Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

Article

Novel Ensemble approach of Deep Learning Neural Network model and Particle swarm optimization (PSO) algorithm for prediction of gully erosion susceptibility

Shahab S. Band[1*], Saeid Janizadeh[3], Subodh Chandra Pal[3], Asish Saha[4], Rabin Chakrabortty[5], Manouchehr Shokri[7] and Amirhosein Mosavi[8,9]

1 Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
2 Future Technology Research Center, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.
3 Department of Watershed Management Engineering and Sciences, Faculty in Natural Resources and Marine Science, Tarbiat Modares University, Tehran, 14115-111 Iran, janizadehsaeid@modares.ac.ir
4 Department of Geography, The University of Burdwan, West Bengal, India, scpal@geo.buruniv.ac.in
5Department of Geography, The University of Burdwan, West Bengal, India, asishsaha01@gmail.com
6Department of Geography, The University of Burdwan, West Bengal, India, rabingeo8@gmail.com
7 Institute of Structural Mechanics, Bauhaus Universität Weimar, 99423 Weimar, Germany; manouchehr.shokri@uni-weimar.de
8 Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc ThangUniversity, Ho Chi Minh City, Vietnam
9 Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam; amirhosein.mosavi@tdtu.edu.vn

Abstract This study aims to evaluate a new approach in modeling gully erosion susceptibility (GES) based on deep learning neural network (DLNN) model, ensemble Particle swarm optimization (PSO) algorithm with DLNN (PSO-DLNN) and comparing these approaches with common artificial neural network (ANN) and support vector machine (SVM) models in Shiran watershed, Iran. For this purpose, 13 independent variables affecting GES in the study area, including altitude, slope, aspect, plan curvature, profile curvature, drainage density, distance from the river, land use, soil, lithology, rainfall, stream power index (SPI), topographic wetness index (TWI), were prepared. Also, 132 gully erosion locations were identified during field visits. Data for modeling were divided into two categories of training (70%) and testing (30%). Receiver operating characteristic (ROC) parameters including sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), and area under curve (AUC) were used to evaluate the performance of the models. The results showed that, the AUC values from ROC with considering testing datasets of PSO-DLNN is 0.89 and which is associated with superb accuracy. The rest of the models are also associated with optimal accuracy and near about PSO-DLNN model; the AUC values from ROC of DLNN, SVM, and ANN for testing datasets are 0.87, 0.85 and 0.84 respectively. The PSO algorithm has updated and optimized the weights of DLNN model, and as a result, the efficiency of this model in predicting GES has increased. Therefore, it can be concluded that the use of DLNN model and its ensemble with the PSO algorithm can be used as a novel and practical method in predicting the susceptibility of gully erosion that helps planners and managers in managing and reducing the risk of this phenomenon.

Keywords: Gully erosion susceptibility; deep learning neural network (DLNN), particle swarm optimization (PSO), Shiran watershed

© 2020 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

1. Introduction

Biodiversity in the area depends to a large extent on and supports the most vital natural resources in the soil, also providing basic human needs such as food, fresh air and clean water [1]. Therefore, human survival largely depends on the soil component. Although soil erosion in the form of gully erosion is a serious global problem, it continues to pose a threat to soil and water resources, particularly in arid and semi-arid regions of Iran [2,3]. Among the several types of water induced erosion, gully erosion is more intense form of soil erosion[4], and one of the most complex geomorphic phenomena on the earth surface [5]. Such kind of erosional activities also change shape of the earth's landform and produced a rugged topography which is not suitable for production activity, construction of communication network etc. Thus, water-induced soil erosion is the main cause of the destruction of agricultural land, vegetation, ecosystems and, finally, responsible for devastating land degradation phenomenon. It has been estimated that the annual rate of global soil erosion is approximately 75 billion tons [6]. From an international perspective, Iran ranks second in terms of land losses and the annual rate of soil erosion is close to 2 to 2.5 billion tons [7]. It has also been predicted that Iran's average soil erosion rate is 30-32 tons/ha/year, which is 4.3 times the world average (FAO, 1984). In Iran, soil erosion has been estimated to have caused almost 10 trillion economic losses (NGD, 2017), and is a national threat [8]. Thus, it is indeed necessary to protect the soil from erosion and to avoid the phenomenon of land degradation worldwide. The main causes of intensive water related gully erosion and its development is to long hot-dry season followed by extreme wet period. Therefore, extreme rainfall causes utmost surface runoff over the infiltration capacity and easily transported looses soil particles into the downward slope. Thus, soil erosion related to water in Iran is a major barrier to sustainable development in the areas of agriculture, watershed management and other activities related to resource development [9]. Hence, preparation of GES map is essential for sustainable management, development and conservation of most vital natural resources on the earth surface i.e. soil from intense gully formation and development.

. Before preparing GES map, it is necessary to understand about a gully, its morphological characteristics, causes of occurrences, conditioning factors and its ultimate impact on land surface. A gully can be defined as a deep, narrow channel with a depth of more than 30 cm, usually produced by a surface and subsurface runoff after heavy downpour with a temporary flow of water within that channel [8]. Gullies are generally transported large amount of sediment from the high slope or plateau of the unprotected soil surface i.e. less vegetation areas to the down-slope areas of a watershed. It is also an established fact that within 5% area of a watershed, near about 10% to 94% sediment moves down wards due to gully erosion [10]. According to Poesen [11] different factors affecting gully erosion, and this factors are classified into two categories i.e. (a) anthropogenic activities such as excessive use of farm land, over grazing, unplanned way of road constructions, deforestation etc., and (b) physical conditions such as topography, climate, vegetation cover, mineral composition in the soil etc. Basically, depending upon the depth, gullies are classified into three types i.e. if depth is <0.3 m then it is called grove, if depth varies between 0.3 m to 2 m it is called shallow gully and known as deep gully if depth is >2 m [12]. Intensive gully erosion causes a number of environmental problems, such as accumulation of sediment in rivers and devastating floods, as it removes fertile soils that have had a serious impact on agricultural fields and minimizes soil water storage capacity, destroys roads and ultimately produces badlands [13–15]. It is also well known fact that the similar factors are not responsible for the occurrences of gullies in several places on the world. Gullies are generally formed and developed based on the local topographical, climatological and hydrological characteristics. Therefore, different gully prone areas and associated factors need to be identified by mapping the susceptibility of gully erosion. Not only this, suitable prediction model along with identification of respective favorable gully erosion conditioning factors (GECFs) are also essential for bias less prediction result. With the passage of time several methods such as statistical, machine learning (ML) and ensemble algorithm have been used for mapping GES, with the combination of remote sensing-geographic information system. Thus, GES mapping, using aforesaid mentioned newly developed methods can helps land use planners to maintain soil and water

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

resources in a sustainable and accurate manner. In addition, the potential of the respective region will ultimately increase in the manner in which suitable measures are taken.

In recent times, ML algorithms have been widely used to spatial prediction of several natural hazards such as flooding, landslide [16], wildfire [17] etc. Several researchers throughout the world have been carried out GES mapping by using statistical as well as ML algorithms. Some of the widely used statistical methods to predict GES mapping are frequency ratio [7], logistic regression [18], weight of evidence (WoE) [19], index of entropy (IoE) [5] etc. Beside statistical methods, different ML algorithm have also been widely used to predict GES mapping such as artificial neural network (ANN) [20], support vector machine (SVM) [21], random forest (RF) (Hosseinalizadeh et al. 2019), multi-layer perception approach (MLPC) [22], classification and regression tree (CART) [23], boosted regression tree (BRT) [7], particle swarm optimization (PSO) [24], multivariate adaptive regression spline (MARS) [5], maximum entropy [25] etc. Beside this, Ensemble Models has also been widely used for its novelties and capabilities in the comprehensive analysis of GES mapping [26]. Basically, ensemble models are applied for high precision and predictive analysis of any kind of natural hazards susceptibility mapping. In another words, presentation of a ML models are significantly enhanced by using ensemble models. Along with machine learning models different ensemble models have also been used for gully erosion modeling [21].

In very recent times, the Deep Neural Learning Network (DLNN) is a striking ML algorithm and has been widely used by people in several research groups. This method was proposed for the first time in 2006 including the different key features of ML as well as artificial intelligence (AI). Basically, the DLNN algorithm consists of fully convolutional neural networks (CNNs), deep belief networks (DBNs), stacked auto-encoders (SAEs) networks, etc. [27]. In addition to this, Adaptive moment estimation (Adam) and Rectified Linear Unit (ReLU) algorithm was used to train and activation purposes in every learning unit of a DLNN model [28]. Generally, DLNN algorithm has been used in different fields such as feature extraction and transformation through supervised and unsupervised processes, recognition of pattern and their classification [29]. On the other hand, Particle swarm optimization (PSO) algorithm is an extended part of AI, in amalgamation of the conventional ML techniques. Basically, PSO algorithm is based on swarm intelligence and it is a straightforward with efficient universal optimization techniques [27]. PSO is used to feature selection of a dataset through the optimization techniques.

Although, deep learning (DL) and traditional ML algorithm have some basic differences and the are DL algorithm needs big data size to performed and analyze in a better way and in the case of ML they performed in the way of established rules. The algorithm of DL requires lots of matrices operation function than in ML algorithm to perform better [30]. In the case of problem solving method, DL algorithm done it through end to end problems solving and in ML case, it breaks down into multiple sub-problems. Therefore, the DL algorithm is much better than the traditional ML algorithm for mapping the GES zone. Thus, the most advantage of using the DLNN algorithm is that this model is capable of building a high-level feature from a raw dataset scientifically, and is also capable of delivering forecasting results using time series data. In addition to this, DLNN consist of different topology than the general neural network of a single hidden layer, thus, more than one hidden layers present in this algorithm. For this reason, in various research areas, DL algorithm has better performance than the conventional ML algorithm [31]. In the case of PSO, it is also used to conquer the problems of local optima through feature selection methods. Basically, the PSO determines the quality of datasets features through multi-objective fitness function [32]. As a result, the output layer of different hidden layers was optimized by the PSO algorithm to obtain more accurate predictions [33].

Therefore, the present research work has been carried out to predict GES mapping in Shirahan watershed, which is tremendously affected by water induced gully erosion. To fulfill our research objective here we used thirteen suitable GECFs with a total of 132 gully head-cut points (each for gully and non-gully), which is splitting into 70/30 ratio for training and testing dataset. Furthermore, for modelling the GES mapping in a creative way here we used DL as well as conventional ML algorithm. In this study, we used DLNN, PSO, artificial neural network (ANN) and support vector

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

machine (SVM) algorithm. According to the several literature survey on GES mapping and best of our knowledge it has been noticed that DLNN model was not used in GES assessment so far; thus, this study was carried out to potential application of DLNN model for GES mapping. This study also an attempt has been made to use the PSO algorithm in optimizing the parameters of the deep learning model (DLNN) in the training phase and to introduce a new approach of ensemble the PSO and DLNN in GES modeling. Not only this, a comparison was carried out among the ensemble of PSO-DLNN to conventional ANN and SVM algorithm. Thus, the application of DL and the ensemble of PSO-DLNN approach for GES mapping is the novelty in this research study as the result of this approach has improved the prediction accuracy than any single ML algorithm. Thereafter, all of the output result was validated through sensitivity (SST), specificity (SPF), positive predictive values (PPV), negative predictive values (NPV), receiver operating characteristics-area under curve (ROC-AUC), likelihood ratio, F-measures and maximum probability of correct decision (MPCD) statistical analysis. Thus, DL and the ensemble of PSO-DLNN methods will help to forecast, control gully creation and their development in Shiran watershed, Iran.

