Physics and Chemistry of the Earth 129 (2023) 103295

Contents lists available at ScienceDirect

Physics and Chemistry of the Earth

journal homepage: www.elsevier.com/locate/pce

Gully erosion susceptibility mapping using four machine learning methods in Luzinzi watershed, eastern Democratic Republic of Congo

Géant Basimine Chuma [a,b,*], Yannick Mugumaarhahama [a,d], Jean Mubalama Mond [a], Espoir Mukengere Bagula [a], Adrien Byamungu Ndeko [a], Prince Baraka Lucungu [c], Katcho Karume [a,e], Gustave Nachigera Mushagalusa [a], Serge Schmitz [b]

[a] Faculty of Agriculture and Environmental Sciences, Université Evangélique en Afrique (UEA), Bukavu, Democratic Republic of the Congo
[b] Laplec UR SPHERES, Institut de Géographie, University of Liege, Clos Mercator 3 4000, Liège, Belgium
[c] Department of Natural Resources Management, University of Kinshasa, Kishansa, Democratic Republic of the Congo
[d] Environmental Management, Institut Superieur de Developpement Rural (ISDR), Kaziba, Democratic Republic of the Congo
[e] Centre de Recherche en Géothermie, Bukavu, South Kivu, Democratic Republic of the Congo

ARTICLE INFO

Keywords:
Gully erosion
GIS
Machine learning
Luzinzi watershed
Gully driving factors

ABSTRACT

Soil erosion by gullying causes severe soil degradation, which in turn leads to severe socio-economic and environmental damages in tropical and subtropical regions. To mitigate these negative effects and guarantee sustainable management of natural resources, gullies must be prevented. Gully management strategies start by devising adequate assessment tools and identification of driving factors and control measures. To achieve this, machine learning methods are essential tools to assist in the identification of driving factors to implement site-specific control measures. This study aimed at assessing the effectiveness of four machine learning methods (Random Forest (RF), Maximum of Entropy (MaxEnt), Artificial Neural Network (ANN), and Boosted Regression Tree (BRT)) to identify gully's driving factors, and predict gully erosion susceptibility in the Luzinzi watershed, in Walungu territory, eastern Democratic Republic of the Congo (DRC). In this study, gullies were first identified through multiple field surveys and then digitized using a very high-resolution image (CNI/airbus) from Google Earth. Overall, 270 gullies were identified, of which 70% (189) were randomly selected to train the four machine learning methods using topographical, hydrological, and environmental factors hypothesized to be gully-related conditioning factors. The remaining 30% (81 gullies) were used for testing studied methods using the threshold-independent area under the receiver operating characteristic (AUROC) and the true skill statistic (TSS) as performance measures. The results showed that RF and MaxEnt algorithms outperformed other methods; performance assessment results showed that the RF model with AUROC = 0.82 (82%) and MaxEnt (0.804: 80.4%) had higher prediction accuracies than BRT: 0.69 (69%) and ANN: 0.55 (55%). TSS results indicated that RF and MaxEnt are best methods in predicting gully susceptibility in Luzinzi watershed. On the other hand, the conditioning factors such as Digital Elevation Model (DEM), Normalized Difference Water Index (NDWI), Normalized Difference Vegetation Index (NDVI), slope, distance to roads, distance to rivers, and Stream Power Index (SPI) played key roles in the gully occurrence. Given the significance of these factors in gullies' occurrence, as shown in this study, policy-makers must adopt strategies that consider these factors to lower the risk of gully occurrence and related consequences at the watershed scale in eastern DRC.

1. Introduction

Soil erosion is widely recognized as a drastic soil destructive process, which undermines activities in about one billion hectares worldwide (Borrelli et al., 2020). Gully erosion is one of the land degradation processes and serves as the most intricate soil erosion phenomenon (Poesen et al., 1998; Valentin et al., 2005; Borrelli et al., 2020). It is the most important form of erosion in the river basin areas and comprises a wide variety of small processes, including head-cut, fluting, piping, continuous cracking progress, and mass flow (Poesen et al., 1998). Thus, gully erosion can be perceived as a major problem for natural resource management, leading to land degradation and economic losses worldwide (Valentin et al., 2005). For instance, this phenomenon causes

* Corresponding author. Faculty of Agriculture and Environmental Sciences, Université Evangélique en Afrique (UEA), Bukavu, Democratic Republic of the Congo.
E-mail address: geant.ch@gmail.com (G.B. Chuma).

https://doi.org/10.1016/j.pce.2022.103295
Received 29 December 2021; Received in revised form 4 October 2022; Accepted 31 October 2022
Available online 18 November 2022
1474-7065/

Abbreviation

GESM	Gully Erosion Susceptibility Mapping
GIS	Geographic Information System
RF	Random Forest
MLR	Multinomial Logistic Regression
CART	Classification and Regression Trees
NDVI	Normalized Difference Vegetation Index
nNDVI	Non-Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
SPI	Stream Power Index
WTI	Wetness Topographic Index
LS	Topographic factor
DEM	Digital Elevation Model
TRI	Topographic Terrain Index
MLA	Machine Learning Algorithm
BRT	Boosted Regression Tree
ANN	Artificial Neural Network
MaxEnt	Maximum of Entropy
RS	Remote Sensing
AUC	Area Under the Curve
TSS	True Skill Statistic
ED	Euclidean Distance

various damages to roads, natural resources, and agriculture. As gully erosion is a threshold phenomenon, various studies focused on characterizing the topographic and hydraulic conditions to forecast and assess the starting gullies susceptibility mapping (Poesen et al., 1998; Borrelli et al., 2020). Both natural and anthropogenic factors are mentioned as the driving forces behind the gullying process. Anthropogenic activities favoring gullying include unplanned agriculture (Ollobarren Del Barrio et al., 2018), grazing activity (Higaki et al., 2020), infrastructural development (Imwangana et al., 2015), and deforestation (Chuma et al., 2021a, 2021b). Generally, gully erosion and its areal extension is a natural process fully controlled by external forces and shaped by internal settings.

Many natural factors have been reported as gullying controlling factors. Geo-environmental factors controlling critical conditions for gully erosion occurrence and development were primarily associated with topography, lithology, rainfall, soil, and land use. Other studies stated that surface runoff is one of the principal factors contributing to the occurrence of gully erosion (Poesen, 2011). The velocity and volume of concentrated surface flow are also controlled by land use and topographic attributes (e.g., contributing drainage area, slope steepness, and slope curvature, etc.) (Gayen et al., 2019; Le Roux and van der Waal, 2020). Thus, a conceptual model should combine these elements to refine the mapping and monitoring process. However, in areas where accessibility and data availability are major constraints, the most important determinants are to be recommended for eventual mapping and monitoring.

Several models have been developed for assessing the gully erosion rate (Poesen et al., 2003; Bingner et al., 2016). Among these are the following physically-based models: Chemicals, Runoff, and Erosion from Agricultural Management Systems (CREAMS) (USDA and Knisel, 1980), Water Erosion Prediction Project (WEPP; Flanagan and Nearing, 1995), the Annualized Agricultural Non-Point Source (AnnAGNPS) and the methods developed (Sidorchuk, 1999; Bingner and Theurer, 2001). The Ephemeral Gully Erosion Model (EGEM) (Woodward, 1999), the Ephemeral Gully Erosion Estimator model (EphGEE) (Dabney et al., 2015) utilize algorithms similar to those of the Chemicals, Runoff, and Erosion from the Agricultural Management Systems model (CREAMS) (Bingner et al., 2016). However, the above-mentioned models do not predict the spatial distribution of gullies, which is an essential aspect for evaluating the impact of environmental changes on the occurrence of gullies and for planning erosion-control measures (Conoscenti et al., 2018). Additionally, the spatial distribution of gully erosion is an important factor in watershed management (Zakerinejad and Maerker, 2015). Some models allow an investigator to produce a gully erosion susceptibility map (GESM) or assess the spatial probability of gully occurrence by defining statistical relationships between geo-environmental conditioning factors and the spatial distribution of gullies.

Mapping areas susceptible to gully formation is not a new concept, most case studies are based in the northern hemisphere and use multiple regression models (MRM) relying on various topographic, soil type, land cover and land use characteristics (Zhu and Xu, 2021; Borrelli et al., 2022). The multiple regression models tend to suffer from a limited sample design, subjectivity during factor rating, and a large percentage of variability is usually unexplained (Kheir et al., 2007). Multivariate Adaptive Regression Splines (MARS) algorithm combined with Geographic Information System (GIS) techniques are currently considered for gully erosion susceptibility mapping evaluation (Amiri and Pourghasemi, 2020). Such techniques provide the ability to describe soil erosion processes, through evaluating spatial distribution of the gully erosion forms (the consequences) compared to some predictors (geological and environmental factors). Currently linked to statistical methods, these tools are enormously used to evaluate erosion susceptibility mapping (Sidorchuk, 1999; Nachtergaele et al., 2001; Nalivan et al., 2022). With the development of tools such as Machine Learning Algorithms (MLA), the refining and investigation of the potential distribution of these processes are being undertaken. Many studies have been conducted successfully by researchers on gully erosion susceptibility mapping using modern statistical models such as Random Forest (Gayen et al., 2020; Jiang et al., 2021), Regression models (Garosi et al., 2018; Wei *et al.*, 2021), Decision Tree (Arabameri et al., 2021), Maximum of Entropy (MaxEnt) (Arabameri et al., 2020a; Arabameri et al., 2020b), etc.

In the Democratic Republic of Congo (DRC), gully erosion is considered as an important source of land degradation in both urban and rural areas (Imwangana, 2014; Chuma et al., 2021a). Today, the majority of the country's population is affected by rapid soil degradation problems that have undermined agricultural productivity and acreage covered by agriculture by washing nutrients from topsoil in rural areas (Heri-Kazi and Bielders, 2021; Chuma et al., 2022). Although advanced soil erosion assessment models are in development, they have not been tested in data-scarce regions such as eastern DRC, and more specifically in South-Kivu province, to assess their effectiveness and the choice of an effective algorithm for refined recommendations to decision-makers. Thus, to minimize the adverse effects of gullying at a watershed scale, there is a need to determine the magnitude and spatial distribution of gully erosion using freely available data in data-scarce regions such as eastern DRC.

In this study, we hypothesized that arbitrary selection of a model to study the susceptibility of soils to gully erosion without having tested any in the region could lead to erroneous or unreliable conclusions. The choice of a model should depend on a comparison of models' outcomes to select the most suitable for analyzing, evaluating and predicting the local impact of gully erosion pathways. Earlier gully erosion mapping was based on the direct field survey, which was an expensive and time-consuming process (Poesen et al., 2003); and the outputs were mainly micro-scale mapping. Currently, remote sensing (RS) and GIS, along with other earth observatory techniques are extensively applied to map the GESM in larger areas within a short period and at the lowest cost (Zabihi et al., 2018). Several statistical techniques have been employed along with RS and GIS techniques for the gully erosion potential mapping such as frequency ratio, weights of evidence, bi-variate statistical models, and probabilistic approach, but they were special-purpose inventory.

