Elsevier Editorial System(tm) for Journal of Environmental Management
Manuscript Draft

Manuscript Number: JEMA-D-18-04956R1

Title: Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS

Article Type: Research Article

Keywords: Soil erosion; gullying; GIS; statistical model; data mining model

Corresponding Author: Professor Biswajeet Pradhan, PhD

Corresponding Author's Institution: University Technology Sydney

First Author: Alireza Arabameri, Ph.D.

Order of Authors: Alireza Arabameri, Ph.D.; Biswajeet Pradhan, PhD; Khalil Rezaei, Ph.D.

Abstract: Every year, gully erosion causes substantial damage to agricultural land, residential areas and infrastructure, such as roads. Gully erosion assessment and mapping can facilitate decision making in environmental management and soil conservation. Thus, this research aims to propose a new model by combining the geographically weighted regression (GWR) technique with the certainty factor (CF) and random forest (RF) models to produce gully erosion zonation mapping. The proposed model was implemented in the Mahabia watershed of Iran, which is highly sensitive to gully erosion. Firstly, dependent and independent variables, including a gully erosion inventory map (GEIM) and gully-related causal factors (GRCFs), were prepared using several data sources. Secondly, the GEIM was randomly divided into two groups: training (70%) and validation (30%) datasets. Thirdly, tolerance and variance inflation factor indicators were used for multicollinearity analysis. The results of the analysis corroborated that no collinearity exists amongst GRCFs. A total of 12 topographic, hydrologic, geologic, climatologic, environmental and soil-related GRCFs and 150 gully locations were used for modelling. The watershed was divided into eight homogeneous units because the importance level of the parameters in different parts of the watershed is not the same. For this purpose, coefficients of elevation, distance to stream and distance to road parameters were used. These coefficients were obtained by extracting bi-square kernel and AIC via the GWR method. Subsequently, the RF-CF integrated model was applied in each unit. Finally, with the units combined, the final gully erosion susceptibility map was obtained. On the basis of the RF model, distance to stream, distance to road and land use/land cover exhibited a high influence on gully formation. Validation results using area under curve indicated that new GWR-CF-RF approach has a higher predictive accuracy 0.967 (96.7%) than the individual models of CF 0.763 (76.3%) and RF 0.776 (77.6%) and the CF-RF integrated model 0.897 (89.7%). Thus, the results of this research can be used by local managers and planners for environmental management.

Research Data Related to this Submission

There are no linked research data sets for this submission. The following reason is given:
Data will be made available on request

Title page

Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS

Alireza Arabameri[1], Biswajeet Pradhan[2,3]*, Khalil Rezaei[4]

[1]Department of Geomorphology, Tarbiat Modares University, Tehran 36581-17994, Iran

[2]Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW 2007, Australia

[3]Department of Energy and Mineral Resources Engineering, Choongmu-gwan, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea

[4]Faculty of Earth Sciences, Kharazmi University, Tehran 14911-15719, Iran

*Corresponding author: Biswajeet.Pradhan@uts.edu.au; biswajeet24@gmail.com

***Cover Letter**

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

18 November 2018

Distinguished
Professor
Biswajeet Pradhan
Faculty of Engineering & IT
City Campus
PO Box 123 Broadway
NSW 2007 Australia
T: +61 2 9514 7937
Biswajeet.Pradhan@uts.edu.au
www.uts.edu.au

Prof. Lixiao Zhang
Editor
Journal of Environmental Management

Dear Professor Zhang,

RE: Submission of revised manuscript JEMA-D-18-04956 R1

We would like to submit the revised paper JEMA-D-18-03727 R2. titled "Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS" for publication in the JEMA.

The authors appreciate the time and effort by the editor and reviewers in reviewing this manuscript and would like to thank them for their constructive comments and suggestions. Please see our detailed response to each comment below. In addition, English corrections were made by a native English speaker. We hope that the final version meets your expectations. According to Editor' suggestion, we have tried to reduce the length of the manuscript. However, reviewers requested to add more explanations into the manuscript.

I kindly request you to consider this manuscript for publication in JEMA.

Distinguished Professor Biswajeet Pradhan
Faculty of Engineering and IT

University of Technology Sydney

Email. Biswajeet.Pradhan@uts.edu.au

Response to Editor and Reviewers

Ms. Ref. No.: JEMA-D-18-04956.R1
Title: Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS
Journal of Environmental Management

Dear Dr. Pradhan,
Following this message are the reviews of the above-referenced manuscript. We'll be glad to consider this paper for publication after it's been revised in accordance with the reviewers' comments.

Due to space limitations in the printed journal, we are requesting that all authors reduce the length of their papers by at least 10% if possible. If your paper includes large tables or datasets, it is preferred that these be published as supplementary material in Science Direct rather than in print. Further information is provided at the end of this message.

With the revised manuscript, please provide a detailed response to the reviewers' comments, indicating how each comment is addressed in the revised manuscript. If you disagree with any of the reviewers' comments, please address them in a rebuttal.

Yours sincerely,
Lixiao Zhang, Ph.D
Associate Editor
Journal of Environmental Management

Dear Prof. Zhang,

The authors appreciate the time and effort by the editor and reviewers in reviewing this manuscript and would like to thank them for their constructive comments and suggestions. Please see our detailed response to each comment below. In this revised version, we considered all the editor comments. We appreciate these valuable information that have strengthened the quality of the paper. Responses to each comment are also provided below. We believe that the responses and the revision made would meet the expectations In addition, English corrections were made by a native English speaker. We hope that the final version meets your expectations.

Sincerely,

Professor Biswajeet Pradhan,
(Corresponding author)
University of Technology Sydney

****

List of changes in the revised paper:

This document explains the changes made in the revised manuscript while dealing with the comments raised by the editor. Editor comments are marked in **black**, author's response is shown in **blue** while the changes in the manuscript are marked in **red** to keep track of changes.

COMMENTS FROM THE EDITOR

Following this message are the reviews of the above-referenced manuscript. We'll be glad to consider this paper for publication after it's been revised in accordance with the reviewers' comments. Due to space limitations in the printed journal, we are requesting that all authors reduce the length of their papers by at least 10% if possible. If your paper includes large tables or datasets, it is preferred that these be published as supplementary material in Science Direct rather than in print. Further information is provided at the end of this message.

Response:

Dear Editor, as per your suggestion we deleted some un-important texts (10%) from the revised manuscript. We track changed all the corrections in the MS so you can see the changes point-by-point. I hope that the revised manuscript revised manuscript would meet to your satisfaction.

Reviewer #1:

This paper introduced a new model by combining the geographically weighted regression (GWR) technique with the certainty factor (CF) and random forest (RF) models to produce gully erosion zonation mapping, and apply the model in the Mahabia watershed of Iran. The results showed GWR-RF-CF model obtained higher predictive accuracy than others and can be used in development planning. Although the model is not ground-breaking, the algorithm is really applicable and innovative.

Response:

Dear Reviewer #1: Thank you so much for your positive opinion about our manuscript. We addressed all the comments one-by-one here and in the MS (track changed in red).

(1) Abstract. The quantitative precisions of the three models are better to be compared.

Response:

Thank you. The quantitative precisions of the three models are added in page 2 lines 33-35.

(2). Introduction. There are many abbreviations only appear once in the third paragraph, please use the full name directly. Please separate the research objective to an independent paragraph from the literature review.

Response:

Thanks. Introduction is edited as per your comment. The third paragraph was deleted because we found it unnecessary. The purpose of this study is mentioned in a separated paragraph.

(3) Methods. How to unify the spatial resolution of the data? What's the validation result of Landsat 8 image? When there are more factors in training, the simulation precision would be higher. Thus, why the selected 12 factors are necessary? Please only provide the basic formula and the formula with the author's own contribution. Many expanded formular is usual and not need to be listed.

Response:

Thank you. Spatial resolution of the data was unified based on DEM spatial resolution (page 8 lines 232-235). Validation results of Landsat 8 image for extraction of LU/LC is added in page 8 line 220-227. For selection of parameters we used collinearity test (page 14 lines 391-398) and parameters that have collinearity were not used in the final model. Unnecessary formula were deleted.

(4) Result. The abbreviations in the figures should list the full name in the figure title. In Table 3, what's the scientific basis of using such thresholds?

Response:

Thank you so much. Full name of abbreviations are added in figures. Maximum and minimum values, number of classes and classification models of GRCFs are added in table 2.

(5) Discussion. The literature review in this section is not sufficient. What's the factor and precision using RF and CF in other studies? Maybe more comparison can better prove the advantage. In addition, please discuss the uncertainty of the methods and

the future direction.

Response:

Thanks. The literature review is improved in discussion section (page 18 lines 522-526, page 18 lines 554-561. A detailed discussed on uncertainty is added in page 19 lines 577-594.

Reviewer #2

(1) The introduction needs to convey a sense of the research problem in the case study area. The author explains most of the introduction part to cover general aspects of erosion. General / global problems and definitions of gully erosion is discussed from line 45 in page 2 to line 38 in page 3 rather than problems and severity of problems in the case study area. The line 48 in page 3 groups all the models used for gully erosion but it does not incorporates GWR, whereas line 32 in page 4 reports usage of GWR for gully erosion?

Response:

First, we would like to express our thanks to your constructive and positive feedback about our research. The introduction section is improved as per your comment. The irrelevant and unimportant texts (paragraph 3) were deleted. In addition, the problem of gully erosion in the study area is added in page 4 line 112-118. Although GWR has been used in other studies such as landslide susceptibility mapping, however so far it has not been used for gully erosion which is a good novelty in this research.