2. Materials and methods

2.1. Description of the study area

Shirahan watershed is located between the length of 20° 57' to 28° 57' and the width of 51° 25' to 51° 26' and is located in the central part of Hormozgan province and south of Bandar Jask city (Figure 1). The area of this area is 138 km^2, the minimum height of the area is 2 meter and the maximum height is 214 meters above sea level. According to the statistics recorded in Jask Synoptic Station with a statistical period of 28 years (1989-2017), rainfall in this region is very heavy in the form of rain and more than 50% of it occurs in winter. According to the information of the above station, the average annual rainfall is 116.75 mm, the maximum annual rainfall is 320 mm and the minimum is 27 mm. The climate of the region is hot with dry amber method and hot dry with Domartans. Soil texture is generally silt/loam and loam. In this area, the percentage of clay has increased with increasing depth. However, changes in the percentage of sand and silt do not follow a specific trend and have high fluctuations. In this area, the horizon of 75-75 cm has the highest degree of salinity (Table 1). Pictures of ditch erosion are shown in Figure 2. In order to study the geometric features and physical and chemical properties of the soil, 20 ditches were sampled in the study area. Studies have shown that further expansion of ditch erosion in salt marshes, which are located in the plain type, the general plan of ditches is compound and their cross-sectional shape is trapezoidal. The average depth of ditches is 2.7 meters; the average width is 10.3 meters. Laboratory studies have been used to evaluate the soil characteristics of gully Shirahan. Meanwhile, soil samples were removed from the soil surface to a depth of 290 cm and sent to the laboratory of Bandar Abbas Agricultural and Natural Resources Research Center for soil testing. Laboratory results showed that the soil texture in the area up to the depth under study is loose. The physical and chemical properties of the soil at 6 different soil depths are shown in the ditches studied in Table 1. Some of the field photographs of gully in the present study area of Shirahan watershed have been shown in Figure 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

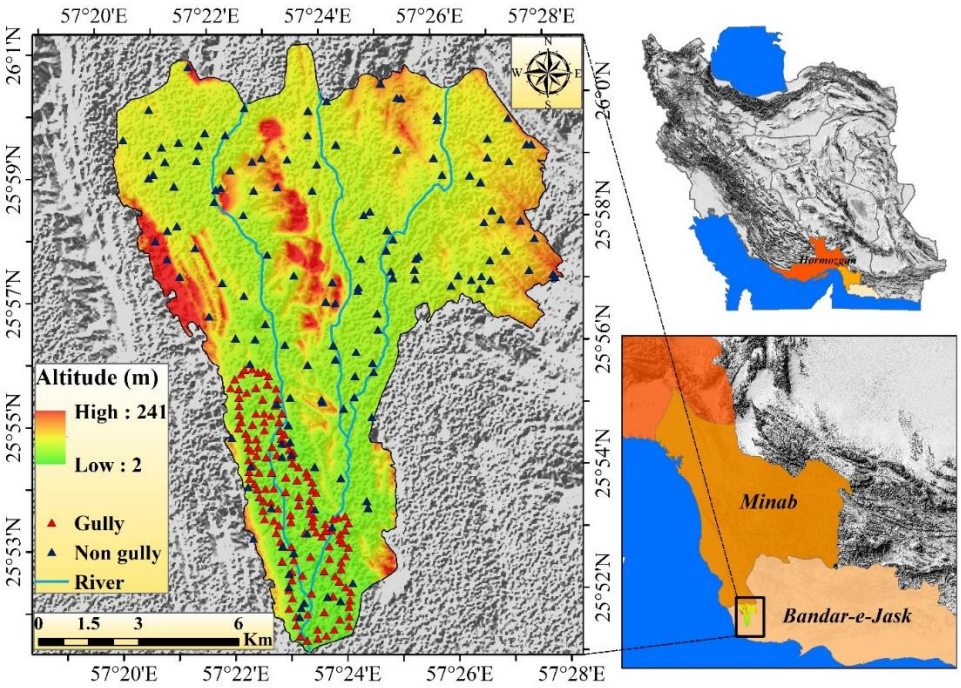


Figure 1. Location of the study area in Iran.

Table 1. Physical and chemical properties of soil in the gullies of Shirahan watershed.

Features	Soil depth (cm)					
	0-30	30-75	75-130	130-180	18-250	250-290
PH	8.06	7.59	8.19	7.69	8.32	7.38
EC (mmhos/cm)	2.26	34.6	2.23	33.9	2.4	33.2
Na (Meq/lit)	8.82	285	8.87	285	8.8	248
Ca+Mg (Meq/lit)	13.6	64.4	13.7	62.4	13.4	63.2
SAR	3.4	50.2	3.4	50.2	3.4	50.2
Clay (%)	24	26	26	27	29	28
Silt (%)	58	30	60	32	56	30
Sand (%)	18	44	14	41	15	42
Soil texture	Silt-Loam	Loam	Silt-Loam	Loam/clay-loam	Silty-Clay-Loam	Clay-Loam

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

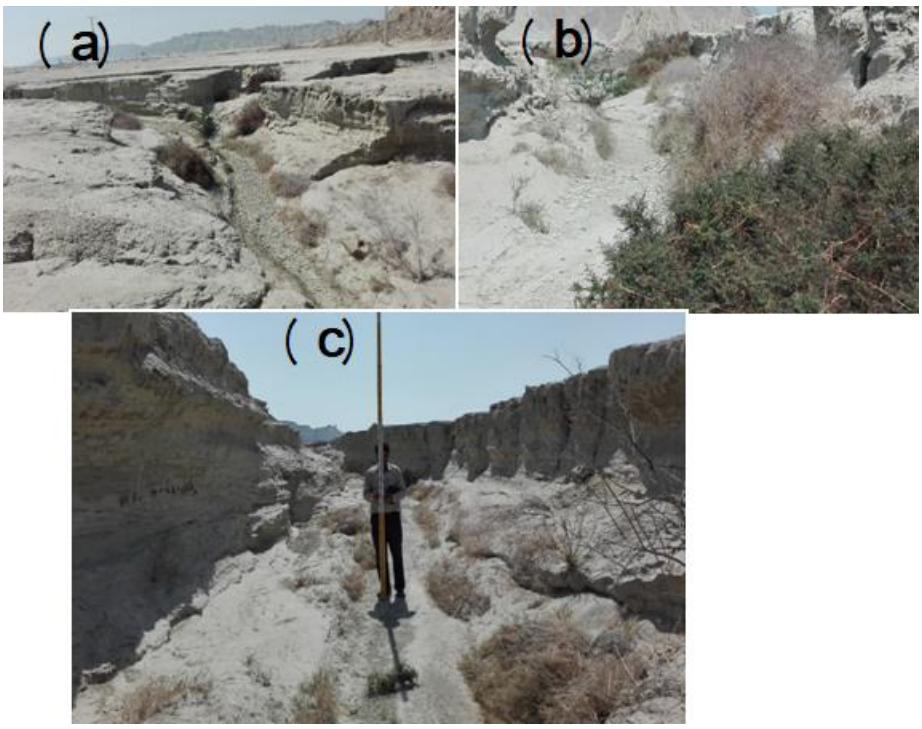


Figure 2. Images of the gully erosion in the Shirahan watershed.

2.2. Methodology

The methodological approach in this research work has been discussed in the following section and respective flowchart is presented in Figure 3.

In the very early, a gully erosion inventory map was prepared based on the 132 gully head-cut points (for each gully and non-gully). These gully head-cut points were identified based on field visits and information of Administration of Natural Resource of Hormozghan Province. Along with this, the non-gully points were randomly selected throughout the basin area with the help of GIS environment. Beside this, a total of thirteen (13) of gully erosion conditioning factors (GECFs) i.e. target variables were considered for modeling GES based on the local topographical and climatological factors in association with several literature studies. These GECFs are altitude, aspect, slope, plan curvature, profile curvature, drainage density (DD), distance from river, land use, lithology, Soil, rainfall, stream power index (SPI) and topographic wetness index (TWI). Thereafter, Multi-collinearity analysis of variance inflation factor (VIF) and tolerance (TOL) techniques was used among different GECFs to know the linear relationship among the variables. Afterward, modelling of GES was done by using SVM, ANN, DLNN machine learning (ML) algorithm and a novel ensemble by PSO-DLNN. Lastly, several GES model's result was validated through ROC curve analysis to know the accuracy assessment.

Here, it is also mentioned that the methodology of the present research work has been carried out to solve classification problems through aforesaid ML and DL algorithms for prediction GES mapping. Beside this, several target variables used in this study are a combination of logic, discrete and continuous variables. During the processing of all of these variables data, it has been recognized among the variables that which one is logic, discrete or a continuous one, in the SPSS 25 statistical software designed by International Business Machines (IBM). In this study we also analysis the both i.e. which one is already affected by gully erosion susceptible zone, by using the presence of gully head-cut points and also compute the GES zones based on the gully, non-gully had-cut points along with several conditioning factors for sustainable management of the gully affected areas.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

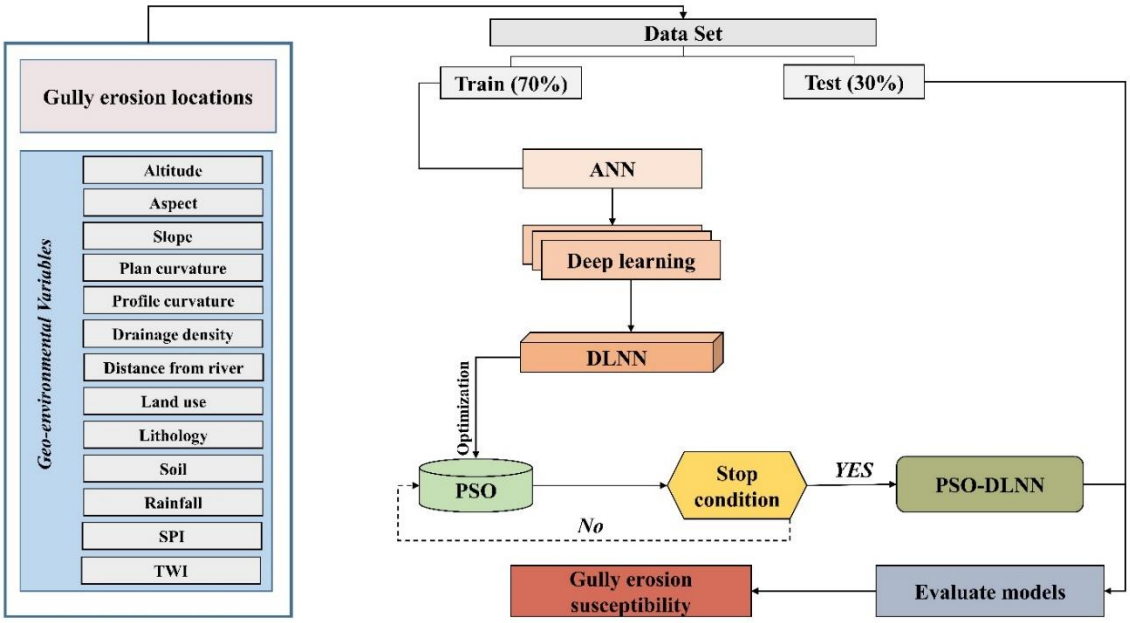


Figure 3. Methodological flowchart of PSO-DLNN in gully erosion susceptibility.

2.3. Dataset preparation for spatial modeling

In this study, the gully erosion inventory map (Figure 1.)was prepared based on field visits and information of Administration of Natural Resource of Hormozghan Province, which resulted in a total of 132 gully points. In order to determine the non-gully points, GIS software was used and 132 points were randomly selected. The DEM map was obtained with a pixel size of 12.5 meters from the ALOSPALSAR sensor. The topographical factors such as slope map, direction curve, Plan curvature, Profile curvature were prepared based on DEM in the GIS environment. The map of the distance from the river based on the Euclidean extension was obtained in GIS software. Drainage density map was prepared using line density extension. SAGAGIS software was used to map TWI and SPI. The soil type map of the region was obtained based on the map prepared by the Administration of Natural Resources of Hormozghan Province. The lithological map was prepared based on the geological map of 1: 100,000 of the country's mapping organization. Land use maps were prepared based on Landsat satellite images, OLI measurement, and using the maximum probability algorithm in the ENVI software environment. The precipitation map of the constituency was prepared from the statistics of 4 climatological around the constituency with a statistical period of 28 years (1989-2017) and based on the IDW interpolation method. The details about the data source in this research work were presented in Table 2.

Table 2. Details about the data sources of several factors used in this study

Parameters	Data source	Time (year)	Spatial resolution/scale

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

Altitude, slope, aspect, profile curvature, plan curvature, drainage density (DD), distance from river, stream power index (SPI), topographic wetness index (TWI)	ALOS PALSAR DEM (Alaska Satellite Facility)	2012	12.5 m
Rainfall	Iran Meteorological Organization (IMO) (http://www.weather.ir/)	1989 to 2017	
Lithology	Geological Survey of Iran (GSI) (http://www.gsi.ir/)	2019	1:1000000
Land use	Landsat OLI 8 satellite image (USGS)	2019	30 m
Soil texture	Soil and Water Research Institute (http://www.iran.swri.com)	2019	1:1000000

A total of 13 GECFs have been selected for GES mapping in this research work and these are altitude, aspect, slope, plan curvature, profile curvature, drainage density (DD), distance from river, land use, lithology, Soil, rainfall, stream power index (SPI) and topographic wetness index (TWI) (Figure 4a - m).