Soil erosion is a severe geo-environmental problem in various parts of DRC in general, and the eastern highland parts in particular. Soil erosion by water is one of the most land resources' constraints in Walungu, Kabare, and Kalehe territories. For decades, the water erosion process seems to have accentuated (Heri-Kazi, 2020; Chuma et al., 2021a, 2022). These territories are being seriously affected by soil erosion through gullies, rills, and shifting agriculture aggravation (Heri-Kazi, 2020). One of the decisions was to manage the land degradation process at the watershed scale to emphasize management and decision-making. Soil water management has been studied at a watershed scale in several case studies (Chuma, 2019; Heri-Kazi, 2020; Chuma et al., 2021a, 2021b, 2022) in South-Kivu; only a few focused on the gullying process (Chuma et al., 2021a, 2021b). For the Luzinzi watershed, the gully erosion process has been studied by Chuma et al. (2021a) by characterizing and determining driving factors. The Luzinzi watershed has experienced gully erosion that causes many challenges and which led organizations and local communities to reassess the weirdness of adopted constant genesis strategies. The gully in Luzinzi watershed leads to soil loss, the imposition of high production costs, reduced agricultural land potential, road cut, water pollution, and thus, exacerbated soil erosion pathways influence on farming system efficiency. Consequently, describing the best fitting and robust model to assess the sensitivity of watersheds to the development of gully pathways seems necessary for sustainable ecosystem management.

The overall objective of this study is to contribute to gully soil erosion estimation in eastern DRC by assessing the factors contributing to its expansion at the watershed scale. Specifically, the objectives of this study were to (1) select the important conditioning factors in gully erosion using the multi-collinearity analysis; (2) develop ANN, MaxEnt, BRT and RF machine learning models to predict the areas prone to gully erosion; (3) use validation method to deal with the randomness adverse effects on the resulting GESM using machine learning models, and (4) evaluate the capability and robustness of these models by comparing their performance using area under the curve (AUC), True Skill Statistic (TSS), and DeLong test.

The performance and accuracy of each algorithm were assessed for the best gully zoning algorithm selection. The outcomes of this study will help decision-makers, government agencies, and the private sector for sustainable land-use management practices in South-Kivu watersheds and to inspire studies in regions of the world sharing similar struggles and ecological conditions. The novelty of this study is the fact that it is for the first time that GESM is used for South-Kivu province and DRC in general. Before this study, there was no reliable data on gully erosion vulnerability in the region. Conclusions from this study will be of great help to soil scientists in tropical areas of Africa where such adopted methods are rarely applied and their effectiveness less demonstrated.

2. Materials and methods

2.1. Study area

Luzinzi watershed is located in Walungu territory, in the South-Kivu province, eastern DRC. This watershed is located between 28°46′ and 28°48′ East (Longitude) and 2°45′ and 2°58′ South (Latitude); it has a surface area of ~154 km^2 (Fig. 1). Luzinzi extends into four chiefdoms comprising Kaziba, Ngweshe (north and east), Luhwindja (west and south-west), and Bafulero (south-east). A large part of the Luzinzi watershed is found in Kaziba chiefdom, which is one of the main production areas for firewood, timber, embers, and other wood products. In its northern part, it extends to one of the most important gold mining sites of the province (Birhenjira et al., 2014; Chuma et al., 2021a).

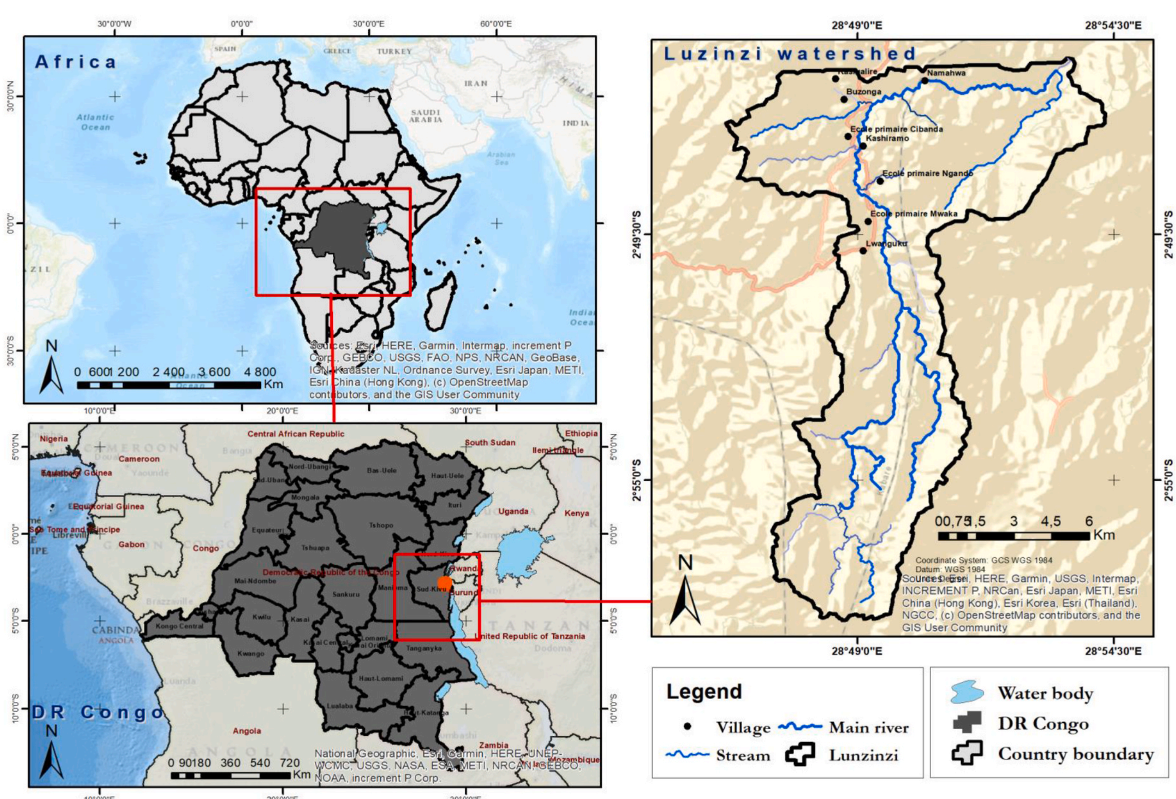

Fig. 1. Luzinzi watershed in Walungu territory, South-Kivu province, Eastern D.R. Congo.

2.2. Watershed characteristics

The Luzinzi watershed was recently characterized by Chuma et al. (2021a). Indeed, Luzinzi is part of the great Ruzizi watershed; it extends over the highlands of the central African Graben to the east of the Congo River. Some morphological and hydrological characteristics of this watershed showed that Luzinzi presents a width and length of the overland flow of ~8.86 and ~39.05 km, respectively. The stream length ratio (RL), drainage density (Dd), slope index, and average watershed slope (%) are ~320, 0.466, 0.038 km/km^2, and 56.67%, respectively. The highest elevation point is at ~3435 m above sea level (masl) while the lowest downstream point is at ~852 masl. A high relief ratio was obtained after calculation (Chuma et al., 2021a).

The region is dominated (>70%) with high elevation (generally >2000 m), while lowlands, inland valleys, and wetlands are found in the center of the watershed. These areas are currently used for agricultural

Fig. 2. Some gullies found in the Luzinzi watershed during fieldworks (**a**: soil excavation in a pasture after heavy rains with concentrated runoff, **b**: creation of small gullies in an agricultural farm, **c**: gully more or less stable with the regrowth of some grasses in the bottom, **d:** creation of a gully in a steeply sloping area formerly protected by rural waste deposits, **e** and **f**: large gullies already expanding in the area and extending over large areas of about 10 m).

purposes. According to FAO/UNESCO World Reference Base (WRB) soil classification system taxonomy, the dominant soil unities are Ferralsols, Cambisols, and Acrisols, which are susceptible to soil erosion (low structural stability, low organic matter content, and thus high erodibility) (Dewitte et al., 2013; Zádorová et al., 2021). The main talweg length was ~56.67 while the moving time and Gravelius index (considered as Compactness Coefficient) were ~6.46 h and ~1.9, respectively (Chuma et al., 2021a).

According to the climate, Luzinzi is under a humid tropical climate influenced by elevation and characterized by two seasons: rainy season (from September to May) and dry season (June to August). The available weather data showed an annual average temperature and precipitation of ~19 °C in the south and ~10 °C in the north, and ~1200–~1700 mm, respectively. Luzinzi watershed hydrography is rich in terms of rivers, springs, and streams; the watershed is mainly drained by the Luzinzi River (after which the watershed is named), while other rivers include Mugaba, Ntyazo, Nachibundo, Mugusa, Kazinzi, Kiko, Chishi, and Shaliro flows (Chuma et al., 2021a).

2.3. Characterization of gullies located in the Luzinzi watershed

Spatio-temporal characterization of gullies in the Luzinzi watershed was previously conducted by Chuma et al. (2021a). From 2010 to 2020, gullies increased in terms of number, length, width, surface, and volume. In total, ~270 gullies were identified in 2020 while there were only 38 in 2011. These gullies had a minimum length of ~23.8 m (in 2011) while the largest had a maximum length value estimated at ~143.0 m in 2020 (Chuma et al., 2021a). In this study, we used geographic coordinates of 270 gully head cuts that were digitized and validated with extensive field observations. To characterize these gullies, the “add polygon” function of Google Earth Pro was used for gully digitization in the selected areas of Luzinzi. Then, gully polygons were saved as Keyhole Markup Language files (KML), converted into Shapefile (“.shp file format”) using the conversion analysis tool from ArcGIS 10.7 Esri¯™. Some of the gullies observed in the watershed during the fieldworks are presented by Fig. 2.

2.4. Methods

The present study has employed different thematic layers which were transformed into the spatial database using ArcGIS 10.7 Esri¯™ environment. Four Machine Learning Algorithms (MLA) have been used to analyze the relationship between gully head cut locations and gully erosion driving factors. Thereafter, gully erosion potential maps were generated, prepared, and reclassified into five threshold levels: “very low”, “low”, “medium”, “high” and “very high” susceptibility, respectively.

2.4.1. Environmental factors hypothesized to condition gully erosion

It is essential to select gully erosion conditioning factors for creating a gully erosion susceptibility map (GESM). As the gully erosion process is controlled by both erosivity of the runoff water and the soil cover erodibility, these data are primordial for the mapping process. Unfortunately, in data-scarce regions such as eastern DRC, other factors have to be integrated into the modeling process. In such regard, several geo-environmental attributes can be considered (Agnesi et al., 2011; Javidan et al., 2020). It is essential to determine the gully erosion conditioning factors to perform gully susceptibility mapping (Conoscenti et al., 2014; Dube et al., 2014).

Based on the literature and freely available data, 16 factors were selected for use. These included the soil, hydrological and vegetation indexes, Euclidean distances from rivers and villages, topographic indexes, slope, slope aspect, plan curvature, and elevation (Table 1). Data on soil, climate, and geomorphology at the watershed scale were not available and thus they were not included in this study.

To obtain these conditioning factors, a digital elevation model (DEM) was prepared using the topographic database of The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and a Landsat 8 OLI/TIRS for topographic, vegetation, and soil indexes calculation (both were rescaled at 30 m spatial resolution). Table 1 presents the predisposing factors, their calculation methods, and references while Fig. 3 illustrates the methodological flowchart of the approach that was used for GESM analysis. As shown in Fig. 3, the flowchart consists of four steps: (1) preparing thematic layers or geo-environmental conditioning factors, (2) stepwise determination of effective factors, (3) gully erosion susceptibility modeling using the four MLA, and (4) validation of the used susceptibility maps.

Table 1
Description of conditioning factors used in this study.