(2) The description about the study area is to be revised in a manner how physio-climate-socio-cultural setup of the area has influenced the gully erosion. For instance, author reported study area receives 90 mm rainfall in a year but not mentioning about the annual distribution of rainfall/intensity of rainfall which causes gully erosion.

Response:

Thanks. The detailed description about the study area is improved and important factors related to gully occurrence in the study area are added in page 5 lines 146 – 152.

(2) Methodology part needs to present in a logical/scientific manner. The section 2.2 should be renamed in order to avoid confusions, because this section is further explained in the sections 2.3 and 2.4. The author need to strengthen the content

reported. The line 42-44 in page 5 denotes sources of satellite data in a vague manner (e.g. Google earth image - Is there?) Instead of saying landsat 8 image mention proper the satellite and sensor with source in parenthesis (website). Line 13-22 in page 5 should be revised thoroughly. The equation 1 should be expanded (Line 9 in page 7) and equation 12 to be mentioned. On what basis the author reported GWR and RF models are very recent? In text, it is referred with the literatures of year 2001 (Line 8 in page 8 and line 47 in page 9). What is CART? At least expand for the first time.

Response:

Thanks. Section 2.2 is renamed. Source of data and its website are added in page 8 lines 211, 215, 216 and 217. The equation 1 is explained in page 7 line 202. Equation 12 is mentioned. So far GWR technique has not been used in gully erosion mapping and used of RF in gully erosion is new. CART model is explained in page 4 line 91.

(3) The results should be an interpretation of model outcomes and it should not just mention the data (E.g. line 18 to 55 in page 16). Likewise the results section, the discussion part should have strong arguments on models employed. The discussion is very vague and does not reflect the major outcomes of the approach. (e.g line 50-59 in page 17: "Via remote sensing, data can be collected for vast areas in a short time"; "it can save time and reduce costs in researches"; "use of GIS in data analysis is essential")

Response:

Thanks. The results explained in pages 14 and 15 lines 400-442. Discussion part is largely improved in pages 18-20 lines 510 – 515, 521-525, 552 – 560, 576-593.

(4). The conclusions at the end of the paper should correspond to the questions posed at the beginning of the paper. It should not introduce new models/data other than discussed earlier. For e.g. LIM training dataset in line 18 in page 19. No where reported what is LIM? The last half of conclusion is not a conclusion of the study.

Response:

Thanks. The conclusion part is improved in pages 21-22. LIM is corrected in page 21 lines 607 and 608.

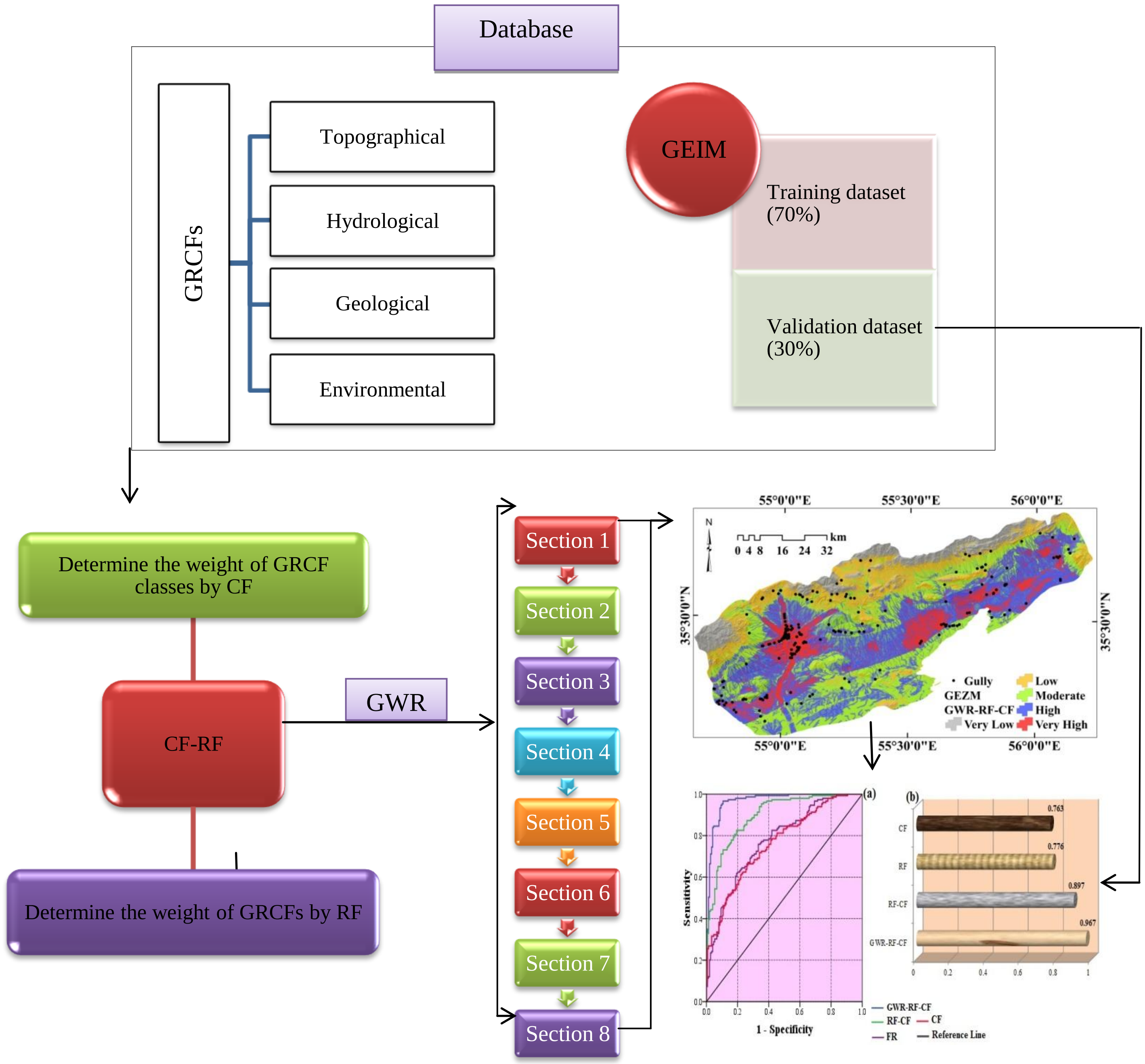
(5). Graphical abstract and highlights section does not convey the purpose. The author should highlight the simplified methodology, notable results and a line of applicability rather than just split the title/method. Technical writing needs to be strengthened. GRCF and GWR need to be expanded very first (Line 12&33 in page 4). It is expanded in the abstract and later in page 5 (line 46) but not for the first time in the main paper. Typographic/Spelling /grammar mistakes are prevalent throughout the text (For instance, the word "used" in first line of the paper in highlights section). The figures should be self explanatory or it should be described briefly. E.g. Fig.12. Fig. 7a and 7b can be combined.

Response:

Thanks. Graphical abstract and highlights are corrected. The conclusion part is improved in pages 21-22. LIM is corrected in page 21 lines 607 and 608. GRCF and GWR are explained in the first use in page 4 line 96 and 105. English of manuscript is largely improved. Figures are described briefly.

Finally, we appreciate your time to read our manuscript. We have tried to address all the comments one-by-one and we believe that they improved the quality and clarity of the manuscript. We hope the revised version of the manuscript meets your expectations.

Graphical Abstracts



Highlights

1- Three approaches ((a) CF and RF; (b) integrated CF-RF; and (c) combined GWR-CF-RF) used for GEZM.
2- A new methodological framework (GWR-RF-CF) was introduced for GEZM.
3- Geographically Weighted Regression was used to create several homogenous units.
4- GWR-CF-RF has higher prediction accuracy than other employed models.
5- GWR-CF-RF as a new approach can be used by decision makers for GEZM.

Gully erosion zonation mapping using integrated geographically weighted regression with certainty factor and random forest models in GIS

Alireza Arabameri[1], Biswajeet Pradhan[2,3]*, Khalil Rezaei[4]

[1]Department of Geomorphology, Tarbiat Modares University, Tehran 36581-17994, Iran

[2]Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW 2007, Australia

[3]Department of Energy and Mineral Resources Engineering, Choongmu-gwan, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea

[4]Faculty of Earth Sciences, Kharazmi University, Tehran 14911-15719, Iran

*Corresponding author: Biswajeet.Pradhan@uts.edu.au; biswajeet24@gmail.com

Abstract

Every year, gully erosion causes substantial damage to agricultural land, residential areas and infrastructure, such as roads. Gully erosion assessment and mapping can facilitate decision making in environmental management and soil conservation. Thus, this research aims to propose a new model by combining the geographically weighted regression (GWR) technique with the certainty factor (CF) and random forest (RF) models to produce gully erosion zonation mapping. The proposed model was implemented in the Mahabia watershed of Iran, which is highly sensitive to gully erosion. Firstly, dependent and independent variables, including a gully erosion inventory map (GEIM) and gully-related causal factors (GRCFs), were prepared using several data sources. Secondly, the GEIM was randomly divided into two groups: training (70%) and validation (30%) datasets. Thirdly, tolerance and variance inflation factor indicators were used for multicollinearity analysis. The results of the analysis corroborated that no collinearity exists amongst GRCFs. A total of 12 topographic, hydrologic, geologic, climatologic, environmental and soil-related GRCFs and 150 gully locations were used for modelling. The watershed was divided into eight homogeneous units because the importance level of the parameters in different parts of the watershed is not the same. For this purpose, coefficients of elevation, distance to stream and distance to road parameters were used. These coefficients were obtained by extracting bi-square kernel and AIC via the GWR method. Subsequently, the RF-CF integrated

model was applied in each unit. Finally, with the units combined, the final gully erosion susceptibility map was obtained. On the basis of the RF model, distance to stream, distance to road and land use/land cover exhibited a high influence on gully formation. Validation results using area under curve indicated that new GWR-CF-RF approach has a higher predictive accuracy 0.967 (96.7%) than the individual models of CF 0.763 (76.3%) and RF 0.776 (77.6%) and the CF-RF integrated model 0.897 (89.7%). Thus, the results of this research can be used by local managers and planners for environmental management.