The altitude map of the present study area has been ranges from 2 m to 241 m (Figure 4a). Altitude is an important factor for occurrence of gullies due to influences on rainfall-runoff processes and largely employed on GES mapping [34]. Slope Aspect indirectly affects the occurrence of gully erosion as it affects the reception of sunlight, vegetation cover and humidity [35]. Here, slope aspect map have nine classes such as flat, N, NE, E, SE, S, SW, W and NW (Figure 4b). Slope angle influences on pattern of runoff and infiltration rate. Therefore, depending on slope, erosional rate also varies from place to place i.e. high slope areas have high erosion rate and vice-versa. The slope map has been shown in (Figure 4c) and value ranges from 0 to 362.74%. In a particular direction the rate of gradient changes is known as curvature. In which, plan and profile curvature generally represent topographic characteristics of an area. The value of plan curvature ranges from -30.27 to 24.08 (Figure 4d) and profile curvature from -29.63 to 30.93 (Figure 4e). DD directly impact on occurrences of gully erosion. Horton (1932) following equation was used to calculate DD. In this study, DD value ranges between 0 to 2.27 km/km^2 (Figure 4f).

$$DD = \frac{\sum_{i=1}^{n} S_i}{a} \quad (1)$$

Where, $\sum_{i=1}^{n} S_i$ is the length of drainage in km and 'a' indicates total area of drainage basin in km^2.

Distance from river also influences on occurrences of gully erosion as it greatly impact on wetting capacity of surface area and associated erosional activities. The value of distance from river ranges between 0 to 4680.17m (Figure 4g). The land use type of the area is very much responsible for the occurrence of gully erosion. Bare or less vegetated areas of land surface are highly prone to gully erosion. In this study, four types of land use were recognized i.e. agricultural land, rangeland, rock surface and salt land (Figure 4h). Lithological factor of an area is highly responsible for erosional

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

activities such as development of gully [36]. The present study area of Shirahan watershed consists of five types of lithological units (Figure 4i). Soil map of the study area have been shown in (Figure 4j) and it classified into two categories i.e. entisols/aridisols and badlands. Rainfall is the most important factor for the formation of gully and its development mainly in the arid and semi-arid areas. High intensity, with short rainfall duration, is the most devastating for gullies. Here, 28 years rainfalls data have been used to prepare rainfall map (Figure 4k) and it ranges from 125 to 175 mm. SPI indicates stream's erosional capacity [26]. SPI value was calculated by using following equations. And the value ranges from 0 to 2.625 in this research work (Figure 4l).

$$SPI = As * tan\beta \qquad (2)$$

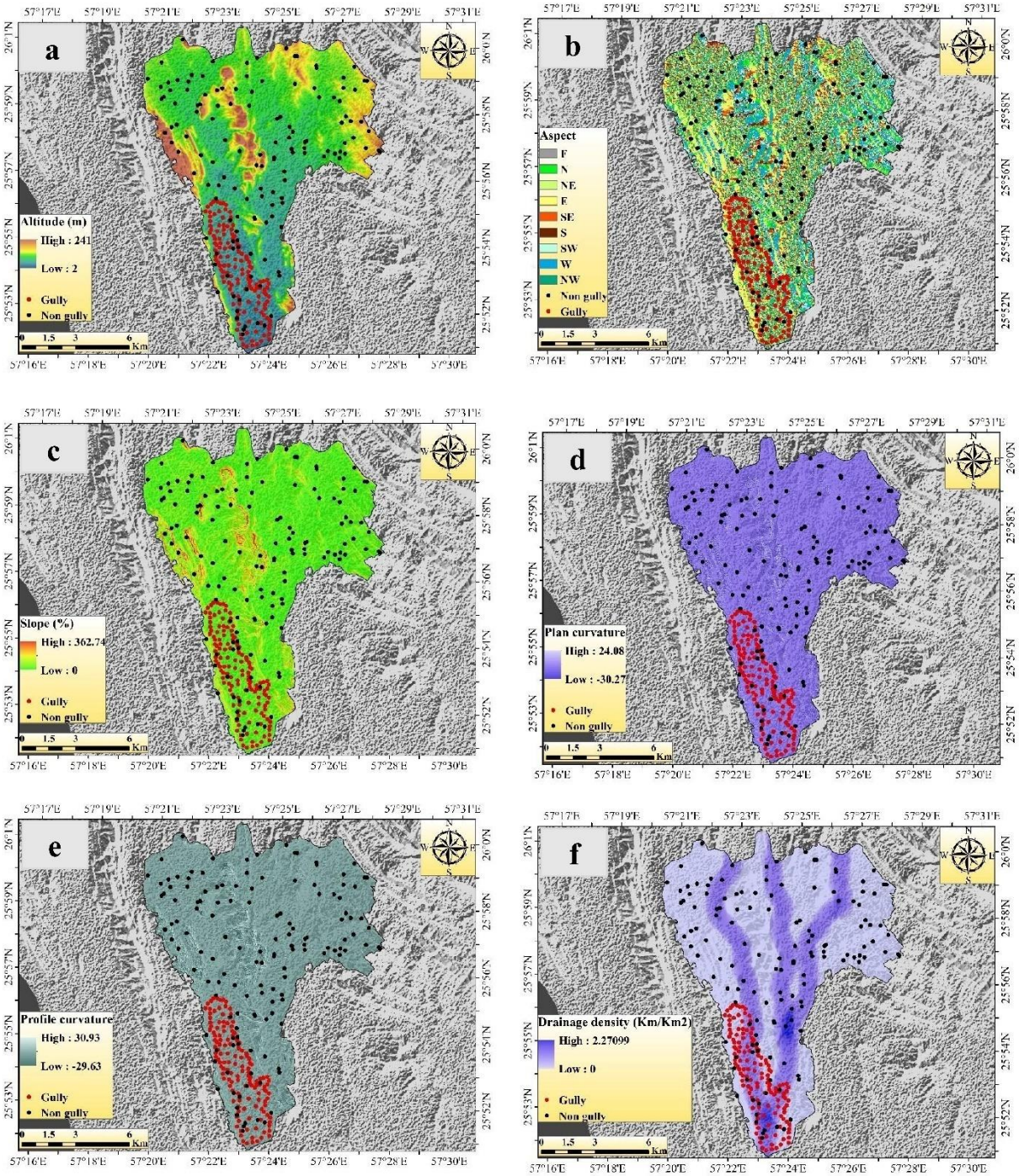
Where, As represents the upslope contributing area and β represents the slope angle.

TWI determines transport capacity along with flow velocity [7] and it is an essential factor to identifying gully erosion prone areas [37]. The following equation was used to calculate TWI value and it ranges from 0.14 to 18.86 (Figure 4m).

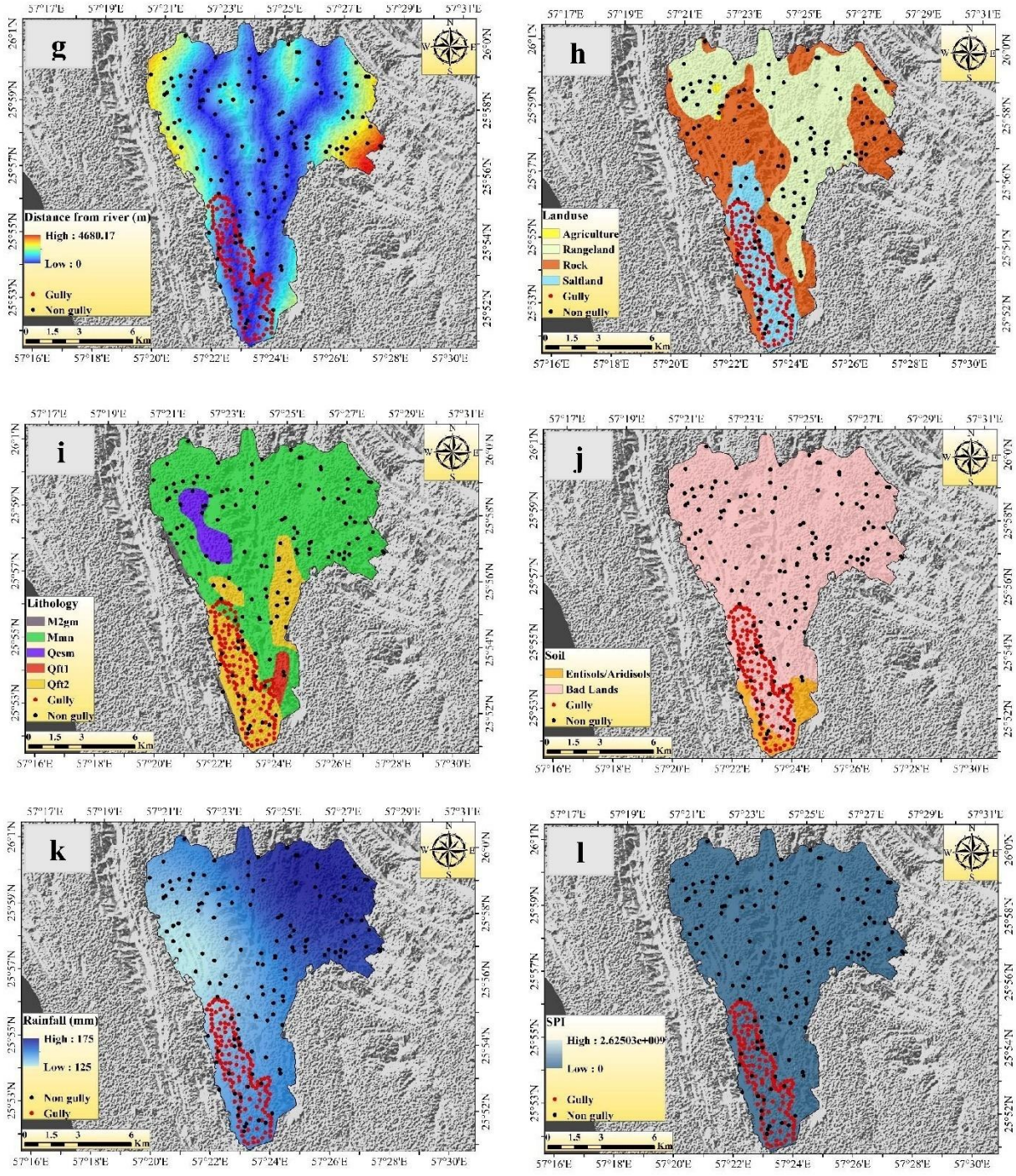
$$TWI = In\left(\frac{A_s}{tan\beta}\right) \qquad (3)$$

Where, A_s represents area of catchment in m^2 and β represents gradient of the slope in radians

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

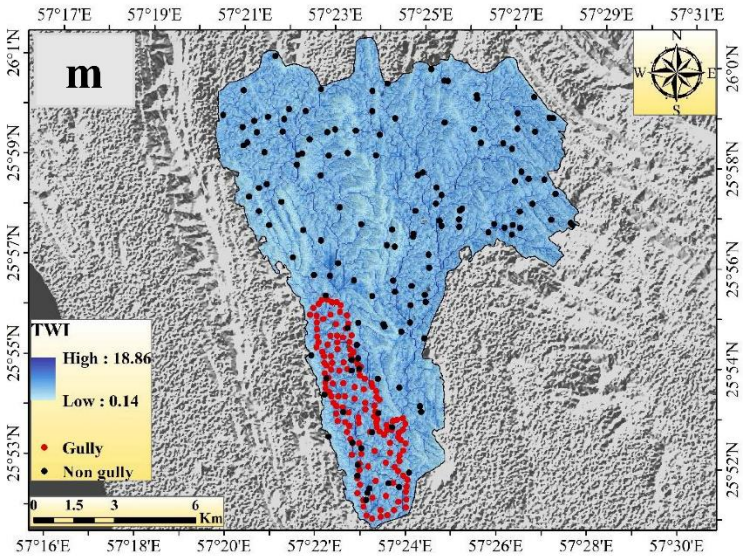


Figure 4. Gully erosion conditioning factors: a)altitude, b)slope, c)aspect, d)plan curvature, e)profile curvature, f) drainage density, g)distance from river, h) land use. i) soil, j) lithology, k) rainfall, , l) SPI, m) TWI.

2.4. Multi-collinearity analysis

Multi-collinearity analysis always gives the perfect outcome to evaluate the linear dependency of different geo-environmental factors in a ML model [15,38]. Basically, it is a statistical analysis and found among the two variables of high correlation in a multiple regression study. Thus, it is very much essential to analyze the multi-collinearity of a model to get better result through removing the high multi-collinearity factors and minimize the bias of the model [39]. Several researchers throughout the world have been used multi-collinearity analysis in different fields such as GSE mapping [22], flood [40], landslide susceptibility mapping [41] etc. Multi-collinearity can be analyzed through variance inflation factor (VIF) and tolerance (TOL) [42]. As a general rule, if TOL value is <0.10 or 0.20 and VIF value is >5 or 10, then the result indicate high multi-collinearity among the variables [43]. The following equations were used to calculate TOL and VIF in a dataset.

$$TOL = 1 - R_j^2 \quad (4)$$

$$VIF = \frac{1}{TOL} \quad (5)$$

Where, R_j^2 indicates regression value of j on other different variables in a dataset.

