

Ensemble hybrid machine learning methods for gully erosion susceptibility mapping: K-fold cross validation approach



Jagabandhu Roy, Sunil Saha *

Department of Geography, University of Gour Banga, Malda, 732103, West Bengal, India

ARTICLE INFO

Keywords:

K-fold cross-validation
Gully erosion susceptibility
Radial basis function neural network
Hybrid ensemble algorithms
R-Index

ABSTRACT

Gully erosion is one of the important problems creating barrier to agricultural development. The present research used the radial basis function neural network (RBFnn) and its ensemble with random sub-space (RSS) and rotation forest (RTF) ensemble Meta classifiers for the spatial mapping of gully erosion susceptibility (GES) in Hinglo river basin. 120 gullies were marked and grouped into four-fold. A total of 23 factors including topographical, hydrological, lithological, and soil physio-chemical properties were effectively used. GES maps were built by RBFnn, RSS-RBFnn, and RTF-RBFnn models. The very high susceptibility zone of RBFnn, RTF-RBFnn and RSS-RBFnn models covered 6.75%, 6.72% and 6.57% in Fold-1, 6.21%, 6.10% and 6.09% in Fold-2, 6.26%, 6.13% and 6.05% in Fold-3 and 7%, 6.975% and 6.42% in Fold-4 of the basin. Receiver operating characteristics (ROC) curve and statistical techniques such as mean-absolute-error (MAE), root-mean-absolute-error (RMSE) and relative gully density area (R-index) methods were used for evaluating the GES maps. The results of the ROC, MAE, RMSE and R-index methods showed that the models of susceptibility to gully erosion have excellent predictive efficiency. The simulation results based on machine learning are satisfactory and outstanding and could be used to forecast the areas vulnerable to gully erosion.

1. Introduction

Soil and water are severely threatened by soil erosion, which is a worldwide big environmental issue (Arabameri et al., 2020a). Long-term erosion effects are visible, but short-term erosion effects may not be apparent (Singh and Singh, 2018). Over the past decades, the impact of soil erosion has increased rapidly (Gayen et al., 2019). Due to the formation of dissolution and alkalinity in rangelands, farming lands, and forest areas the rate of gully erosion in these area are excessive (García-Ruiz, 2010). Gully erosion breaks down the soil ecosystem and degrades the quality of the river and wetlands water (Vanmaercke et al., 2016; Debanshi and Pal, 2020).

As per the Hungarian classification, gully erosion is linear erosion process (Kertész, 2009). Gullies are two types that are permanent gullies and ephemeral gullies (Casali et al., 1999). A permanent gully is defined as the broad and deep channels eroded by the concentrated flow that removes surface soil and parent material not removable via normal tillage operations. Ephemeral gullies, on the other hand, are created by the concentrated overland flow that can be remedied through regular tillage activities (Casali et al., 1999). Gully erosion occurs when the surface runoff is concentrated into a channel and results in the formation

of rills that grow over time into deep trenches on the ground (Karuma et al., 2014; Debanshi and Pal, 2020). Several geo-environmental variables, including climate, landscape, soil, geology, and land use, are the key factors influencing the growth and occurrence of the gully (Guerra et al., 2018). Several researchers used lithology, land use, slope, aspect, plan curvature, stream power index, topographical wetness index and length-slope factor as gully erosion predisposing factors in different parts of the world (Conforti et al., 2011; Conoscentiet al., 2018; Cominoet al., 2016). The integration of rainfall, runoff, and infiltration affects the soil erosivity that often results in gully erosion (Lal, 2001). Gully erosion is also the three-dimensional in nature affected by the broad range of environmental factors (Zhang et al., 2015). Normally, a gully have a steep-sided or vertical headwall, a width larger than 0.30 m, and a depth greater than 0.60 m (Brice, 1966). The factors of lithology, soil quality, topography, climate, vegetation, and land use are important controlling factors for the formation of gullies (Ogboma et al., 2011). Several physio-chemical factors of soil like texture, soil volume, clay, sand pH, electrical conductivity, sodium absorption ratio (SAR), sodium, calcium, manganese, bulk density influence the formation of gullies (AsghariSaraskanroud et al., 2017; Hosseinalizadeh et al., 2019). To various erosive agents and external forces, physio-chemical factors assist to detach and

* Corresponding author.

E-mail addresses: jagabandhuroy1991@gmail.com (J. Roy), sunilgeo.88@gmail.com (S. Saha).

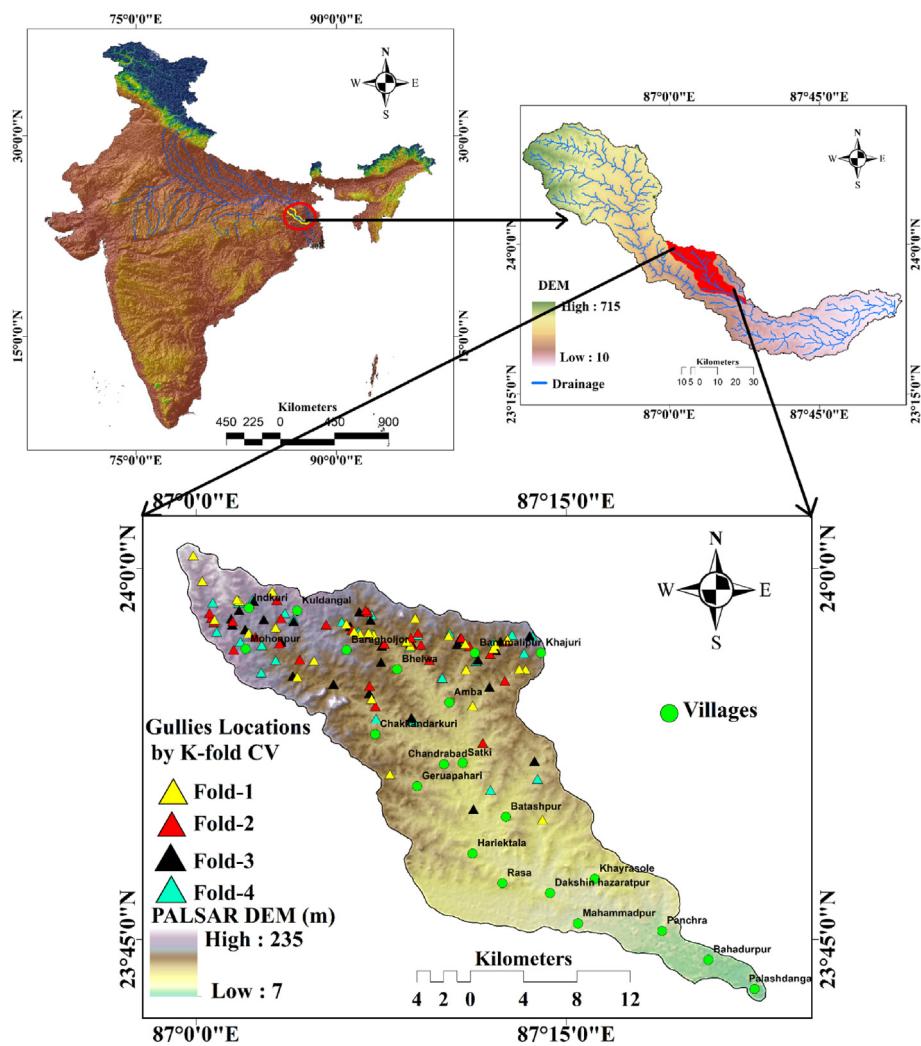


Fig. 1. Location of the study area.

transport the soil (Dondofema et al., 2008). Electrical Conductivity (EC) and SAR interactions increase the sensitivity of soil to gully erosion (Shahrivar et al., 2012). Besides, soil physical and chemical properties can not only increase the susceptibility to gully erosion but can also reduce the growth of vegetation in the soil. The literatures of Nandi and Luffman (2012); Battaglia et al. (2002) assessed gully erosion based on soil physio-chemical factors. However, the integration of numerous environmental factors using appropriate methods for gully susceptibility erosion modeling (GESM) is essential for protecting the soil and water (Shit et al., 2015).

In the present decade, machine learning ensemble approaches have been applied to the assessment of natural hazards in different parts of the world. The machine learning-based models are more effective and precise than traditional conventional methods. The ensemble techniques e.g. functional classifier and its ensemble with meta, tree classifier, multi-criteria decision method and its ensemble with bivariate and multivariate statistical models, etc. were used in mapping the gully erosion susceptibility, landslide susceptibility, flood susceptibility, land subsidence susceptibility, etc. (Pham et al., 2019; Gayen et al., 2019; Hosseinalizadeh et al., 2019; Garosi et al., 2019; Taheri et al., 2019). Present GESM is regarded as a resource management method for the prevention and maintenance of soil erosion (Roy and Saha, 2019; Gayen and Saha, 2017). Some researchers e.g. Pham et al. (2019), applied the REPT approaches and made an ensemble with hybrid machine learning meta classifiers Bagging, MultiBoost, Rotation Forest, Random Subspace to map the landslide susceptibility. The results of the models showed that

the meta classifiers increased the predictive capability of the REPTree model. Pham et al. (2017) also applied the MLP and its ensemble with Bagging, Dagging, Random subspace, rotation forest classifiers for LSM modeling, and found similar results. Chen et al. (2017) were used ANN, ME, and SVM for landslide susceptibility assessment. The ensemble ANN-SVM has the highest predictive performance for the assessment of LSM in its study. Hembram et al. (2021) used a similar approach for mapping the susceptibility to gully erosion. For mapping, the susceptibility to gully erosion of the present study area novel ensemble methods of radial basis functions neural network (RBFnn) and random sub-space (RSS) and rotation forest (RTF) hybrid ensemble meta classifiers were used. Geographical information system (GIS) and remote sensing (RS) integrating with the machine learning algorithms have created a good basement for mapping the different natural hazards. RS and GIS are reliable and efficient technologies that induce meaningful results in the prediction of small and medium-scale gully erosion (Zerihun et al., 2018). For this phenomenon, many geospatial and geo-statistic methods integrating with the GIS were used for natural hazard modeling (Choubin et al., 2019).

The k-fold cross-validation (CV) approach is one of the statistical validation approaches used for mapping the various natural hazards. The K-fold may be in several forms, namely two-fold, three-fold, four-fold, etc. Arabameri et al. (2020b) used a four-fold cross-validation approach to map the susceptibility of land subsidence. Ghorbanzadeh et al. (2018) used the 4-fold validation method for mapping land subsidence and vulnerability to forest fires. In this study, we have used the

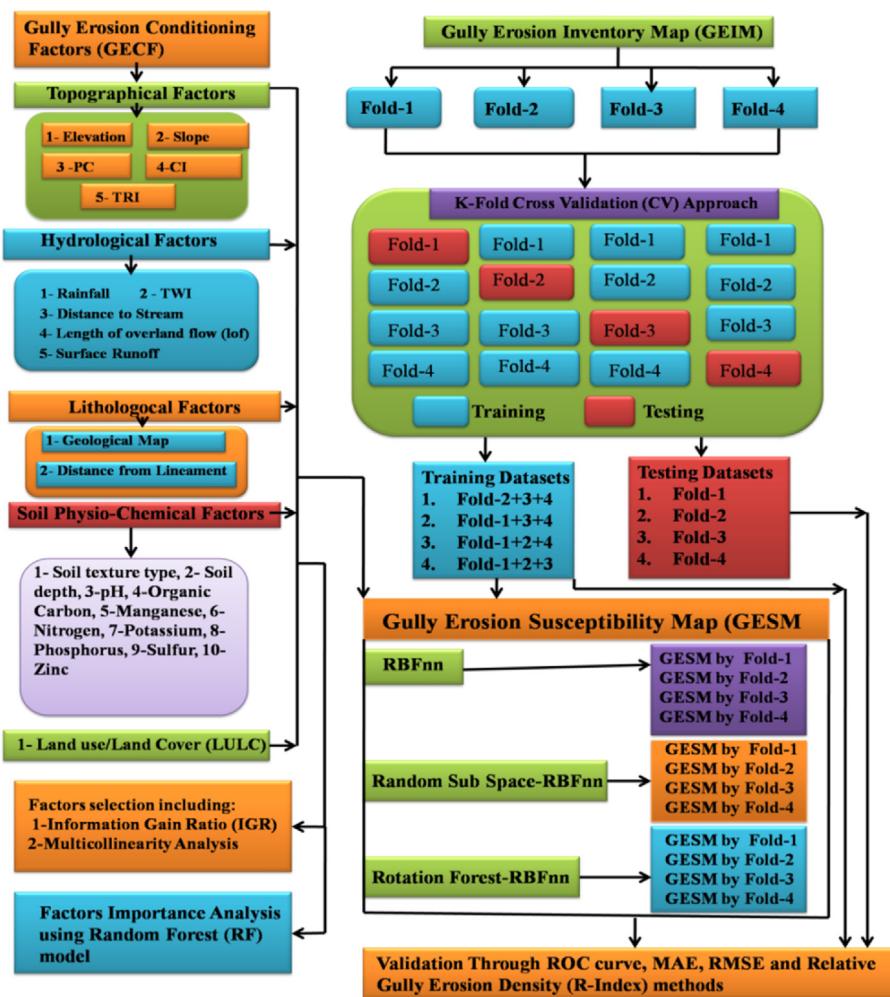


Fig. 2. Flowchart showing the methodology of the present work.