Keywords: Soil erosion; gullying; GIS; statistical model; data mining model

1. Introduction

Soil erosion is a global problem that seriously threatens soil and water resources (Comino et al., 2016; Swarnkar et al., 2018; Arabameri et al., 2017b). The short-term, destructive effects of erosion may not be noticeable, but, in the long run, its effects will be clearly visible (Singh and Singh, 2018). The consequences of soil erosion could be extremely dangerous that they have caused the destruction of some civilisations in the past (Boardman and Favis-Mortlock, 1998). Therefore, decision-makers must take the necessary measures to reduce the damage caused by erosion (Alam et al., 2016). In Iran, soil erosion is a major problem, especially in agriculture, natural resources and the environment. Approximately 125 million hectares of the country's 165 million hectares of land are exposed to water erosion (Refahi, 2009).

Gully erosion is the most destructive type of water erosion due to the dissolution and alkalinity of formations in forests, rangelands and agricultural lands (Lesschen et al., 2007). Gully erosion occurs in the arid and semiarid regions of the world due to the overexploitation of water and soil and improper environmental management (Ezechi, 2000). Gully erosion leads to soil and ecosystem destruction, quantitative and qualitative reduction of groundwater, soil degradation and destruction of infrastructure, such as roads and bridges (Paolo et al., 2014; Boardman, 2014; Dube et al., 2014; Conoscenti et al., 2014; Torri et al., 2014). The process that leads to gullying and its development on the surface begins with water flow and sheet erosion. Subsequently, the formation of surface grooves are affected by rill erosion with the same slope relative to the domain, and continues with depth of grooves over time. Thus, grooves with a depth of several centimeters turn into gullies with depths of several meters with slopes less than the domain (Keller, 2011; Goodwin et al., 2017; Barnes et al., 2016; Dymond et al., 2016).

Given that various environmental factors affect gully erosion, understanding and recognising the relationship between these parameters and the occurrence of gullies and the prediction of gully erosion-prone areas are important strategies for water and soil resource management (Shit et al., 2015). In recent years, various methods have been used by researchers around the world for gully erosion zonation mapping (GEZM). These models can be divided into three groups: (i) knowledge-based models, such as the analytical hierarchy process (AHP) (Zakerinejad and Maerker, 2014); (ii) bivariate and multivariate statistical models, such as weights-of-evidence (WOE) (Zabihi et al., 2018), logistic regression (LR) (Dewitte et al., 2015), maximum entropy (ME) (Rahmati et al., 2017), information value (IV) (Conforti et al., 2011), conditional analysis (CA) (Conoscenti et al., 2013) and frequency ratio (FR) (Rahmati et al., 2016) and (iii) data-mining models, such as multivariate adaptive regression splines (MARS) (Conoscenti et al., 2018), random forest (RF) (Arabameri et al., 2018), support vector machine (SVM) (Pourghasemi et al., 2017), classification and regression trees (CART) (Märker et al., 2011) and acritical neural networks (ANN) (Pourghasemi et al., 2017).

The main disadvantage of bivariate statistical methods is the lack of calculation of the gully-related causal factors (GRCFs) importance on the occurrence of gully erosion. By contrast, the disadvantage of data-mining methods lies on their inability to calculate the spatial relationship between GRCFs and gully locations. Therefore, in this research, an integrated CF bivariate statistical model and RF data-mining model are used for GEZM. Literature review indicates that the CF model has not been used in the field of quantitative spatial correlation between GRCFs and gully locations in the GEZM. Moreover, the RF model has been used to determine the importance of GRCFs because of its higher ability and accuracy compared with other data-mining models (Arabameri et al., 2018). The main gap in the previous studies of GEZM is that the importance of GRCFs across the area is considered the same, whereas different degrees of parameter influence may occur in an area, such that, with the change of location, the effect of parameters can be changed (Brunsdon et al., 1998; Feuillet et al., 2014). To solve this problem, Nakaya (2002) Fotheringham et al. (2003) used the geographically weighted regression (GWR) technique. This method is used to segment the study area into several predictive areas with spatial autocorrelation and appropriate sizes. Thereafter, the RF-CF integrated model is applied in each segment. Finally, GEZM was obtained using the combination of segments. The GWR technique (Sabokbar et al., 2014; Yu et al., 2016), CF (Chen et al., 2015; Fan et al., 2017; Chen et al., 2017) and RF (Kim et al., 2018; Taalab et al., 2018) methods are useful and effective for environmental management and are used by various researchers in environmental risk management.

In the study area, gully erosion have caused several problems such as destruction of: natural ecosystem, human infrastructure such as power and gas transmission lines, roads and bridges, land degradation, reduction of amount and quality of surface and underground waters, sediment accumulation in canals, destruction of natural landscape faces and reduction of soil fertility. On the other hand, this phenomenon has caused other problems in the study area such as groundwater pollution, flooding, and desertification. Therefore, in order to reduce its damages and take preventive measures, in the present research, a novel approach that is derived from the integration of the GWR technique with the CF bivariate statistical and RF data-mining methods is used to predict gully erosion-prone areas in the study area.

2. Materials and Methods

2.1. Study area

Mahabia watershed (35°12′18″N–35°37′13″N, 54°44′27″E–56°13′35″E), covering a total area of 5,757 km^2, is located in the northeastern part of the Semnan province and in the southeastern part of Shahroud city in Iran (Fig. 1). The study area has a mean annual precipitation of 90 mm and it receives approximately 80% of its annual rainfall from November to April, a mean annual temperature of 25.7 °C and has an arid climate (IRIMO, 2012). The elevation of the study area ranges from 683 m to 2,297 m. Topographically, the central and southern parts of the study area have a flat curvature, and the north section, which is located in the mountainous area, has convex and concave curvatures. Approximately 41.33% of the study area has a flat topography; 34% is concave, and 24.66% is convex. The average slope in the study area is 3.84°, with slope gradient slowly reducing from the north to the south. Quaternary sediments, including unconsolidated windblown sand deposit including sand dunes (Qs,d,) clay flat (Qcf), salt flat (Qsf), stream channel, braided channel and flood plain deposits (Qal), high-level piedmont fan and valley terrace deposits (Qft1), low-level piedmont fan and valley terrace deposits (Qft2) and salt lake, covering more than half of the study area (GSDI, 2012). Land use/land cover classes of the study area include agriculture (0.04%), orchard (0.027%), bare land (3.87%), kavir (19.45%), salt land–kavir (1.19%), poor range (56.16%), rock (16.22%), salt lake (1.63%), salt land (1.36%) and wet land (0.015%). Soil types in the Mahabia watershed comprises of badlands (7.42%), playa (1.23%), rock outcrops/entisols (24.18%), rocky lands (2.5%), salt flats (18.76%), aridisols (6.82%) and aridisols and entisols (39.05%). The gully erosion, as one of the most destructive types of water-related erosion, annually imposes economic damages to the inhabitants of the region. The gullies of the study area are mainly created in plains and low slope areas with high water concentration and drainage density. High precipitation and flood scenarios are the general characteristic of arid areas (in some cases, 80% of annual rain falls in a few hours), presence of gypsum and salt minerals due to high evaporation, overgrazing and destroying vegetation, caused the formation of piping and gullying in the studied area.

2.2. Flowchart of research

As shown in Figure 2, the methodology used in this study consists of nine main steps: (1) preparation of basic data, including PALSAR DEM, geology map (1:100000), topography map (1:50000), Landsat 8 image, Google Earth images and GPS points of gully locations; (2) determination and extraction of gully-related causal factors (GRCFs) using several resources; (3) determination of the location of gullies and preparation of GEIM and randomly dividing them into two groups of training and validation; (4) application of the RF model and determination of the importance of GRCFs; (5) application of the CF model and determination of the spatial relation between GRCFs and location of gullies; (6) segmentation of the study area into several homogeneous units using the GWR technique and determination of the weight of the GRCFs in each homogeneous unit; (7) application of the RF-CF model in each segment and combination of segments; (8) preparation of GEZMs using three approaches: (i) individual models of CF and RF, (ii) combination of CF statistical model and RF data-driven model and (iii) combination of RF-CF integrated model with GWR technique and (9) validation of GEZMs using AUC, SCAI and FR indicators.

2.3. Database preparation

A gully erosion inventory map (GEIM) was used to analyse the relationship between gullies and GECFs quantitatively and was also the basis of LSZM (Pourghasemi et al., 2017). In this study, 215 gullies were identified using extensive field surveys and satellite image interpretations. Amongst them, 70% (150 gully locations) were selected for modelling, and 30% (65 gully locations) were considered for validation purposes (Arabameri et al., 2018). In Figure 3, several gullies that were identified in the study area are shown.