2.5. Machine learning method used in modeling the gully erosion

2.5.1. Support Vector Machine (SVM)

SVM is a very popular machine learning algorithm and was introduced by Vapnik and Chervonenkis in 1963. Several researchers throughout the world have used this machine learning classifier in the field of different natural hazards prediction such as in GES mapping [44], landslide prediction [45], flood susceptibility mapping [46], etc. SVM is basically implemented to solve the regression analysis and multifaceted classifier problems [47]. Vapnik (1995) stated that SVM is based on the principle of structural risk minimization and statistical learning, and it is a supervised machine learning model. SVM is very much affective to reduce the error of complexity of a linear computation and model over fitting [49]. Two types of statistical induced problems are engaged in SVM modeling. The first one is linear separating hyper plane by using statistical data and second one is converting non-linear data into linearly separable data [50]. Generally, the data processing in SVM of non-linear

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

relationship is done through kernel function [51]. In addition to this, discretely two classes have been generated in a SVM modeling by optimal hyper plane, in which one class indicate above the hyper plane assigned as 1 and another one, located below the hyper plane assigned as 0 i.e. in this case gully erosion and non gully erosion respectively [52]. SVM has been developed for regression estimation and particularly give attention on solution of inverse problems. The novelty of SVM model is that it has attempted to relocate the idea through kernel techniques for work out the inner products of an unsupervised learning. Beside this, it can also be applied for singular components where distribution of data is not well defined. Therefore, large class of function can be applied for nonlinearity mapping with high feature space by using this kernel trick. The hyper plane in a SVM can be calculated by using following equations.

$$Min \sum_{i=1}^{n} \varphi_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \varphi_i \varphi_j y_i y_j (x_i, x_j) \quad (6)$$

Subject to

$$Min \sum_{i=1}^{n} \varphi_i y_j = 0 and 0 \leq \alpha_i \leq D \quad (7)$$

Where, $x = x_i$, i = 1, 2,… n is input variables of vector, $y = y_i$, $j = 1, 2, \ldots n$ is output variables of vector and φ_i represent as Lagrange multipliers.

Finally, decision function of SVM can be classified as

$$f(x) = sgn\left(\sum_{i=j}^{n} y_i \varphi_i K(x_i, x_j) + a \right) \quad (8)$$

Where, a represent as bias which indicate linear distance of hyper plane from the origin, $K(x_i, x_j)$ represent kernel functions i.e. polynomial (POL) and radial basis function (RBF) and, these can be expressed as follow [53].

$$K_{POL}(x_i, x_j) = ((x * y) + 1)^d \quad (9)$$

$$K_{RBF}(x_i, x_j) = e^{-y||x - x_i||^2} \quad (10)$$

2.5.2. Artificial Neural Network (ANN)

ANN is a popular machine learning algorithm that simulates neural networks of a human brain and can works in a specific way [54,55]. Basically, it is used to analyze and predict non-linear statistical dataset by using different algorithms [56]. ANN has been widely used in the case of pattern recognition and classification studies [57]. Therefore, classifications of a landscape in different ordinal areas of GES zone are treated as a classification problem. Different types of algorithms have been used in ANN modeling, among them Multilayer Perceptron (MLP) is most popular, based on its outcome result and frequency uses by the researchers [58]. In order to run and analyze ANN algorithms, some basic knowledge was needed to understand the structure of input data and the relationship between the variables [59]. ANN model with MLP algorithm consist of three layers such as input layer, hidden layer and output layer. Schematic diagram of feed-forward artificial neural network model have been shown in Figure 5. In this research work, the input layers are training points for the erosion of the gully and the various GECFs which have finally been connected to the output layer. Input nodes help to predict and analyze the model structure through input and hidden layers and, ultimately, to evaluate the output layer result [60,61]. This output layer gives us the GES map. The output layer consists of Boolean value of 0 and 1, in which 0 represents non-gully erosion and 1 represents gully erosion. Basically, feed-forward of ANN algorithm model deals with three

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

stages such as feed-forward of input data, calculation and back propagation of related errors and their adjustments [59].

The novelty of ANN model is that it has the capability to learn the model through non-linear and complex relationship. Thus, model's uniqueness is evaluated on the basis of observation the coherence of network dynamics than the other models. It has also the ability of model generalization and can predict unseen data within the model through understanding the hidden relationship.

The algorithms of ANN were elaborate in the following equations by Hagan et al. 1996.

$$net_j^l(t) = \sum_{i=o}^{p} (y_i^{i-1}(t) w_{ji}^l(t)) \quad (11)$$

The net input of jth neuron of layer l and I iteration

$$y_j^l(t) = f(net_j^{(l)}(t) \quad (12)$$

$$f(net) = \frac{1}{1 + e^{(-net)}} \quad (13)$$

$$e_j(t) = c_j(t) - a_j(t) \quad (14)$$

$$\delta_j^l(t) = e_j^l(t)\, a_j(t)\left[1 - a_j x(t)\right] \quad (15)$$

δ factor for neuron jth in the output layer ith

$$\delta_j^l(t) = y_j^l(t)\left[1 - y_j(t)\right] \sum \delta_j^l(t) w_{kj}^{(l+1)}(t) \quad (16)$$

δ factor for neuron jth in the hidden layer ith

$$w_{ji}^l(t+1) = w_{ji}^l(t) + \alpha\left[w_{ji}^l(t) - w_{ji}^l(t-1)\right] + n\delta_j^{(l)}(t) y_j^{(l-1)}(t) \quad (17)$$

Where, α is the momentum rate and n is the learning rate.

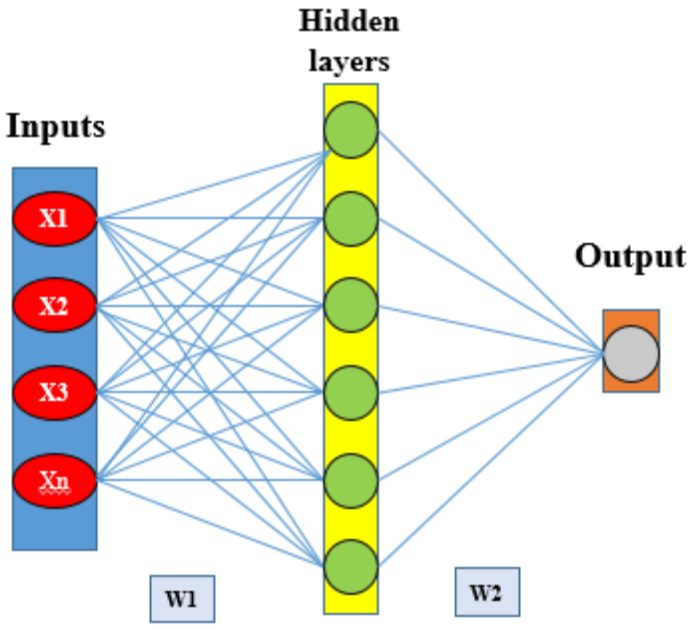


Figure 5. Schematic of feed-forward artificial neural network.

2.5.3. Deep Learning Neural Network (DLNN)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

The DLNN is a well accepted machine learning model among the research groups of people throughout the world. This ML model has a prominent advantage in constructing a high level feature in an appropriate way by using the raw dataset [28]. Basically, DLNN consist of three layers i.e. an input layer, several hidden layers and result in an output layer [62]. The speculative configuration of DLNN model used for GES mapping in this research work has been shown in Figure 6. The general structure of the DLNN model is to run in such a way that the input layer receives signals which are basically different GECFs, this information is a process and analysis in several hidden layers, and finally the output model's result is presented in the last layer, i.e. the output layer. The output layer has two possible labels i.e. first one is negative labels (non-gully erosion) and second one is positive labels (gully erosion). These classification results have been obtained from last hidden layer and shown in output layer [63].

DLNN have some specific compensation over the traditional ML algorithm and thus, in the field of prediction analysis, use of DLNN model has been given much more emphasis. Therefore, DLNN has give some novelty performance over the other ML models and these are maximum utilization of unstructured data through relevant insights understanding of training dataset, robust enough to recognize the novel data and can developed additional learning model through adding more layers in the neural network system.

According to Kim (2017) the following mathematical equation were used in a DLNN machine learning model.

$$h(x) = \begin{cases} x & if \quad x > 0 \\ 0 & if \quad x \leq 0 \end{cases} = \max(0, x) \tag{18}$$

Where, x represents input signal and h indicates activation function.

Based on the ReLU activation function, it can be as expressed as follows:

$$h'(x) = \begin{cases} 1 & if \quad x > 0 \\ 0 & if \quad x \leq 0 \end{cases} \tag{19}$$

The cost function is the differences between experiential and predicted class outputs. Loss function (L) of a cross entropy are used to pattern recognition and expressed as follows:

$$L = -\frac{1}{N_D} \sum_{n=1}^{N_D} T \, 1n(Y) + (1 - T) 1n(1 - Y) \tag{20}$$

Where, N_D represents number of training dataset, T indicates observed class outputs and Y indicates predicted class outputs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

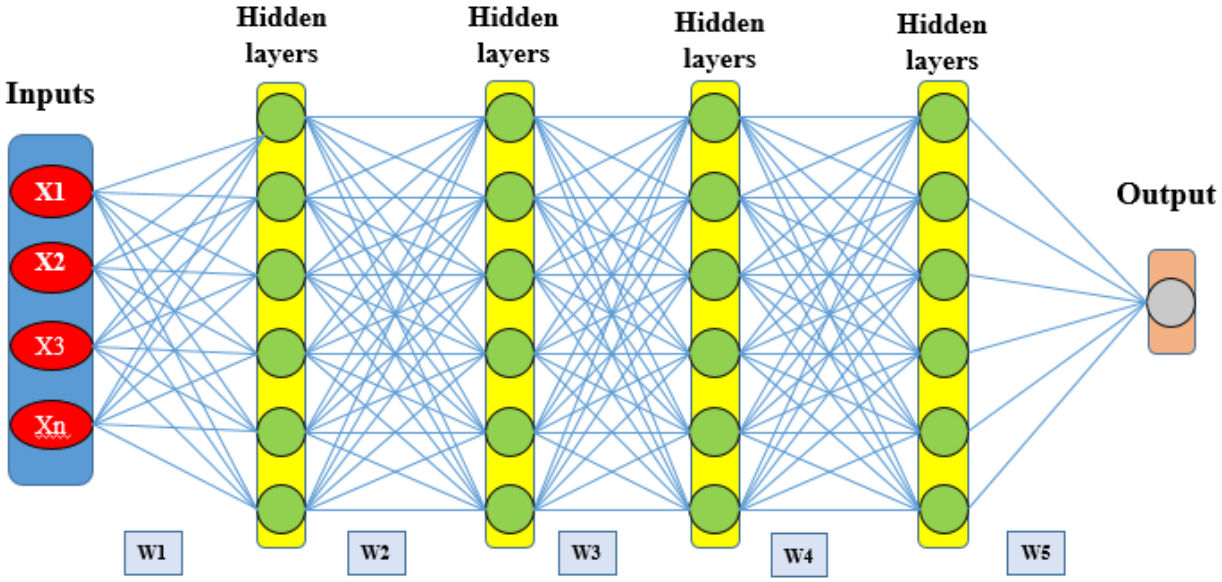


Figure 6. Schematic of deep learning neural network.

2.5.4. Particle swarm optimization (PSO)

The algorithm of PSO is a meta-heuristic and originally developed by an American social psychologist Kennedy [64]. In our research work, we are faced with a number of non-linear problems and, in order to find the correct solution, the PSO method has been developed and widely used. The PSO algorithm was stimulated to locate the best possible food route for bird and fish intelligence. Here, birds are the particles and try to find a solution to the problem. Particles are always tried to find out best possible solution to a problem through n-dimensional space, in which n represents each and every problem's different parameters [65]. Optimization of position and velocity is the basic principle of each particle.