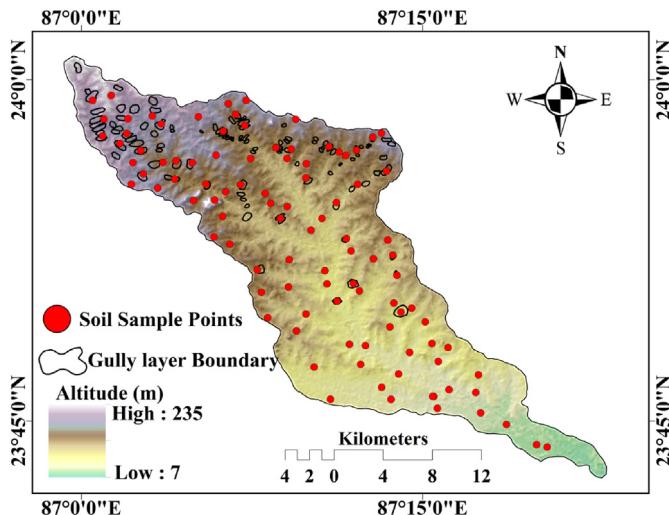


Fig. 3. Location of gully and soil sample points.

four-fold validation approach for mapping the gully erosion susceptibility and selecting the best ensemble method.

The study was aimed to detect gully erosion-prone areas in Eastern India's Hinglo river basin using the hybrid machine learning ensemble approaches namely RBFnn, RSS, and RTF. For GESM the K-fold cross-

validation approach was applied. The receiver operating characteristic (ROC), mean absolute error, root mean square error (RMSE), and relative gully density (R-index) methods were selected for the evaluation of the GESMs.

2. Study area

Geographically, the Hinglo river basin occupies an area of approximately 442.95 sq. km and extends from 23° 42' 7.09"N to 24° 0' 56.78"N latitudes and 86° 59' 32.68"E to 87° 23' 31.91"E longitudes (Fig. 1). The height of basin ranges from 7 m to 235 m from the mean sea level. The climatic condition of the basin is largely influenced by the South-West Monsoon. The wet and rainy south-west Indian monsoon is very strong, storming compared to the winter Indian Monsoon. According to observation data from the Indian Metrological Department (IMD), the average rainfall is 1326 mm.

Granite-gneiss formation, Barakar formation, ironstone shale formation, quartzite, and younger alluvium geological segments made up the research area (GSI, 1985). The greater part of the study area is occupied by the granite-gneiss geological formation. The newer geological alluvium formation is highly fertile and suitable for agriculture, and rice, master, wheat, maize, and sugarcane are the major crops cultivated in this basin. The agricultural activities are the basic and main economic activity of the river bank dwellers. Based on USDA's soil texture classification given in 1985, the basin consists of fine loamy mixed haplustalf, fine loamy mixed with plustalf, clay, clay loam, sandy, sandy loam, and loam soil textures (NATMO, 2001). From a morphological viewpoint the

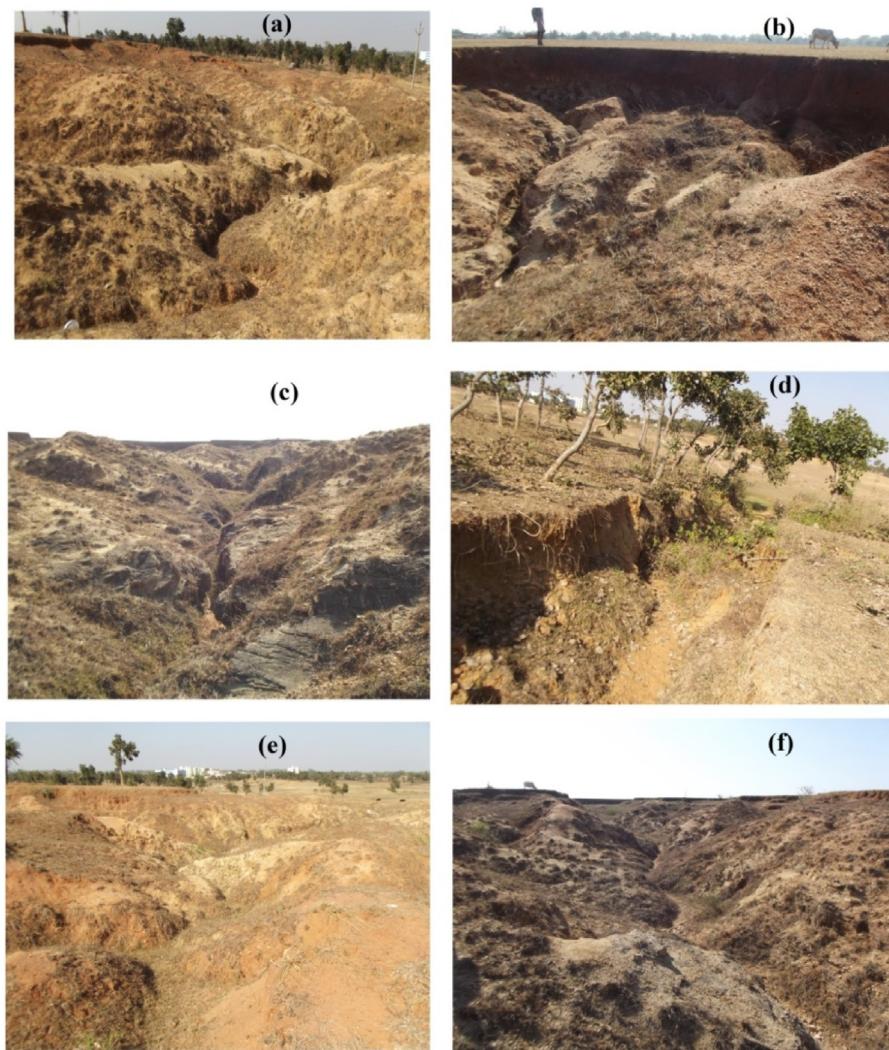


Fig. 4. Photographs are showing the gully distribution: (a) Pindergaria ($23^{\circ} 59' 14''\text{N}$, $87^{\circ} 00' 22''\text{E}$), (b) Dainghati ($23^{\circ} 57' 54''\text{N}$, $87^{\circ} 00' 45''\text{E}$), (c) Jagannathpur ($23^{\circ} 58' 42''\text{N}$, $87^{\circ} 01' 47''\text{E}$), (d) Hesaltanr ($23^{\circ} 56' 12''\text{N}$, $87^{\circ} 07' 27''\text{E}$), (e) Sima ($23^{\circ} 55' 53''\text{N}$, $87^{\circ} 13' 17''\text{E}$), (f) Kadma ($23^{\circ} 57' 5''\text{N}$, $87^{\circ} 13' 35''\text{E}$).

research area's upper section is characterized by higher slope. The maximum slope is 35° which is found in the catchment's northwest portion. The upper portion of the basin is facing a huge problem of gully erosion (Ghosh and Shah, 2015). The main erosive processes that affect the landscape in the study area are related to runoff waters. In locations where there is no vegetative cover and in cultivated fields, overland flow processes are highly active. In the upper catchment area large portion is covered by the lateritic soil. Bare surface, presence of lateritic soil and runoff are combinedly inducing the gully erosion in this study area. In these circumstances for sustainable land management and reducing the gully erosion some protective measures should be taken into consideration. For the sustainable management of the basin first, it is essential to map the potential area of gully erosion in the basin and then strategies should be formulated giving the priority on level of gully erosion potential. Viewing the gully erosion problem we mapped the area susceptibility to gully erosion using the novel ensemble method.

3. Materials and methods

In this research following steps were carried out (Fig. 2).

1. Data regarding gully erosion and gully erosion conditioning factors (GECFs) were collected;

2. Thematic layers of GECFs were prepared and gully erosion datasets were classified into four folds using K-fold cross validation (CV) approach in GIS platform,
3. GECFs were selected using multi-collinearity assessment and Information gain ratio (IGR)
4. An artificial intelligence i.e. Radial Basis Function (RBF) neural network and its ensemble with Meta classifiers i.e. Random sub space (RSS) and Rotation forest (RTF) models were applied to produce gully erosion susceptibility maps (GESMs).
5. Significance of GECFs was assessed using random forest (RF) model,
6. Models performance was analyzed using ROC, statistical methods like MAE, RMSE and relative gully erosion density (R-Index) methods.

3.1. Data sources

The ALOS PALSAR DEM was obtained from the Alaska satellite facility department. Precipitation data was obtained from the India Metrological Department of the nearest rainfall stations. The current study area's geological map was collected from the Geological Survey of India. The Landsat 8OLI/TIRS was downloaded from the United States Geological Survey (dated April 2108, path = 143, row = 39) for the extraction of the land use map and NDVI. Soil texture map was obtained from the department of NSSLUP (National Soil Survey and Land Use

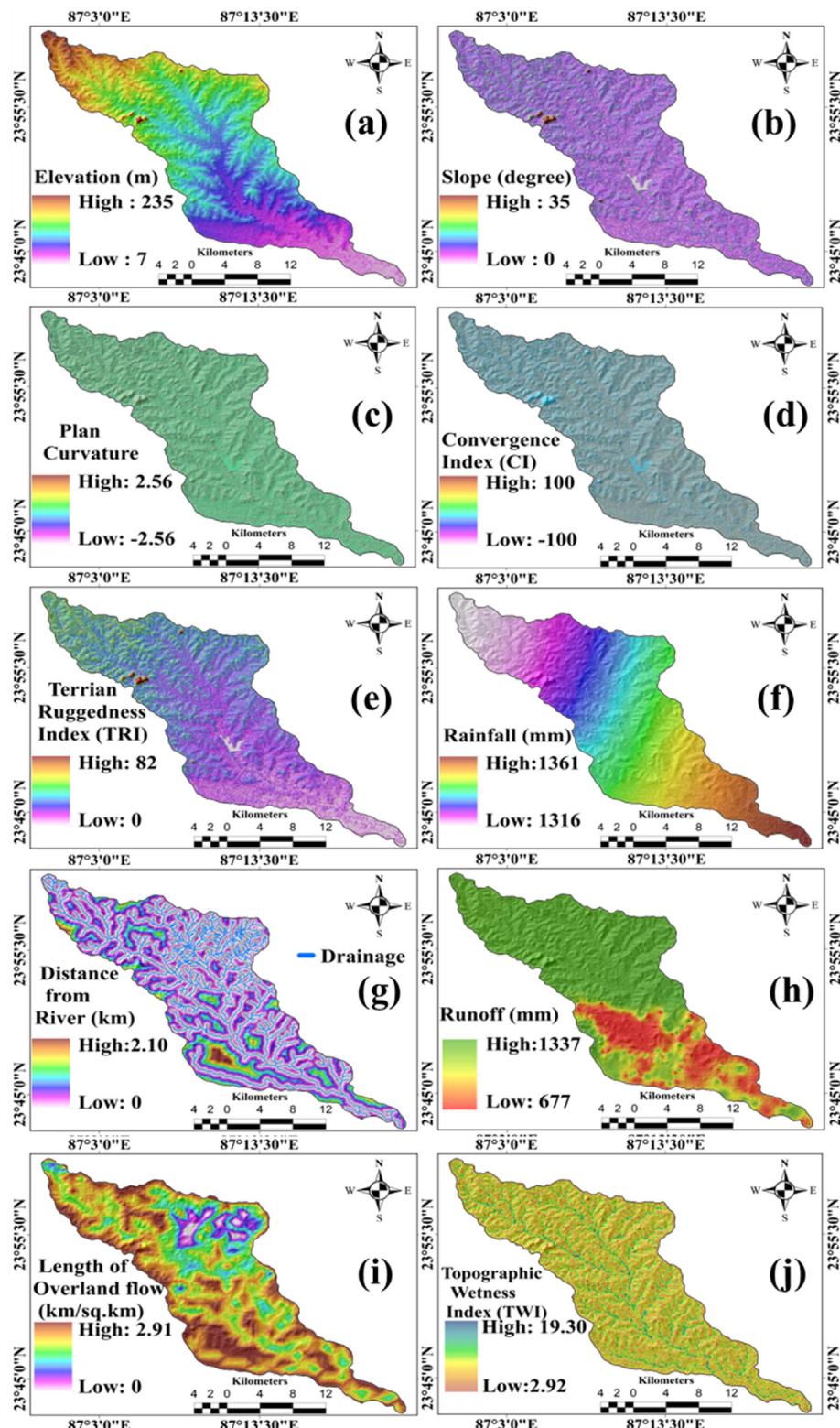


Fig. 5. Topographical and Hydrological gully erosion conditioning factors: (a) Elevation, (b) Slope, (c) Plan curvature, (d) Convergence index, (e) Terrain ruggedness index, (f) Rainfall, (g) Distance from River, (h) Surface Runoff, (i) Length of overland flow, (j) Topographical wetness index.

Planning Bureau). The primary data like soil samples were collected through field survey for assessing the physiochemical properties of the study area. Measurement of width, depth and areal coverage were also done through field survey. The soil chemical properties such as pH, manganese, phosphorous, potassium, iron, organic carbon, sulfur, zinc were measured through laboratory analysis.