In the occurrence of gully erosion, as a threshold-dependent process, various parameters, such as geology, topography, hydrology, soil characteristics, climate and human activities, are involved (Poesen et al., 1998; Conforti et al., 2014; Conoscenti et al., 2014; Go´mez-Gutie´rrez et al., 2015; Arabameri et al., 2018). Therefore, selecting the effective parameters is essential to identify areas prone to gully erosion. In the present study, an extensive literature review (Rahmati et al., 2017; Pourghasemi et al., 2017; Arabameri et al., 2018), the features of the study area and a collinearity test amongst 12 GRCFs (elevation, slope, plan curvature, topographic wetness index (TWI), stream power index (SPI), distance to stream, drainage density, distance to road, lithology, land use/land cover, soil type and rainfall) were considered for modelling (Fig 4a-l).

Topography, due to the determination of the erosive power of the flow, plays an important role in the initiation and development of gullies. This parameter also affects geology (Wade, 1935), climate (Bonacina, 1945) and vegetation cover (Zakharov, 1940); therefore, the quality of topographic data has an extraordinary impact on the accuracy of the GEZM (Hancock and Evans, 2005; Go´mez-Gutie´rrez et al., 2015). In this research, PALSAR DEM with a spatial resolution of 12.5 meters was used to extract topographic data, including elevation, slope, plan curvature, TWI and SPI in ArcGIS10.5. Eqs. (1) and (2) are used for the calculation of TWI and SPI, which indicate the spatial distribution of areas of saturation and erosive power of flowing runoff, respectively (Moore et al., 1991):

$$\mathrm{SPI} = \ln\ (\mathrm{As} \times \tan\beta), \qquad (1)$$

$$\mathrm{TWI} = \mathrm{In}\ (\mathrm{A_S}/\tan\beta), \qquad (2)$$

where As is the upstream contributing area, and β is and the slope gradient.

The extraction of the stream network from the PALSAR DEM was performed in the ArcGIS10.5 to produce distance to stream and drainage density. For this purpose, the holes were filled in the DEM using the 'Fill' tool, and the direction and accumulation of streams were obtained using the 'Flow Direction' and 'Flow Accumulation' tools. The threshold of 1000 cells was considered in the extraction of the stream network. After extraction, the distance to stream and drainage was calculated using the 'Euclidean Distance' and 'Line Density' tools. For the production of distance to road, the extraction and digitisation of the roads from topographic maps of 1: 50000 was obtained from the National Geographic Organization of Iran (www.ngo-org.ir). (NGOI, 2008) and Google Earth satellite images (15/08/2018) were used, and then, using the 'Euclidean Distance' tool, distance to road was calculated. On the basis of the separation and digitisation of the polygons in the lithological units from the geological map (scale of 1: 100000) prepared by Geological Society of Iran (GSI) (http://www.gsi.ir/), a lithology map was prepared in the ArcGIS10.1 and was classified into 11 groups (Table 1). A Landsat 8 OLI/TIRS image (path/162, row/35) (14/07/2018) (https://earthexplorer.usgs.gov/) was used for the preparation of the land use/land cover (LU/LC) map in ENVI4.8, and a supervised (maximum likelihood) algorithm was used for this purpose. 465 GCP (ground control point) were used for the validation of the LU/LC map with the Kappa coefficient. The Kappa coefficient for the final map was estimated using Eq. 3 (Lo and Yeung, 2002):

$$K = {\{N\sum_{i=1}^{r}(X_{ii}) - N\sum_{i=1}^{r}(X_{i+}.X_{+i})\}}/{N^2} - \sum_{i=1}^{r}(X_{i+}.X_{+i}) \qquad (3)$$

where, r is number of rows in error matrix; Xii is number of observations in row i and column i, X i + is total of observations in row I, X+I is total of observations in column i and N is total number of observations included in the matrix. The Kappa coefficient of generated map was found to be 99.12%.

A soil type map (1:250000) was obtained from the Semnan Agricultural and Natural Resources Research Centre and was classified into three categories. For the preparation of the annual rainfall data, the rainfall statistics of Toroud, Moalleman, Dameghan, Shahroud and Biarjamand during a 30-year period (from 1987 to 2017) were used; using the Kriging method, the rainfall map was prepared in ArcGIS10.5 (IRIMO, 2012). In the final step, all layers were unified in the 12.5m pixel size based on PALSAR DEM spatial resolution and the UTM

Zone39N geographic coordinate system. Maximum and minimum values, number of classes and classification models of GRCFs are shown in table 2.

2.4. Modelling approach

2.4.1. GWR technique

One of the new methods for achieving higher accuracy in spatial analysis is the GWR method, which is highly effective when a spatial correlation exists between independent variables, (Fotheringham et al., 2001). The GWR model is highly important in the modelling of heterogeneous spatial processes (Wheeler et al., 2014). It has attracted the attention of many researchers because of its precise performance at the time of exploring local changes (Chalkias et al., 2014). This model is an extended mode of general regression and is applied to the gain of equations of regression for each spatial area separately (Celik et al., 2016). In this model, the coefficients are estimated on the basis of the geographical coordinates of the measurement points; the coefficients of the model are estimated for each point in the region, and the values and symbols of the coefficients vary at different points. If the relationship between the independent and dependent variables in a part of a region is positive but negative in the other parts, this model can provide valid spatial relationships (Zhang & Griffith, 2000). This model is in the form of Eq. 4:

$$y_i = \beta_{0\,(u_i\,,v_i)} + \sum_{K=1}^{Q} \beta_{K\ (u_i\,,v_i)}\, x_{ik} + \in_i \qquad {}_{i=1,2,\dots,l}\ , \qquad (4)$$

where $(u_i\,, vi)$ represents the coordinates of an ith point in space i=1,2,3,…L and Q and L are the regression coefficients and the number of points, respectively. y_i is the dependent variable in position i, x_{ik} is the quantity of the kth explanatory variable in position i, $\beta_{K\ (u_i\,,v_i)}$ is the local regression coefficients for the kth explanatory variable in position I, $\beta_{0\,(u_i\,,v_i)}$ is the intercept parameter in position I and $\in_i$ indicates a random error.

If the components of the weight of the observation points are entered in the regression equation, the vector relation of the evaluation parameters becomes a GWR relation. According to the Tobler law in geography, the adjacent points are more similar spatially. In the estimation of the point properties, points near them are more weighted than the farther points (Tobler, 1970). The coefficients of the model are obtained by the following equation:

$$\hat{\beta}(u) = (X^T W_i X)^{-1} X^T W_i Y, \qquad (5)$$

and its variance is:

$$var\ (\hat{\beta}) = \left(X^T\, W_i{}^{-1} X\right)^{-1}, \qquad (6)$$

where i=1,2,3, … p, X is an independent variable, Y is a dependent variable, W_i is the weight matrix (n×n), in which the inlines of the matrix, except for the diagonal elements, are zero and whose diagonal elements are the geographical weights.

$$W_i = \begin{bmatrix} W_{i1} & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & W_{in} \end{bmatrix},$$

The selection of Wi depends on the selected kernel function, which can either be fixed or adaptive kernels.

AIC is used to measure the relative efficiency of the model and the selection of the criteria. Eq. (7) is used for the calculation of AIC:

$$AIC = -2Ln\ l\left(\hat{\theta}_L\, , x\right) + 2q, \quad (7)$$

where $l\left(\hat{\theta}_L\, , x\right)$ is the maximised likelihood of the parameter vector ɵ, x is a random sample, $\hat{\theta}_L$ is the maximum likelihood estimate of ɵ and ɵ is the number of the unknown parameters. The low values of this model indicate that the estimated value of the model is closer to reality (Wang, et al., 2005)

2.4.2. RF model

The use of decision trees has become common due to its simplicity, interpretation capability and ability to measure with covariance scales at different measurement levels (including nominal variables) (Duda et al., 2001). The RF method is a modern nonparametric technique, which has been one of the best learning algorithms in recent years. This model can be used for a large dataset and can make highly accurate classifications. Unlike classical models such as regressions that use only one model, the RF method uses hundreds and thousands of decision trees such that they can have better elicitation from the variables (Duda et al., 2001). The RF method is generalised from a classification tree method in the CART model although, in this model, each set of decision trees performs the classification (vote) and the class that wins the most votes is selected as the result of the model. Different from the CART model, the decision trees in the RF model do not prune, and whilst the diversity of trees is high, they preserve the accuracy of prediction by selecting the best separator amongst the subsets. This model offers a scale of importance of variables that is used to determine the impact of each factor (Breiman and Adele, 2013). The RF model has less errors than the decision tree method due to the use of several trees. Meanwhile, the correlation of trees in the RF model is low due to the random sampling of predictors in each node and the heterogeneity of trees. The best way to determine the number of needed trees is to compare forest predictions with the predictions from its subsets, and if the prediction of the subset is as accurate as the forest prediction, the number of trees is sufficient (Payet and Todorovic, 2008).

2.4.3. Certainty factor (CF)

The CF model was firstly proposed by Shortliffe and Buchanan (1975) and improved by Heckeman (1986). To apply the CF model in the case of gully erosion, one must determine the dependent and independent variables. Therefore, in this study, gully erosion was considered a dependent variable, and its effective factors were considered as independent variables. After environmental analysis and determination of the number of gully events in each of the classes of independent variables in ArcGIS10.5 final map using CF model obtained.