Therefore, let us, $x_i^t = (x_{i1}^t, x_{i2}^t, \dots, x_{in}^t) and \;\; v_i^t = (v_{i1}^t, v_{i2}^t, \dots, v_{in}^t)$ are the position and velocity of changing position designed for i^{th} particle in t^{th} iteration accordingly. The following equations are used for the i^{th} particle's position and velocity in $(t+1)^{th}$ iteration.

$$v_i^{t+1} = \omega . v_i^t + c_1 . r_1 . (p_i^t - x_i^t) + c_2 . r_2 . (g_i^t - x_i^t) \;\; with - v_{max} \le v_i^{t+1} \le v_{max} \qquad (21)$$

$$x_i^{t+1} = (x_i^t + v_i^{t+1}) \qquad (22)$$

Where, x_i^t represents previous i^{th} position; p_i^t represents most excellent found position; g_i^t represents particle's best position; r_1 and r_2 represents random numbers within 0 and 1; ω is weights of inertia; c_1 is coefficient and c_2 represents social coefficient. There are numerous methods to particle's weight assignment [66,67], among them standard 2011 PSO had been widely used and calculate by following equation.

$$\omega = \frac{1}{2 \;\; In \;\; 2} and \;\; c_1 = c_2 = 0.5 + In \;\; 2 \qquad (23)$$

Therefore, it is supposed to be believed that the concentration of all particle swarms in a point and space has been achieved when problem to be solved. The intelligence based PSO algorithm has been widely used in high efficiency swarm paralleling and optimization property. Beside this, by using multi-objective fitness function, PSO determines the quality of several features in a dataset. Finally, ensemble of particle swarm optimization (PSO) and deep learning neural network (DLNN) structure have been shown in Figure 7. Therefore, this ensemble method is the novel approach in this research study for GES mapping with utmost accuracy.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

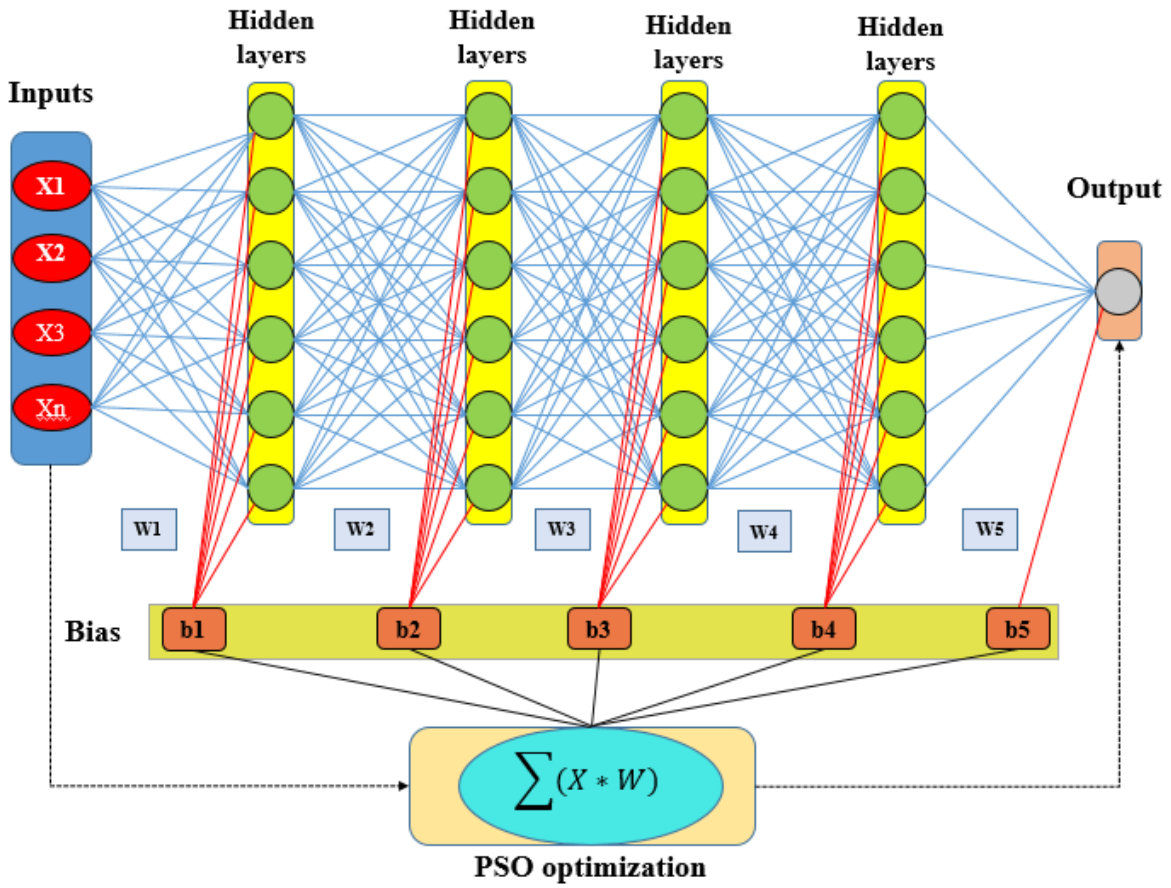


Figure 7. Schematic of ensemble particle swarm optimization and deep learning neural network.

2.6. Methods of Validation and accuracy assessment

GES maps were prepared based on prediction performance of training and validation dataset by using different machine learning models. Therefore, it is very much necessary to evaluate the model performance to get validity of results. In the present research work statistical indices along with area under the receiver operating characteristic (AUROC) curve were used to predict accuracy of ML and ensemble models.

2.6.1. Statistical indices

In this study sensitivity (SST), specificity (SPF), positive predictive values (PPV) and negative predictive values (NPV) were used to evaluate the predictive results. Four types of possible consequences were used to analyze these statistical indices and these are true positive (TP), true negative (TN), false positive (FP) and false negative (FN). TP, when gully pixels are correctly classified as gully and FP when gully pixels are incorrectly classified as gully. On the other hand, if gully pixels are correctly or incorrectly classified as non-gully then they are TN and FN respectively [38]. If higher values are found among these statistical indices then model gives better results and vice-versa [23]. The following equations were used to calculate the value of these four statistical indices.

$$PPV = \frac{TP}{FP + TP} \quad (24)$$

$$NPV = \frac{TP}{TP + FN} \quad (25)$$

$$SST = \frac{TP}{TP + FN} \quad (26)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

$$SPF = \frac{TN}{FP + TN} \quad (27)$$

2.6.2. ROC curve

ROC curve is one of the most widely used tools for analyzing the performance validation of the ML model. ROC curve have two dimensions i.e. events and non-events phenomenon [68]. Basically, this curve plotted on 'X' and 'Y' co-ordinates known as sensitivity and 1-specificity respectively and represents true positive and false positive. The optimum value in both cases i.e. in sensitivity (detected gullies) and specificity (detected non-gullies) are 1 [3]. The value of ROC-AUC ranges from 0.5 to 1; in which 0.5 indicates poor performance and 1 indicates very good performance. Beside this, in a proper way it can be classified into five classes i.e., poor (0.5-0.6), moderate (0.6-0.7), good (0.7-0.8), very good (0.8-0.9) and excellent (0.9-1) [69]. The following equation has been used to compute the ROC-AUC.

$$S_{AUC} = \sum_{k=1}^{n} (X_{k+1} - X_k)\left(S_k + 1 - S_{k+1} - \frac{S_k}{2}\right) \quad (28)$$

Where, S_{AUC} indicates area under curve, X_k indicates 1-specificity and S_k indicates the sensitivity of the receiver operating characteristic (ROC) curve.

Besides the above validation methods, here we also used Likelihood Ratio (LR), F-measure and Maximum Probability of Correct Decision (MPCD) analysis for better understand the accuracy assessment of the result. In this study, LR model is the relationship between the distribution of gully head-cut points and related GECFs. Therefore, LR model emphasized the ratio of the probability of events and non-events phenomena of the gully occurrences. In this method, if the ratio is higher than 1, there is a high relationship among the gully erosion and associated factors. On the other side, if ratio is less than 1, low relationship is found between the gully erosion and associated factors. Thus, the linear relationship of LR can be expressed as follows:

$$GESI = \sum Fr \quad (29)$$

Where, GESI represents gully erosion susceptibility index and Fr represent the rating of several factors range.

F-measure is a popular validation method in the field of classification and information retrieval communities. Basically, F-measure balances between precision and recall. The following equation was used to calculate the F-measure in this study.

$$F - measure = 2 \times TP/2 \times TP + FP + FN \quad (30)$$

In a classification performance, MPCD is a probabilistic based measure. It is a sensitive method for recognition of class than just to estimate the proportion of guesses. The following equation was used to calculate the MPCD.

$$MPCD = (1 - \alpha)(1 - \beta) \quad (31)$$

Where, α is $\frac{FP}{FP+TN}$ and β is $\frac{FN}{FN+TP}$

3. Results

3.1. Multi-collinearity analysis:

To be maintained in view of the VIF and TOL limits; 13 gully erosion conditioning parameters have been selected for gully erosion modeling. The co-linear factors (i.e., distance from road, geomorphology and bulk density) were excluded from this analysis. These three factors; distance from road (TOL 0.028 and VIF 35.65), geomorphology (TOL 0.032 and VIF 31.63) and bulk density (TOL 0.022 and VIF 45.23) are associated with co-linearity problems. The ranges of VIF of the selected

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

parameters are 1.06 to 3.04. In the case of TOL, the range varies among the selected conditioning factors are 0.33 to 0.94 (Table 3). Among the 13 GECFs, altitude is the highest VIF value of 3.04 and lowest TOL value of 0.33. On the other side, aspect factors consist of highest TOL value of 0.94 and lowest VIF value of 1.06. Therefore, this indicates no multi-collinearity has been found between the thirteen conditioning factors of gully erosion used in this study.

Table3. Multi-collinearity analysis to determine the linearity of the independent variables

Variables	VIF	Tolerance
Altitude	3.04	0.33
Slope	1.34	0.75
Aspect	1.06	0.94
Plan curvature	1.83	0.55
Profile curvature	1.82	0.55
Distance from river	2.93	0.34
Drainage density	2.07	0.48
Rainfall	1.41	0.71
Land use	1.81	0.55
Lithology	2.07	0.48
Soil	1.11	0.90
SPI	1.58	0.63
TWI	1.94	0.52

3.2. Gully erosion susceptibility modeling:

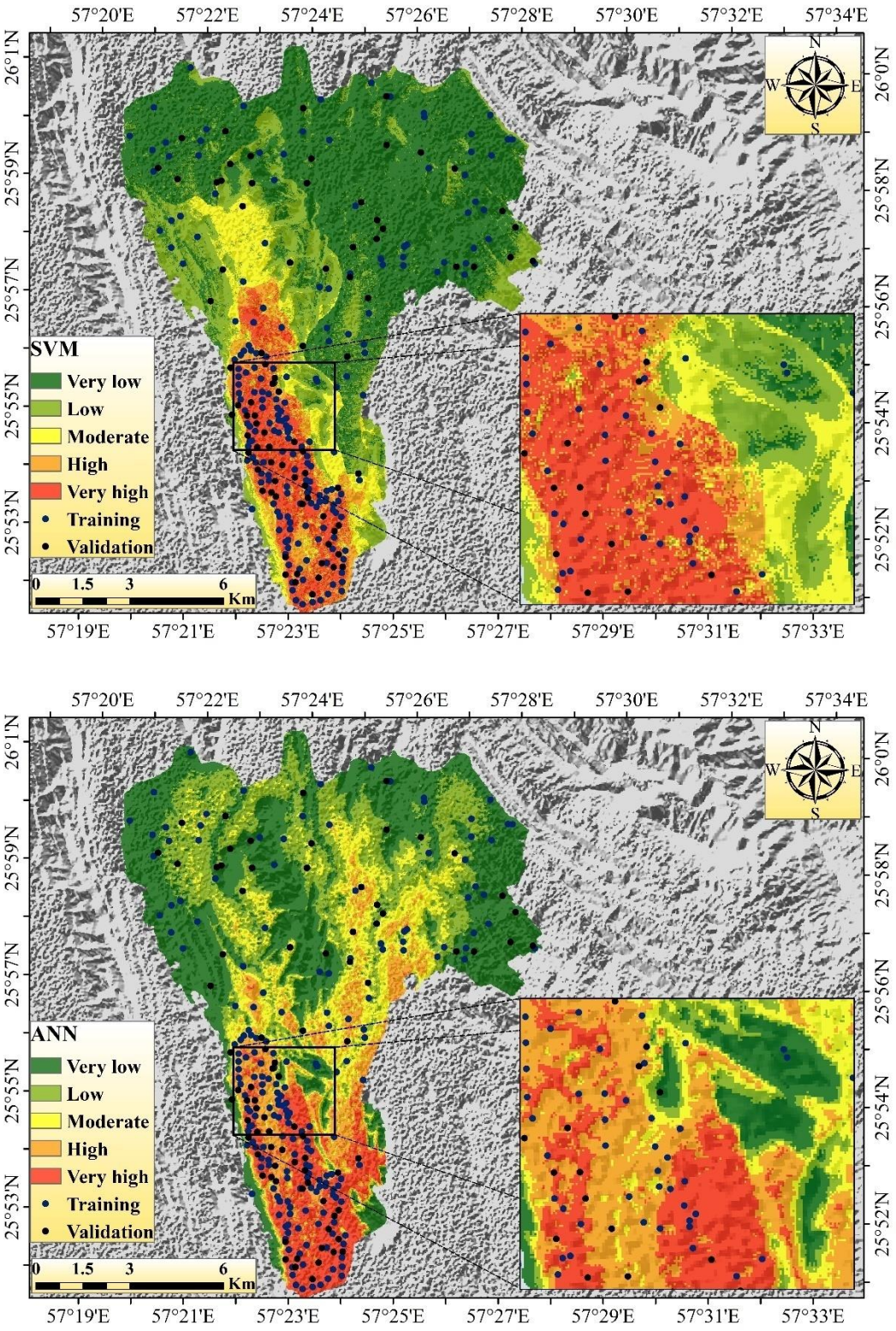
In SVM model, the very low GES areas are mainly concentrated in the eastern and northern portion of this region. The low GES areas are mainly found in the middle and western part of this region. The moderate susceptibility areas are mainly concentrated in the middle and southern part of this region (Figure 8a). The very high and high GES areas are mainly limited in the southern portion of the watershed. The areal coverage of very low, low, moderate, high and very high gully erosion susceptible areas in SVM model are 65.86 (52.08%), 28.92 (22.87%), 10.7 (8.46%), 8.0 (6.33%) and 12.97 Km^2 (10.26%) respectively (Table4).