3.2. Inventory map of gully erosion (GEIM)

GEIM is the basic prerequisite for the gully erosion assessment and modeling. The GEIM represents the location of gullies that were mapped in GIS. The GEIM was prepared using field investigation with a global positioning system (GPS) and Google Earth images. The high-resolution

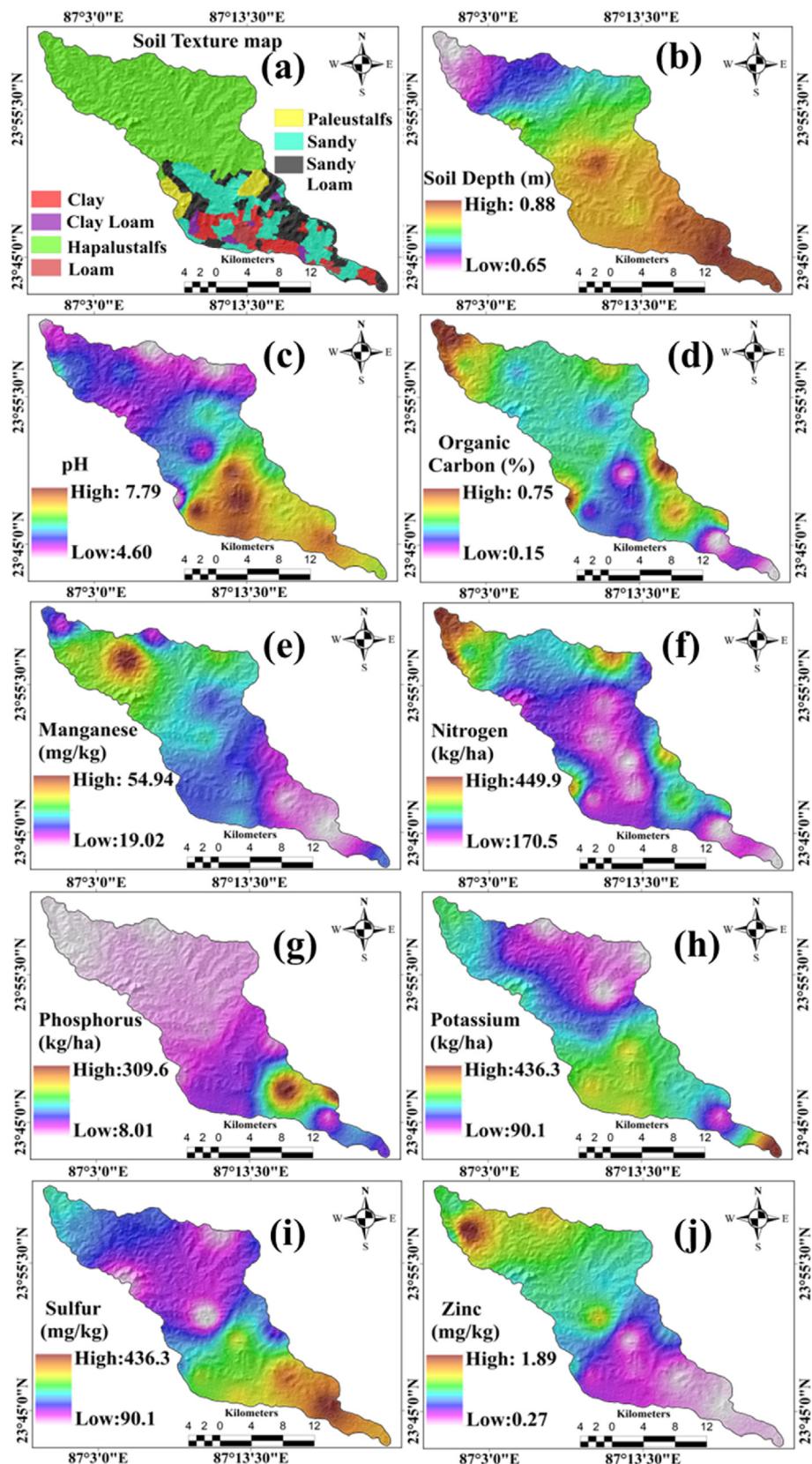


Fig. 6. Soil physio-chemical gully erosion conditioning factors: (a) Soil texture map, (b) Soil depth, (c) pH, (d) Organic Carbon, (e) Manganese, (f) Nitrogen, (g) Phosphorous, (h) Potassium, (i) Sulfur, (j) Zinc.

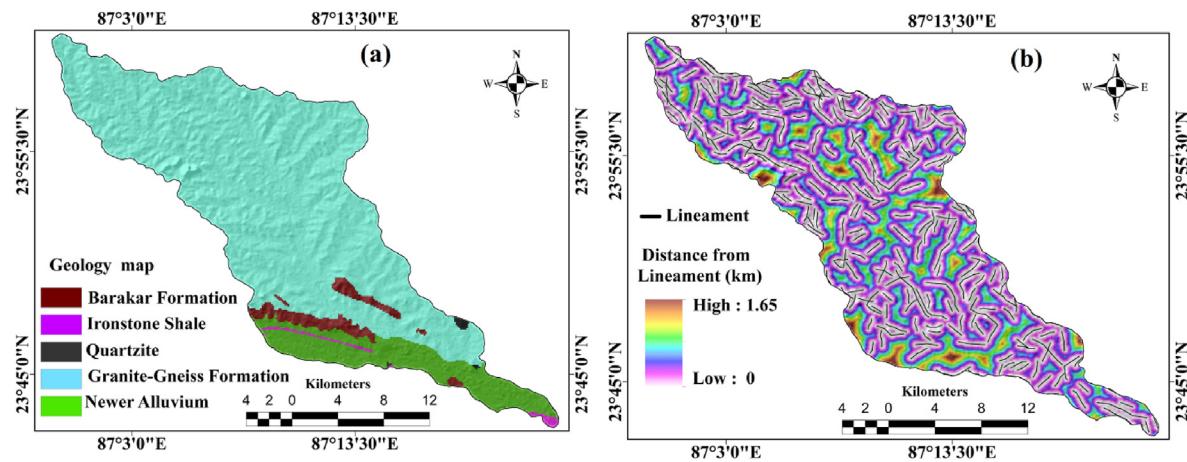


Fig. 7. Lithological gully erosion conditioning factors: (a) Geology map, (b) Distance from Lineament.

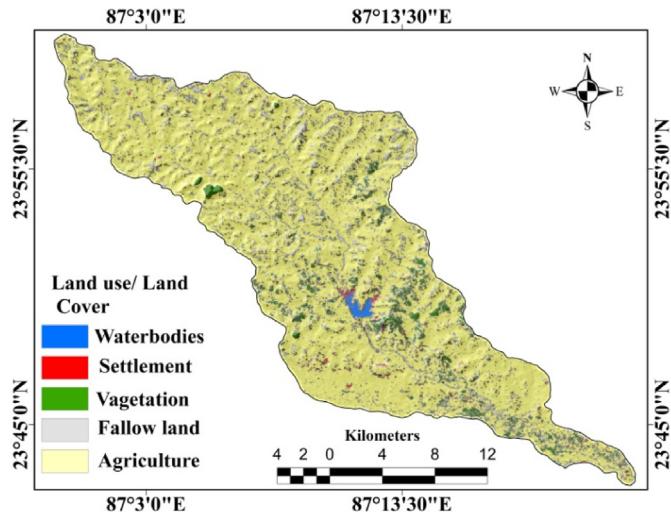


Fig. 8. Land use/land cover (LULC) map of the study area.

Table 1
Multi-collinearity results of Gully Erosion Conditioning Factors.

Factors	Collinearity analysis		Factors	Collinearity analysis	
	TOL	VIF		TOL	VIF
Elevation	0.383	2.611	Zinc	0.662	1.511
Slope	0.61	1.639	sulfur	0.552	1.812
PC	0.738	1.355	Nitrogen	0.339	2.95
CI	0.745	1.342	Manganese	0.228	4.386
TRI	0.536	1.866	Potassium	0.231	4.329
TWI	0.909	1.1	Phosphorus	0.208	4.808
Lof	0.427	2.342	OC	0.335	2.985
Dist. to river	0.604	1.656	pH	0.266	3.759
Dist. to lineament	0.968	1.033	Soil Depth	0.242	4.132
Rainfall	0.41	2.439	Soil Type	0.662	1.511
Geology	0.893	1.12	Surface Runoff	0.235	4.112
LULC	0.946	1.057			

imagery can be used to obtain the information about gully erosion in remote areas where human cannot reach. The Google Earth image is of high resolution and has good capability to detect the gullies in the remote area (Arabameri et al., 2020b). First, gullies were identified through the Google earth images and then for the locational verification and measurement of width, depth and areal coverage field survey was done with global positioning system (GPS). In the study area, a total of 120 gullies

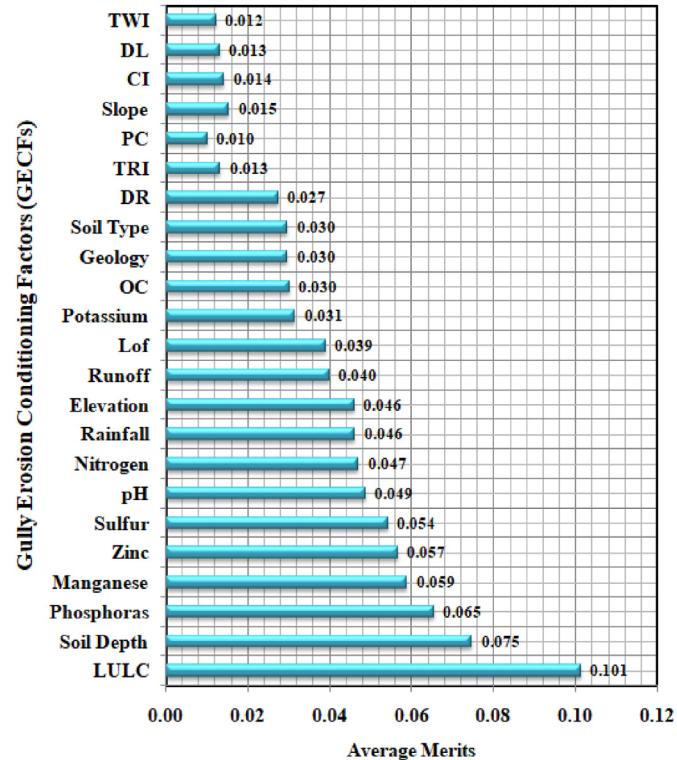


Fig. 9. Graph is showing the average merit of GECFs calculated by IGR.

were identified (Fig. 3). The gully polygons were converted to points (head-cut point). Equal number of non-gully points was also selected randomly. Then the points were used in training and testing the models. A k-fold CV framework (Fig. 2) was chosen to remove the negative impacts of randomness on the efficiency of machine learning approaches. The four-fold CV was used to divide the GEIM into four folds (F_1 , F_2 , F_3 and F_4) for modelling GESM (Fig. 2). Then, the models were run four times. For instance, the 't' model was run with F_1 , F_2 and F_3 datasets without allowing the F_4 fold and then model was evaluated with the F_4 subset. For each time 75% of the selected gully and non-gully points were used for training the models and 25% were used for validating the trained models. During field survey geometry of some gully was measured. The gully's maximum length is 782 m, and the gully's shortest length is 387 m. The maximum depth is 6.5 m, and the minimum depth is 2.5 m. The maximum width is 9.2 m, and the minimum width is 3.5 m.

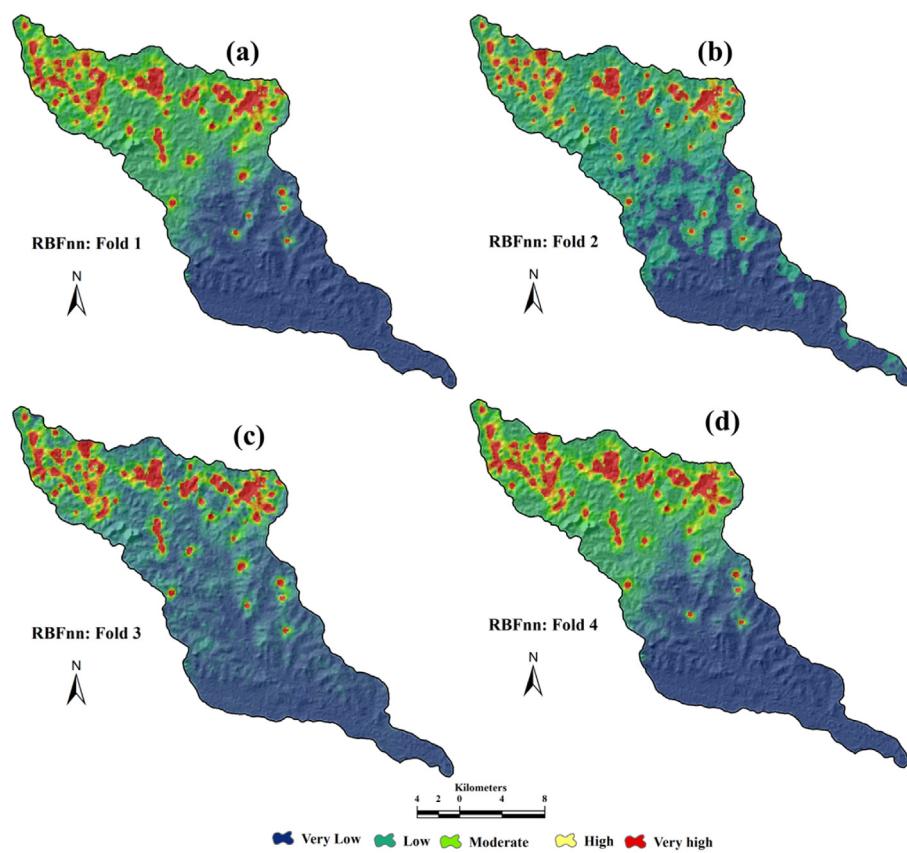


Fig. 10. GESM constructed by RBFnn: (a) fold-1, (b) fold-2, (c) fold-3 and (d) fold-4.