2.4.4. Validation processes

The GEZM validation process is important; without this, the study lacks scientific credibility. In this research, a validation dataset was used for this purpose. The prediction accuracy of the final maps was determined in SPSS21 using the ROC method, which is the most common quantitative method in spatial modelling and in the prediction of natural phenomena (Chen et al., 2017). AUC indicates the accuracy of the final map in predicting areas susceptible to gully erosion. The values of this parameter fluctuated from 0.5 to 1; the closer the value is to one, the higher the prediction accuracy of the model (Yesilnacar, 2005). To evaluate the classification accuracy of the models, FR and SCAI were employed (Ilinca and Gheuca, 2011). FR is the ratio of the gully surface area in each class to the surface area of that class (Pradhan and Lee, 2010). SCAI is the ratio of surface area of class to the gully surface area of that class (Yilmaz, 2009). Low FR and high SCAI values in an extremely low susceptibility class and vice versa indicate acceptable accuracy in classification.

3. Results

3.1. Segmentation of the study area using the GWR technique

The GWR allows us to have different regression equations in different parts of the study area and improves modelling implementation by reducing spatial correlation (Zhang et al., 2014). According to Tobler's theory of proximity and similarity, observations closer to a particular location should receive a higher score than the observations that are farther away (Miller, 2004). Therefore, we can use this technique to estimate the parameters for the model. For the segmentation of the study area, the coefficients of the GWR model were calculated to explore the spatial variations of the relationships between the study area and environmental factors. The natural break method is a common classification method based on the inherent nature of the data (Jensen, 1967). By contrast, the GWR coefficient values can be used to describe the spatial correlation of factors. Therefore, we divided the study area into several sections; in each section, the values of the GWR coefficients are similar. Moreover, the number of parameters used for segmentation has a great influence on the plotting results. Thus, if the number of parameters is substantial, we will have several small pieces in the plot, which will cause problems in the development of the sample for training and validation and in obtaining a satisfactory prediction. By contrast, if the number of parameters is extremely low, a few large pieces will exist in the plot, which means that we cannot reduce spatial correlation in each region effectively, affecting the predicted results negatively. In this research, the most important GRCFs used for fragmentation were elevation, distance to stream and distance to road; the coefficients of their GWR weights were obtained by extracting bi-square kernel and AIC in the GWR method and was divided into 8 pieces based on the study area (Fig. 5).

3.2. Multicollinearity analysis

Before the analysis of the models, a collinearity was verified amongst the GRCFs using tolerance (TOL) and variance inflation factor (VIF), which are the most important indexes for measuring multicollinearity. This process was performed because the multicollinearity amongst GRCFs reduces the prediction accuracy of the models. When the value of TOL is less than 0.1 and the value of VIF is greater than 10, a collinearity exists amongst the parameters (Bui et al., 2011). The results of the multicollinearity test (Table 3) affirmed that no collinearity exists amongst the GRCFs (independent variables).

3.3. Certainty factor (CF) model

The values of the CF model vary from −1 to 1; positive values represent an increasing certainty in the occurrence of an event, negative values indicate a decreasing certainty in the occurrence of an event and close-zero values indicate information about the variable is not enough (Dou et al. 2014). The results of the CF model (Table 34) confirmed that, in the elevation parameter, class <829m with the highest value of CF (0.248) has a strong relationship with the occurrence of gullies in the study area. This result is in line with Hongchun et al. (2014). In slope parameters, the relationship between the slope parameter and gully locations with CF method showed that with the increase in slope, the sensitivity of the areas to the occurrence of the gully decreases, and the low slopes are more prone to gullying. Thus, compared with the other of the slope classes, the <2.4° class with CF= 0.104 has the greatest impact on the occurrence of a gully. The reason is that low slopes cause runoff accumulation and gully development, which is consistent with the results of Le Roux and Sumner (2012). In the plan curvature parameter, flat areas (CF= 0.125) showed a higher correlation with gullies compared with the concave (CF= 0.070) and convex (CF= -0.30) topography. This result is in line with Conforti et al. (2010). TWI results corroborated that the >11.47 class with the highest TWI has the greatest impact on the occurrence of the gully (CF= 0.693). Gully erosion occurs when the flow power exceeds the soil shear stress, therefore, a higher TWI in an area indicates a higher erosive power of flow. This result is in line with Go´mez-Gutie´rrez et al. (2015). On the basis of the SPI factor, the <14.6, class with highest SPI had a strong relation with gullies (CF= 0.797). In distance to stream, the <176m class had a positive value (CF= 0.586). Therefore, gullying probability is higher in this class. By contrast, other classes with more distance from the stream had negative values: 176–

415 m (CF= −0.06), 415–704 m (CF=−0.113), 704–1107 m (CF= −0.207) and >1,107 m (CF= −0.261). This result is in line with Dube et al. (2014). The values of CF in drainage density indicate that with increasing drainage density, the sensitivity of the area to gully erosion has increased; thus, all classes had negative values, except the >1.54 km/km^2 class (CF= 0.445). This result is in line with Arabameri et al. (2018). In distance to road, the <500 m (CF= 0.859) and 500–1000 m (CF= 0.425) classes have positive values. By contrast, the 1000–1500 m (CF= −0.952), 1500–2000 m (CF= −0.270) and >2000 m classes have negative values. These results contend that the above parameter played a major role in the occurrence of gullies in the study area. In the case of lithology, PlQc (Fluvial conglomerate, piedmont conglomerate and sandstone) with CF= 0.907 had the most sensitivity to gully erosion. In the LU/LC parameter, only bare land (0.658) and kavir (0.478) classes have positive values, and other classes including agriculture, orchard, salt land-kavir, poor range, rock, salt lake, salt land and wetland have negative values. Sensitivity to gully erosion is high in the bare land and kavir classes due to lack of vegetation and existence of soluble sediments (gypsum and salt), respectively. For soil type values, only the entisols/aridisols class had positive value (CF= 0.276), which indicates that this soil type has high susceptibility to erosion. In rainfall class, the classes of 66.27–85.06 mm (CF= 0.644) and 85.06–103.19 mm (CF= 0.054) have positive values. After determining the weight of the classes, GEZM was obtained using Eq. (11) in the CF method, whose values are between -7.796 and 6.277. Finally, the GEZM was categorised into five classes from very low to very high susceptibility classes using the natural break method (Fig. 6a). According to results (Figs. 7a and 7b), the highest area of the Mahabia watershed (33.92%) belongs to the moderate sensitivity class, whereas the lowest (4.57%) belongs to the very high sensitivity class.

3.4. RF model

The RF model with an out-of-bag error rate = 28.57% (Fig. 8) was applied in R using the caret package. This scenario implies that the accuracy of the model is equal to 71.43%. Table 5 exhibits the results of the confusion matrix. According to results, from 110 nongully locations observed, 81 (73.63%) are predicted as nongullies, and 29 (26.36%) are predicted as gullies; from 100 observed gully locations, 31 (31%) are predicted as nongullies, and 69 (69%) are predicted as gullies. Determining the importance of GRCFs using the RF method (Fig. 9) showed that amongst 12 GRCFs, the distance to stream (22.29), distance to road (18.66) and land use/land cover (17.32) had the highest impact on the occurrence of gullies in the study area. By contrast, plan curvature (0.73), slope (2.24) and TWI (2.84) had the least impact on gullying. The resultant GEZM by the RF model was classified into five classes (very low to very high) using the natural break method (Fig. 6b). The resultant map shows that 9.71 % of study area and 45.11% of the gullies (Fig. 7b) belong to the very high susceptibility class (Fig. 7a), whereas 9.28% of the study area and 0.45% of the gullies belong to the very low sensitivity class.

3.5. CF-RF integrated model

The RF model does not consider the weight of GRCF classes, and the CF model does not calculate the significance of each GRCF. To eliminate these limitations and to increase the performance of the models, the two models were integrated in ArcGIS10.5 using Eq. 8:

$$GEZM_{CF-RF} = (W_{CF} Distance\ to\ stream \times 22.29) + (W_{CF} Distance\ to\ road \times 18.66) + (W_{CF} LU/LC \times 17.32) + (W_{CF} Rainfall \times 10.93) + (W_{CF} Drinage\ density \times 9.34) + (W_{CF} Elevation \times 0.067) + (W_{CF} Lithology \times 5.17) + (W_{CF} Soil\ type \times 5.07) + (W_{CF} SPI \times 2.97) \times (W_{CF} TWI \times 2.84) \times (W_{CF} Slope \times 2.24) + (W_{CF} Plan\ curvature \times 0.73) \quad (8)$$

The resultant map based on natural breaks was divided into five classes (Arabameri et al., 2017a) including very low, low, moderate, high and very high (Figure 6c). The GEZM was prepared using the integrated model, which showed that 23.77% and 9.35% of the area belong to the high and very high sensitivity classes, respectively, whereas 9.79%, 20.89% and 36.17% belong to very low, low and moderate classes. By contrast, 71.15% of the gullies belong to high and very high classes.

