 2018. Presenting logistic regression-based landslide susceptibility results. *Eng Geol*. 244:14–24.
- López-Vicente M, Calvo-Seas E, Álvarez S, Cerdà A. 2020. Effectiveness of cover crops to reduce loss of soil organic matter in a rainfed vineyard. *Land*. 9(7):230.
- Lu H, Tian P, He L. 2019. Evaluating the global potential of aquifer thermal energy storage and determining the potential worldwide hotspots driven by socio-economic, geo-hydrologic and climatic conditions. *Renew Sustain Energy Rev*. 112:788–796.
- Lu J, Liao X, Li S, Ouyang H, Chen K, Huang B. 2019. An effective ABC-SVM approach for surface roughness prediction in manufacturing processes. *Complexity*. 2019:e3094670.
- Lv Z, Qiao L. 2020. Deep belief network and linear perceptron based cognitive computing for collaborative robots. *Appl Soft Comput*. 92:106300.
- McHugh ML. 2012. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 22(3): 276–282.
- Mekonnen M, Keesstra SD, Baartman JEM, Stroosnijder L, Maroulis J. 2017. Reducing sediment connectivity through man-made and natural sediment sinks in the Minizr Catchment, Northwest Ethiopia. *Land Degrad Dev*. 28(2):708–717.
- Meliho M, Khattabi A, Mhammdi N. 2018. A GIS-based approach for gully erosion susceptibility modelling using bivariate statistics methods in the Ourika watershed. *Environ Earth Sci*. 77(18):655.
- Mirjalili S, Song Dong J, Sadiq AS, Faris H. 2020. Genetic algorithm: theory, literature review, and application in image reconstruction. In: Mirjalili S, Song Dong J, Lewis A, editors. *Nature-inspired optimizers: theories, literature reviews and applications* [Internet]. Cham:

- Springer International Publishing; p. 69–85. [accessed 2020 Oct 13]. https://doi.org/10.1007/978-3-030-12127-3_5
- Moayedi H, Tien Bui D, Gör M, Pradhan B, Jaafari A. 2019. The feasibility of three prediction techniques of the artificial neural network, adaptive neuro-fuzzy inference system, and hybrid particle swarm optimization for assessing the safety factor of cohesive slopes. *IJGI*. 8(9):391.
- Ngoc Thach N, Bao-Toan Ngo D, Xuan-Canh P, Hong-Thi N, Hang Thi B, Nhat-Duc H, Dieu TB. 2018. Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: a comparative study. *Ecol Inf*. 46:74–85.
- Nguyen H, Mehrabi M, Kalantar B, Moayedi H, Abdullahi MM. 2019. Potential of hybrid evolutionary approaches for assessment of geo-hazard landslide susceptibility mapping. *Geomatics Nat Hazards Risk*. 10(1):1667–1693.
- Novara A, Cerda A, Barone E, Gristina L. 2021. Cover crop management and water conservation in vineyard and olive orchards. *Soil Tillage Res*. 208:104896.
- Novara A, Keesstra S, Cerdà A, Pereira P, Gristina L. 2016. Understanding the role of soil erosion on CO₂-C loss using (13)c isotopic signatures in abandoned Mediterranean agricultural land. *Sci Total Environ*. 550:330–336.
- Nyssen J, Poesen J, Moeyersons J, Luyten E, Veyret-Picot M, Deckers J, Haile M, Govers G. 2002. Impact of road building on gully erosion risk: a case study from the Northern Ethiopian Highlands. *Earth Surf Process Landforms*. 27(12):1267–1283.
- Nyssen J, Poesen J, Veyret -Picot M, Moeyersons J, Haile M, Deckers J, Dewit J, Naudts J, Tekla K, Govers G. 2006. Assessment of gully erosion rates through interviews and measurements: a case study from northern Ethiopia. *Earth Surf Process Landforms*. 31(2):167–185.
- Oh H-J, Lee S, Hong S-M. 2017. Landslide susceptibility assessment using frequency ratio technique with iterative random sampling. *J Sens*. 2017:e3730913.
- Pal SC, Arabameri A, Blaschke T, Chowdhuri I, Saha A, Chakraborty R, Lee S, Band SS. 2020. Ensemble of machine-learning methods for predicting gully erosion susceptibility. *Remote Sens*. 12(22):3675.