In ANN, the areal coverage for very low, low, moderate, high and very high gully erosion susceptible areas are 55.76 (44.10%), 26.85 (21.23%), 16.85 (13.33%), 13.48 (10.66%) and 13.51 Km^2 (10.68%) respectively. According to the GES map of ANN model, the maximum portion of the area occupied by very low (44.10%) to low (21.23%), susceptibility classes, while very high (10.68%), high (10.66 %) and moderate (13.33%) susceptibility classes covered rest of the studied region respectively. In this model, the very high, high and moderate susceptible areas are mainly concentrated in the southern, middle and eastern portion of the watershed (Figure 8b). Rest of the portion of this watershed is associated with very low to low GES zones.

In the case of DLNN, the areal coverage for very low, low, moderate, high and very high gully erosion susceptible zones are 96.34 (76.19%), 5.85 (4.63%), 2.73 (2.16%), 3.17 (2.51%) and 18.36 km^2 (14.52%) respectively. According to the GES map of DLNN model, the maximum portion of the area occupied by very low (76.19%) to low (28.73%) susceptibility classes, while very high (14.52%), high (2.51%) and moderate (2.16%) susceptibility classes occupied rest of the studied region respectively. In this model, the very high to moderate susceptible areas are mainly concentrated in southern and middle portion of the watershed and rest of the portion are associated with very low to low susceptible zones (Figure 8c).

In PSO-DLNN model, the areal coverage of low, low, moderate, high and very high gully erosion susceptible zones are 94.58 (74.80%), 8.15 (6.45%), 4.03 (3.19%), 6.31 (4.99%), 13.38 (10.58%) km^2 respectively. According to the GES map of PSO-DLNN model, the major portion of the area occupied by very low (74.80%) to low (6.45%) susceptibility classes, while very high (10.58%), high (4.99%) and moderate (3.19%) susceptibility classes covered rest of the studied region respectively. Very high, high and moderate gully erosion susceptible zones are mainly occupied in southern portion of the watershed And rest of the portion are associated with very low to low susceptible zones (Figure 8d).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

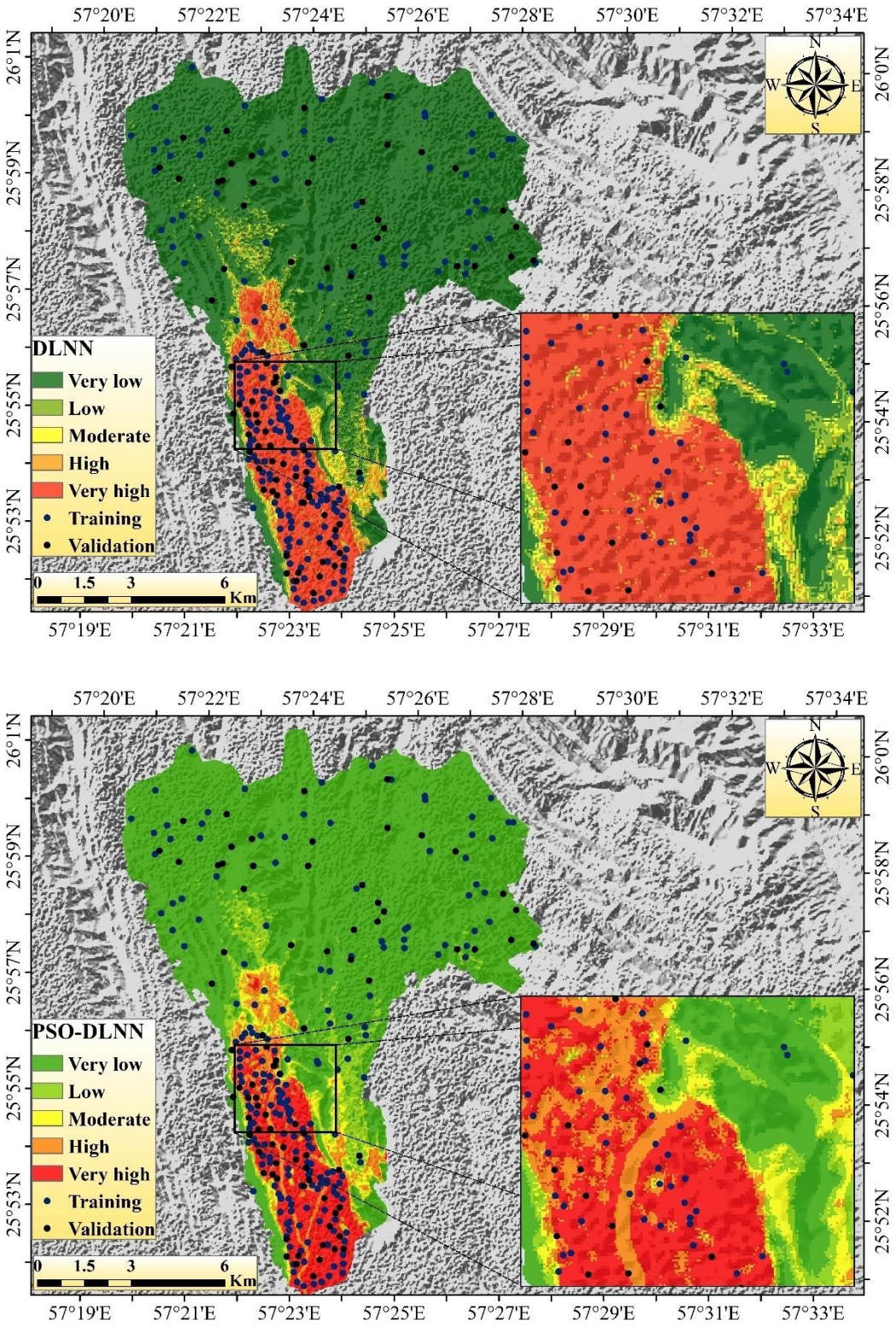


Figure 8. Head-cut gully erosion map using the five models: SVM (a), ANN (b), DLNN (c) and PSO-DLNN (d).

Table4. Gully erosion susceptibility classes' area.

Models	Area	Susceptibility Class				
		Very low	Low	Moderate	High	Very high
SVM	Km^2	65.86	28.92	10.7	8	12.97
	%	52.08	22.87	8.46	6.33	10.26
ANN	Km^2	55.76	26.85	16.85	13.48	13.51

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

DLNN	%	44.10	21.23	13.33	10.66	10.68
	Km^2	96.34	5.85	2.73	3.17	18.36
PSO-DLNN	%	76.19	4.63	2.16	2.51	14.52
	Km^2	94.58	8.15	4.03	6.31	13.38
	%	74.80	6.45	3.19	4.99	10.58

3.3. Validation of the models

The PSO-DLNN is the most optimal model in this analysis which is associated with maximum accuracy. The AUC values from ROC with considering testing datasets of PSO-DLNN is 0.89 and which is associated with superb accuracy. Rest of the models also associated with optimal accuracy and near about PSO-DLNN model; the AUC values from ROC of DLNN, SVM and ANN for testing datasets are 0.87, 0.85 and 0.84 respectively (Figure 9). Apart from this various statistical indices were considered for estimating the optimal capacity of all the models for GES modeling. The values of sensitivity in PSO-DLNN, DLNN, SVM and ANN for training datasets are 0.98, 0.95, 0.99 and 0.99 respectively. And the same values for the validation datasets in PSO-DLNN, DLNN, SVM and ANN are 0.95, 0.90, 0.82 and 0.95 respectively. The values of specificity for the training datasets in PSO-DLNN, DLNN, SVM and ANN are 0.85, 0.82, 0.86 and 0.87 are respectively. In the case of validation datasets, the values of specificity in PSO-DLNN, DLNN, SVM and ANN are 0.74, 0.74, 0.69 and 0.67 respectively. The values of PPV in the case of training datasets in PSO-DLNN, DLNN, SVM and ANN are 0.87, 0.85, 0.88 and 0.89 respectively. When we consider the validation datasets, the values of PPV in PSO-DLNN, DLNN, SVM and ANN are 0.77, 0.77, 0.71 and 0.73 (Table5). In PSO-DLNN, DLNN, SVM and ANN models, the values of NPV for the training datasets are 0.97, 0.94, 0.99 and 0.99 respectively. In the case of validation datasets, the values of NPV of PSO-DLNN, DLNN, SVM and ANN are 0.94, 0.89, 0.81 and 0.93 respectively. F-measure in validation datasets for PSO-DLNN, DLNN, SVM and ANN models are 0.66, 0.635, 0.63 and 0.64 respectively.

The details about the DLNN and its associated parameters are shown in Table 6. The details about the combination of PSO and DLNN and its associated parameters are shown in Table7. The values of population, iteration, phi, phi1, phi2, W, C1, C2 and best cost are 50, 500, 4.1, 2.05, 2.05, 0.73, 1.49, 1.49 and 0.26. The objective cost function of PSO-DLNN model is shown in Figure 10.

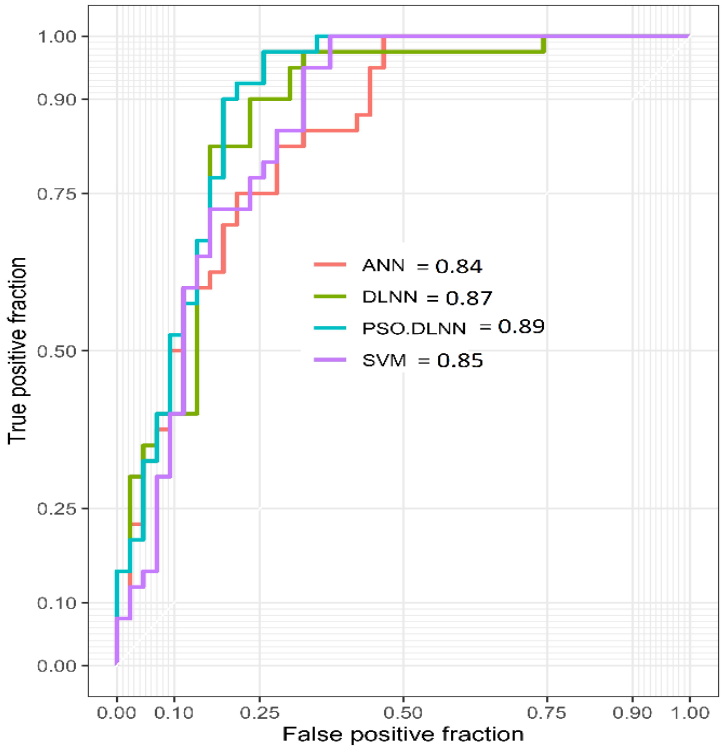


Figure 9. The ROC curve analysis for four head-cut gully erosion models using the testing dataset.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

Table5. Predictive capability of HCGES models using train and test dataset.

Models	Stage	Parameters					
		Sensitivity	Specificity	PPV	NPV	AUC	F-measure
SVM	Train	0.99	0.87	0.89	0.99	0.94	0.84
	Validation	0.95	0.67	0.73	0.93	0.85	0.63
ANN	Train	0.99	0.86	0.88	0.99	0.94	0.83
	Validation	0.82	0.69	0.71	0.81	0.84	0.64
DLNN	Train	0.95	0.82	0.85	0.94	0.91	0.82
	Validation	0.90	0.74	0.77	0.89	0.87	0.65
PSO-DLNN	Train	0.98	0.85	0.87	0.97	0.93	0.84
	Validation	0.95	0.74	0.77	0.94	0.89	0.66

Table 6. Result of optimal parameters in DLNN model

Parameters	**Optimum**
Input number of units	13
Output	2
Activation Function	ReLU
Activation	'softmax'
Function	Sigmoid
reluLeak	0.01
eta	0.8
Hidden layer unit	3,3
Iteration	200

Table7. Parameters used in PSO algorithms in combined DLNN.