No.	Factor	Source/Tool	Reference
1	NDVI	NDVI=(NIR-RED)/(NIR + RED)	Rouse et al. (1973)
2	NDWI	NDWI= (NIR-SWIR)/(NIR + SWIR)	McFeeters (1996)
3	TWI	WI = ln [As/tan (β)]	Moore et al. (1991)
4	LS	LS= (FA × CS/22.13)$^{0.4}$ × (sin σ/0.0896)$^{1.3}$	Moore et al. (1991)
5	Slope (degree)	Based on ArcGIS 10.7	Moore et al. (1991)
6	SPI	SPI= As × tan σ	Conforti et al. (2011)
7	Plan curvature	U.S Geological Survey	Arabameri et al. (2021)
8	nNDVI	Mean of annual NDVI of the watershed	Rigge et al. (2013)
9	Elevation (m)	DEM	Chuma et al. (2021a)
10	Streams	Extract from ArcGIS 10.7	Theobald et al. (2005)
11	Streams order	Extract from ArcGIS	Theobald et al. (2005)
12	ED roads	Euclidean Distance from roads with a 500 m distance in ArcGIS 10.7	Gedeon et al. (2021)
13	ED rivers	Euclidean Distance from rivers with a 500 m distance in ArcGIS 10.7	Gideon et al. (2021)
14	ED villages	Euclidean Distance from settlements with a 500 m distance in ArcGIS 10.7	Gideon et al. (2021)
15	Profile	Extract from ArcGIS 10.7	Theobald et al. (2005)
16	TRI	Terrain Ruggedness Index	Riley et al. (1999)

Legend: SPI: Stream power index, LS: Slope length, TWI: Topographic wetness index, NDVI: Normalized Difference Vegetation Index, NDWI: Normalized Difference Water Index in Agriculture, nNDVI: Mean Normalized Difference Vegetation Index, ED: Euclidean Distance, TRI: Terrain Ruggedness Index (TRI).

All the thematic layers were converted into the spatial database using ArcGIS 10.7 Esri USA software. A 30 m resolution DEM was employed to prepare the elevation, slope, slope length, slope aspect, and plan curvature maps, SPI, LS, TWI, while Landsat 8 OLI/TIRS satellite images were used to create the vegetation and water indexes (NDVI, nNDVI, and NDWI). Existing shapefiles of roads, rivers, and villages were used to calculate the distance from lineament maps using Euclidean Distance (ED) spatial analysis from ArcGIS 10.7. A drainage map was developed from the topographical maps and calculated according to the formula presented in Table 1. Maps generated using these conditioning factors are presented by Fig. 4.

2.4.2. Multicollinearity analysis

To determine relationships among variables, independent linear relationship with the correlation matrix followed by the variance inflation factor (VIF) index were employed. The VIF determines whether a linear relationship exists between independent variables or not (Ozdemir and Altural, 2013). VIF shows also how much one independent variable is increased for establishing a linear relationship between

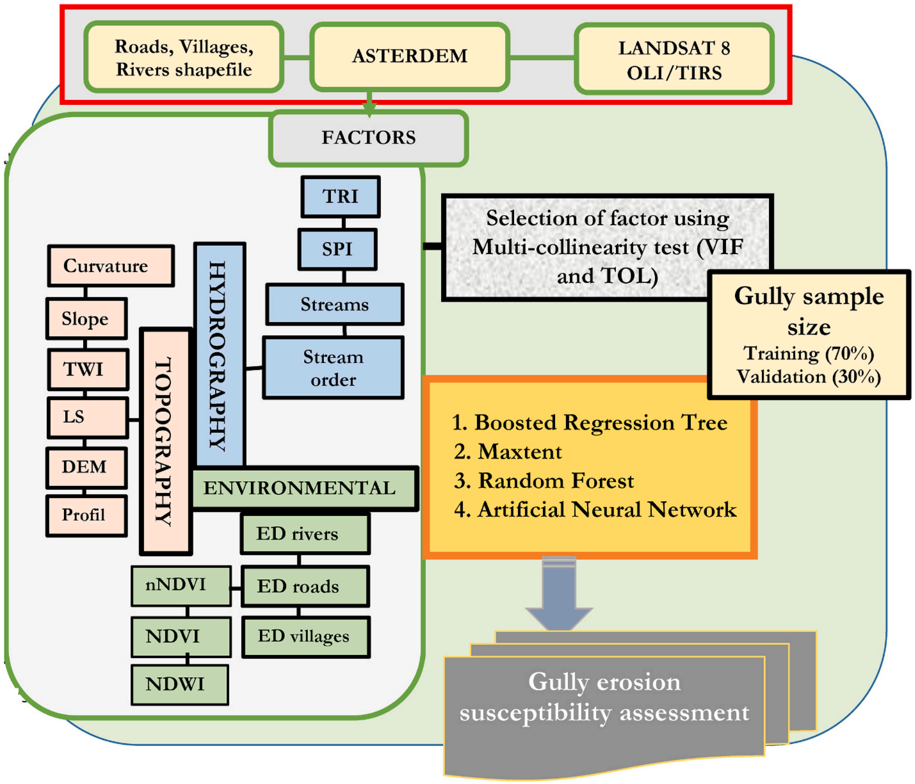

Fig. 3. Methodology flowchart.

independent variables. According to Yu et al. (2015), if VIF is greater than 10 and Tolerance (TOL) is less than 0.1, it indicates high multiple collinearities.

2.4.3. Overview of machine learning methods used

According to Hong et al. (2016), machine learning methods perform better than statistical models. In this study, four machine learning methods were used for GESM. The selected machine learning methods are among those that have been used in previous studies for constructing accurate GESM. We provide below a brief overview of each method.

a) Random Forest (RF)

Random Forest (RF) is a machine-learning algorithm developed by Breiman in 2001. RF is a regression tree-based model that is one of the most accurate and effective methods (Rahmati et al., 2017). See Breiman (2001) and Svetnik et al. (2003) for a thorough understanding of RF algorithm. RF has been widely used in numerous environmental applications such as water resource management and natural hazard management. Kuhnert et al. (2010) and Rahmati et al. (2017) are examples of studies that used RF for GESM. In this study, we used the 'RandomForest' package (Liaw and Wiener, 2002) in R 4.2.0 software to fit the RF models.

b) Maximum of Entropy (MaxEnt)

MaxEnt is a machine learning and intelligence-based model that is based on the entropy maximization principle (Woodbury et al., 1995). This principle forms on information theory and statistics with the concept of best approximation of an unknown probability distribution (Phillips et al., 2006; Sivia and Skilling, 2006). MaxEnt can handle data from a variety of measurement scales without relying on parametric statistical assumptions. MaxEnt has been widely used in studies related to environmental analysis and ecological niche modeling (e.g., Townsend Peterson et al., 2007; Moreno et al., 2011; Warren and Seifert, 2011) as well as landslide susceptibility (e.g., Conforti et al., 2011; Felicísimo et al., 2013; Park, 2015; Kornejady et al., 2017). In this study, the 'maxnet' R package was used in R 4.2.0 software to fit the MaxEnt model with some modified parameters (Phillips and Dudík, 2008). See Phillips et al. (2006) and Phillips and Dudík (2008) for a thorough understanding of the MaxEnt algorithm.

c) Boosted Regression Tree (BRT)

BRT is a combination of two techniques: decision tree algorithms and boosting methods. With BRT, a regression tree is used to explain discrete data (Breiman, 2001). Logistic function is reconstructed by splitting the predictors' domain and using the observations from each split (Therneau et al., 2014). The splitting method is repeated iteratively until the optimal split that minimizes the error or loss function is identified (Lombardo et al., 2015). When the gain in minimizing the error becomes insignificant, the procedure ends. BRT has shown good predictive performance in predicting gully erosion susceptibility (Ahmadpour et al., 2021; Garosi et al., 2018). In this study, BRT model was fitted using 'Dismo' package (Ridgway et al., 2008), and using settings as recommended by Elith et al. (2008). For in-depth understanding of BRT, see De'Ath (2007).

d) Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is one of the most popular techniques used in GESM (Tsangaratos and Benardos, 2014; Rahmati et al., 2017). The most widely applied machine-learning technique; ANN can be efficiently used in non-linear phenomena such as parameter retrieval. It is currently one of the most used non-parametric classification techniques and does not depend on any assumption of generally distributed data (Lu and Weng, 2007; Liou et al., 2001). The ANN algorithm functions like a human brain or nervous system containing nerve fibers with many interconnections through other axons. It can learn and produce meaningful results from examples, even when the input data is having

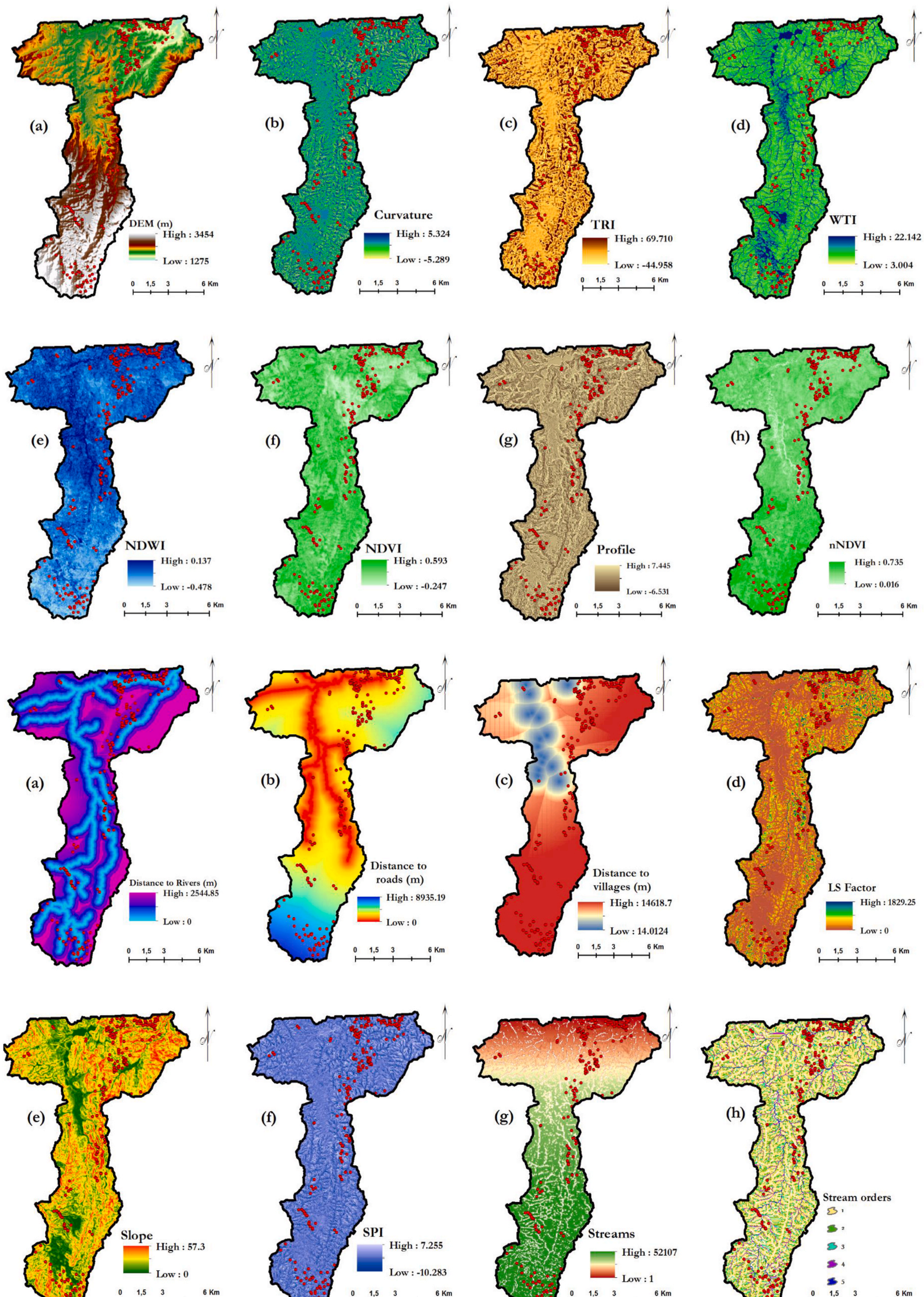

Fig. 4. Conditioning factors used for gully erosion susceptibility mapping (GESM) in Luzinzi watershed.

errors or complexity and is incomplete. Therefore, it can simulate exactly like the human nervous system. However, the ANN has one input layer, at least one hidden layer, and one output layer. Each layer is formed by neurons (like brain nerves) (Dixon and Candade, 2008; Schuman and Birdwell, 2013). In this study, the 'nnet' R package was used in R 4.2.0 software to fit the ANN model.