- Pal SC, Chakraborty R. 2019. Simulating the impact of climate change on soil erosion in subtropical monsoon dominated watershed based on RUSLE, SCS runoff and MIROC5 climatic model. *Adv Space Res*. 64(2):352–377.
- Peng S, Zhang Z, Liu E, Liu W, Qiao W. 2020. A new hybrid algorithm model for prediction of internal corrosion rate of multiphase pipeline. *J Nat Gas Sci Eng*. 64:103716.
- Pimentel D. 2006. Soil erosion: a food and environmental threat. *Environ Dev Sustain*. 8(1): 119–137.
- Poesen J, Nachtergaele J, Verstraeten G, Valentin C. 2003. Gully erosion and environmental change: importance and research needs. *CATENA*. 50(2–4):91–133.
- Poesen J, Vandaele K, van Wesemael B. 1998. Gully erosion: importance and model implications. In: Boardman J, Favis-Mortlock D, editors. *Modelling soil erosion by water*. Berlin, Heidelberg: Springer; p. 285–311.
- Qian J, Feng S, Tao T, Hu Y, Li Y, Chen Q, Zuo C. 2020. Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3D shape measurement. *APL Photonics*. 5(4):046105.
- Qu S, Han Y, Wu Z, Raza H. 2020. Consensus modeling with asymmetric cost based on data-driven robust optimization. *Group Decis Negot*. 19:1–38.
- Rahmati O, Haghizadeh A, Pourghasemi HR, Noormohamadi F. 2016. Gully erosion susceptibility mapping: the role of GIS-based bivariate statistical models and their comparison. *Nat Hazards*. 82(2):1231–1258.
- Rahmati O, Pourghasemi HR, Melesse AM. 2016. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. *Catena*. 137:360–372.

- Rahmati O, Tahmasebipour N, Haghizadeh A, Pourghasemi HR, Feizizadeh B. 2017. 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.
- Rodrigo-Comino J, Keesstra S, Cerdà A. 2018. Soil erosion as an environmental concern in vineyards: the case study of Celler del Roure, Eastern Spain, by means of rainfall simulation experiments. *Beverages.* 4(2):31.
- Rodrigo-Comino J, Terol E, Mora G, Giménez-Morera A, Cerdà A. 2020. *Vicia sativa* Roth. Can reduce soil and water losses in recently planted vineyards (*Vitis vinifera* L.). *Earth Syst Environ.* 4(4):827–842.
- Rojas R, Kahunde S, Peeters L, Batelaan O, Feyen L, Dassargues A. 2010. Application of a multi-model approach to account for conceptual model and scenario uncertainties in ground-water modelling. *J Hydrol.* 394(3–4):416–435.
- Roy P, Chandra Pal S, Arabameri A, Chakraborty R, Pradhan B, Chowdhuri I, Lee S, Tien Bui D. 2020. Novel ensemble of multivariate adaptive regression spline with spatial logistic regression and boosted regression tree for gully erosion susceptibility. *Remote Sens.* 12(20):3284.
- 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.
- Sayed GI, Soliman M, Hassanien AE. 2018. Modified optimal foraging algorithm for parameters optimization of support vector machine. In: Hassanien AE, Tolba MF, Elhoseny M, Mostafa M, editors. *The international conference on advanced machine learning technologies and applications (aMLTA2018)*. Cham: Springer International Publishing; p. 23–32.
- Shi K, Wang J, Tang Y, Zhong S. 2020. Reliable asynchronous sampled-data filtering of T–S fuzzy uncertain delayed neural networks with stochastic switched topologies. *Fuzzy Sets Syst.* 381:1–25.
- Shit PK, Nandi AS, Bhunia GS. 2015. Soil erosion risk mapping using RUSLE model on Jhargram Sub-Division at West Bengal in India. *Model Earth Syst Environ.* 1(3):28.
- Shahkamari S, Haghizadeh A, Zeinivand H, Tahmasebipour N, Rahmati O. 2018. Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. *Geocarto Int.* 33(9):927–941.
- Slimane AB, Raclot D, Evrard O, Sanaa M, Lefevre I, Bissonnais YL. 2016. Relative contribution of rill/interrill and gully/channel erosion to small reservoir siltation in Mediterranean environments. *Land Degrad Dev.* 27(3):785–797.
- Tehrany MS, Pradhan B, Mansor S, Ahmad N. 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *CATENA.* 125:91–101.