Parameters	**Number**
Population	50
Iteration	500
phi	4.1
phi1	2.05
Phi2	2.05
W	0.73
C1	1.49
C2	1.49
Best Cost	0.26

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

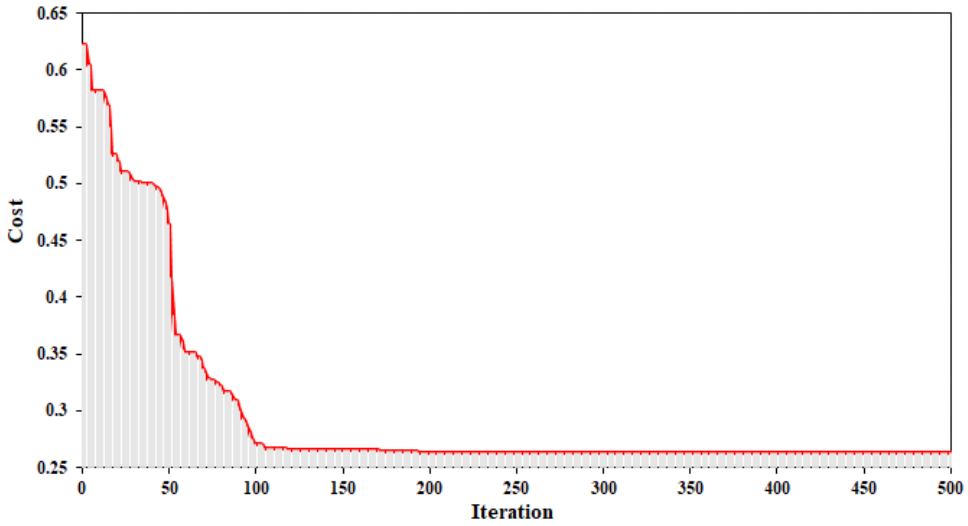


Figure 10. Convergence graphs of the objective cost function (MSE) in PSO-DLNN model.

3.4. Variable importance

The conditioning factor for GES modeling for this region has been selected considering the different kinds of literature. The most important parameters for the creation and development of gullies in this region are land use, altitude, lithology, rainfall and distance from road etc. The relative importance of land use, altitude, lithology, rainfall and distance from road for GES models are 100, 97.94, 59.51, 46.94 and 29.48 respectively. Rest of the factors (i.e., profile curvature, TWI, plan curvature, slope, soil, drainage density, SPI and aspect) are associated with moderate to very low relative importance for GES. The relative importance of profile curvature, TWI, plan curvature, slope, soil, drainage density, SPI and aspect for gulling are 16.22, 14.37, 11.31, 7.89, 7.1, 6.91, 5.12 and 0 (Table8). Here, apart from the topographical and geo-hydrological characteristics, the impact of anthropogenic activities accelerates the rate of land degradation in the form of gullies.

Table8. Variable importance analysis based on PSO-DLNN model.

Variables	Importance
Altitude	97.94
Aspect	0
Slope	7.89
Plane curvature	11.31
Profile curvature	16.22
Drainage density	6.91
Distance from river	29.48
Land use	100
Lithology	59.51
Soil	7.1
Rainfall	46.94
SPI	5.12
TWI	`14.37

4. Discussion

Land degradation in various forms of soil erosion can accelerate the extensive damages and it has an adverse impact on society and livelihoods of the people throughout the world [70]. There are various forms of erosion i.e. sheet erosion, formation of rills, formation and development of gullies and ravines etc. [71]. Of these, the formation and development of gullies and their associated erosion is the most destructive element of land degradation in worldwide [2]. Although it is a natural process of erosion, this process can greatly accelerate anthropogenic activities and have a serious impact on

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

the ecosystem [72]. From this type of erosion, the agricultural activities have not only effected but have also been associated with damages of the infrastructure created by the human. On the one hand, it is responsible for removing the top soil, but on the other hand, it is responsible for the creation and accumulation of sediment in the lower catchment area [73]. The life span of the reservoir will cause serious damage to the sediment deposition resulting from this type of erosion [74,75].

The Shirahan watershed in Iran has recently faced severe gullies erosion, which is responsible for large-scale erosion and the main barrier to sustainable land management practices. So, identifying vulnerable regions with the most optimal model is therefore very optimistic that appropriate soil and water conservation measures will be put in place. For this purpose, we considered the SVM, ANN, DLNN and PSO-DNN for estimating the GES of this region with maximum possible accuracy and to suggest the most suitable model. The erosion of the gully is controlled by various causal factors and can be considered important for gulling by determining the importance of these factors. Apart from the topographic and hydro-geomorphic attributes, land use is the most important variable for gully erosion which indicates the larger anthropogenic impact on the development of gullies. Other factors, like altitude, lithology, rainfall and the distance from the river, are very optimistic too about gully erosion and promote gulling. The transformation of land use is a crucial element and is responsible for large-scale erosion [76]. Alteration in land use influences landscape ecology functions on behalf, on far-reaching implications for natural ecosystems and land reclamation [77]. The character and volume of the surface runoff may change directly with the changing pattern of land use in the region. From this perspective, the nature of erosion in the form of gullies can have a significant impact on the impact of rainfall and its associated runoff characteristics in a changing environment. This type of outcome is similar to some of the findings of the study of a number of researchers in a diversified discipline. This finding has been highlighted by many other contributions in which morphological and geological properties are assigned as the determinants of the highest possible location of GES [25,78]. Other research outcomes suggest that environmental and hydrological parameters are very significant and responsible for gulling.

All predicted models are associated with higher accuracy, but the PSO-DLNN is the most optimal and the AUC of this model is 0.89. The efficiency of all predicted models is excellent and the AUC values for DLNN, SVM and ANN are 0.87, 0.85 and 0.84, respectively. Apart from this, considering various statistical indices, the PSO-DLNN is the best model among the other models used in this study. According to the PSO-DLNN model, 18.76% of the total area is associated with a moderate to very high susceptible area of gully erosion. The southern portion of this watershed mainly associated with higher gully occurrences. The complex geo-hydrological characteristic of this region is favorable for large scale erosion in the form of gullies.

Deep learning framework is associated with higher accuracy than conventional ANN and SVM ML methods. This model can handle a larger number of samples and even a large amount of big data and estimate the results with optimum accuracy. The traditional ML algorithm is not capable of handling this large number of samples, and the outcome from this perspective is less optimal compared to the deep learning framework. Significant progress in DLN-dependent deep learning (DL) systems has significantly increased the consistency of machine learning for various purposes. While the standardized features of multi-layer NNs are well established, the main advantage of DL is its structured method of self-governing training of DNN-layer organizations. The benefits of structured data and expertise descriptions were recognized prior to the recent increase in DNN's interest. This definition is widespread in the physical sciences where the proposed method is popular in both specific theoretical structures and complicated system implementations in practice.

First, PSO produces an arbitrary remedy and then discovers accurate solutions with an incremental optimum fitness attribute. This type of methodology has already been used primarily for Back Propagation (BP) genetic algorithms, due to the efficiency of simple installation, fast response and accuracy of predictions. It also demonstrated dominance in the resolution of complex applications and was initially implemented in the context of DL. The best function of the PSO algorithm is to combine the various particles that are interlinked to each other in order to achieve an optimum position. The same technique alerts the position, the velocity and the highest accuracy of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

each particle, which are dictated by the basic concepts used to enhance the problem. Particularly in comparison to other optimization algorithms, the advantage of the PSO algorithm is that the PSO technique usually involves a quick and important search procedure, is easy to perform, and can find the globally optimal path that is closest to the concrete ideas.

5. Conclusion

Choosing the most efficient machine learning algorithm is necessary to decrease the inconsistencies connected with predicting the susceptibility of gully erosions. The main objectives in most of the susceptibility modeling are to identify the optimal model according to its predictive capabilities. The identification of key parameters for the formation and development of the gully is necessary in order to estimate the susceptibility mapping of the spatial distribution of the gully erosion. Therefore, to control damages in the future, it is important to make an appropriate selection of the model in order to manage areas that are prone to degrading the gully. The primary objective of this research is to estimate the optimal model with maximum predictive capability. For this reason, various ML algorithms like ANN, SVM, DLNN and PSO were considered for estimating the GES zone with optimal capacity. The PSO-DLNN is the best-fitted model and is associated with higher AUC values (0.89). Here, all the datasets were randomly partitioned with 70/30 ratio as training and validation datasets. Topographical, hydrological and environmental factors were most dominated and influential factors in susceptibility modeling. The role of land use is higher in susceptibility modeling than any other component. Most of the region of this watershed is associated with a very low to low susceptible zone while 15.57% area is associated with a very high susceptible zone. This study region must take the appropriate planning initiative to reduce the level of vulnerability and to protect this type of precious resource. The role of the future researcher is to develop the PSO-DLNN algorithm with incorporating some new components or to develop the same algorithm with slight modifications. This would be a great contribution to the research community and as well as to society. Apart from this, the selection of inappropriate parameters can reduce the efficiency of the predicted models. So, the selection of more appropriate variables for susceptibility modeling is one of the important tasks of the researchers.

Declaration of competing interest

There is no conflict of interest regarding this work.

Acknowledgments

References

1. Keesstra, S.; Mol, G.; De Leeuw, J.; Okx, J.; De Cleen, M.; Visser, S.; others Soil-related sustainable development goals: Four concepts to make land degradation neutrality and restoration work. *Land* **2018**, *7*, 133.

2. Lal, R. Soil degradation by erosion. *L. Degrad. Dev.* **2001**, *12*, 519–539.

3. Conoscenti, C.; Angileri, S.; Cappadonia, C.; Rotigliano, E.; Agnesi, V.; Märker, M. Gully erosion susceptibility assessment by means of GIS-based logistic regression: a case of Sicily (Italy).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

Geomorphology **2014**, *204*, 399–411, doi:10.1016/j.geomorph.2013.08.021.

4. Vanwalleghem, T.; Poesen, J.; Van Den Eeckhaut, M.; Nachtergaelel, J.; Deckers, J. Reconstructing rainfall and land-use conditions leading to the development of old gullies. *The Holocene* **2005**, *15*, 378–386.

5. Zabihi, M.; Pourghasemi, H.R.; Motevalli, A.; Zakeri, M.A. Gully erosion modeling using GIS-based data mining techniques in northern Iran: a comparison between boosted regression tree and multivariate adaptive regression spline. In *Natural Hazards GIS-Based Spatial Modeling Using Data Mining Techniques*; Springer, 2019; pp. 1–26.

6. Pimentel, D.; Burgess, M. Soil erosion threatens food production. *Agriculture* **2013**, *3*, 443–463.

7. Arabameri, A.; Rezaei, K.; Pourghasemi, H.R.; Lee, S.; Yamani, M. GIS-based gully erosion susceptibility mapping: a comparison among three data-driven models and AHP knowledge-based technique. *Environ. Earth Sci.* **2018**, *77*, 1–22, doi:10.1007/s12665-018-7808-5.

8. Arabameri, A.; Pradhan, B.; Rezaei, K. Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS. *J. Environ. Manage.* **2019**, *232*, 928–942.

9. Vaezi, A.R.; Abbasi, M.; Bussi, G.; Keesstra, S. Modeling sediment yield in semi-arid pasture micro-catchments, NW Iran. *L. Degrad. Dev.* **2017**, *28*, 1274–1286.

10. Poesen, J.; Nachtergaele, J.; Verstraeten, G.; Valentin, C. Gully erosion and environmental change: importance and research needs. *Catena* **2003**, *50*, 91–133.

11. Poesen, J. Soil erosion in the Anthropocene: Research needs. *Earth Surf. Process. Landforms* **2018**, *43*, 64–84.

12. Valentin, C.; Poesen, J.; Li, Y. Gully erosion: impacts, factors and control. *Catena* **2005**, *63*, 132–153,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

doi:10.1016/j.catena.2005.06.001.

13. Chaplot, V. Impact of terrain attributes, parent material and soil types on gully erosion. *Geomorphology* **2013**, *186*, 1–11.

14. Angileri, S.E.; Conoscenti, C.; Hochschild, V.; Märker, M.; Rotigliano, E.; Agnesi, V. Water erosion susceptibility mapping by applying stochastic gradient treeboost to the Imera Meridionale river basin (Sicily, Italy). *Geomorphology* **2016**, *262*, 61–76.

15. Saha, S.; Roy, J.; Arabameri, A.; Blaschke, T.; Tien Bui, D. Machine learning-based gully erosion susceptibility mapping: A case study of Eastern India. *Sensors* **2020**, *20*, 1313.

16. Moradi, H.R.; Avand, M.T.; Janizadeh, S. landslide susceptibility survey using modeling methods. In; 2019; pp. 259–276.