2.4.4. Model training

We used the coordinates of the sites where gully erosions have been observed to train these machine learning methods, which is a common approach used in modeling species distribution (Guisan et al., 2017). This approach involves extracting the values of each conditioning factor that is hypothesized to influence the phenomenon under investigation using the coordinates of the sites of presence. The values of the associated conditioning factors are also extracted once pseudo-absence points are randomly picked. Both types of data are then used to train the different machine learning methods to build a prediction function (hidden for machine learning) and to determine the importance of each conditioning factor in the occurrence of the event under investigation. The trained models are then used to predict the occurrence probability of the studied phenomenon over the whole study area. See Guisan et al. (2017) for in-depth understanding of the modeling process.

Following the recommendation of Barbet-Massin et al. (2012), 'randomPoints' R function of the 'dismo' package was used to select randomly 270 background points that were used in the modeling together with gully occurrence points. Next, as suggested by Barbet-Massin et al. (2012), we performed 10 replications of each machine learning method. For each replication, independent (presence and background) data were used to train the model and performance was assessed. Averaging the 10 replications yielded the maps and performance measures provided in this paper. R 4.1.2 and R Studio were used for data analysis.

2.4.5. Validation of gully erosion susceptibility map (GESM) produced with each MLA

According to Naghibi et al. (2016) and Chen et al. (2019), validation is a fundamental step in the model-building approach for the scientific significance of the research. In this study, the Area Under the Receiver Operating characteristic curve (AUC the ROC) was used to implement the model evaluation. It provides an alternative technique for the assessment of the accuracy of ordinal score models. The AUC (Area Under Curve) represents values between 0.5 and 1, and the higher the value it represents, the better the performance of the model. AUC has been demonstrated to be independent of prevalence and is perceived to be an accurate measure of ordinal score model performance. The DeLong's test was used for pairwise comparison of the differences in the models' AUCs (Supplementary material 1).

Sensitivity and Specificity: Sensitivity represents the proportion of correctly predicted gully presence records and thus the quantification of omission errors. In calculation, Sensitivity equals $A/\ (A+C)$ where A notes the number of correctly predicted presence cells and C the number of cells in which gully was found, but absence is predicted by the model. Specificity represents the proportion of correctly predicted absences and thus the quantification of commission errors. In calculation, Specificity equals $D/B+D$ where B denotes the number of cells in which gully was not found but presence is predicted by the model, and D is the number of cells correctly predicting absence. It is noteworthy that compared across models, sensitivity and specificity are independent from one another, as well as being independent of prevalence, which represents the proportion of sites where gully was recorded as present.

True Skill Statistic (TSS): The TSS is independent of prevalence and equals to $(AD-BC)(A+C)(B+D)$. In this formula, A is the number of correctly predicted presence cells and, B is the number of cells in which gully was not found but presence is predicted by the model, C the number of cells in which gully was found and D the number of cells correctly predicting absence. Allouche et al. (2006) have shown that TSS is an intuitive method of performance measurement of models in which predictions are expressed as presence-absence maps. It was further found that TSS gives results showing significant correlation with those of the threshold-independent AUC statistic (Allouche et al., 2006). R 4.2.4 and R Studio software helped in data analyses and graph presentation.

3. Results

3.1. Multicollinearity analysis and variable description

Results from the correlation matrix and after multicollinearity (VIF and TOL calculations) test showed that amongst the 16 conditioning factors initially selected for GESM, all were not collinear, which means no collinearity existed amongst the selected factors (independent variables) (Supplementary material). All the values of VIF and TOL were above 0.2 and 5–10, respectively. This process was performed because multicollinearity amongst factors reduces the prediction accuracy of the models. When the value of TOL is less than 0.1 and the value of VIF is greater than 10, collinearity exists amongst the parameters. Table 2 presents the gully conditioning factor classes, number of pixels, and corresponding percentages in the Luzinzi watershed. More than half (62.12%) of the watershed is located in an area with a value 5–10 for TWI, while TRI of <0.5% represents ~48.5% of the watershed. According to NDVI, more than half (~55.9%) of the Luzinzi watershed is covered by vegetation with NDVI varying from 0.35 to 0.75 every year. The watershed area can be considered as being located near rivers (<100 m: 36.8%), roads (<1000 m: 44.8%), villages (<1000 m: 46.3%) while the dominant slopes are 10–20% (~30.5%) and 20–30% (~31.4%). These slopes lead to an LS factor mostly >30 (~81.7%). The dominant altitude is located in the northern and north-eastern part of the watershed with values < 1500 masl (~43.3%), while the one with higher values > 2500 m (~4.6%) is mostly located in the southern and south-western part of the watershed.

3.2. Variables importance

a) Boosted Regression Tree (BRT) and Artificial Neural Network (ANN)

Results for the contribution or importance of factors for each MLA are presented in Fig. 5. For the 16 geo-environmental factors, only six contributed in the BRT model, comprising the elevation (83.3%), NDWI (8.9%), NDVI (4.3%), Distance to rivers (1.7%), plan curvature (1%), and profile (<1%). For ANN, up to 10 factors intervened in the model; these comprised plan curvature (40.3%), elevation (DEM: 27.9%), distance to roads (27.6%), slope length (1.5%), NDVI (1.3%), and NDWI (1%) while profile, distance to rivers, slope, and SPI contributed each up to 1%. For BRT, the elevation, NDVI, and NDWI contributed up to ~96.5% while for ANN, curvature, elevation, and distance to roads contributed up to ~95.8% (Fig. 5).

b) Random Forest (RF) and Maximum Entropy (MaxEnt)

According to Random Forest (RF), 11 factors were contributors while only seven were taken into account for the MaxEnt model. The factors that mainly contributed in the susceptibility analysis for RF comprised 76% for DEM (elevation), 11% for NDWI, 8.4% for NDVI, 2.2% for distance to roads while all other factors contributed with less than 1%. The first four factors contributed up to 97.6%. For the MaxEnt model, all the factors have highly contributed to the prediction. Major contributors were the elevation (25.1%), the NDWI (20.4%), the slope (15.3%), the NDVI (12.3%), the distance to roads (10%), the distance to rivers (5.6%), and the SPI (3.4%). These seven factors contributed up to ~92.1% (Fig. 5).

After reclassification of the 16 factors used for GESM, Table 2 was generated. Each one was reclassified into five classes.

Table 2
Selected gully conditioning factors in the Luzinzi watershed.

N°	Factors	Classes	Number of pixels	Percentage (%)
1	**TWI**	*<5*	48764	30.99
		5–10	97773	62.13
		10–15	9008	5.72
		>15	1817	1.15
2	**TRI**	*<0.5*	76265	48.56
		0.5–5.0	14179	9.03
		5.0–10.0	42268	26.91
		>10.0	24349	15.50
3	**nNDVI**	*<0.35*	64039	41.03
		0.35–0.75	87284	55.93
		>0.75	4671	2.99
		<0.1	74	0.05
4	**Profile**	*F*	1292	0.82
		N	6636	4.22
		NE	17878	11.36
		E	28823	18.32
		SE	46632	29.63
		S	30307	19.26
		SW	17418	11.07
		W	7187	4.57
		NW	1189	0.76
5	**ED Rivers (m)**	*<100*	57916	36.83
		100–200	16044	10.20
		200–300	14463	9.20
		>300	68819	43.77
6	**ED Roads (m)**	*<1000*	70148	44.58
		2000	30670	19.49
		3000	20420	12.98
		>3000	36124	22.96
7	**Slope (%)**	*<10*	26814	17.04
		10–20	47947	30.47
		20–30	49001	31.14
		>30	33600	21.35
8	**SPI**	*<5*	21358	13.57
		5–10	29390	18.68
		10–15	61671	39.19
		>15	44942	28.56
9	**Stream order**	1st	94822	69.67
		2nd	27674	20.33
		3–4th	8892	6.53
		>4th	4713	3.46
10	**Streams**	*<10*	27164	25.03
		10–20	26997	24.88
		20–30	27460	25.31
		>30	26891	24.78
11	**NDWI**	*<-0.5*	40231	25.78
		−0.5–0.10	53697	34.41
		0.10–0.15	58471	37.47
		>0.15	3668	2.35
12	**NDVI**	*<0.1*	23788	15.24
		0.1–0.35	43782	28.05
		0.35–0.5	88036	56.41
		>0.5	461	0.30
13	**LS**	*<5*	7954	5.05
		5–15	19329	12.28
		15–30	1444	0.92
		>30	128635	81.74
14	**ED Villages (m)**	*<1000*	72880	46.33
		1000–2000	15318	9.74
		2000–3000	18519	11.77
		>3000	50590	32.16
15	**Elevation (m)**	*<1500*	68270	43.38
		1500–2000	36164	22.98
		2000–2500	45629	29.00
		>2500	7299	4.64
16	**Curvature**	*Flat*	65019	45.96
		Convex	55588	39.30
		Concave	20852	14.74

Legend: SPI: Stream power index, LS: Slope length, TWI: Topographic wetness index, NDVI: Normalized Difference Vegetation Index, NDWI: Normalized Difference Water Index in Agriculture, nNDVI: Mean Normalized Difference Vegetation Index, ED: Euclidean Distance, TRI: Terrain Ruggedness Index (TRI).

3.3. Gully susceptibility modelling

Results obtained for the GESM using four MLA in the Luzinzi watershed are presented in Fig. 6. The reclassification maps obtained into five susceptibility classes are presented in Fig. 7.

According to results, the Random Forest algorithm (RF) divided Luzinzi into five susceptibility classes varying from moderate (46%), high (36%), very high (11%), low (3%), and very low (4%) susceptible. MaxEnt presented the same trend with 52, 9, 28, 6, and 5%, respectively. BRT divided the watershed into only two classes: very high (90%) and high (10%) susceptibility, while ANN classified the watershed surface into high (88%), very high (12%), moderate (<1%), and low (<1%) susceptibility areas (Figs. 6 and 7).

3.4. Model accuracy assessment

Results from Fig. 6 and Table 3 present the GESM obtained from the four MLA. Table 3 shows the descriptive statistics of the two parameters (AUC and TSS) according to MLA used. Two models presented high values in terms of AUC: Random Forest (0.82: 82%) and MaxEnt (0.80: 80.4%); these values corresponded to TSS of 0.56 and 0.50, respectively. BRT and ANN presented low values in terms of AUC (0.69: 69% and 0.55: 55%, respectively) that corresponded to the TSS values of 0.38 and 0.11, respectively. Maximum and minimum variations of AUC for RF were 0.81–0.83 while for MaxEnt these ranged from 0.78 to 0.85, ANN (0.50–0.59), and BRT (0.68–0.71). DeLong test confirmed as well the effectiveness of RF and MaxEnt for the GESM (Table 3, Supplementary material 1).

4. Discussion

4.1. Conditioning factors related to gully erosion susceptibility mapping in Luzinzi watershed

Geomorphic hazards such as gully erosion are caused by imbalanced systems from natural and anthropogenic factors (Arabameri et al., 2020a). Changes in equilibrium among these geo-environmental factors lead to an increase in such process. Though gully erosion seems to be a complex process, its causes are inferred using new deterministic and statistical spatial models; one of the new approaches in machine learning algorithms. Thus, discovering cause-and-effect relationships is the key to identifying appropriate prevention and management techniques for gully development (Arabameri et al., 2020a). In this study, four MLA and 16 driving factors were used to assess GESM performance at a watershed scale. For all these models, contributing geo-environmental factors varied from one model to another; it varied from 5 to 10, for MaxEnt (7), RF (11), BRT (5), and ANN (10). It can be inferred from these results that five factors were often top-ranked and thus contributed significantly to the GESM mapping, these included elevation (m), NDVI, NDWI, ED to roads and rivers as well as the slope. Other factors were specific to the model and contributed little to the gully susceptibility mapping in the watershed. This is the case for SPI, TRI, Profile, Curvature, and LS. However, for all of these models, elevation, NDVI, NDWI, and ED to road contributed significantly.