3.6. GWR-CF-RF new approach

The results of determining the weight of the GRCFs in each of the eight sections are shown in Fig. 10, which indicates that in different sections of a watershed, the effect of GRCFs is different. Thus, in section 1, drainage density (7.02), lithology (6.86) and rainfall (6.44), in section 2, distance to road (15.32), LU/LC (10.01) and lithology (5.57), in section 3, distance to stream (11.29), SPI (2.97) and slope (0.34), in section 4, soil type (6.64), drainage density (6.55) and TWI (6.26), in section 5, distance to stream (2.51), elevation (2.03) and rainfall (1.19), in section 6, TWI (5.85), distance to stream (4.07), and rainfall (0.78), in section 7, soil type (2.05), and LU/LC (0.24), and finally in section 8, rainfall (4.32), drainage density (3.93), and soil type (3.51) have high impact in gully occurrence. After applying of CF-RF combination model in each of sections, by combination of the sections, the GEZM was prepared and classified according to the natural break to 5 classes including very high, high, moderate, low and very low (figure 6d). The results of this map affirm that 9.25%, 15.31%, 29.64%, 32.07% and 13.7% of study area located in very low to very high sensitivity classes (figure 7a), respectively. In contrast, 0.46%, 6.04%, 13.48%, 26.51%, and 53.48% of gullies (figure 7b) located in this classes respectively.

3.7. Validation of GEZMs

The results of the GEZM validation using ROC (Figure 11a) and its AUC (Figure 11b) showed that amongst the four models, the GWR-CF-RF has a higher predictive accuracy 0.967 (96.7%) than the individual CF 0.763 (76.3%) and RF 0.776 (77.6%) models and the CF-RF integrated model 0.897 (89.7%) model. The results of the FR and SCAI indicators (Figure 12) showed that the classification accuracy in all four models is acceptable.

4. Discussion

Gully erosion affects the environment in two dimensions: firstly, by destroying the surface and subsurface horizons of the soil, causing high sediment production and degradation of the production bed, and secondly, by exacerbating the discharge of surface runoffs and reducing groundwater recharge (Prosser, 1996; Kou et al., 2015). Given the extraordinary importance and extensive damage caused by gully erosion, substantial research has been conducted on the causes of gullying and the identification of erosion-susceptible areas to manage this phenomenon and take protective measures to reduce its damage (Conoscenti et al., 2014; Conforti et al., 2014; Arabameri et al., 2018). In this research, three approaches, along with remote sensing data and the GIS technique, have been used for the modelling of gully erosion and preparation of GEZM: (1) using the CF bivariate statistical model and RF data-mining model, (2) integrating the CF bivariate statistical model with the RF data-mining model and (3) using the GWR technique along with the CF-RF integrated model.

The use of remote sensing science because of its advantages including the ability to collect data from vast and inaccessible areas as well as saving time and money is one of the most reliable and efficient ways for Earth observation such as geomorphology, geology, agriculture, natural resources, etc (Karlsson et al., 2017). The high quality of analyzing and managing a large amount of data and the possibility of analyzing them with advanced methods in GIS has made it a powerful tool for managing environmental hazards (Ahlmer et al., 2018).

CF is one of the statistical models with spatial prediction capabilities in various environmental sciences. One of the advantages of using this model is to increase the certainty of prediction, because the modelling process is based on the occurrence of previous gullies, which reduces the uncertainty of modelling (Dou et al., 2014). The disadvantage of this model is the non-calculation of parameter importance in event occurrence.

Wang et al (2015) compared CF and index of entropy models for landslide susceptibility mapping in the China using 15 landslide conditioning factors, including altitude, slope, aspect, plan curvature, general curvature, profile curvature, distance to faults, distance to rivers, distance to roads, the sediment transport index (STI), SPI, TWI, geomorphology, lithology, and rainfall. Their results indicated that CF (0.843) has higher prediction accuracy than the entropy method (0.822). RF is a data-mining model that has several advantages, such as high accuracy in spatial prediction, as well as high ability to determine important variables in prediction (Breiman and Adele, 2013). This model suffers from the non-computation of spatial relationship between parameters affecting events.

The use of the GWR technique to divide the area into homogeneous units can increase the accuracy of modelling and prediction, because the effective parameters in the occurrence of a phenomenon do not have the same importance in different parts of an area (Yu et al., 2016). Nowadays, remote sensing data and GIS technique are widely used by researchers worldwide for hazard assessment and environmental management because of their advantages (Arabameri et al., 2017; Raouf et al., 2017; Nwankwo and Nwankwoala, 2018; Sharma and Kumar Mahajan, 2018; Rahim et al., 2018). Given that the study area is large, data collection using only field surveys is difficult and somewhat impossible. Via remote sensing, data can be collected for vast areas in a short time. Therefore, it can save time and reduce costs in researches. Moreover, by providing repetitious coverage, we can monitor phenomena. Given the occurrence of gullies, many parameters involved in gully occurrence should be investigated and analysed, and the use of GIS in data analysis is essential.

The results of the RF model corroborate that distance to stream, distance to road and LU/LC parameters have the most impact on gullying in the study area (Reid and Dune, 1996; Conoscenti et al. 2014; Nyssen et al., 2002; Malik, 2008; Vanmaercke et al., 2016). In most cases, the gullies are linked to the stream networks, and the streams cause gullying in areas where conditions are suitable (Conoscenti et al., 2014). Linear infrastructure, such as roads, through the concentration of surface runoffs, the transfer of concentrated runoff to other watersheds and an increase in watershed size, causes the gullying process (Nyssen et al., 2002; Malik, 2008). Results of the model verification showed that: (i) The RF data-mining model has a higher predictive accuracy than the CF statistical model that is line with (Rahmati et al., 2017; Kim et al., 2018; Taalab et al., 2018; Arabameri et al., 2018a). Arabameri et al (2018a) used three data-mining models, including RF, MARS and BRT (boosted regression tree) for gully erosion assessment in the Shahroud watershed (northeastern part of Iran), For this purpose, 12 gully erosion conditioning factors including elevation, slope, aspect, plan curvature, convergence index, TWI, lithology, land use/land cover (LU/LC), distance to rivers, distance to roads, drainage density, and NDVI are used and their results indicate that the RF model with AUC=0.927 has a higher prediction accuracy than MARS and BRT. Rahmati et al (2017) applied different models of FR, support vector machine, artificial neural network, and boosted regression tree for gully erosion susceptibility mapping in Iran. For this purpose, 12 geo-environmental factors of altitude, slope, aspect, curvature, soil texture, lithology, distance to streams, drainage density, TWI, distance to road, and land use are used. Based on their results, RF showed an excellent prediction accuracy; (ii) The integration of CF bivariate statistical model with the RF data-mining model eliminated their disadvantages and increased their efficiency and accuracy. This result is in line with (Rahmati et al., 2017; Pourghasemi et al., 2017; Kornejady et al., 2019). Kornejady et al. (2019) used the integration of FR bivariate statistical model and RF data-mining model for landslide susceptibility assessment in the Chehel-Chai watershed, Golestan Province and stated that the integrated model with AUC=0.831 has a high ability to identify susceptible areas to landslide occurrence;(iii) Given that the study area is divided into several homogeneous units, and the importance of the GRCFs in each unit is calculated, the prediction accuracy of the GWR-RF-CF approach is higher than the individual and integrated models (Yu et al., 2016). Yu et al. (2016) used the GWR technique along with the SVM model and PSO (particle swarm optimisation) algorithm for landslide susceptibility mapping in Wanzhou in the Three Gorges Area in China and stated that GWR has improved the prediction accuracy of the model. Thus, the AUC of SVM (0.817) and PSO-SVM (0.867) has improved by using the GWR-PSO-SVM technique (0.971). Results corroborate that the methodological framework introduced in this research has a reasonably good accuracy in the prediction of areas prone to gully erosion.

The prediction accuracy of future events, such as gully erosion that occurs frequently in the study area and causes damages to agricultural land, infrastructure and the natural ecosystem depend on several factors such as uncertainty in imput data and models and heterogeneity of the study area. Uncertainty can be defined as an occurrence of a phenomenon that is beyond the control of the researcher. Determining the gully erosion susceptibility is always faced with uncertainties due to incomplete knowledge of the studied physical system, the structure of the model, and the temporal and spatial variability. Because the data on which mathematical models are based is not usually sufficient and the algorithm chosen for modelling is not exactly the same as what happens in the nature. Even in very complex models with proper validation, the input data of the model has uncertainties (Rojas and Kahunde, 2010). If the conceptual model and the initial data are accurate and the validation is done correctly, the uncertainty in future predictions will be minimized (Rojas and Kahunde, 2010). The methodological framework introduced in this research uses accurate remote sensing data and integrated (RF-CF) model for reducing the uncertainty in prediction. Additionally, the use of GWR technique solves the heterogeneity issue in the study area by dividing the study area into several homogeneous units and determining the importance of the parameters in each unit. The GWR technique has resolved the problem of changes in the importance of parameters in different parts of a region.