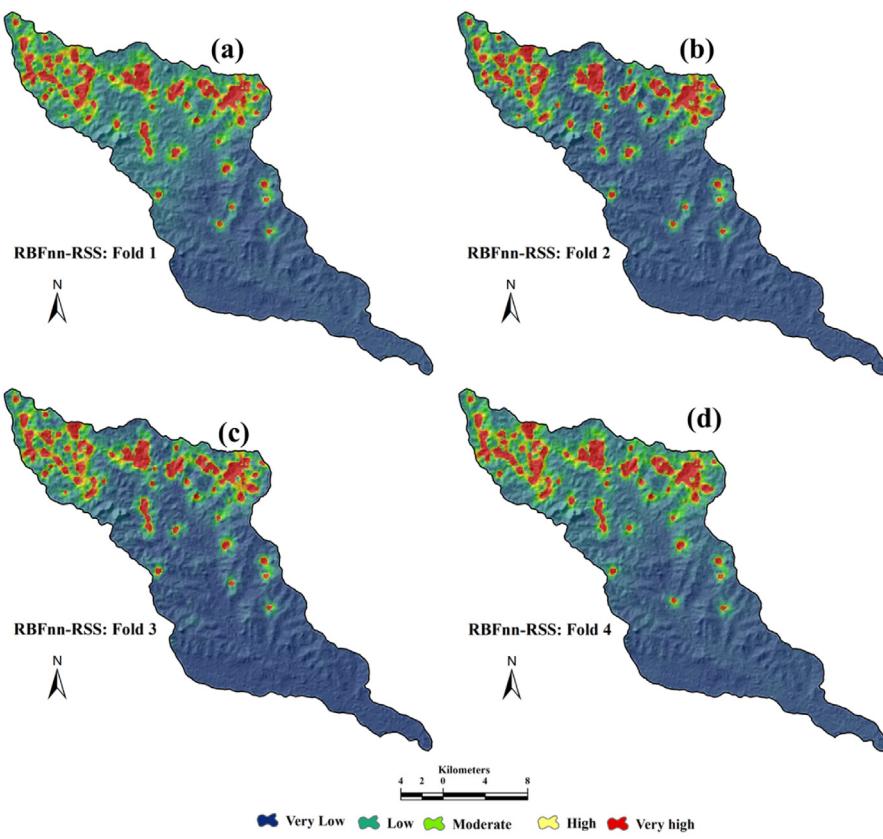


Fig. 11. GESM constructed by RSS-RBFnn: (a) fold-1, (b) fold-2, (c) fold-3 and (d) fold-4.

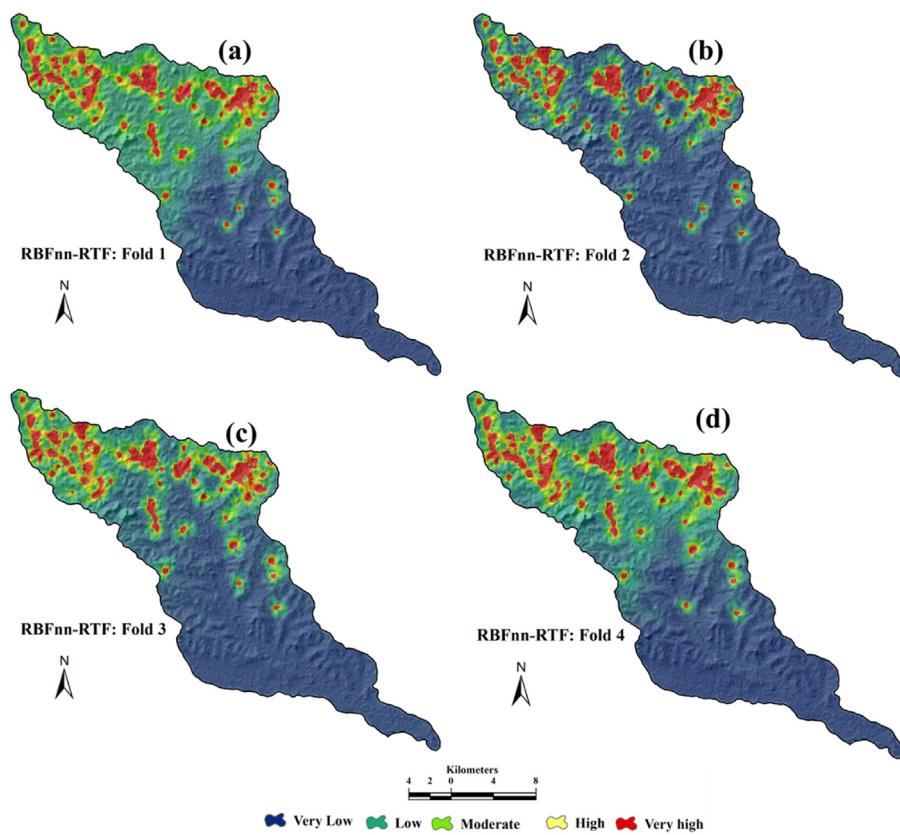


Fig. 12. GESM constructed by RTF-RBFnn: (a) fold-1, (b) fold-2, (c) fold-3 and (d) fold-4.

Table 2
Values of AUC of ROC curve, MAE and RMSE methods.

Statistical techniques	Training datasets			Validation datasets		
	RBF	RSS-RBF	RTF-RBF	RBF	RSS-RBF	RTF-RBF
Fold-1						
AUC	0.911	0.924	0.920	0.910	0.920	0.915
MAE	0.070	0.041	0.058	0.080	0.071	0.075
RMSE	0.265	0.203	0.241	0.283	0.266	0.274
Fold-2						
AUC	0.923	0.939	0.925	0.895	0.910	0.904
MAE	0.047	0.042	0.042	0.040	0.039	0.040
RMSE	0.217	0.205	0.205	0.199	0.197	0.200
Fold-3						
AUC	0.928	0.936	0.924	0.903	0.913	0.909
MAE	0.030	0.025	0.025	0.090	0.075	0.080
RMSE	0.174	0.159	0.158	0.299	0.273	0.284
Fold-4						
AUC	0.901	0.938	0.887	0.908	0.940	0.923
MAE	0.053	0.031	0.041	0.093	0.082	0.083
RMSE	0.229	0.176	0.202	0.305	0.286	0.288

Some valuable field photos were taken during the field measurement and survey, as shown in Fig. 4.

3.3. Preparation of effective factors

In the present study various environmental factors for modeling the gully erosion susceptibility (GES) including topographical, hydrological, lithological, soil physical and chemical characteristics were selected considering the previous literatures. These factors were constructed in the GIS platform as spatial datasets for modeling through the different ensembles methods.

3.3.1. Topographical and Hydrological factors

Gully formation is regulated by topographic factors (Shit et al., 2013). The topographic factors affect the erosive power of runoff, possible discharge, flow velocity, and transport efficiency (Claps and Rossi, 1994). Two types of topographic attributes exist, i.e. the primary and secondary attributes. The primary topographical attributes are altitude, slope, slope aspect, catchment area, the curvature of plane and profile and secondary topographical attributes are SPI, STI, CI, TRI, TPI, and TWI (Garosi et al., 2019). All these topographical factors were derived from PALSAR DEM using the SAGA GIS tool. Altitude is an important gully conditioning factor that influences the formation of the gully (Gayen et al., 2020). In the present study, the altitude of the basin is derived from PALSAR DEM with a resolution of 12.5 m*12.5 m. Therefore, elevation varies from 7 m to 235 m as per the PALSAR DEM (Fig. 5a). Drainage development and surface water flow are determined by the degree of slope, which is considered to be the main explanatory factor for gully formation (Mararakanye, N., 2016; Hembram et al., 2020). In this research, the slope map is extracted from PALSAR DEM and the maximum slope of the basin is 31° (Fig. 5b). The plan and profile curvatures on local terrains influence overland flow, surface runoff, and subsequently gully formation (Burian et al., 2015). A curvature of the plane can be defined as the hypothetical line crossing a specific cell on the contour line (Evans and Cox, 1999). The plan curvature for this basin was extracted from DEM with the SAGA GIS tool (Fig. 5c). Convergence index (CI) shows the relief structure as a set of channels and ridges. This reflects the agreement between the slope orientation of the surrounding cells and the theoretical orientation of the matrix. The value of CI ranges from 100 to -100 respectively (Fig. 5d). Terrain ruggedness index (TRI) influences the initiation of a gully. TRI is the form of the terrain and it impacts on water flows which determine the rate of gully erosion (Claps and Rossi, 1994) TRI ranges from 0 to 82.56 (Fig. 5e).

Precipitation data was collected from the Indian Metrological Department for the various stations. Based on IDW interpolation method,

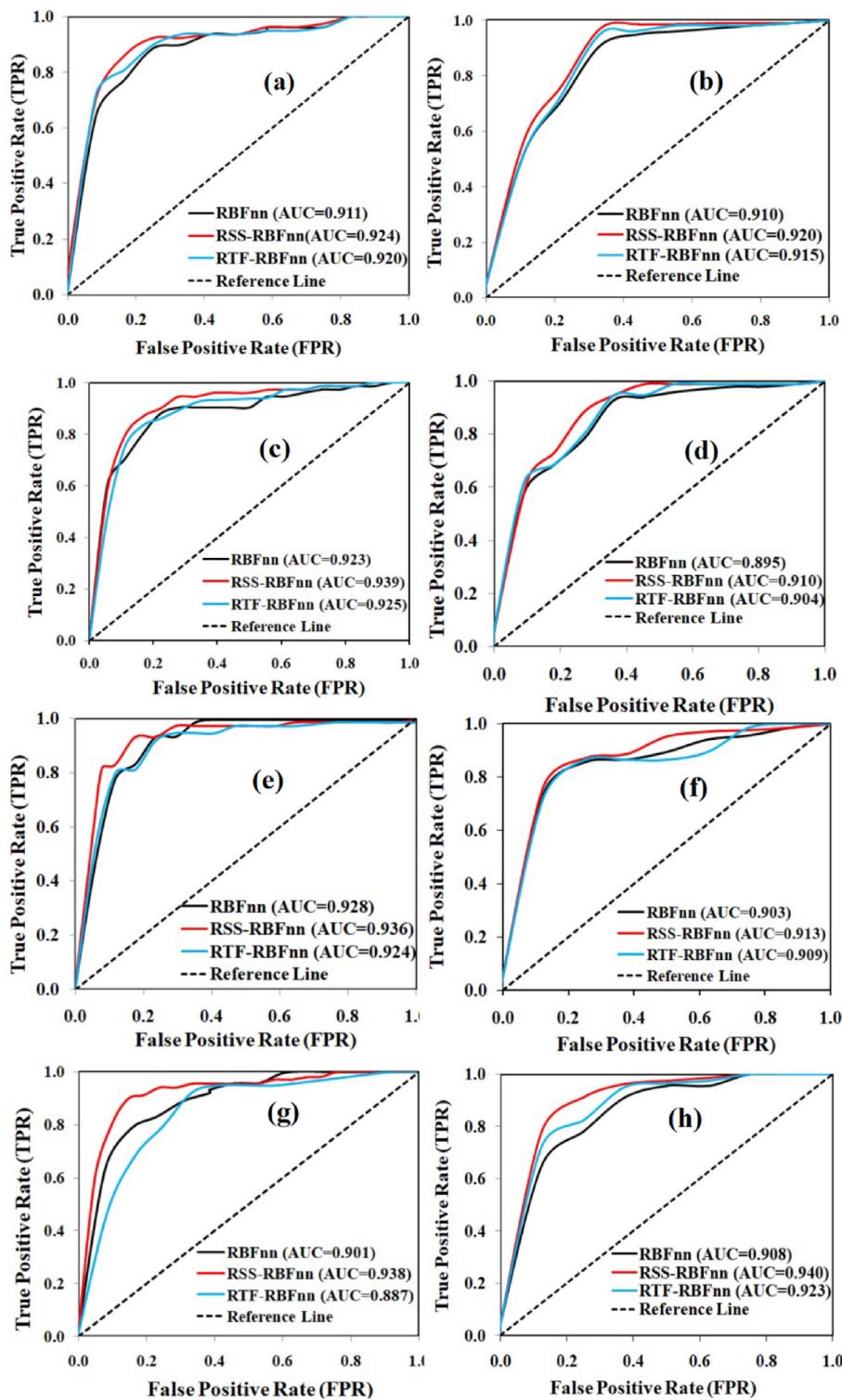


Fig. 13. Validation of results using area under the curve of the receiver operating characteristic: using training datasets (success rate curve) of (a) Fold-1, (c) Fold-2, (e) Fold-3 and (g) Fold-4; using validation datasets (prediction rate curve) of (b) Fold-1, (d) Fold-2, (f) Fold-3 and (h) Fold-4.

the rainfall map of the basin was generated. The average rainfall of the basin is 1326 mm for the last five years (Fig. 5f). The topographic wetness index (TWI) depicts the soil water content and saturation level. TWI was calculated using Eq. (1) where it considered specific catchment area (A_s) and the slope factor (β) (Mohamedou et al., 2017) (Fig. 5j).

$$TWI = \ln\left(\frac{A_s}{\tan\beta}\right) \quad (1)$$

The Hinglo River networks were derived from open series topographical maps collected from SOI. In GIS setting, distance to stream map was prepared using Euclidean distance buffering method. The maximum distance to stream in this basin is 2.10 km (Fig. 5g). Length of overland flow was calculated using the following Eq. (2) as developed by Horton.

$$Lof = 1/2Dd \quad (2)$$

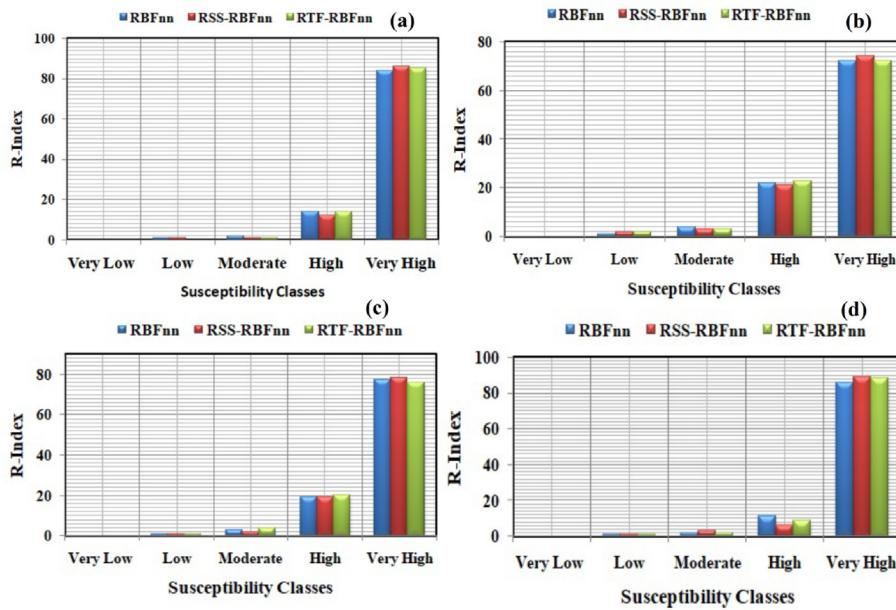


Fig. 14. Graphical representation of R-index of GESM models: (a) Fold-1, (b) Fold-2, (c) Fold-3 and (d) Fold-4.