Altitude mainly controls gully erosion and as a result, determines the distribution of ditches and it affects vegetation and rainfall. It controls the flow of drains and the amount of drainage and has a significant impact on soil moisture content and the slope of the hillside (Zhu et al., 2014; Gómez-Gutiérrez et al., 2015; Rahmati et al., 2017). The ‘elevation’ parameter along with the local slope is considered by many researchers (Gómez-Gutiérrez et al., 2015) as a major gully erosion-affecting factor. The elevation of the Luzinzi watershed varied from 1200 to 3500 m above the mean sea level. The slope played an essential role in the formation of gullies, since the majority of the watershed is located in areas with steep (20–30%: ~31%) to very steep (>30%: ~21%) slopes. Other factors such as slope aspect, LS factor, and

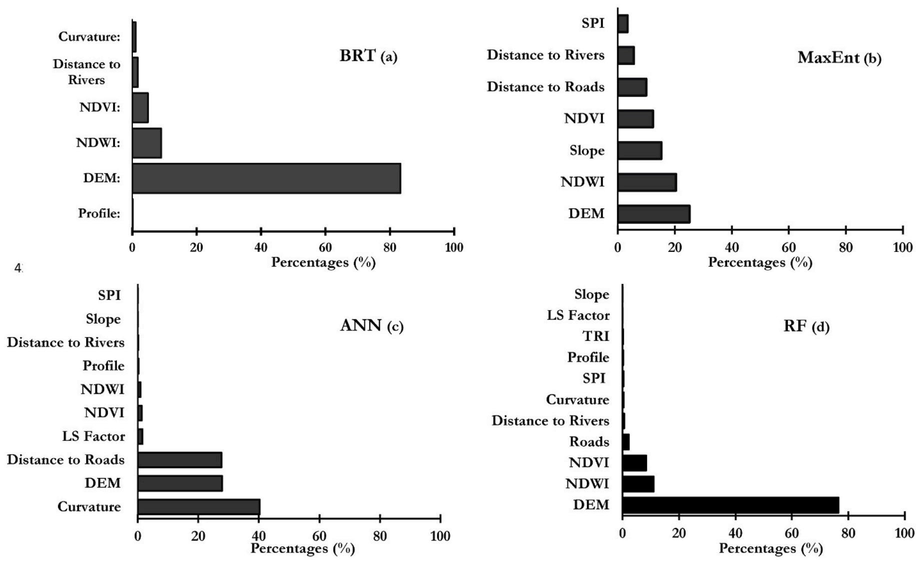

Fig. 5. Factors' contributions for GESM using BRT (a), MaxEnt (b), ANN (c), and RF (d) algorithms in Luzinzi watershed. LS: Slope length, TWI: Topographic wetness index, NDVI: Normalized Difference Vegetation Index, NDWI: Normalized Difference Water Index in Agriculture, nNDVI: Mean Normalized Difference Vegetation Index, ED: Euclidean Distance, TRI: Terrain Ruggedness Index, RF: Random Forest, MaxEnt: Maximum Entropy, BRT: Boosted Regression Tree, and ANN: Artificial Neural Network.

curvature have also contributed to the GESM. Directional land segments have been associated with exposure to sunlight, winds, lineament, and rainfall. The slope profile contributed in GESM for ANN and RF models. The vegetation cover of the soil (assessed by NDVI) seems to be an important factor in mapping gullies in the basin, regardless of the used model. Evidence links between land use and gullies expansion have been studied in many regions; these include case studies from Spain (Lesschen et al., 2007), Brazil (De Oliveira et al., 2020), Ethiopia (Nyssen et al., 2006), DRC (Chuma et al., 2021a) and South Africa (Podwojewski et al., 2020). Many studies showed the relation between NDVI and land use (Griffith et al., 2002; Mekonnen et al., 2018). Other studies reported that LULC plays an important role in the formation, expansion, or retreat of gullies. Lesschen et al. (2007) reported that the potential vulnerability of lands in Spain was increased by different land use and land cover scenarios. Morgan and Mngomezulu (2003) showed that changes in terms of land use and land cover (LULC) and particularly the conversion of forest to agricultural lands accelerated erosion; this is much pronounced with factors like slope and rainfall. In the Luzinzi watershed, these factors were studied and the effect of LULC on gully occurrence and expansion was ascertained (Chuma et al., 2021a).

Another important factor is the NDWI. It is a common index used in remote sensing studies as the Normalized Difference Water Index (NDWI) (McFeeters, 1996). This index looks at the difference between the green and near-infrared bands, as they are strongly absorbed by water bodies making the delineation. However, NDWI is sensitive to built-up land, resulting in risks of over-estimation. Most of the time, gullies are associated with the drainage network, which enables the release of eroded substances from the upstream areas (Dube et al., 2016). Uchida (2015) suggested using several indexes, not only NDWI or NDVI, but also NDBSI (Normalized Difference Bare Soil Index), and NDSI (Normalized Difference Soil Index) representing the soil surface conditions and LSWI (Land Surface Water Index). The author also mentioned the importance of LS in determining gully soil losses (using the USLE) and in mapping gullies. The slope aspect can indirectly affect the erosion process through sunlight exposure, evapotranspiration, moisture storage, vegetation type, and vegetation distribution. In general, the effect of the plan curvature on the incidence of gully erosion occurs through divergence and convergence of water on the slope. Hence, the slope shape was chosen due to its effect on the creation and expansion of the gullies (Gómez-Gutiérrez et al., 2015). Important effective factors for gully erosion susceptibility can, therefore, lead to the best models to assess gully susceptibility. However, Gómez-Gutiérrez et al. (2015) used the slope, catchment area, and NDVI as threshold and the accuracy indicated a negligible overestimation. This suggests that other parameters needed to be incorporated instead. Our analyses showed that since Luzinzi watershed is characterized with high altitude (m), low NDVI (low vegetation cover) coupled with proximity to roads and rivers, its soils are more likely susceptible to the gully erosion.

These findings have been supported by many other studies in which a location vegetation and man activities coupled with hydrologic properties are assigned as determinants of the highest gully erosion susceptibility (Imwangana, 2014; Ahmadpour et al., 2021). Other results have suggested that distance to streams and roads, drainage density, and NDVI are significant factors that promote favorable conditions for gullying (Arabameri et al., 2020a). Hosseinalizadeh et al. (2019) reported that land use, slope degree, and slope length (LS) are the major gully drivers in the arid regions such as Iran. The same conclusion was made by Pourghasemi et al. (2019) at a watershed scale with an AHP model. These authors showed that LULC, drainage density, and elevation models are the most important predictors of gully occurrence. It was, therefore, assumed that these factors would have effects because according to Taruvinga (2008), the spectral signature of gullies varies with whether they contain vegetation or are un-vegetated. Additionally, the spectral signature of bare soil (un-vegetated gully) is dependent on the moisture content, soil type, texture, and structure (Aggarwal, 2004). Torkashvand and Alipour (2009) used supervised classification in plain physiography of Iran and found out that the accuracy decreases where there is another land use such as cultivation due to similarity in spectral characteristics.

Another important factor in gully erosion susceptibility mapping in the Luzinzi watershed was the distance to roads and rivers. The closer the roads, the more erosion susceptible the location was. Many studies using RF corroborate that distance to stream, distance to road, and LULC parameters have the most impact on gullying in investigated areas (Conoscenti et al., 2014; Vanmaercke et al., 2016). Therefore, gullies are linked to the stream networks and the streams cause gullying in areas where conditions are suitable. These studies coupled with the findings of Imwangana (2014) allowed concluding that linear infrastructure, such as roads, rivers through the concentration of surface runoffs, the transfer of concentrated runoff to other watersheds, and an increase in watershed size, causes the gullying process. The TWI has been used widely to describe the effect of topography on the location and size of saturated areas of flow surface generation (Moore et al., 1991). The power of runoff erosion, discharge potential, and carrying capacity are modeled using the mentioned factor. TWI estimates the probability of water accumulation in soil due to slope and upstream catchment area, and

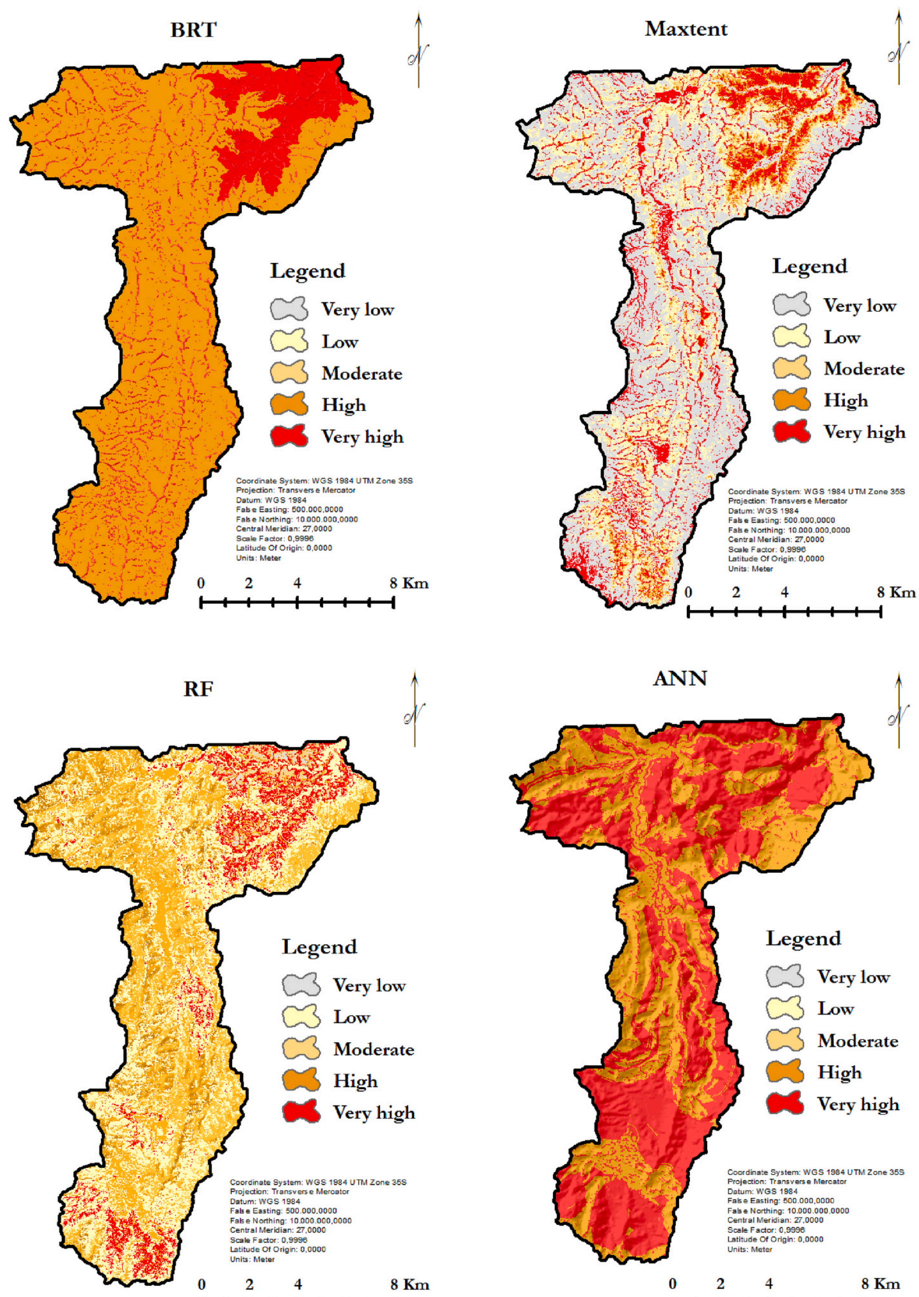

Fig. 6. Gully erosion susceptibility mapping (GESM) using four machine learning algorithms: Random Forest (RF), Maximum Entropy (MaxEnt), Boosted Regression Tree (BRT), and Artificial Neural Network (ANN).

therefore, it is an important factor for assessing prone areas to gully erosion. Stream Power Index (SPI) is one of the most important factors controlling slope erosion processes; regions with high stream power have high erosion potential. It is a measure of the erosive power of water flow based on the assumption that discharge is commensurate to a particular catchment area (Conforti et al., 2011). Arabameri et al. (2020a) used 12 gully erosion conditioning factors including elevation, slope, aspect, plan curvature, convergence index, TWI, lithology, land use/land cover (LU/LC), distance to rivers, distance to roads, drainage density, and NDVI and indicated that the RF model (with AUC = 0.927) had the highest prediction accuracy as compared to MARS and BRT.