5. Conclusion

. Predicting areas that are susceptible to gullying is useful in implementing protective measures and reducing possible damages. Therefore, in this research, the GWR technique and integration of statistical and data-mining methods are used for the assessment of the gully erosion sensitivity of the Mahabia watershed. For this purpose, 12 GRCFs including elevation, slope, plan curvature, TWI, SPI, drainage density, distance to stream, distance to road, lithology, LU/LC, soil type and rainfall and the gully erosion inventory map (GEIM) training dataset are used for GEZM. Moreover, the GEIM validation dataset is used for validation. Field observations in the Mahabia watershed indicate that gully density is not the same in all parts of the area; thus, the high slopes have a low density due to rocky outcrops and unsaturation of the soils, whereas the low slope areas have a high density. According to the obtained results and field surveys, the most important factors in gully occurrence in the study area can be divided into two groups: natural and man-made. The natural factors include lithology and soil and land gradients, whereas the human factors include land management, road construction, and vegetation eliminate and overgrazing. One of the most important protective measures for controlling and reducing gully processes in the study area can be planting on the edges of and around the gullies. The most important effects of plants on decreasing gully erosion are as follows: (i) increasing the topsoil shear strength through their roots, (ii) slowing down extreme rainfall runoff and trapping sediments, (iii) drying out and reducing the soil saturation by evapotranspiration process and (iv) adjusting overland flow and infiltration patterns, thus affecting subsurface drainage. Based on validation results, GWR technique has an effective role in increasing the prediction accuracy of the integrated CF-RF model. In summary, results indicate that combination of GWR techniques with integrated models along with remote sensing and GIS techniques as a powerful and efficient tool can be used for prediction of gully erosion with low uncertainty and resonable prediction accuracy. Given that the methodology presented in this study has high efficiency and precision in identifying areas susceptible to gully erosion, using it in areas with similar climatic and topographic conditions is recommended. The results of this research can be used by decision makers and managers in soil and water conservation, and development planning, such as road construction, determining the direction of the electricity and gas transmission lines.

Acknowledgement

This research is funded by the UTS under Grants 321740.2232335 and 321740.2232357.

References

Arabameri, A., Pourghasemi, H.R., Yamani, M. 2017a. Applying different scenarios for landslide spatial modeling using computational intelligence methods. Environmental Earth Sciences. 76, 832. https://doi.org/10.1007/s12665-017-7177-5

Arabameri, A., Pourghasemi, H.R., Cerda, A. 2017b. Erodibility prioritization of sub-watersheds using morphometric parameters analysis and its mapping: A comparison among TOPSIS, VIKOR, SAW, and CF multi-criteria decision making models. Science of the Total Environment. 613–614, 1385–1400. https://doi.org/10.1016/j.scitotenv.2017.09.210

Arabameri, A., Pradhan, B., Pourghasemi, H.R., Rezaei, K., Kerle, N. 2018a. Spatial Modelling of Gully Erosion Using GIS and R Programing: A Comparison among Three Data Mining Algorithms. Applied sciences. 8(8), 1369. https://doi.org/10.3390/app8081369

Arabameri, A., Pradhan, B., Rezaei , K., Yamani, M., Pourghasemi, H.R., Lombardo, L., 2018b. Spatial modelling of gully erosion using Evidential Belief Function, Logistic Regression and a new ensemble EBF–LR algorithm. Land Degradation and Development. https://doi.org/10.1002/ldr.3151

Arabameri, A., Rezaei , K., Pradhan, B., Pourghasemi, H.R., Lee, S., Yamani, M. 2018c. GIS-based gully erosion susceptibility mapping: a comparison among three data-driven models and AHP knowledge-based technique. Environmental Earth Sciences. 77, 628.

Ahlmer, A.-K., Cavalli, M., Hansson, K., Koutsouris, A. J., Crema, S., Kalantari, Z. 2018. Soil moisture remote-sensing applications for identification of flood-prone areas along transport infrastructure. Environmental Earth Sciences. 77(14), 533.

Alam, M., Hussain, R.R., Islam, A.S. 2016. Impact assessment of rainfall-vegetation on sedimentation and predicting erosion-prone region by GIS and RS. Geomat Nat Haz Risk. 7,667–679. https://doi.org/10.1080/19475705.2014.942387

Barnes, N., Luffman, I., Nandi, A., 2016. Gully erosion and freeze-thaw processes in clayrich soils, northeast Tennessee, USA. GeoResJ 9–12, 67–76.

Ben Slimane, A., Raclot, D., Evrard, O., Sanaa, M., Lefevre, I., Le Bissonnais, Y. 2016. Relative Contribution of Rill/Interrill and Gully/Channel Erosion to Small Reservoir Siltation in Mediterranean Environments. Land Degradation & Development. 27 (3), 785-797. https://doi.org/10.1002/ldr.2387

Bonacina, L.C.W. 1945. Orographic rainfall and its place in the hydrology of the globe. Q J R Meteorol Soc. 71, 41–55. https://doi.org/10.1002/qj.49707130705

Boardman, J., Favis-Mortlock, D. 1998. Modelling Soil Erosion by Water. First edition, Springer-Verlag Berlin Heidelberg. 10.1007/978-3-642-58913-3

Boardman, J. 2014. How old are the gullies (dongas) of the sneeuberg uplands, eastern karoo, South Africa. Catena. 113, 79-85. https://doi.org/10.1016/j.catena.2013.09.012

Breiman, L., Adele, C. 2013. Breiman and Cutler's random forests for classification and regression. MASS 4, 6-17. 10.1023/A:1010933404324

Brunsdon, C., Fotheringham, S., Charlton, M. 1998. Spatial nonstationarity and autoregressive models. Environ. Plann. 30, 957–973. https://doi.org/10.1068/a300957

Celik, M., Kazar, B.M., Shekhar, S., Boley, D. 2016. Parameter estimation for the spatial autoregression Model: A Rigorous Approach. 10.3844/jmssp.2013.169.174

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

Conforti, M., Pascale, S., Robustelli, G., Sdao, F. 2014. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). Catena. 113, 236–250. https://doi.org/10.1016/j.catena.2013.08.006

Conoscenti, C., Agnesi, 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 Degradation and Development. 29 (3), 724-736. https://doi.org/10.1002/ldr.2772

Conoscenti, C., Angileri, S., Cappadonia, C., Rotigliano, E., Agnesi, V., Märker, M. 2014. Gully erosion susceptibility assessment by means of GIS-based logistic regression: a case of Sicily (Italy). Geomorphology. 204 (1), 399–411. https://doi.org/10.1016/j.geomorph.2013.08.021

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

Chen, Z., Liang, S., Ke, Y., Yang, Z., Zhao, H. 2017. Landslide susceptibility assessment using evidential belief function, certainty factor and frequency ratio model at Baxie River basin, NW China. Geocato International. https://doi.org/10.1080/10106049.2017.1404143

Chalkias, C., Kalogirou, S., Ferentinou, M. 2014. Landslide susceptibility, Peloponnese peninsula in south Greece. J. Maps. 10, 211–222. https://doi.org/10.1080/17445647.2014.884022

Dewitte, O., Daoudi, M., Bosco, C., Eeckhaut, M. 2015. Predicting the susceptibility to gully initiation in data-poor regions. Geomorphology. 228, 101–115. https://doi.org/10.1016/j.geomorph.2014.08.010

Dou, J., Oguchi, T., Hayakawa, Y.S., Uchiyama, S., Saito, H., Paudel, U. 2014. GIS-based landslide susceptibility mapping using a certainty factor model and its validation in the Chuetsu area, central Japan; In: Sassa K., Canuti P., Yin Y. (eds) Landslide Science for a Safer Geoenvironment. Springer, Cham. https://doi.org/10.1007/978-3-319-05050-8_65

Dube, E., Nhapi, I., Murwira, A., Gumindoga, W., Goldin, J., Mashauri, D, A. 2014. Potential of weight of evidence modeling for gully erosion hazard assessment in Mbire District-Zimbabwe, Physics and Chemistry of the Earth. 67(69), 145-152. https://doi.org/10.1016/j.pce.2014.02.002

Duda, R.O., Hart, P.E., Stork, D.G. 2001. Pattern classification. 2nd Edition. Wiley.

Dymond, J.R., Herzig, A., Basher, L., Betts, H.D., Marden, M., Phillips, C.J., Ausseil, A.G.E., Palmer, D.J., Clark, M., Roygard, J. 2016. Development of a New Zealand SedNet model for assessment of catchment-wide soil-conservation works. Geomorphology. 257, 85–93. https://doi.org/10.1016/j.geomorph.2015.12.022

Ezechi, J. I. 2000. The influnce of runoff, lithology and water table on the dimensions and rate of gullying Processes in Eastern, Nigeria. Elsevier, Catena Verlag. Cremlingen (Web-Site).

Fan, W., Wei, X.S., Cao, Y.B., Zheng, B. 2017. Landslide susceptibility assessment using the certainty factor and analytic hierarchy process. Journal of Mountain Science. 14 (5), 906-925. https://doi.org/10.1007/s11629-016-4068-2

Feuillet, T., Coquin, J., Mercier, D., Cossart, E., Decaulne, A., Jónsson, H.P., Sæmundsson, B. 2014. Focusing on the spatial non-stationarity of landslide predisposing factors in Northern Iceland: Do paraglacial factors vary over space? Prog. Phys. Geogr. 2014. 38, 354–377. https://doi.org/10.1177/0309133314528944

Fotheringham, A.S., Charlton, M.E., and Brunsdon, C. 2001. Spatial variations in school performance, a local analysis using geographically weighted regression. Geographical and Environmental Modelling. 5(1), 43-66. https://doi.org/10.1080/13615930120032617

Fotheringham, A.S., Brunsdon, C., Charlton, M. 2003. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. First edition, John Wiley & Sons: New York, USA.

Geological Survey Department of Iran (GSDI), 2012. Available online: http://www.mazan daranmet.ir.