Where, Dd is drainage density which is length of streams per sq. km. The spatial distribution of Lof ranges from 0 to 2.91 respectively (Fig. 5i). Using surface curve number (SCN) method (Soil Conservation Service, 1985), the annual runoff was estimated in GIS platform. The surface runoff value of this study ranges 667 mm–1337 mm respectively (Fig. 5h).

3.3.2. Soil physio-chemical factors

Physical and chemical properties of soil also play an important role in the degradation of soil and the initiation of gullies. The soil texture map of the study area was obtained from the National Bureau of Soil Survey and Land Use Planning. The basin is composed of various types of soil according to the USDA classification, including fine loamy mixed hapludalfs, and plutalfs, clay, clay loam, sandy, sandy loam, and loam respectively (Fig. 6a). The soil depth was prepared using the IDW method in ArcGIS environment (Fig. 6b). Soil chemical factors were measured from soil samples, including pH, boron, copper, manganese, zinc, iron, phosphorous, and potassium, organic carbon, nitrogen, and sulfur. A narrow stretch of 0.5 m × 0.5 m was chosen, and around 1 kg soil from the 0–20 cm depth from each sample was collected. A total of 106 soil samples were collected (Fig. 3). After collecting the soil samples analogous standard laboratory analytical methods were used to measure the soil chemical properties. The thematic maps of chemical parameters i.e. pH, OC, manganese, nitrogen, phosphorous, potassium, sulfur, and zinc were produced in the GIS setting (Fig. 6).

3.3.3. Lithological factors

A geological map was developed using the digitized method from the geological map no. 73 m (scale 1:50,000) obtained from the Geological Survey of India (GSI, 1985). The research area is covered by the granite-gneiss, barakar formation, ironstone shale, quartzite, and newer alluvium (Fig. 7a). The larger part of the basin is covered by the granite-gneiss. In the upper part of the basin is covered by the laterite soil and bared. As results most of the gullies were found in this part. The lineament of the present study was derived from the panchromatic band-8 of Landsat 8OLI/TIRS. The distance to the lineament map was prepared in a GIS platform using the Euclidean distance buffering method. The maximum distance to lineament in this study area is 1.65 km (Fig. 7b).

3.3.4. Land use/land cover (LULC)

Another important factor that widely regulates the formation of gullies is the LULC (Galang et al., 2007). The rangeland and barren lands are generally the most vulnerable to gully erosion due to the greatest impact of precipitation and comparatively higher surface runoff than the vegetation-covered areas. The vegetation-covered area can reduce the erosive effect of flowing water (Maugnard et al., 2014). In general, there is a negative association between the density of vegetation and the erosion rate (Collins et al., 2004). The supervised LULC map was obtained from Landsat 8OLI/TIRS in the present study. The five main land-use types were identified comprising agricultural land, water bodies, fallow land, vegetation-covered area, and settlement or built-up areas (Fig. 8). The maximum parts of the basin are covered by the agricultural land, followed by the vegetation, water bodies, built-up, and fallow land area.

3.4. Selection of GECFs

For modeling the gully erosion susceptibility in the Hinglo river basin first GECFs were chosen based on the previous literatures. Thereafter, effectiveness of these factors was assessed using multi-collinearity test and Information gain Ratio (IGR) method before using for training the models.

3.4.1. Multi-collinearity assessment

In multiple regression models, linearly related two or more explanatory variables are called multi-collinearity. To remove the highly correlated and inappropriate factors among the various geo-environmental factors for mapping various natural hazards, tolerance (TOI) and variance inflation factor (VIF) were widely used (Saha 2017; Roy and Saha 2019; Arabmaeri et al. 2020; Sardooi et al., 2019; Yu et al., 2015). The threshold values of VIF and TOL are <5 and >0.1, above these values factor has the collinearity problem (Saha et al., 2022).

3.4.2. Information Gain Ratio (IGR)

The predictive ability of the selected 24 conditioning factors of gully erosion was tested. Weak and inappropriate variables should have to be excluded from GECFs. The selection of effective factors can provide a

Table 3

R-index values of GESM models in different folds.

Models	GES class	pixels	% of pixels	No of gully pixels	% of gullies	R-Index
Fold 1						
RBFnn	Very Low	1292196	45.59	0	0	0
	Low	804506	28.38	3	2.5	1
	Moderate	385100	13.59	4	3.33	2
	High	161607	5.7	14	11.67	14
	Very	191281	6.75	99	82.5	84
	High					
RSS-RBFnn	Very Low	1456340	51.38	0	0	0
	Low	729494	25.73	3	2.5	1
	Moderate	320788	11.32	2	1.67	1
	High	141861	5	11	9.17	12
	Very	186207	6.57	104	86.67	86
	High					
RTF-RBFnn	Very Low	1283592	45.28	0	0	0
	Low	811787	28.64	2	1.67	0
	Moderate	385430	13.6	3	2.5	1
	High	163262	5.76	14	11.67	14
	Very	190619	6.72	101	84.17	85
	High					
Fold-2						
RBFnn	Very Low	1878946	66.28	2	1.67	0
	Low	448750	15.83	4	3.33	1
	Moderate	222720	7.86	7	5.83	4
	High	108106	3.81	17	14.17	22
	Very	176168	6.21	90	75	72
	High					
RSS-RBFnn	Very Low	2008342	70.85	3	2.5	0
	Low	354874	12.52	5	4.17	2
	Moderate	199334	7.03	5	4.17	3
	High	99612	3.51	15	12.5	21
	Very	172528	6.09	92	76.67	74
	High					
RTF-RBFnn	Very Low	2009335	70.88	2	1.67	0
	Low	353219	12.46	6	5	2
	Moderate	199224	7.03	4	3.33	3
	High	100053	3.53	17	14.17	23
	Very	172859	6.1	91	75.83	72
	High					
Fold-3						
RBFnn	Very Low	1580662	55.76	1	0.83	0
	Low	673786	23.77	3	2.5	1
	Moderate	282620	9.97	6	5	3
	High	120130	4.24	16	13.33	19
	Very	177492	6.26	94	78.33	77
	High					
RSS-RBFnn	Very Low	1876298	66.19	2	1.67	0
	Low	451397	15.92	2	1.67	1
	Moderate	232869	8.21	4	3.33	2
	High	101708	3.59	14	11.67	19
	Very	172418	6.08	98	81.67	78
	High					
RTF-RBFnn	Very Low	1902222	67.11	2	1.67	0
	Low	414553	14.62	2	1.67	1
	Moderate	237833	8.39	6	5	4
	High	106231	3.75	15	12.5	20
	Very	173852	6.13	95	79.17	76
	High					
Fold-4						
RBFnn	Very Low	1266604	44.68	1	0.83	0
	Low	822487	29.02	4	3.33	1
	Moderate	381128	13.45	4	3.33	2
	High	166130	5.86	11	9.17	11
	Very	198341	7	100	83.33	86
	High					
RSS-RBFnn	Very Low	1781099	62.83	4	3.33	0
	Low	510855	18.02	2	1.67	1
	Moderate	241694	8.53	4	3.33	3
	High	119137	4.2	5	4.17	6
	Very	181905	6.42	105	87.5	89
	High					
RTF-RBFnn	Very Low	1232958	43.5	1	0.83	0
	Low	865067	30.52	4	3.33	1

Table 3 (continued)

Models	GES class	pixels	% of pixels	No of gully pixels	% of gullies	R-Index
	Moderate	372634	13.15	5	4.17	2
	High	166461	5.87	9	7.5	9
	Very	197569	6.97	101	84.17	88
	High					

precise and proper prediction of model results (Ngo et al., 2018.). The predictive capacity of various data mining methods such as Fuzzy-Rough sets (Liu, 2007.), Relief-F (Park et al., 2019), and Information Gain Ratio (Tien Bui et al., 2019.) was applied by various researchers. Information Gain (IG) depends on the theory of information which uses to measure the significance of GECF variables. It was considered as the standard technique for quantifying the predictive capability of GECFs in data mining approaches (Svoray et al., 2012). However, IG has a natural error that tends to favour attributes with many possible values and can, therefore, lead to low predictability of the resulting models (Al-Abadi and Al-Najar 2020). To remove this problem, Quinlan (1993) developed an IGR method where a higher IGR value indicates a higher or more predictive ability of the factor. It has a particular static formula that is mentioned in the following equation to obtain the GECFs' IGR values in this study.

The training data S is composed of n input samples, n (L_i, S) is the number of samples in the training data S belonging to the class L_i (Gullies, non-gullies). The information (entropy) requires for classification S is calculated using Eq. (3).

$$\text{Info}(S) = - \sum_{i=1}^2 \frac{n(L_i, S)}{|S|} \log_2 \frac{n(L_i, S)}{|S|} \quad (3)$$

Using Eq. (4) the amount of information required to break S into subsets (S_1, S_2, \dots, S_m) with respect to gully determining factor A is calculated.

$$\text{Info}(S, A) = \sum_{j=1}^m \frac{|S_j|}{|S|} \text{Info}(S) \quad (4)$$

The IGR was determined using Eq. (5) for a given gully determining factors A .

$$\text{Information Gain Ratio}(S, A) = \frac{\text{Info}(S) - \text{Info}(S, A)}{\text{SplitInfo}(S, A)} \quad (5)$$

Where, SplitInfo represents potentially information generated by dividing the training data S into m subsets. SplitInfo is computed with Eq. (6).

$$\text{SplitInfo}(S, A) = - \sum_{j=1}^m \frac{|S_j|}{|S|} \log_2 \frac{|S_j|}{|S|} \quad (6)$$

3.5. Gully erosion susceptibility mapping models

3.5.1. Base classifier: Radius basis function neural network (RBFnn)

Radial Basis Function Neural Network (RBFnn) is a familiar nonlinear neural network. The RBFnn is characterized as a neural network with hidden layers. In addition, RBF consists of three layers, namely the input layer, the hidden layer, and the output layer. For each unit, the input layer converts data or vector elements into hidden layers. Each unit in the hidden layer then activates according to the associated RBFnn. The output layer eventually calculates a linear combination of the hidden unit activations. The performance of the RBFnn model learning for the input pattern x is as follows in the classification case (Yavari et al., 2019) (Eq. (7)):

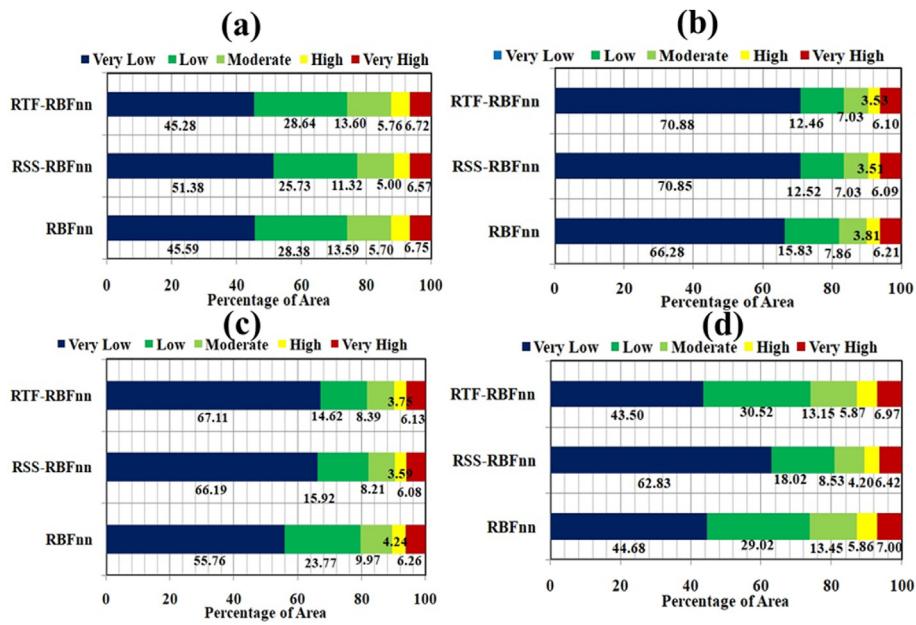


Fig. 15. Graphs showing the percentage distribution of GESMs: a. Fold-1, b. Fold-2, c. Fold-3, and d. Fold-4.

Table 4
Values of Mean Decrease Gini of GECFs using RF model.

Factors	Mean Decrease Gini	Factors	Mean Decrease Gini
Elevation	9.76	Zinc	9.136
Slope	6.876	Sulfur	6.169
PC	2.559	Nitrogen	7.843
CI	4.512	Manganese	10.3
TRI	7.302	Potassium	7.436
TWI	10.758	Phosphorus	9.46
Lof	11.715	OC	6.453
DR	4.45	pH	7.472
DL	4.645	Soil Depth	9.421
Rainfall	7.594	Soil Type	2.91
Geology	2.013	Surface Runoff	8.071
LULC	22.149		

$$f_i(x) = \sum_{k=1}^m w_{ki} \theta(\|x - \alpha_k\|) \quad (7)$$

where m and w_{ki} are the numbers and integrated weights between hidden and output layer while α_k and θ are RBFnn centers and Gaussian function. Random selection is used on the training data set to identify the key secret unit centers. In addition, the primary value of all variance parameters (s) in the network is set to the absolute Euclidean squared distance between any pair of cluster centers.