These environmental factors also influence other phenomena such as cities classification (Rahmana et al., 2019); planning in integrated water resource management (Kafy et al., 2021a); in dynamic of land cover changes and in predicting seasonal urban thermal fields (Kafy et al., 2021b), etc. They must, therefore, be considered for a variety of environmental phenomena including gullying.

4.2. Machine learning algorithms as tools for gully erosion susceptibility mapping

From our results, MaxEnt and RF presented the best AUC (0.82 and 0.80, respectively) and TSS (0.56 and 0.50, respectively). DeLong test supported the same trends. The accuracy assessment is presented in Table 3. These models can, thus, be recommended for GESM at the watershed scale in eastern DRC. The use of an entropy model such as MaxEnt to develop susceptibility classifications for the gully erosion process is relatively new but has been gaining popularity in the fields of geosciences and geomorphology (Azareh et al., 2019; Avand et al., 2019). These models do not require assumptions about the

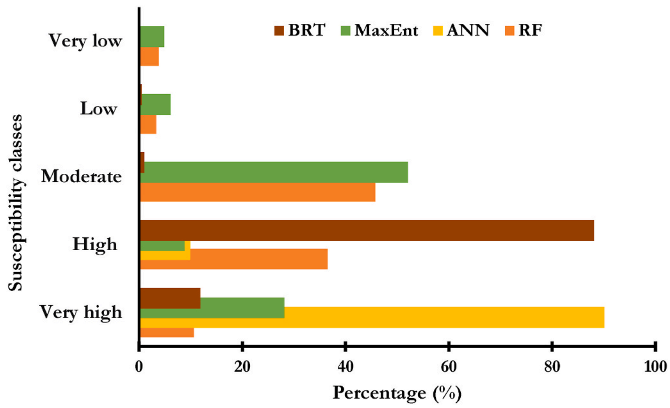

Fig. 7. Susceptibility classes variation according to the four machine learning algorithms used: Random Forest (RF), Maximum Entropy (MaxEnt), Boosted Regression Tree (BRT), and Artificial Neural Network (ANN).

appropriateness in distribution of explanatory variables, and thus, several properties can be used and tested. They examine quite well the statistical relationships between independent and dependent variables and provide metrics for the significance of the variables. Different researches have been done to generate GIS-based gully erosion susceptibility maps (GESMs). Svoray et al. (2012) used different machine learning models, such as SVM, decision tree (DT), SV, and ANN, for predicting gully initiation, and then, they compared their results with the AHP (analytic hierarchy process) and TT (topographic threshold) methods. The results of this study indicated that machine learning models provide a better predictive ability of gully susceptibility mapping than the use of both TT and AHP methods.

Rahmati et al. (2016) investigated GESM in Iran using bivariate statistical models including weights-of-evidence and index entropy models and stated that the index of entropy model with (AUC = 78.11%) had better accuracy in comparison to the weights-of-evidence model (AUC = 70.07%). These models have been tested in many regions at different scales. These models can be divided into three groups: the first can be called "knowledge-based models", such as the analytical hierarchy process (AHP)(Rahmati et al., 2016); the seconds are bivariate and multivariate statistical models, such as weights-of-evidence (WOE), logistic regression (LR) (Ahmadpour et al., 2021), maximum entropy (MaxEnt) (Rahmati et al., 2017), information value, conditional analysis (CA), and frequency ratio (FR) (Rahmati et al., 2016) and the thirds are called data-mining models, such as multivariate adaptive regression splines (MARS) (Conoscenti et al., 2018), Random Forest (RF) (Ahmadpour et al., 2021), support vector machine (SVM), classification and regression trees (CART) and acritical neural networks (ANN) (Pourghasemi et al., 2019).

However, natural events such as floods, earthquakes and avalanches, landslides, gully erosion, and soil erosion are often difficult to predict because they are uncertain with potentially detrimental consequences. Governments and research institutions worldwide have attempted for years to assess natural and man-made hazards to predict risk (Pradhan et al., 2010), but a few studies have investigated the prediction of gully erosion and its susceptibility mapping and particularly in data scarce regions. In all those models, the susceptibility technique is used to assess the relative probability of occurrence of erosion at a particular location compared to other locations under the influence of triggering factors known as geo-environmental factors. In all GESM analyses, soil, geological, and climate data are almost integrated and their effect is currently significant. However, in data scarce regions such as eastern DRC, such data are missing. The very low resolution of available geology and climate data at the watershed scale do not facilitate their integration into the model. However, according to the Luzinzi watershed location (Mitumba mountain chains with humid tropical to equatorial climate types), soil erodibility and rainfall erosivity could be high as mentioned by some estimations made by Chuma et al. (2021a) and Kulimushi et al. (2021) in the region. Though soil, geological, and climate data were not available, indexes such as SPI (Stream Power Index), TWI, TRI seemed to contribute and explain those missing factors. One of the most important factors in quantitative geomorphology and morphometry in determining the distribution of soil water content material is the TWI (Moore and Burch, 1986).

Other indexes such as the NDWI usefulness for drought monitoring and early warning has been demonstrated in different studies. It is sensitive to changes in liquid water content and spongy mesophyll of vegetation canopies (Ceccato et al., 2002; Gu et al., 2007). NDWI gives information on the deficit of rainfall/soil moisture to determine whether the variation in the vegetation response (signal) is linked with a drought event (Gu et al., 2007). In data-scare regions, higher accuracy is achieved by manual digitizing, unfortunately, it is extremely laborious and time-consuming with lots of subjectivity during gully interpretations. Such a study will be expensive to repeat for a large area such as eastern DRC territories. It is, therefore, necessary to adopt new methodologies such those presented in this study since they would be cost-effective to repeat and thus would help for monitoring purposes.

5. Conclusion

From the results, we concluded that RF and MaxEnt present high accuracy levels and can be adopted for GESM at the watershed scale. The four most important and influential factors among the 16 geo-environmental factors used for GESM are the elevation, NDWI, NDVI, and the Euclidean distance to roads. According to the two models, the investigated watershed is mostly dominated by moderate to very highly susceptible areas. The increasing nature of gully erosion-affected areas

Table 3
Model accuracy assessment for GESM in Luzinzi watershed, the Area Under the Curve (AUC), and the True Skill Statistic (TSS) determined for the four MLA.

	RF		ANN		MaxEnt		BRT	
Description	AUC	TSS	AUC	TSS	AUC	TSS	AUC	TSS
Mean	0.82[A]	0.56	0.55[D]	0.11	0.80[B]	0.50	0.69[C]	0.38
Median	0.82	0.57	0.52	0.06	0.81	0.51	0.69	0.38
Mode	0.83	0.57	0.59	0.21	0.78	0.48	0.70	0.38
Standard Deviation	**0.01**	**0.02**	**0.03**	**0.07**	**0.02**	**0.04**	**0.01**	**0.01**
Kurstosis	−1.50	−1.54	−1.48	−1.49	−1.35	−1.02	−0.69	−1.13
Asymmetry Coefficient	0.32	−0.49	0.72	0.72	0.34	−0.23	−0.28	0.50
Plage	0.02	0.04	0.09	0.18	0.06	0.15	0.03	0.03
Sum	81.89	55.95	54.61	10.76	80.40	50.04	68.96	38.31
Maximum	**0.83**	**0.58**	**0.59**	**0.21**	**0.85**	**0.59**	**0.71**	**0.40**
Minimum	**0.81**	**0.54**	**0.50**	**0.02**	**0.78**	**0.44**	**0.68**	**0.38**

Legend: AUC: Area Under the Curve, TSS: True Skill Statistic, GESM: Gully Erosion Susceptibility Mapping, MLA: Machine Learning Algorithm, RF: Random Forest, ANN: Artificial Neural Network, MaxEnt: Maximum Entropy, BRT: Boosted Regression Tree. The mean AUC values linked with different letters are statistically different at α = 0.05 (DeLong's test).

reduces the productivity of land and induces road cuts.

Identifying gully erosion susceptibility in cultivated watersheds such as Luzinzi is important for the managers and decision-makers, government, planners, and private agencies for effective land resource management, land-use planning, and environmental protection at the watershed scale starting with Luzinzi watershed and later extending to neighboring watersheds in the region. This study demonstrated the effectiveness of machine learning models for gully erosion susceptibility mapping using GIS and remote sensing tools. It also identified prominent geo-environmental factors that can be used in data-scarce regions such as eastern DRC and the sub-Saharan Africa in general. Future research should be focused on integrating soil property factors (both chemical and physical) while establishing a multi-temporal database to continuously update conditioning factors to improve mapping. A future study should also verify if the accuracy and precision of the same four machine-learning models tested in this study are maintained in time, with the spatial distribution of gullies at different and in lower density in the hydrographic basin. Nevertheless, the outcome of this study can help managers in Luzinzi watershed to mitigate the soil erosion problem and prevent future gully erosion by taking preventive measures. Conclusions are also valuable for other tropical areas sharing similar climatic and topographic conditions.

Funding

Not applicable.

Declaration of competing interest

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.

Data availability

Data will be made available on request.

Acknowledgment

We would like to thank anonymous reviewers for comments and suggestions to refine the quality of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pce.2022.103295.