17. Watson, G.L.; Telesca, D.; Reid, C.E.; Pfister, G.G.; Jerrett, M. Machine learning models accurately predict ozone exposure during wildfire events. *Environ. Pollut.* **2019**, *254*, 112792.

18. Arabameri, A.; Pradhan, B.; Rezaei, K.; Yamani, M.; Pourghasemi, H.R.; Lombardo, L. Spatial modelling of gully erosion using evidential belief function, logistic regression, and a new ensemble of evidential belief function--logistic regression algorithm. *L. Degrad. Dev.* **2018**, *29*, 4035–4049.

19. Dube, F.; Nhapi, I.; Murwira, A.; Gumindoga, W.; Goldin, J.; Mashauri, D.A. Potential of weight of evidence modelling for gully erosion hazard assessment in Mbire District – Zimbabwe. *J. Phys. Chem. EAR* **2014**, doi:10.1016/j.pce.2014.02.002.

20. Pourghasemi, H..; Youse, S.; Kornejady, A.; Cerdà, A. Science of the Total Environment Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. **2017**, *609*, 764–775, doi:10.1016/j.scitotenv.2017.07.198.

21. Pourghasemi, H.R.; Yousefi, S.; Kornejady, A.; Cerdà, A. Performance assessment of individual and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

ensemble data-mining techniques for gully erosion modeling. *Sci. Total Environ.* **2017**, *609*, 764–775.

22. Roy, P.; Chakrabortty, R.; Chowdhuri, I.; Malik, S.; Das, B.; Pal, S.C. Development of Different Machine Learning Ensemble Classifier for Gully Erosion Susceptibility in Gandheswari Watershed of West Bengal, India. In *Machine Learning for Intelligent Decision Science*; Springer, 2020; pp. 1–26.

23. Gayen, A.; Pourghasemi, H.R. Spatial Modeling of Gully Erosion: A New Ensemble of CART and GLM Data-Mining Algorithms. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*; Elsevier, 2019; pp. 653–669.

24. Yunkai, L.; Yingjie, T.; Zhiyun, O.; Lingyan, W.; Tingwu, X.; Peiling, Y.; Huanxun, Z. Analysis of soil erosion characteristics in small watersheds with particle swarm optimization, support vector machine, and artificial neuronal networks. *Environ. Earth Sci.* **2010**, *60*, 1559–1568.

25. Saha, A.; Ghosh, M.; Pal, S.C. Understanding the Morphology and Development of a Rill-Gully: An Empirical Study of Khoai Badland, West Bengal, India. In *Gully Erosion Studies from India and Surrounding Regions*; Springer, 2020; pp. 147–161.

26. Avand, M.; Janizadeh, S.; Naghibi, S.A.; Pourghasemi, H.R.; Khosrobeigi Bozchaloei, S.; Blaschke, T. A Comparative Assessment of Random Forest and k-Nearest Neighbor Classifiers for Gully Erosion Susceptibility Mapping. *Water* **2019**, *11*, 2076.

27. Shi, S.; Xu, G. Novel performance prediction model of a biofilm system treating domestic wastewater based on stacked denoising auto-encoders deep learning network. *Chem. Eng. J.* **2018**, *347*, 280–290.

28. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sens.* **2019**, *11*, 196.

29. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural networks* **2015**, *61*, 85–117.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

30. Coelho, I.M.; Coelho, V.N.; Luz, E.J. da S.; Ochi, L.S.; Guimarães, F.G.; Rios, E. A GPU deep learning metaheuristic based model for time series forecasting. *Appl. Energy* **2017**, *201*, 412–418.

31. Hong, H.; Pradhan, B.; Sameen, M.I.; Kalantar, B.; Zhu, A.; Chen, W. Improving the accuracy of landslide susceptibility model using a novel region-partitioning approach. *Landslides* **2018**, *15*, 753–772.

32. Berlin, S.J.; John, M. Particle swarm optimization with deep learning for human action recognition. *Multimed. Tools Appl.* **2020**, 1–23.

33. Poli, R.; Kennedy, J.; Blackwell, T. Particle swarm optimization. *Swarm Intell.* **2007**, *1*, 33–57.

34. Conoscenti, C.; Angileri, S.; Cappadonia, C.; Rotigliano, E.; Agnesi, V.; Märker, M. Geomorphology Gully erosion susceptibility assessment by means of GIS-based logistic regression : A case of Sicily (Italy). *Geomorphology* **2013**, doi:10.1016/j.geomorph.2013.08.021.

35. Conforti, M.; Aucelli, P.P.C.; Robustelli, G.; Scarciglia, F. Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). *Nat. hazards* **2011**, *56*, 881–898.

36. El Maaoui, M.A.; Felfoul, M.S.; Boussema, M.R.; Snane, M.H. Sediment yield from irregularly shaped gullies located on the Fortuna lithologic formation in semi-arid area of Tunisia. *Catena* **2012**, *93*, 97–104.

37. Arabameri, A.; Pradhan, B.; Rezaei, K.; Conoscenti, C. Gully erosion susceptibility mapping using GIS-based multi-criteria decision analysis techniques. *Catena* **2019**, *180*, 282–297.

38. Kalantar, B.; Ueda, N.; Saeidi, V.; Ahmadi, K.; Halin, A.A.; Shabani, F. Landslide Susceptibility Mapping: Machine and Ensemble Learning Based on Remote Sensing Big Data. *Remote Sens.* **2020**, *12*, 1737.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

39. Wang, G.; Chen, X.; Chen, W. Spatial Prediction of Landslide Susceptibility Based on GIS and Discriminant Functions. *ISPRS Int. J. Geo-Information* **2020**, *9*, 144.

40. Chowdhuri, I.; Pal, S.C.; Chakrabortty, R. Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India. *Adv. Sp. Res.* **2020**, *65*, 1466–1489.

41. Youssef, A.M.; Pourghasemi, H.R. Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geosci. Front.* **2020**.

42. Green, I.R.A.; Stephenson, D. Criteria for comparison of single event models. *Hydrol. Sci. J.* **1986**, *31*, 395–411.

43. Kutner, M.H.; Nachtsheim, C.J.; Neter, J.; Li, W.; others *Applied linear statistical models*; McGraw-Hill Irwin New York, 2005; Vol. 5;.

44. Gayen, A.; Pourghasemi, H.R.; Saha, S.; Keesstra, S.; Bai, S. Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Sci. Total Environ.* **2019**, *668*, 124–138.

45. Hong, H.; Liu, J.; Zhu, A.-X.; Shahabi, H.; Pham, B.T.; Chen, W.; Pradhan, B.; Bui, D.T. A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China). *Environ. Earth Sci.* **2017**, *76*, 652.

46. Tehrany, M.S.; Pradhan, B.; Mansor, S.; Ahmad, N. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena* **2015**, *125*, 91–101.

47. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

48. Vapnik, V.N. The nature of statistical learning. *Theory* **1995**.

49. Abedini, M.; Ghasemian, B.; Shirzadi, A.; Bui, D.T. A comparative study of support vector machine and logistic model tree classifiers for shallow landslide susceptibility modeling. *Environ. Earth Sci.* **2019**, *78*, 560.

50. Yao, X.; Tham, L.G.; Dai, F.C. Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582.

51. Naghibi, S.A.; Moghaddam, D.D.; Kalantar, B.; Pradhan, B.; Kisi, O. A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. *J. Hydrol.* **2017**, *548*, 471–483.

52. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365.

53. Kavzoglu, T.; Colkesen, I. A kernel functions analysis for support vector machines for land cover classification. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 352–359.

54. Haykin, S. Neural Networks: A Comprehensive Foundation. 2nd Edn Prentice Hall. *Englewood Cliffs* **1999**.

55. Cherkassky, V.; Krasnopolsky, V.; Solomatine, D.P.; Valdes, J. Computational intelligence in earth sciences and environmental applications: Issues and challenges. *Neural Networks* **2006**, *19*, 113–121.

56. Saha, A.K.; Gupta, R.P.; Arora, M.K. GIS-based landslide hazard zonation in the Bhagirathi (Ganga) valley, Himalayas. *Int. J. Remote Sens.* **2002**, *23*, 357–369.

57. Kawabata, D.; Bandibas, J. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). *Geomorphology* **2009**, *113*, 97–109.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

58. Kosko, B. *Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence.*; 1992;

59. Mandal, S.; Mondal, S. Machine Learning Models and Spatial Distribution of Landslide Susceptibility. In *Geoinformatics and Modelling of Landslide Susceptibility and Risk*; Springer, 2019; pp. 165–175.

60. Falaschi, F.; Giacomelli, F.; Federici, P.R.; Puccinelli, A.; Avanzi, G.; Pochini, A.; Ribolini, A. Logistic regression versus artificial neural networks: landslide susceptibility evaluation in a sample area of the Serchio River valley, Italy. *Nat. hazards* **2009**, *50*, 551–569.

61. Chen, W.; Pourghasemi, H.R.; Zhao, Z. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto Int.* **2017**, *32*, 367–385.

62. Kim, P. Matlab deep learning. *With Mach. Learn. Neural Networks Artif. Intell.* **2017**, *130*, 21.

63. Lewis, N.D.C. *Deep learning made easy with R: A gentle introduction for data science*; AusCov, 2016;

64. Kennedy, J.; Eberhart, R.C.; Shi, Y. The particle swarm. *Swarm Intell.* **2001**, 287–325.

65. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the Proceedings of ICNN'95 - International Conference on Neural Networks; 1995; Vol. 4, pp. 1942–1948.

66. Clerc, M.; Kennedy, J. The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Trans. Evol. Comput.* **2002**, *6*, 58–73.

67. Olsson, A.E. *Particle swarm optimization: theory, techniques and applications*; Nova Science Publishers, Inc., 2010;

68. Frattini, P.; Crosta, G.; Carrara, A. Techniques for evaluating the performance of landslide susceptibility models. *Eng. Geol.* **2010**, *111*, 62–72.

69. Yesilnacar, E.; Topal, T. Landslide susceptibility mapping: a comparison of logistic regression and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

neural networks methods in a medium scale study, Hendek region (Turkey). *Eng. Geol.* **2005**, *79*, 251–266.

70. Biot, Y.; Blaikie, P.M.; Jackson, C.; Palmer-Jones, R. *Rethinking research on land degradation in developing countries*; The World Bank, 1995;

71. Sirviö, T.; Rebeiro-Hargrave, A.; Pellikka, P. Geoinformation in gully erosion studies in the Taita Hills, SE-Kenya, preliminary results. In Proceedings of the Proceedings of the 5th AARSE conference (African Association of Remote Sensing of the Environment); 2004; pp. 18–21.

72. Dotterweich, M.; Rodzik, J.; Zgłobicki, W.; Schmitt, A.; Schmidtchen, G.; Bork, H.-R. High resolution gully erosion and sedimentation processes, and land use changes since the Bronze Age and future trajectories in the Kazimierz Dolny area (Nał{\k{e}}czów Plateau, SE-Poland). *Catena* **2012**, *95*, 50–62.

73. Pal, S.C.; Chakrabortty, R. Modeling of water induced surface soil erosion and the potential risk zone prediction in a sub-tropical watershed of Eastern India. *Model. Earth Syst. Environ.* **2019**, *5*, 369–393.

74. Chakrabortty, R.; Pal, S.C.; Chowdhuri, I.; Malik, S.; Das, B. Assessing the Importance of Static and Dynamic Causative Factors on Erosion Potentiality Using SWAT, EBF with Uncertainty and Plausibility, Logistic Regression and Novel Ensemble Model in a Sub-tropical Environment. *J. Indian Soc. Remote Sens.* **2020**, 1–25.

75. Pal, S.C.; Chakrabortty, R. Simulating the impact of climate change on soil erosion in sub-tropical monsoon dominated watershed based on RUSLE, SCS runoff and MIROC5 climatic model. *Adv. Sp. Res.* **2019**, *64*, 352–377.

76. Borrelli, P.; Robinson, D.A.; Fleischer, L.R.; Lugato, E.; Ballabio, C.; Alewell, C.; Meusburger, K.; Modugno, S.; Schütt, B.; Ferro, V.; et al. An assessment of the global impact of 21st century land use

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 September 2020 doi:10.20944/preprints202009.0516.v1

change on soil erosion. *Nat. Commun.* **2017**, *8*, 1–13.

77. Peng, T.; Wang, S. Effects of land use, land cover and rainfall regimes on the surface runoff and soil loss on karst slopes in southwest China. *Catena* **2012**, *90*, 53–62.

78. Arabameri, A.; Pradhan, B.; Pourghasemi, H.R.; Rezaei, K.; Kerle, N. Spatial modelling of gully erosion using GIS and R programing: A comparison among three data mining algorithms. *Appl. Sci.* **2018**, *8*, 1369.