3.5.2. Hybrid ensemble models

3.5.2.1. Random sub space (RSS) classifier. The random sub-space classifier (RSS) is an important hybrid ensemble and a parallel learning algorithm. Ho (1998) introduced Random sub-space (RSS). For this algorithm, multiple decisions of the classifier were combined using the optimization of the sub-set. Such function space subsets are chosen randomly from the training classifiers (gully inventory datasets). Comparatively, the ensemble approach of random sub-space (RSS) is distinguished from the others by an ensemble algorithm since it consists of multiple sample numbers (Pham et al., 2020). The classification of the original feature space was done in the first stage by the q dimensional training of subsets L. In this study, the RBFnn as base classifier was applied in this algorithm for each of these subsets. In the end, the integration of the base classifier has extracted from the weighted majority

vote that is stated in the following (Shirzadi et al., 2017) Eq. (8).

$$\beta(x) = \operatorname{argmax}_{y \in \{-1, 1\}} \sum_b \delta_{\operatorname{sgn}(C^b(x)), y} \quad (8)$$

Where, Kronecker symbol is $\delta_{i,j}$. It stems from a generalization of the symbol Jacobi to all entities (Pham et al., 2019). $y \in \{-1, 1\}$ is considered as the presence of gullies and non-gullies classifier while combined classifiers is $C^b(x)$, $b = 1, 2, \dots, B$. The vote by sample majority helps to get the rule of final decision.

3.5.2.2. Rotation forest (RTF) classifier. Rodriguez et al. (2006) suggested rotation forest (RTF) which is one of the common hybrid ensemble techniques. It is regarded as an important technique for strengthening the weaker classifiers (Ozcift 2012). The RTF analyzes large multivariate datasets using the Principal Component Analysis (PCA) to reduce their dimensionality (Jolliffe 2002) and splitting the original training datasets into sub-sets which are then used to train the classifiers. It has a large application in the various branches and fields e.g. medical (Ozcift and Gulten, 2012) and remote sensing data classification (Xia et al., 2014; Kavzoglu and Colkesen, 2013) as the effective and powerful machine learning ensemble technique. RTF has also been used in hazard modeling particularly for landslide susceptibility modeling (Pham et al., 2017). Rodriguez et al. (2006) suggested rotation forest (RTF) which is one of the common hybrid ensemble techniques. It is regarded as an important technique for strengthening the weaker classifiers (Ozcift 2012). The RTF analyzes large multivariate datasets using the Principal Component Analysis (PCA) to reduce their dimensionality (Jolliffe 2002) and splitting the original training datasets into sub-sets which are then used to train the classifiers. It has a large application in the various branches and fields e.g. medical (Ozcift and Gulten, 2012) and remote sensing data classification (Xia et al., 2014; Kavzoglu and Colkesen, 2013) as an effective and powerful machine learning ensemble technique. RTF has also been used in hazard modeling particularly for landslide susceptibility modeling (Pham et al., 2017).

3.6. Measuring the relative importance of the GECFs by random forest (RF) model

RF machine learning technique is a revised form of classification and regression tree (CART). Breiman (2001) introduced RF. It can be used as

an efficient model for solving problems of classification and regression (Kuhnert et al., 2010). It is a significant overhaul of the bootstrap aggregation and belongs to the ensemble model family (Jaafari and Pourghasemi, 2019). The RF algorithm was performed in two stages: 1) the RF model uses a bootstrap sampling technique to set up training sets around 2/3 of all observations randomly and creates a tree for each training set (Youssef et al., 2016). Approximately 1/3 of all observations that are not used during the construction of the training set will be used during the bootstrap sampling as a test set and known as the out-of-bag sample (OOB) which can be used to determine misclassification errors and estimate anticipated predictive accuracy (Youssef et al., 2016). In fact, error in OOB can be used as a generalization error measure. 2) The nodes of each tree shall be divided according to the best explanatory variables selected from the explanatory variables input randomly selected subset. The random selection of explanatory variables at each node reduces the impact between any pair of trees in the forest; therefore, raising the forest error rate. Further details on the RF model can be found in Breiman (2001), Palczewska et al. (2014), and Oshiro et al. (2012). In our analysis, the package “random forest” in R 3.5.1 program was used to predict the relative importance of the GECFs.

3.7. Validation techniques

ROC curve, statistical methods such as MAE, RMSE, and relative gully density (R-Index) methods were applied for the justification and accuracy assessment of the GESMs.

3.7.1. ROC curve

Receiver Operating Characteristics (ROC) curve is a well-known approach used for performance analysis of models (Arabameri et al., 2020b; Gayen et al., 2020). The area under the curve (AUC) deals with the theoretical accuracy of the models (Youssef et al., 2016). This technique was, therefore, applied in the various natural hazards mapping, e.g. gully erosion susceptibility (Saha et al., 2020; Arabameri et al., 2020b; Debanshi and Pal, 2020); land subsidence susceptibility (Ghorbanzadeh et al., 2018), groundwater potential (Saha 2017), landslide susceptibility (Meena et al., 2019), rill-interrill susceptibility (Bosino et al., 2020) and also flood susceptibility (Arabameri et al., 2020a) mapping. The ROC curve has the specific cut-off values which classify model performance (Hembram et al., 2021). The AUC was designed by the true positive rate (sensitivity) against false positive rate (1-specificity) following equations (11)–(13).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (11)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (12)$$

$$AUC = \frac{(\sum TP + \sum TN)}{(P + N)} \quad (13)$$

Based on the four types of classification such as true positive (TP), true negative (TN), false positive (FP) and False Negative (FN) area under the curve of ROC (AUCROC) was calculated. The true positive (TP) and the True negative (TN) are the number of pixels properly classified by the models, while false positives (FP) and false negatives (FN) are the number of pixels incorrectly classified by the modes (Arabameri et al., 2020a). Researchers considered if the AUC value is less than 0.5 the performance is erroneous of models (He et al., 2019). The values of AUCROC were categorized into different categories by Fressard et al. (2014), e.g. AUCROC<0.7 as bad, AUCROC = 0.7 to 0.8 as average, AUCROC = 0.8 to 0.9 as good and AUCROC = 0.9 to 1 as excellent results.

3.7.2. Statistical techniques

For this analysis, statistical techniques such as MAE and RMSE were applied for the accuracy evaluation of the models. Garosi et al. (2019a,b); Saha et al. (2020) were used these methods for analyzing the performance of gully erosion modeling. MAE is defined as the sum of the difference between predicted values and actual values. The square root of MAE is known as RMSE. The calculation of the MAE and RMSE was done following equations (14) and (15).

$$MAE = \frac{1}{N} \sum_{i=1}^n |Y_{pred.} - Y_{act.}| \quad (14)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{pred.} - Y_{act.})^2}{N}} \quad (15)$$

Where, N is number of samples observation. $X_{pred.}$ donates the predicted and $X_{act.}$ indicates actual values. Can et al. (2005) set the cutoff value for RMSE is 0.5. RMSE<0.5 indicates good performance, while RMSE >0.5 represents the model as wrong.

3.7.3. R-index

For this analysis, the relative gully erosion density (R-index) method was used for evaluating the relationship between the GESM and temporal gully locations. In field survey the gully sample was collected using GPS. A total of 120 gullies in the study region are demarcated. The R-index was developed by Baeza and Corominas (2001). The R-index was calculated using Eq. (16).

$$R = (ni / Ni) / \sum (ni / Ni) \times 100 \quad (16)$$

Where, ni is the percentage of the area susceptible to gully in each class of GESM and Ni is the percentage of gully in each susceptibility class. If the R-index values are higher for very high and high susceptibility classes, performance of the model will be treated as good.

4. Results

4.1. Multi-collinearity evaluation

The results of the multi-collinearity test (Table 1) are showing that the GECFs have no collinearity problems because all the selected factors maintained the threshold limits of TOL and VIF values (>1 and < 5). In the present study, distance to lineament has the highest TOI value (0.968) and the lowest VIF value (1.033). Contrarily, phosphorus has the lowest TOI value (0.208) and the highest VIF value (4.808). The two or more variables in the present analysis are not strongly associated. Hence all GECFs are suitable for modeling gully erosion.

4.2. Application of Information Gain Ratio (IGR)

The IGR is an important machine learning method, applied to judge the efficacy of GECFs. According to the IGR method, all GECFs are suitable for GES modeling. Among these factors, the land use/land cover has the highest IGR value (AM = 0.101) followed by TWI, DL, CI, slope, PC, TRI, DR, surface runoff, soil type, geology, OC, Potassium, Lof, elevation, rainfall, Nitrogen, pH, Sulfur, Zinc, Manganese, Phosphorous, and soil depth respectively (Fig. 9).

4.3. Analysis of GESMs

GESMs were produced using RBFnn, RSS-RBFnn, and RTF-RBFnn models. Each fold has three GESMs like Fold-1 has RBFnn and RSS-RBFnn and RTF-RBFnn models and a total of 12 GESMs were

generated in GIS settings. The GESMs were then categorized into five groups with the help of Jenk's natural break classification system, namely very low, low, medium, high, and very high classes of susceptibility. The areal distributions of GESMs are shown in Figs. 10–12. The areas of the basin covered by the very high susceptibility class in RBFnn RTF-RBFnn RSS-RBFnn are 6.75%, 6.72% and 6.57% for Fold-1. In the case of fold-2, very high susceptibility covered 6.21% (RBFnn), 6.09% (RSS-RBFnn) and 6.10% (RTF-RBFnn) areas of the basin and on the other hand, 6.23% (RBFnn), 6.05% (RSS-RBFnn), and 6.13% (RTF-RBFnn) areas are occupied by the very high susceptibility class for Fold-3. For Fold-4 the very high susceptibility class covered 7% (RBFnn), 6.42% (RSS-RBFnn), and 6.97% (RTF-RBFnn) of the catchment respectively. GESMs results showed that the very high susceptibility region of gully erosion is located in the catchment's northwest and middle parts due to the concentration of lateritic soil and the presence of barren land. However, a very small portion of the basin is covered by the very high susceptibility class.

4.4. Validation of GESMs

ROC curve, statistical methods, as well as relative gully erosion density (R-index) methods were used for judging the model accuracy. In the present analysis, AUC was obtained using training and validation gully erosion datasets. The curve produced using training gully locations is known as the succession rate curve (SRC), and validation data is known as the prediction curve rate (PRC). The ROC results are shown in Table 2. AUC values for GESMs are calculated using the succession curve rate (SCR) and the prediction curve rate (PRC). The RSS-RBFnn hybrid ensemble was achieved the highest AUC values for both SCR and PRC in the case of all the folds (Fig. 13 and Table 2) followed by the RBFnn and RTF-RBFnn respectively. After combining the hybrid meta classifier (RSS) with the RBFnn the level of accuracy of the RBFnn model has increased for all the folds. Although, AUC of the success rate of the RTF-RBFnn model for fold 3 and fold 4 has decreased. Besides, these the level of accuracy has also increased for RTF-RBFnn model.

MAE and RMSE were also calculated for validating the produced models of four folds datasets. The reliable and important models have the RMSE and MAE values < 0.5. The MAE values of RBFnn, RSS-RBFnn and RTF-RBFnn models for the training datasets are 0.070, 0.041, 0.058 in Fold-1, 0.047, 0.042, 0.042 in Fold-2, 0.030, 0.025, 0.025 in Fold-3, 0.053, 0.031, 0.041 in Fold-4 and for the validation datasets are 0.080, 0.071, 0.075 in Fold-1, 0.040, 0.039, 0.040 in Fold-2, and 0.090, 0.075, 0.080 in Fold-3, and 0.093, 0.082, 0.083 in Fold-4 respectively (Table 2). The RSS-RBFnn model has the lowest value of RMSE for both the datasets (training and validation) in all the folds and followed by the RTF-RBFnn and RBFnn respectively. In the current study, the MAE and RMSE values of all the models for both training and validation datasets are below the cut off value (<0.5). Therefore, the performances of all the models are excellent.

The results of the R-index are shown in Fig. 14 and Table 3. The very high gully erosion susceptibility class of RBFnn, RSS-RBFnn and RTF-RBFnn achieved the highest R-index values i.e. 84, 86, 85 in Fold-1, 72, 74, 72 in Fold-2, 77, 78, 76 in Fold-3, and 86, 89, 88 in Fold-4 (Table 3). If the R-index values of models are increased from very low to very high susceptibility, the models will more accurate and appropriate for gully erosion susceptibility assessment. In the current analysis, all models' R-Index values increased from very low to very high susceptibility class (Fig. 14). Hence the R-index also justified the selected models as excellent for mapping the gully erosion susceptibility.

5. Discussions

In this analysis, the construction of GESM was done using the RBFnn and its ensemble with RSS and RTF meta classifiers. K-fold cross-validation (CV) method for GESM modeling was implemented in the present analysis and the inventory map of the gully was then categorized

into four-fold and a total of 12 GESMs were made. In the current research, to obtain the right and optimal predictive performance of models, the machine learning-based ensemble techniques were applied for the gully erosion modeling. RBFnn and its ensemble with RSS and RT ensemble Meta classifiers were used for the spatial mapping of gully erosion susceptibility (GES). In the present study, the accuracy of the base classifier RBFnn has increased by nearly 2%–3% after ensemble with the hybrid meta classifiers as like the works done by Pham et al. (2017, 2019); Chen et al. (2017). GESMs were assessed using the methods of ROC curve, RMSE, MAE, and R-index. Garosi et al. (2019); Arabameri et al. (2020b); Chen et al. (2019); Pham et al. (2017; 2019) used the ROC, RMSE and MAE curves to assess the accuracy of different hazard models. Meena et al. (2019) used the R-index for assessment of the susceptibility to landslides. This similar method was applied in this research to measure the accuracy of GES models and to select the best model. The R-index method helps to detect the density of the pixels of gullies in susceptibility classes.