References

Aggarwal, S., 2004. Principles of remote sensing. In: Sivakumar, M.V.K., Roy, P.S., Harmsen, K., Saha, S.K. (Eds.), Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, Dehra Dun, India, 7–11 July 2003. World Meteorological Organisation, Switzerland, pp. 23–38.
Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43 (6), 1223–1232.
Amiri, M., Pourghasemi, H.R., 2020. Mapping and preparing a susceptibility map of gully erosion using the MARS Model. In: Gully Erosion Studies from India and Surrounding Regions. Springer, Cham, pp. 405–413.
Arabameri, A., Asadi Nalivan, O., Saha, S., Roy, J., Pradhan, B., Tiefenbacher, J.P., Thi Ngo, P.T., 2020a. Novel ensemble approaches of machine learning techniques in modeling the gully erosion susceptibility. Rem. Sens. 12 (11), 1890.
Arabameri, A., Chen, W., Loche, M., Zhao, X., Li, Y., Lombardo, L., Cerda, A., Pradhan, B., Bui, D.T., 2020b. Comparison of machine learning models for gully erosion susceptibility mapping. Geosci. Front. 11 (5), 1609–1620.
Agnesi, V., Angileri, S., Cappadonia, C., Conoscenti, C., Rotigliano, E., 2011. Multi parametric gis analysis to assess gully erosion susceptibility: A test in southern sicily, italy. Landf. Anal. 17, 15–20.
Ahmadpour, H., Bazrafshan, O., Rafiei-Sardooi, E., Zamani, H., Panagopoulos, T., 2021. Gully erosion susceptibility assessment in the Kondoran watershed using machine learning algorithms and the Boruta feature selection. Sustainability 13 (18), 10110. https://doi.org/10.3390/su131810110.
Arabameri, A., Sadhasivam, N., Turabieh, H., Mafarja, M., Rezaie, F., Pal, S.C., Santosh, M., 2021. Credal decision tree based novel ensemble models for spatial assessment of gully erosion and sustainable management. Sci. Rep. 11 (1), 1–18.
Avand, M., Janizadeh, S., Naghibi, S.A., Pourghasemi, H.R., Khosrobeigi Bozchaloei, S., Blaschke, T., 2019. A comparative assessment of random forest and k-nearest neighbor classifiers for gully erosion susceptibility mapping. Water 11 (10), 2076.
Azareh, A., Rahmati, O., Rafiei-Sardooi, E., Sankey, J.B., Lee, S., Shahabi, H., Ahmad, B. B., 2019. Modelling gully-erosion susceptibility in a semi-arid region, Iran: investigation of applicability of certainty factor and maximum entropy models. Sci. Total Environ. 655, 684–696.
Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3, 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x.
Bingner, R.L., Theurer, F.D., 2001. AnnAGNPS: estimating sediment yield by particle size for sheet & rill erosion. In: Proceedings of the Sedimentation: Monitoring, Modeling, and Managing, 7th Federal Interagency Sedimentation Conference, pp. 1–7. Reno, NV, 25–29 March 2001.
Bingner, R.L., Wells, R.R., Momm, H.G., Rigby, J.R., Theurer, F.D., 2016. Ephemeral gully channel width and erosion simulation technology. Nat. Hazards 80 (3), 1949–1966.
Birhenjira, E.M., Useni, P.M., Badibanga, K., Myango, M., 2014. Caractérisation de la déformation dans le synclinorium de l'Itombwe (Kibarien superieur) à Kaziba, Sud-Kivu, RD Congo. Eur. Sci. J. 10 (27).
Borrelli, P., Robinson, D.A., Panagos, P., Lugato, E., Yang, J.E., Alewell, C., Wuepper, D., Montanarella, L., Ballabio, C., 2020. Land use and climate change impacts on global soil erosion by water (2015-2070). Proc. Natl. Acad. Sci. USA 117 (36), 21994–22001.
Borrelli, P., Poesen, J., Vanmaercke, M., Ballabio, C., Hervás, J., Maerker, M., et al., 2022. Monitoring gully erosion in the European Union: a novel approach based on the Land Use/Cover Area frame survey (LUCAS). Int. Soil Water Conserv. Res. 10 (1), 17–28.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Ceccato, P., Flasse, S., Gregoire, J.M., 2002. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 2. Validation and applications. Rem. Sens. Environ. 82, 198–207.
Chen, Z., Liang, S., Ke, Y., Yang, Z., Zhao, H., 2019. Landslide susceptibility assessment using evidential belief function, certainty factor and frequency ratio model at Baxie River basin, NW China. Geocarto Int. 34 (4), 348–367.
Chuma, B.G., 2019. Connaissances paysannes et évaluation des techniques de conservation du sol dans les petites exploitations de Kabare nord, est de la RD Congo. Faculté des bioingénieurs, Université catholique de Louvain, p. 82. http://hdl.handle.net/2078.1/thesis:22507.
Chuma, B.G., Mondo, M.J., Ndeko, B.A., Mugumaarhahama, Y.M., Bagula, M.E., Mulalisi, B., Muhaya, V., Kavimba, J., Katcho, K., Mushagalusa, N.G., 2021a. Forest cover affects gully expansion at the tropical watershed scale: case study of Luzinzi in Eastern DR Congo. Trees, Forests and People 4, 100083. https://doi.org/10.1016/j.tfp.2021.100083.
Chuma, B.G., Safina, B.F., Ndeko, B.A., Mugumaarhaham, Y.M., Cirezi, C.N., Mondo, M. J., Bagula, E.M., Katcho, K., Mushagalusa, N.G., Schmitz, S., 2021b. Estimation of soil erosion using RUSLE modeling and geospatial tools in a tea production watershed (Chisheke in Walungu), eastern Democratic Republic of Congo. Model. Earth Syst. Environ. 1–17. https://doi.org/10.1007/s40808-021-01134-3.
Chuma, G.B., Mondo, J.M., Ndeko, A.B., Bagula, E.M., Lucungu, P.B., Bora, F.S., Karume, K., Mushagalusa, G.N., Schmitz, S., Bielders, C.L., 2022. Farmers' knowledge and practices of soil conservation techniques in smallholder farming systems of northern Kabare, east of DR Congo. Environ. Challenges 7, 100516.
Conforti, M., Aucelli, P.P., Robustelli, G., Scarciglia, F., 2011. Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). Nat. Hazards 56 (3), 881–898.
Conoscenti, C., Agnesi, V., Cama, M., Caraballo-Arias, N.A., 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.
Dabney, S.M., Vieira, D.A.N., Yoder, D.C., Langendoen, E.J., Wells, R.R., Ursic, M.E., 2015. Spatially distributed sheet, rill, and ephemeral gully erosion. J. Hydrol. Eng. 20 (6), C4014009.
De Oliveira, C.M.M., Francelino, M.R., de Mendonça, B.A.F., Ramos, I.Q., 2020. Obtaining morphometric variables from gullies using two methods of interpolation laser scanner data: the case study of Vassouras, Brazil. J. Mt. Sci. 17 (12), 3012–3023.
Dewitte, O., Jones, A., Spaargaren, O., Breuning-Madsen, D., Brossard, M., Dampha, A., Deckers, J., Gallali, T., Hallett, S., Jones, R., Kilasara, M., Le Roux, P., Michéli, E., Montanarella, L., Thiombiano, L., Van Ranst, E., Yemefack, M., Zougmore, R., 2013. Harmonisation of the soil map of Africa at the continental scale. Geoderma 211–212, 138–153. https://doi.org/10.1016/j.geoderma.2013.07.007.
De'Ath, G., 2007. Boosted Trees for Ecological modeling and Prediction. Ecology 88, 243–251. https://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2.
Dixon, B., Candade, N., 2008. Multispectral land use classification using neural networks and support vector machines: one or the other, or both? Int. J. Rem. Sens. 29 (4), 1185–1206.