- Tian P, Lu H, Feng W, Guan Y, Xue Y. 2020. Large decrease in streamflow and sediment load of Qinghai–Tibetan Plateau driven by future climate change: a case study in Lhasa River Basin. *CATENA.* 187:104340.
- Tien Bui D, Hoang N-D, Pham T-D, Ngo P-TT, Hoa PV, Minh NQ, Tran X-T, Samui P. 2019. A new intelligence approach based on GIS-based multivariate adaptive regression splines and metaheuristic optimization for predicting flash flood susceptible areas at high-frequency tropical typhoon area. *J Hydrol.* 575:314–326.
- Tien Bui D, Shirzadi A, Shahabi H, Chapi K, Omidavr E, Pham BT, Talebpour Asl D, Khaleidian H, Pradhan B, Panahi M, et al. 2019. A novel ensemble artificial intelligence approach for gully erosion mapping in a semi-arid watershed (Iran). *Sensors.* 19(11):2444.
- van Erkel AR, Pattynama PMT. 1998. Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. *Eur J Radiol.* 27(2):88–94.
- Van Hall RL, Cammeraat LH, Keesstra SD, Zorn M. 2017. Impact of secondary vegetation succession on soil quality in a humid Mediterranean landscape. *Catena.* 149:836–843.
- Vapnik VN. 1998. *Statistical learning theory*. 1st ed. New York: Wiley-Interscience.
- Visser S, Keesstra S, Maas G, De Cleen M. 2019. Soil as a basis to create enabling conditions for transitions towards sustainable land management as a key to achieve the SDGs by 2030. *Sustainability.* 11(23):6792.

- Wang H, Moayedi H, Kok Foong L. 2020. Genetic algorithm hybridized with multilayer perceptron to have an economical slope stability design. *Eng Comput.* 28:1-2.
- Wang YQ, Shao MA. 2013. Spatial variability of soil physical properties in a region of the loess plateau of Pr china subject to wind and water erosion. *Land Degrad Dev.* 24(3):296-304.
- Xiong J, Zhang Q, Liang Q, Zhu H, Li H. 2018. Combining the multi-genetic algorithm and support vector machine for fault diagnosis of bearings. *Shock Vib.* 2018:e3091618.
- Yan J, Pu W, Zhou S, Liu H, Bao Z. 2020. Collaborative detection and power allocation framework for target tracking in multiple radar system. *Inf Fusion.* 55:173-183.
- Yang W, Zhao Y, Wang D, Wu H, Lin A, He L. 2020. Using principal components analysis and IDW interpolation to determine spatial and temporal changes of surface water quality of Xin'anjiang River in Huangshan. *IJERPH.* 17(8):2942.
- Zema DA, Bingner RL, Denisi P, Govers G, Licciardello F, Zimbone SM. 2012. Evaluation of runoff, peak flow and sediment yield for events simulated by the AnnAGNPS model in a belgian agricultural watershed. *Land Degrad Dev.* 23(3):205-215.
- Zhang K, Ruben GB, Li X, Li Z, Yu Z, Xia J, Dong Z. 2020. A comprehensive assessment framework for quantifying climatic and anthropogenic contributions to streamflow changes: a case study in a typical semi-arid North China basin. *Environ Model Software.* 128:104704.
- Zhang K, Wang Q, Chao L, Ye J, Li Z, Yu Z, Yang T, Ju Q. 2019. Ground observation-based analysis of soil moisture spatiotemporal variability across a humid to semi-humid transitional zone in China. *J Hydrol.* 574:903-914.
- Zhao G, Pang B, Xu Z, Peng D, Xu L. 2019. Assessment of urban flood susceptibility using semi-supervised machine learning model. *Sci Total Environ.* 659:940-949.
- Zhu J, Wang X, Chen M, Wu P, Kim MJ. 2019. Integration of BIM and GIS: IFC geometry transformation to shapefile using enhanced open-source approach. *Autom Constr.* 106: 102859.
- Zhu J, Wang X, Wang P, Wu Z, Kim MJ. 2019. Integration of BIM and GIS: geometry from IFC to shapefile using open-source technology. *Autom Constr.* 102:105-119.
- Zhu J, Wu P, Chen M, Kim MJ, Wang X, Fang T. 2020. Automatically processing IFC clipping representation for BIM and GIS integration at the process level. *Appl Sci.* 10(6):2009.
- Zhu Q. 2020. Research on road traffic situation awareness system based on image big data. *IEEE Intell Syst.* 35(1):18-26.