The GESMs were produced using the CV approach and evaluated the accuracy using different methods such as AUC, MAE, RMSE, and R-index. However, a single statistical method is not adequate to validate the models because validation points were randomly selected, validation samples are small in number and one validation method might have high error (Garosi et al., 2018). AUC of ROC, RMSE, and MAE were used for the validation of the models. A higher AUC value indicates good efficiency and excellent capability of spatial prediction (Hosseinalizadeh et al., 2019). In this study, the results of AUC, R-index, MAE, and RMSE revealed that the used method are excellent for modeling gully erosion. The machine learning-based ensemble models deliver better and perfect results than the single model (Arabameri et al., 2020b; Hembram et al., 2021). The integrated models i.e. RSS-RBFnn and RTF-RBFnn were shown better results than RBFnn. The ensemble model, i.e. RSS-RBFnn, has the highest and best predictive output in the current analysis, followed by the RBFnn and RTF-RBFnn models. RSS-RBFnn model is showing nearly 6.5% of the study area has very high gully erosion susceptibility (Fig. 15). The upper part of the basin has the very high susceptibility of gully erosion because most of the parts of the upper catchment area are bare surface and have laterite soil which is favorable for gully formation. According to the results of the study length of overland flow or surface runoff has significant contribution in producing the gullies in the study area and same kind of results can be found in the work of Conforti et al. (2011); Conoscenti et al. (2018); Comino et al. (2016). For reducing the rate of gully erosion in the upper and middle catchment areas of the basin the planners should take strong strategies of afforestation and check dam formation. However, this work will attract the researchers for studying the gully erosion in other regions and provide valuable information to the planners for land use management in this study area.

5.1. Factor significance analysis

In the study for GESM modeling a large number of geo-environmental factors such as topographical, hydrological, lithological, soil physio-chemical parameters were used. The chemical properties retain the cohesive, soil health and soil particle strength and in turn determine the possibility of gully formation (Hosseinalizadeh et al., 2019; AsghariSaraskanroud et al., 2017). The combinations of the physio-chemical topographical, hydrological, lithological, and soil parameters for mapping gully erosion are more precise and effective. Romer and Ferentinou (2016) noted that a large number of databases can enhance the model's performance in achieving the best results and perfect prediction. Hosseinalizadeh et al. (2019a,b) used 24 geo-environmental factors like topographical, hydrological, lithological, soil physio-chemical parameters to map the susceptibility of the gully head cut. The RF model was used to determine the importance of factors. The results of the model are shown in Table 4. The outcome of RF model showed that the LULC (RF = 22.149) has the highest contribution in making the area susceptible to

gully erosion, followed by Lof, elevation, slope, plan curvature, convergence index, terrain ruggedness index, topographic wetness index, length of overland flow, distance to the river, distance to lineament, rainfall, surface runoff, zinc, sulfur, nitrogen, manganese, potassium, phosphorus, organic carbon, pH, soil depth, soil texture, and geology respectively (Table 4).

6. Conclusions

Geomorphic threats such as soil and gully erosion, flooding, flash flood, and land subsidence have been increased because of increasing human interference with nature. Among the various natural hazards, gully erosion is the most important which is creating a huge problem in the agricultural sector and also reducing the economic growth in a country like India where a large portion of the GDP comes from this sector. The only way for managing the gully erosion is the proper mapping of gully erosion susceptible area through sound methods. The GESM is the most powerful tool for planning land use and for environmental management. The selection of effective GECFs is the preconditions for GES mapping and gully erosion management. In the present analysis 24 geo-environmental variables used for modeling GES using the K-fold CV approach. GESMs were constructed in the GIS platform using the RBFnn, RSS, RTF models, and these maps were evaluated by the ROC curve, MAE, RMSE, and relative gully density (R-index) method. All models have perfect to excellent performance for evaluating gully erosion, according to the statistical procedures used for the validation. Among the GES models, the RSS-RBFnn has the best predictive performance for gully erosion modelling. The results of the GESMs showed that the areas prone to gully erosion are located in the catchment's northwest and middle portions, because of the presence of laterite soil, high drainage density, maximum concentration of surface runoff and bare surface. In the study area for reducing the soil erosion in the upper part of the basin immediate suitable measures such as afforestation, agro-forestry, check-dam construction and proper land use management should be taken. Major drawbacks of the present work are preparation of data layers from small and medium scale maps. Among the selected factors specially, soil micronutrient are very difficult to be related via causal effect to the gully development because these factors indirectly involved in inducing gully through changing soil conditioning. However, the produced GESMs will help in implementing sustainable management plans to decision-makers, engineers, and land-use planners.

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.

Funding

No funding was received for this work.

Credit author statement

Jagabandhu Roy: Methodology, Format analysis, Investigation, Writing original draft Software, preparation; **Sunil Saha:** Methodology, Format analysis, Investigation, writing original draft preparation, Writing review and editing.

Acknowledgments

Authors would like to thanks the inhabitants of Basin because they have helped a lot during our field visit. At last, authors would like to acknowledge all of the agencies and individuals specially, Survey of India (SOI), Geological Survey of India (GSI) and USGS for providing the maps and data required for the study.

References

- Al-Abadi, A.M., Al-Najar, N.A., 2020. Comparative assessment of bivariate, multivariate and machine learning models for mapping flood proneness. *Nat. Hazards* 100 (2), 461–491. <https://doi.org/10.1007/s11069-019-03821-y>.
- Arabameri, A., Saha, S., Chen, W., Roy, J., Pradhan, B., Bui, D.T., 2020a. Flash flood susceptibility modelling using functional tree and hybrid ensemble techniques. *J. Hydrol.* 125007. <https://doi.org/10.1016/j.jhydrol.2020.125007>.
- Arabameri, A., Saha, S., Roy, J., Tiefenbacher, J.P., Cerdá, A., Biggs, T., Pradhan, B., Ngo, P.T.T., Collins, A.L., 2020b. A novel ensemble computational intelligence approach for the spatial prediction of land subsidence susceptibility. *Sci. Total Environ.*, 138595. <https://doi.org/10.1016/j.scitotenv.2020.138595>.
- AsghariSaraskanroud, S., Zeinali, B., Mohammadnejad, V., 2017. Analysis physical and chemical properties of soil and morphometric impacts on gully erosion. *Desert* 22 (2), 157–166. <https://doi.org/10.22059/jdesert.2017.64179>.
- Baeza, C., Corominas, J., 2001. Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. *Earth Surf. Process. Landforms* 26, 1251–1263. <https://doi.org/10.1002/esp.263>.
- Battaglia, S., Leoni, L., Sartori, F., 2002. Mineralogical and grain size composition of clays developing calanchi and biancane erosional landforms. *Geomorphology* 49, 153–170.
- Bosino, A., Giordani, P., Quénéhervé, G., Maerker, M., 2020. Assessment of calanchi and rill-interrill erosion susceptibilities using terrain analysis and geostochastics: a case study in the Oltrepo Pavese, Northern Apennines, Italy. *Earth Surf. Process. Landforms* 45 (12), 3025–3041.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Brice, J.C., 1966. Erosion and Sedimentation in the Loessian-Mantled Great Plains, Madison Creek Drainage Basin, Nebraska. *Geological Survey Prof. paper*, pp. 5255–5339.
- Burian, L.A., Mitusov, V., Poesen, J., 2015. Relationships of attributes of gullies with morphometric variables. *Geomorphometry* 1, 111–114.
- Casali, J., López, J.J., Giráldez, J.V., 1999. Ephemeral gully erosion in southern Navarra (Spain). *Catena* 36 (1–2), 65–84. [https://doi.org/10.1016/S0341-8162\(99\)00013-2](https://doi.org/10.1016/S0341-8162(99)00013-2).
- Chen, W., Pourghasemi, H.R., Kornejehad, A., Zhang, N., 2017. Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma* 314–327. <https://doi.org/10.1016/j.geoderma.2017.06.020>.
- Chen, W., Hong, H., Li, S., Shahabi, H., Wang, Y., Wang, X., Ahmad, B.B., 2019. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J. Hydrol.* 575, 864–873. <https://doi.org/10.1016/j.jhydrol.2019.05.089>.
- Choubin, B., Rahmati, O., Tahmasebipour, N., Feizizadeh, B., Pourghasemi, H.R., 2019. Application of fuzzy analytical network process model for analyzing the gully erosion susceptibility. In: *Natural Hazards Gis-Based Spatial Modeling Using Data Mining Techniques*. Springer, Cham, pp. 105–125. https://doi.org/10.1007/978-3-319-73383-8_5.
- Claps, P., Rossi, F., 1994. Conceptually-based univariate stochastic modelling of river runoff. *Excerpta of the Italian Contributions to the Field of Hydraulic Engineering* 8, 95–134.
- Collins, D.B.G., Bras, R.L., Tucker, G.E., 2004. Modeling the effects of vegetation-erosion coupling on landscape evolution. *J. Geophys. Res.: Earth Surf.* 109. <https://doi.org/10.1029/2003JF000028> (F3).
- Comino, J.R., Iserloh, T., Lassu, T., Cerdá, A., Keestra, S., Prosdocimi, M., Brings, C., Marzen, M., Ramos, M., Senciales, J., 2016. Quantitative comparison of initial soil erosion processes and runoff generation in Spanish and German vineyards. *Sci. Total Environ.* 565, 1165–1174. <https://doi.org/10.1016/j.scitotenv.2016.05.163>.
- 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. <https://doi.org/10.1007/s11069-010-9598-2>.
- Conoscenti, C., Agnese, V., Cama, M., Alamaru Caraballo-Arias, N., 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. <https://doi.org/10.1002/lrd.2772>.
- Debanshi, S., Pal, S., 2020. Assessing gully erosion susceptibility in Mayurakshi river basin of eastern India. *Environ. Dev. Sustain.* 22 (2), 883–914. <https://doi.org/10.1007/s10668-018-0224-x>.
- Dondofema, F., Murwira, A., Mhizha, A., 2008. *Identifying Gullies and Determining Their Relationships with Environmental Factors Using GIS in the Zhulube Meso-Catchment. Water and Sustainable Development for Improved Livelihoods, Johannesburg, South Africa*.
- Evans, I.S., Cox, N.J., 1999. Relations between land surface properties: altitude, slope and curvature. In: *Process Modelling and Landform Evolution*. Springer, Berlin, Heidelberg, pp. 13–45. <https://doi.org/10.1007/BFb0009718>.
- Fressard, M., Thiery, Y., Maquaire, O., 2014. Which data for quantitative landslide susceptibility mapping at operational scale? Case study of the Pays d'Auge plateau hillslopes (Normandy, France). *Nat. Hazards Earth Syst. Sci.* 14 (3), 569–588.
- Galang, M.A., Markewitz, D., Morris, L.A., Bussell, P., 2007. Land use change and gully erosion in the Piedmont region of South Carolina. *Int. Soil Water Conserv.* 62 (3), 122–129.
- García-Ruiz, J.M., 2010. The effects of land uses on soil erosion in Spain: a review. *Catena* 81 (1), 1–11. <https://doi.org/10.1016/j.catena.2010.01.001>.
- Garosi, Y., Shekhabadi, 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. <https://doi.org/10.1016/j.geoderma.2018.05.027>.
- Garosi, Y., Shekhabadi, M., Conoscenti, C., Pourghasemi, H.R., Van Oost, K., 2019a. Assessing the performance of GIS-based machine learning models with different