Dube, F., Nhapi, I., Murwira, A., Gumindoga, W., Goldin, J., Mashauri, D.A., 2014. Potential of weight of evidence modelling for gully erosion hazard assessment in Mbire District–Zimbabwe. Phys. Chem. Earth, Parts A/B/C 67, 145–152.
Dube, T., Sibanda, M., Shoko, C., 2016. Examining the variability of small-reservoir water levels in semi-arid environments for integrated water management purposes, using remote sensing. Trans. Roy. Soc. S. Afr. 71 (1), 115–119.
Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77 (4), 802–813.
Felicísimo, Á.M., Cuartero, A., Remondo, J., Quirós, A., 2013. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10 (2), 175–189. https://doi.org/10.1007/s10346-012-0320-1.
Flanagan, D.C., Nearing, M.A., 1995. USDA-Water Erosion Prediction Project: hillslope profile and watershed model documentation. Nserl Rep 10, 1–123.
Garosi, Y., Sheklabadi, M., Pourghasemi, H.R., Besalatpour, A.A., Conoscenti, C., Van Oost, K., 2018. Comparison of differences in resolution and sources of controlling factors for gully erosion susceptibility mapping. Geoderma 330, 65–78.
Gayen, A., Haque, S.M., Saha, S., 2020. Modeling of gully erosion based on random forest using GIS and R. In: Gully Erosion Studies from India and Surrounding Regions. Springer, Cham, pp. 35–44.
Gayen, A., Pourghasemi, H.R., 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. https://doi.org/10.1016/j.scitotenv.2019.02.436.
Gideon, D., Mustafa, F.B., Victor, I., 2021. The application of an expert knowledge-driven approach for assessing gully erosion susceptibility in the subtropical Nigerian savannah. Singapore J. Trop. Geogr. 42 (1), 107–131.
Gómez-Gutiérrez, Á., Conoscenti, C., Angileri, S.E., Rotigliano, E., Schnabel, S., 2015. Using topographical attributes to evaluate gully erosion proneness (susceptibility) in two Mediterranean basins: advantages and limitations. Nat. Hazards 79 (1), 291–314.
Griffith, J.A., Martinko, E.A., Whistler, J.L., Price, K.P., 2002. Interrelationships among landscapes, NDVI, and stream water quality in the US Central Plains. Ecol. Appl. 12 (6), 1702–1718.
Gu, Y., Brown, J.F., Verdin, J.P., Wardlow, B., 2007. A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States. Geophys. Res. Lett. 34.
Heri-Kazi, A.B., Bielders, C.L., 2021. Cropland Characteristics and Extent of Soil Loss by Rill and Gully Erosion in Smallholder Farms in the Kivu Highlands, DR Congo. Geoderma Regional, e00404.
Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. Habitat suitability and distribution models: with applications in R. Cambridge University Press.
Heri-Kazi, B.A., 2020. Caractérisation de l'état de dégradation des terres par l'érosion hydrique dans le Sud-Kivu montagneux à l'Est de la RD Congo. Doctoral Dissertation, UCL-Université Catholique de Louvain, p. 323.
Higaki, D., Karki, K.K., Koiwa, N., Takahashi, M., Ghimire, S.K., 2020. Rehabilitation of gully-dominant hill slopes by using low-cost measures: a case study in Nepal. In: Workshop on World Landslide Forum. Springer, Cham, pp. 347–353.
Hong, H., Naghibi, S.A., Pourghasemi, H.R., Pradhan, B., 2016. GIS-based landslide spatial modeling in Ganzhou City, China. Arabian J. Geosci. 9 (2), 1–26. https://doi.org/10.1007/s12517-015-2094-y.
Hosseinalizadeh, M., Kariminejad, N., Chen, W., Pourghasemi, H.R., Alinejad, M., Behbahani, A.M., Tiefenbacher, J.P., 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.
Imwangana, F.M., Vandecasteele, I., Trefois, P., Ozer, P., Moeyersons, J., 2015. The origin and control of mega-gullies in Kinshasa (DR Congo). Catena 125, 38–49.
Imwangana, M.F., 2014. In: Etude de l'érosion ravinante à Kinshasa : Dynamisme Pluvio-Morphogénique et Développement d'un Outil de Prévision. Ph.D. Thesis. Université de Kinshasa, p. 209.
Javidan, N., Kavian, A., Pourghasemi, H.R., Conoscenti, C., Jafarian, Z., 2020. Data mining technique (maximum entropy model) for mapping gully erosion susceptibility in the gorganrood watershed, Iran. In: Gully Erosion Studies from India and Surrounding Regions. Springer, Cham, pp. 427–448.
Jiang, C., Fan, W., Yu, N., Liu, E., 2021. Spatial modeling of gully head erosion on the Loess Plateau using a certainty factor and random forest model. Sci. Total Environ. 783, 147040.
Kafy, A.A., Raikwar, V., Al Rakib, A., Kona, M.A., Ferdousi, J., 2021a. Geospatial approach for developing an integrated water resource management plan in Rajshahi, Bangladesh. Environ. Challenges 4, 100139. https://doi.org/10.1016/j.envc.2021.100139.
Kafy, A.A., Naim, M.N.H., Subramanyam, G., Ahmed, N.U., Al Rakib, A., Kona, M.A., Sattar, G.S., 2021b. Cellular Automata approach in dynamic modelling of land cover changes using RapidEye images in Dhaka, Bangladesh. Environ. Challenges 4, 100084. https://doi.org/10.1016/j.envc.2021.100084.
Kheir, R.B., Wilson, J., Deng, Y., 2007. Use of terrain variables for mapping gully erosion susceptibility in Lebanon. Earth Surf. Process. Landforms 32 (12), 1770–1782.
Kornejady, A., Ownegh, M., Bahremand, A., 2017. Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. Catena 152, 144–162. https://doi.org/10.1016/j.catena.2017.01.010.
Kuhnert, P.M., Henderson, A.-K., Bartley, R., Herr, A., 2010. Incorporating uncertainty in gully erosion calculations using the random forests modelling approach. Environmetrics 21, 493–509. https://doi.org/10.1002/env.999.
Kulimushi, L.C., Choudhari, P., Mubalama, L.K., Banswe, G.T., 2021. GIS and remote sensing-based assessment of soil erosion risk using RUSLE model in South-Kivu province, eastern, Democratic Republic of Congo. Geomatics, Nat. Hazards Risk 12 (1), 961–987.
Le Roux, J., van der Waal, B., 2020. Gully erosion susceptibility modelling to support avoided degradation planning. S. Afr. Geogr. J. 102 (3), 406–420. https://doi.org/10.1080/03736245.2020.1786444.
Lesschen, J.P., Kok, K., Verburg, P.H., Cammeraat, L.H., 2007. Identification of vulnerable areas for gully erosion under different scenarios of land abandonment in Southeast Spain. Catena 71 (1), 110–121.
Liaw, A., Wiener, M., 2002. Classification and regression by random Forest. R. News 2 (3), 18–22.
Liou, Y.-A., Liu, S.-F., Wang, W.J., 2001. Retrieving soil moisture from simulated brightness temperatures by a neural network. IEEE Trans. Geosci. Rem. Sens. 39, 1662–1673.
Lombardo, L., Cama, M., Conoscenti, C., Märker, M., Rotigliano, E.J., 2015. Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: application to the 2009 storm event in Messina (Sicily, southern Italy). Nat. Hazards 79 (3), 1621–1648. https://doi.org/10.1007/s11069-015-1915-3.
Lu, D., Weng, Q.A., 2007. Survey of image classification methods and techniques for improving classification performance. Int. J. Rem. Sens. 28, 823–870.
McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Rem. Sens. 17 (7), 1425–1432.
Mekonnen, Z., Berie, H.T., Woldeamanuel, T., Asfaw, Z., Kassa, H., 2018. Land use and land cover changes and the link to land degradation in Arsi Negele district, Central Rift Valley, Ethiopia. Remote Sens. Appl.: Society and Environment 12, 1–9.
Moore, I.D., Burch, G.J., 1986. Sediment transport capacity of sheet and rill flow: application of unit stream power theory. Water Resour. Res. 22 (8), 1350–1360.
Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol. Process. 5 (1), 3–30.
Moreno, R., Zamora, R., Molina, J.R., Vasquez, A., Herrera, M.A., 2011. Predictive modeling of microhabitats for endemic birds in South Chilean temperate forests using Maximum entropy (Maxent). Ecol. Inf. 6, 364–370. https://doi.org/10.1016/j.ecoinf.2011.07.003.
Morgan, R.P.C., Mngomezulu, D., 2003. Threshold conditions for initiation of valley-side gullies in the Middle Veld of Swaziland. Catena 50 (2–4), 401–414.
Nachtergaele, J., Poesen, J., Vandekerckhove, L., Oostwoud Wijdenes, D., Roxo, M., 2001. Testing the ephemeral gully erosion model (EGEM) for two Mediterranean environments. Earth Surface Processes and Landforms. J. British Geomorphol. Res. Group 26 (1), 17–30.
Naghibi, S.A., Pourghasemi, H.R., Dixon, B., 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ. Monit. Assess. 188 (1), 1–27.
Nalivan, O.A., Badehian, Z., Sadeghinia, M., Soltani, A., Islami, I., Boustan, A., 2022. A step beyond susceptibility: an adaptation of risk framework for monetary risk estimation of gully erosion. Nat. Hazards 111, 1661–1684. https://doi.org/10.1007/s11069-021-05110-z.
Nyssen, J., Poesen, J., Veyret-Picot, M., Moeyersons, J., Haile, M., Deckers, J., Dewit, J., Naudts, J., Teka, 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.
Ollobarren Del Barrio, P., Campo-Bescós, M.A., Giménez, R., Casalí, J., 2018. Assessment of soil factors controlling ephemeral gully erosion on agricultural fields. Earth Surf. Process. Landforms 43 (9), 1993–2008.
Ozdemir, A., Altural, T., 2013. A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: sultan Mountains, SW Turkey. J. Asian Earth Sci. 64, 180–197.
Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31 (2), 161–175.
Park, N.W., 2015. Using maximum entropy modeling for landslide susceptibility mapping with multiple geoenvironmental data sets. Environ. Earth Sci. 73 (3), 937–949. https://doi.org/10.1007/s12665-014-3442-z.
Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3–4), 231–259.
Podwojewski, P., Janeau, J.L., Caquineau, S., Hughes, J., 2020. Mechanisms of lateral and linear extension of gullies (dongas) in a subhumid grassland of South Africa. Earth Surf. Process. Landforms 45 (13), 3202–3215.
Poesen, J., 2011. Challenges in gully erosion research. Landform Analysis 17, 5–9.
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: Modelling Soil Erosion by Water. Springer, Berlin, Heidelberg, pp. 285–311.
Pourghasemi, H.R., Gayen, A., Panahi, M., Rezaie, F., Blaschke, T., 2019. Multi-hazard probability assessment and mapping in Iran. Sci. Total Environ. 692, 556–571.
Pradhan, B., Sezer, E.A., Gokceoglu, C., Buchroithner, M.F., 2010. Landslide susceptibility mapping by neuro-fuzzy approach in a landslide-prone area (Cameron Highlands, Malaysia). IEEE Trans. Geosci. Rem. Sens. 48 (12), 4164–4177.
Rahmana, R.M., Mohiuddinc, M., Al Kafy, A., Sheele, P.K., Di, L., 2019. Classification of cities in Bangladesh based on remote sensing derived spatial characteristics. J. Urban Manag. 8, 206–224. https://doi.org/10.1016/j.jum.2018.12.001.
Rahmati, O., Haghizadeh, A., Pourghasemi, H.R., 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., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Feizizadeh, B., 2017. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. Geomorphology 298, 118–137.
Ridgway, T., Riginos, C., Davis, J., Hoegh-Guldberg, O., 2008. Genetic connectivity patterns of *Pocillopora verrucosa* in southern African Marine protected areas. Mar. Ecol. Prog. Ser. 354, 161–168.
Rigge, M., Smart, A., Wylie, B., 2013. Optimal placement of off-stream water sources for ephemeral stream recovery. Rangel. Ecol. Manag. 66 (4), 479–486.
Riley, S.J., DeGloria, S.D., Elliot, R., 1999. Index that quantifies topographic heterogeneity. Intermt. J. Sci. 5 (1–4), 23–27.
Rouse Jr., J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation.
Schuman, C.D., Birdwell, J.D., 2013. Dynamic artificial neural networks with affective systems. PLoS One 8, e80455.
Sidorchuk, A., 1999. Dynamic and static models of gully erosion. Catena 37 (3–4), 401–414.
Sivia, D., Skilling, J., 2006. Data analysis: a Bayesian tutorial. Oxford University Press, p. 246.
Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. 43 (6), 1947–1958.
Svoray, T., Michailov, E., Cohen, A., Rokah, L., Sturm, A., 2012. Predicting gully initiation: comparing data mining techniques, analytical hierarchy processes and the topographic threshold. Earth Surf. Process. Landforms 37 (6), 607–619.
Taruvinga, K., 2008. Gully Mapping Using Remote Sensing: Case Study in Kwazulu Natal, South Africa. M.Sc. Thesis, University of Waterloo, Ontario, Canada.
Theobald, D., Norman, J., Peterson, E.E., Ferraz, S., 2005. Functional linkage of watersheds and streams (FLoWS): network-based ArcGIS tools to analyze freshwater ecosystems. In: Proceedings of the Second Annual International Symposium on GIS Spatial Analyses in Fishery and Aquatic Sciences. ESRI Press, Redlands, California.
Therneau, T.M., Atkinson, B., Ripley, B.. rpart: Recursive Partitioning and Regression Trees. R package version. http://CRAN.R-project.org/package=rpart.
Torkashvand, A.M., Alipour, H.R., 2009. Investigation of the possibility to prepare supervised classification map of gully erosion by RS and GIS. World Academy Sci. Eng. Technol. 55, 291–293.
Townsend Peterson, A., Papeş, M., Eaton, M., 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30, 550–560. https://doi.org/10.1111/j.0906-7590.2007.05102.x.
Tsangaratos, P., Benardos, A., 2014. Estimating landslide susceptibility through an artificial neural network classifier. Nat. Hazards 74 (3), 1489–1516.
Uchida, S., 2015. Applicability of Satellite Remote Sensing for Mapping Hazardous State of Land Degradation by Soil Erosion on Agricultural Areas, vol. 24. Procedia Environmental Sciences, pp. 29–34.
USDA, 1980. CREAMS: a field-scale model for chemicals, runoff, and erosion from agricultural management systems. In: Knisel, W.G. (Ed.), Conservation Research Rep. No. 26. Washington, DC.
Valentin, C., Poesen, J., Li, Y., 2005. Gully erosion: impacts, factors and control. Catena 63 (2–3), 132–153.
Vanmaercke, M., Poesen, J., Van Mele, B., Demuzere, M., Bruynseels, A., Golosov, V., Bezerra, J.F.R., Bolysov, S., Dvinskih, A., Frankl, A., Fuseina, Y., 2016. How fast do gully headcuts retreat? Earth Sci. Rev. 154, 336–355.
Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol. Appl. 21, 335–342. https://doi.org/10.1890/10-1171.1.
Woodbury, A.D., Render, F.W., Ulrych, T.J., 1995. Practical probabilistic groundwater modeling. Groundwater 33 (4), 532–538.
Woodward, D.E., 1999. Method to predict cropland ephemeral gully erosion. Catena 37 (3–4), 393–399.
Yu, H., Jiang, S., Land, K.C., 2015. Multicollinearity in hierarchical linear models. Soc. Sci. Res. 53, 118–136.
Zabihi, M., Mirchooli, F., Motevalli, A., Darvishan, A.K., Pourghasemi, H.R., Zakeri, M. A., Sadighi, F., 2018. Spatial modelling of gully erosion in Mazandaran Province, northern Iran. Catena 161, 1–13.
Zakerinejad, R., Maerker, M., 2015. An integrated assessment of soil erosion dynamics with special emphasis on gully erosion in the Mazayjan basin, southwestern Iran. Nat. Hazards 79 (1), 25–50.
Zádorová, T., Skála, J., Žížala, D., Vaněk, A., Penížek, V., 2021. Harmonization of a large-scale national soil database with the world reference Base for soil resources 2014. Geoderma 384, 114819. https://doi.org/10.1016/j.geoderma.2020.114819.
Zhu, H., Tang, G., Qian, K., Liu, H., 2014. Extraction and analysis of gully head of Loess Plateau in China based on digital elevation model. Chin. Geogr. Sci. 24 (3), 328–338.
Zhu, T., Xu, X., 2021. Gully erosion. In: Watershed Erosion Processes. Springer, Cham, pp. 41–67.