- accuracy measures for determining susceptibility to gully erosion. *Sci. Total Environ.* 664, 1117–1132. <https://doi.org/10.1016/j.scitotenv.2019.02.093>.
- Garosi, Y., Shekhabadi, M., Conoscenti, C., Pourghasemi, H.R., Van Oost, K., 2019b. Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. *Sci. Total Environ.* 664, 1117–1132. <https://doi.org/10.1016/j.scitotenv.2019.02.093>.
- Gayen, A., Saha, S., 2017. Application of weights-of-evidence (WoE) and evidential belief function (EBF) models for the delineation of soil erosion vulnerable zones: a study on Pathro river basin, Jharkhand, India. *Model. Earth Syst. Environ.* 3 (3), 1123–1139. <https://doi.org/10.1007/s40808-017-0362-4>.
- Gayen, A., Pourghasemi, H.R., Saha, S., Keessstra, 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>.
- 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. https://doi.org/10.1007/978-3-030-23243-6_3.
- Ghorbanzadeh, O., Rostamzadeh, H., Blaschke, T., Gholaminia, K., Aryal, J., 2018. A new GIS-based data mining technique using an adaptive neuro-fuzzy inference system (ANFIS) and k-fold cross-validation approach for land subsidence susceptibility mapping. *Nat. Hazards* 94 (2), 497–517. <https://doi.org/10.1007/s11069-018-3449-y>.
- Ghosh, K.G., Shah, S., 2015. Identification of soil erosion susceptible areas in Hinglo river basin, Eastern India based on geo-statistics. *Uni. J. Environ. Res. Technol.* 5 (3), 152–164.
- GSI, 1985. Geological Quadrangle Map, Bardhaman Quadrangle (73M), West Bengal Bihar Geological Survey of India, Printing Div. Hyderabad, Govt. of India.
- Guerra, A.J.T., Fullen, M.A., Bezerra, J.F.R., Jorge, M.D.C.O., 2018. Gully erosion and land degradation in Brazil: a case study from São Luís municipality, Maranhão State. In: *Ravine Lands: Greening for Livelihood and Environmental Security*. Springer, Singapore, pp. 195–216. https://doi.org/10.1007/978-981-10-8043-2_8.
- He, Q., Shahabi, H., Shirzadi, A., Li, S., Chen, W., Wang, N., Chai, H., Bian, H., Ma, J., Chen, Y., Wang, X., 2019. Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. *Sci. Total Environ.* 663, 1–15. <https://doi.org/10.1016/j.scitotenv.2019.01.329>.
- Hembram, T.K., Paul, G.C., Saha, S., 2020. Modelling of gully erosion risk using new ensemble of conditional probability and index of entropy in Jainti River basin of Chotanagpur Plateau Fringe Area, India. *Appl. Geomat.* 1–24. <https://doi.org/10.1007/s12518-020-00301-y>.
- Hembram, T.K., Saha, S., Pradhan, B., Abdul Maulud, K.N., Alamri, A.M., 2021. Robustness analysis of machine learning classifiers in predicting spatial gully erosion susceptibility with altered training samples. *Geomatics, Nat. Hazards Risk* 12 (1), 794–828.
- Ho, T.K., 1998, August. Nearest neighbors in random subspaces. In: *Joint IAPR international workshops on statistical techniques in pattern recognition (SPR) and structural and syntactic pattern recognition (SSPR)*. Springer, Berlin, Heidelberg, pp. 640–648.
- Hosseinalizadeh, M., Karimnejad, N., Chen, W., Pourghasemi, H.R., Alinejad, M., Behbahani, A.M., Tiefenbacher, J.P., 2019a. 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. <https://doi.org/10.1016/j.geomorph.2019.01.006>.
- Hosseinalizadeh, M., Karimnejad, N., Chen, W., Pourghasemi, H.R., Alinejad, M., Behbahani, A.M., Tiefenbacher, J.P., 2019b. Gully headcut susceptibility modeling using functional trees, naïve Bayes tree, and random forest models. *Geoderma* 342, 1–11. <https://doi.org/10.1016/j.geoderma.2019.01.050>.
- Jaaafari, A., Pourghasemi, H.R., 2019. Factors influencing regional-scale wildfire probability in Iran: an application of random forest and support vector machine. In: *Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Elsevier, pp. 607–619. <https://doi.org/10.1016/B978-0-12-815226-3.00028-4>.
- Jolliffe, I., 2002. *Principal Component Analysis*. Wiley Online Library.
- Karuma, A.N., Gachene, K., Charles, K., Msanya, B.M., Mtakwa, P.W., Amuri, N., Gicheru, P.T., 2014. Soil Morphology, Physico-Chemical Properties and Classification of Typical Soils of Mwala District, Kenya. <http://hdl.handle.net/123456789/750>.
- Kavzoglu, T., Colkesen, I., 2013. An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. *Int. J. Rem. Sens.* 34, 4224–4241.
- Kertész, Á., 2009. Environmental conditions of gully erosion in Hungary. *Hungarian Geographical Bulletin* 58 (2), 79–89.
- 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 (5), 493–509. <https://doi.org/10.1002/env.999>.
- Lal, R., 2001. Soil degradation by erosion. *Land Degrad. Dev.* 12 (6), 519–539. <https://doi.org/10.1002/ldr.472>.
- Liu, B., 2007. Uncertainty theory. In: *Uncertainty Theory*. Springer, Berlin, Heidelberg, pp. 205–234. https://doi.org/10.1007/978-3-540-73165-8_5.
- Mararakanye, N., 2016. *A Comparative Study of Gully Erosion Contributing Factors in Two Tertiary Catchments in Mpumalanga South Africa* (Doctoral Dissertation. University of Pretoria).
- Maugnard, A., Van Dyck, S., Bielders, C.L., 2014. Assessing the regional and temporal variability of the topographic threshold for ephemeral gully initiation using quantile regression in Wallonia (Belgium). *Geomorphology* 206, 165–177. <https://doi.org/10.1016/j.geomorph.2013.10.007>.
- Meena, S.R., Ghorbanzadeh, O., Blaschke, T., 2019. A comparative study of statistics-based landslide susceptibility models: a case study of the region affected by the gorkha earthquake in Nepal. *ISPRS Int. J. Geo-Inf.* 8 (2), 94. <https://doi.org/10.3390/ijgi8020094>.
- Mohamedou, C., Tokola, T., Eerikäinen, K., 2017. LiDAR-based TWI and terrain attributes in improving parametric predictor for tree growth in southeast Finland. *Int. J. Appl. Earth Obs. Geoinf.* 62, 183–191. <https://doi.org/10.1016/j.jag.2017.06.004>.
- Nandi, A., Luffman, I., 2012. Erosion related changes to physicochemical properties of ultisols distributed on calcareous sedimentary rocks. *J. Sustain. Dev.* 5, 52–68.
- NATMO, 2001. National Atlas and Thematic Mapping Organization, District Planning Map Series (DST), Digital Mapping and Printed Division. Kolkata.
- Ngo, P.T.T., Hoang, N.D., Pradhan, B., Nguyen, Q.K., Tran, X.T., Nguyen, Q.M., Nguyen, V.N., Samui, P., Tien Bui, D., 2018. A novel hybrid swarm optimized multilayer neural network for spatial prediction of flash floods in tropical areas using sentinel-1 SAR imagery and geospatial data. *Sensors* 18 (11), 3704. <https://doi.org/10.3390/s18113704>.
- Ogbonna, J.U., Alozie, M., Nkemdirim, V., Eze, M.U., 2011. GIS Analysis for mapping gully erosion impacts on the geo-formation of the Old Imo State, Nigeria. *ABSU Journal of Environment, Science and Technology* 1, 48–61.
- Oshiro, T.M., Perez, P.S., Baranauskas, J.A., 2012. How many trees in a random forest?. In: *International Workshop on Machine Learning and Data Mining in Pattern Recognition*. Springer, Berlin, Heidelberg, pp. 154–168. https://doi.org/10.1007/978-3-642-31537-4_13.
- Ozciit, A., 2012. SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease. *J. Med. Syst.* 36, 2141–2147.
- Ozciit, A., Gulten, A., 2012. A robust multi-class feature selection strategy based on rotation forest ensemble algorithm for diagnosis of Erythema-Squamous diseases. *J. Med. Syst.* 36, 941–949.
- Palczewska, A., Palczewski, J., Robinson, R.M., Neagu, D., 2014. Interpreting random forest classification models using a feature contribution method. In: *Integration of Reusable Systems*. Springer, Cham, pp. 193–218. https://doi.org/10.1007/978-3-319-04717-1_9.
- Park, S., Hamm, S.Y., Kim, J., 2019. Performance evaluation of the gis-based data-mining techniques decision tree, random forest, and rotation forest for landslide susceptibility modeling. *Sustainability* 11 (20), 5659. <https://doi.org/10.3390/su11205659>.
- Pham, B.T., Bui, D.T., Prakash, I., Dholakia, M.B., 2017. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* 149, 52–63. <https://doi.org/10.1016/j.catena.2016.09.007>.
- Pham, B.T., Prakash, I., Singh, S.K., Shirzadi, A., Shahabi, H., Bui, D.T., 2019. Landslide Susceptibility Modeling Using Reduced Error Pruning Trees and Different Ensemble Techniques: Hybrid Machine Learning Approaches. *Catena* 175, pp. 203–218. <https://doi.org/10.1016/j.catena.2018.12.018>.
- Pham, B.T., Phong, T.V., Nguyen-Thoi, T., Parial, K., Singh, S., Ly, H.B., Nguyen, K.T., Ho, L.S., Le, H.V., Prakash, I., 2020. Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. *Geocarto Int.* 1–23. <https://doi.org/10.1080/10106049.2020.1737972>.
- Quinlan, J.R., 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo.
- Rodriguez, J.J., Kuncheva, L.I., Alonso, C.J., 2006. Rotation forest: a new classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 1619–1630.
- Romer, C., Ferentino, M., 2016. Shallow landslide susceptibility assessment in a semiarid environment—a Quaternary catchment of KwaZulu-Natal, South Africa. *Engineering Geology* 201, 29–44. <https://doi.org/10.1016/j.enggeo.2015.12.013>.
- Roy, J., Saha, S., 2019. GIS-Based gully erosion susceptibility evaluation using frequency ratio, cosine amplitude and logistic regression ensembled with fuzzy logic in Hinglo river basin, India. *Remote Sens. Appl. Soc. Environ.* 15, 100247. <https://doi.org/10.1016/j.rsase.2019.100247>.
- Roy, J., Saha, S., Arabameri, A., Blaschke, T., Bui, D.T., 2019. A novel ensemble approach for landslide susceptibility mapping (LSM) in Darjeeling and Kalimpong Districts, West Bengal, India. *Rem. Sens.* 11 (23), 2866. <https://doi.org/10.3390/rs11232866>.
- Saha, S., 2017. Groundwater potential mapping using analytical hierarchical process: a study on Md. Bazar Block of Birbhum District, West Bengal. *Spat. Inf. Res.* 25 (4), 615–626. <https://doi.org/10.1007/s41324-017-0127-1>.
- Saha, S., Roy, J., Arabameri, A., Blaschke, T., Tien Bui, D., 2020. Machine learning-based gully erosion susceptibility mapping: a case study of eastern India. *Sensors* 20 (5), 1313. <https://doi.org/10.3390/s20051313>.
- Saha, S., Saha, A., Hembra, T.K., Mandal, K., Sarkar, R., Bhardwaj, D., 2022. Prediction of spatial landslide susceptibility applying the novel ensembles of CNN, GLM and random forest in the Indian Himalayan region. *Stoch. Environ. Res. Risk Assess.* 1–20.
- Sardooi, E.R., Azareh, A., Choubin, B., Barkhori, S., Singh, V.P., Shamshirband, S., 2019. Applying the remotely sensed data to identify homogeneous regions of watersheds using a pixel-based classification approach. *Appl. Geogr.* 111, 102071. <https://doi.org/10.1016/j.apgeog.2019.102071>.
- Sahrivar, A., Boon Sung, C., Jusop, S., Rahim, A., Soufi, M., 2012. Roles of SAR and EC in gully erosion development (A case study of KohgiloyehBoyerahmad province, Iran). *Journal of Research in Agricultural Science* 8, 1–12.
- Shirzadi, A., Bui, D.T., Pham, B.T., Solaimani, K., Chapi, K., Kavian, A., Shahabi, H., Revhaug, I., 2017. Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ. Earth Sci.* 76 (2), 60. <https://doi.org/10.1007/s12665-016-6374-y>.
- Shit, P.K., Bhunia, G.S., Maiti, R., 2013. Assessment of factors affecting ephemeral gully development in Badland topography: a case study at Garbheta Badland (pashchim Medinipur. *Int. J. Geosci.* 4, 461. <https://doi.org/10.4236/ijg.2013.420403>, 02.
- Shit, P.K., Nandi, A.S., Bhunia, G.S., 2015. Soil erosion risk mapping using RUSLE model on Jhargram sub-division at West Bengal in India. *Model. Earth Syst. Environ.* 1 (3), 28. <https://doi.org/10.1007/s40808-015-0032-3>.

- Singh, O., Singh, J., 2018. Soil erosion susceptibility assessment of the lower Himachal Himalayan watershed. *J. Geol. Soc. India* 92 (2), 157–165. <https://doi.org/10.1007/s12594-018-0975-x>.
- Soil Conservation Service SCS, 1985. Hydrology, National Engineering Handbook. Soil Conservation Service, Washington, D. C. USDA.**
- 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 Surface Processes and Landforms* 37 (6), 607–619. <https://doi.org/10.1002/esp.2273>.
- Taheri, K., Shahabi, H., Chapi, K., Shirzadi, A., Gutiérrez, F., Khosravi, K., 2019. Sinkhole susceptibility mapping: a comparison between Bayes-based machine learning algorithms. *Land Degrad. Dev.* 30 (7), 730–745. <https://doi.org/10.1002/lrd.3255>.
- Tien Bui, D., Shirzadi, A., Shahabi, H., Chapi, K., Omidvar, E., Pham, B.T., Talebpour Asl, D., Khaledian, H., Pradhan, B., Panahi, M., Bin Ahmad, B., 2019. A novel ensemble artificial intelligence approach for gully erosion mapping in a semi-arid watershed (Iran). *Sensors* 19 (11), 2444. <https://doi.org/10.3390/s19112444>.
- Vannmaercke, M., Poesen, J., Van Mele, B., Demuzere, M., Bruynseels, A., Golosov, V., Bezerra, J.F.R., Bolysov, S., Dvinskikh, A., Frankl, A., Fuseina, Y., 2016. How fast do gully headcuts retreat? *Earth Sci. Rev.* 154, 336–355. <https://doi.org/10.1016/j.earscirev.2016.01.009>.
- Xia, J., Du, P., He, X., Chanussot, J., 2014. Hyperspectral remote sensing image classification based on rotation forest. *Geosci. Rem. Sens. Lett. IEEE* 11, 239–243.
- Yavari, H., Pahlavani, P., Bigdeli, B., 2019. Landslide Hazard Mapping Using a Radial Basis Function Neural Network Model: a Case Study in Semiroom, Isfahan, Iran. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* 13 (5), 839–856. <https://doi.org/10.1007/s10346-015-0614-1>.
- Yu, H., Jiang, S., Land, K.C., 2015. Multicollinearity in hierarchical linear models. *Soc. Sci. Res.* 53, 118–136. <https://doi.org/10.1016/j.ssr.2015.04.008>.
- Zerihun, M., Mohammedyasin, M.S., Sewnet, D., Adem, A.A., Lakew, M., 2018. Assessment of soil erosion using RUSLE, GIS and remote sensing in NW Ethiopia. *Geoderma regional* 83–90. <https://doi.org/10.1016/j.geodrs.2018.01.002>, 12.
- Zhang, S., Li, F., Li, T., Yang, J., Bu, K., Chang, L., Wang, W., Yan, Y., 2015. Remote sensing monitoring of gullies on a regional scale: a case study of Kebai region in Heilongjiang Province, China. *Chin. Geogr. Sci.* 25 (5), 602–611. <https://doi.org/10.1007/s11769-015-0780-z>.