National Geographic Organization of Iran (NGOI), 2008. http://www.ngo-org.ir/

Go´mez-Gutie´rrez, A., Conoscenti, C., Angileri, S.E., Rotigliano, E., Schnabel, S. 2015. Using topographical attributes to evaluate gully erosion proneness (susceptibility) in two mediterranean basins: advantages and limitations. Nat Hazards. 79, S291–S314. https://doi.org/10.1007/s11069-015-1703-0

Goodwin, N.R., Armston, J.D., Muir, J., Stiller, I. 2017. Monitoring gully change: a comparison of airborne and terrestrial laser scanning using a case study from Aratula, Queensland. Geomorphology. 282, 195–208. https://doi.org/10.1016/j.geomorph.2017.01.001

Hancock, G.R., Evans, K.G. 2005. Channel head location and characteristics using digital elevation models. Earth Surf Proc Land. 31, 809–824. https://doi.org/10.1002/esp.1285

Heckeman, 1986. Probabilistic interpretation of MYCIN's certainty factors; In: Uncertainty in artificial intelligence (eds) Kanal L N and Lemmer J F (New York: Elsevier), 298–311.

Ilinca, V., Gheuca, I., 2011. The red lake landslide (Ucigau Mountain, Romania). Carpathian Jour. Earth Environ. 23, 263-272. 10.1007/978-94-007-6757-7

I.R. of Iran Meteorological Organization. 2012. Available online: http://www.mazan daranmet.ir.

Keller .Edward A., 2011. Environmental Geology,University of California, ,Pearson, ISBN-13.978-0-321-64375-9

Karlsson, C. S. J., Kalantari, Z., Mörtberg, U., Olofsson, B., Lyon, S. W. 2017. Natural Hazard Susceptibility Assessment for Road Planning Using Spatial Multi-Criteria Analysis. Environmental Management. 60(5), 823-851.

Kim, J.C., Lee, S., Jung, H.S., Lee, S. 2018. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. Geocato International. 33 (9), 1000-1015. https://doi.org/10.1080/10106049.2017.1323964

Kou, M., Jiao, J., Yin, Q., Wang, N., Wang, Z., Li, Y., Yu, W., Wei, Y., Yan, F., Cao, B. 2016. Successional Trajectory Over 10 Years of Vegetation Restoration of Abandoned Slope Croplands in the Hill-Gully Region of the Loess Plateau, Land Degradation & Development. 27 (4), 919-932. https://doi.org/10.1002/ldr.2356

Kornejady, A., Pourghasemi, H.R., Afzali, S.F. 2019. Presentation of RFFR New Ensemble Model for Landslide Susceptibility Assessment in Iran. In: Pradhan S., Vishal V., Singh T. (eds) Landslides: Theory, Practice and Modelling. Advances in Natural and Technological Hazards Research, vol 50. Springer, Cham

Lesschen, J. P., Kok, K., Verburg, P. H., Cammeraat, L. H. 2007. Identification of Vulnerable Areas for Gully Erosion under Different Scenarios of Land Abandonment in Southeast Spain, CATENA. 71(1), 110-121. https://doi.org/10.1016/j.catena.2006.05.014

Lo, C.P., Yeung, A.K.W., 2002. Concepts and Techniques of Geographic Information System. New Jersey: Pearson Education Inc.

Märker, M., Pelacani, S., Schröder, B. 2011. A functional entity approach to predict soil erosion processes in a small Plio-Pleistocene Mediterranean catchment in Northern Chianti, Italy. Geomorphology. 125, 530–540. https://doi.org/10.1016/j.geomorph.2010.10.022

Malik, I. 2008. Dating of small gully formation and establishing erosion rates in old gullies under forest by means of anatomical changes in exposed tree roots (Southern Poland). Geomorphology. 93(3), 421-36. https://doi.org/10.1016/j.geomorph.2007.03.007

Moore, I.D., Grayson, R.B., Ladson, A.R. 1991. Digital terrain modeling: a review of hydrological, geomorphological and biological applications. Hydrol. Process. 5, 3–30. https://doi.org/10.1002/hyp.3360050103

Morgan, R. P. C. 1990. Erosion and Soil Conservation. third edition, Blackwell publishing.

Nakaya T. 2003. Local Spatial Interaction Modelling Based on the Geographically Weighted Regression Approach. In: Boots B., Okabe A., Thomas R. (eds) Modelling Geographical Systems. The GeoJournal Library, vol 70. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-2296-4_4

Nyssen, J., Poesen, J., Moeyersons, J., Luyten, E., Veyret Picot, M., Deckers, J., Mitiku, H., Govers, G. 2002. Impact of road building on gully erosion risk, a case study from the northern Ethiopian highlands. Earth Surface Processes and Landforms. 27 (12), 1267– 1283. https://doi.org/10.1002/esp.404

Nwankwo, C., Nwankwoala, H.O. 2018. Gully Erosion Susceptibility Mapping in Ikwuano Local Government Area of Abia State, Nigeria Using GIS Techniques. Earth Sciences Malaysia (ESMY). 2(1), 8-15. 10.26480/esmy.01.2018.08.15

Paolo, P., Desmond, W. E., Antonina, C. 2014. Using 137 CS and 210 Pbex measurements and conventional surveys to investigate the relative contributions of inter rill/rill and gully erosion to soil loss from a small cultivated catchment in Sicily, Soil and Tillage Research. 135 (1), 18-27. http://dx.doi.org/10.1016/j.still.2013.08.013

Payet, N., Todorovic, S. 2008. Random forest random field. Technical report. School of Electrical Engineering and Computer Science Oregon State University. USA.

Pradhan, B., Lee, S. 2010. Landslide susceptibility assessment and factor effect analysis:back propagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environmental Modelling & Software. 25, 747-759. https://doi.org/10.1016/j.envsoft.2009.10.016

Prosser, I.P. 1996. Thresholds of channel initiation in historical and Holocene times, southeastern Australia. Advances in hillslope processes. 2, 687-708.

Poesen, J., Vandaele, K., van Wesemael, B. 1998. Gully erosion: importance and model implications. In: Boardman, J., Favis-Mortlock, D. (Eds.), Modelling Soil Erosion by Water. Springer, Berlin. https://doi.org/10.1007/978-3-642-58913-3_22

Pourghasemi, H.R., Yousefi, S., Kornejady, A., Cerdà, A. 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modelling. Science of the Total Environment. 609, 764–775. 10.1016/j.scitotenv.2017.07.198

Raouf, A., Peng, Y., Shah, T.I. 2017. Integrated use of aerial photographs and LiDAR images for landslide and soil erosion analysis: a case study of Wakamow Valley, Moose Jaw, Canada. Urb Sci. 1, 20–28. https://doi.org/10.3390/urbansci1020020

Rahim, A., Maria Ali, S., Aslam, M. 2018. GIS Based Landslide Susceptibility Mapping with Application of Analytical Hierarchy Process in District Ghizer, Gilgit Baltistan Pakistan, Journal of Geoscience and Environment Protection, 2018, 6, 34-49.

Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H.R., Feizizadeh, B. 2017. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. Geomorphology. 298, 118-137. https://doi.org/10.1016/j.geomorph.2017.09.006

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

Refahi, H. 2009. Soil erosion by water & conservation. Tehran University Press. 10-202. (In Farsi with English Summary)

Rojas, R., Kahunde, S. 2010. Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling. Journal of Hydrology. 394(3), 416-435.

Comino, J.R., Iserloh, T., Lassu, T., Cerdà, 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

Reid, L.M., Dune, T. 1996. Rapid Evaluation of Sediment Budgets. First edition, Catena Verlag Gmbh, Reiskirchen, Germany.

Sabokbar, H.F., Roodposhti, M.S., Tazik, E. 2014. Landslide susceptibility mapping using geographically-weighted principal component analysis. Geomorphology. 226, 15–24. https://doi.org/10.1016/j.geomorph.2014.07.026

Shortliffe, E.H., Buchanan, B.G. 1975. A model of inexact reasoning in medicine; Math. Biosci. 23(3), 351–379. https://doi.org/10.1016/0025-5564(75)90047-4

Sharma, S., Kumar Mahajan, A. 2018. Comparative evaluation of GIS-based landslide susceptibility mapping using statistical and heuristic approach for Dharamshala region of Kangra Valley, India. Geoenvironmental Disasters. 5, 4. https://doi.org/10.1186/s40677-018-0097-1

Shit, P. K., Paira, R., Bhunia, G., Maiti, R. 2015. Modeling of potential gully erosion hazard using geo-spatial technology at Garbheta block, West Bengal in India. Modeling Earth Systems and Environment. 1(1-2), 1-16. https://doi.org/10.1007/s40808-015-0001-x

Singh, O., Singh, J. 2018. Soil Erosion Susceptibility Assessment of the Lower Himachal Himalayan Watershed. Journal of the Geological Society of India. 92 (2), 157-165. 10.1007/s12594-018-0975-x

Swarnkar, S., Malini, A., Tripathi, S., Sinha, R. 2018. Assessment of uncertainties in soil erosion and sediment yield estimates at ungauged basins: an application to the Garra River basin, India. Hydrol. Earth Syst. Sci. 22, 2471–2485. https://doi.org/10.5194/hess-22-2471-2018

Tobler, W.R. 1970. A Computer movie simulating urban growth in the Detroit region. Economic Geography. 46(2), 234-240. 10.2307/143141

Torri, D., Poesen, J. 2014. A review of topographic threshold conditions for gully head development in different environments, Earth Science Reviews. 130 (1), 73-85. https://doi.org/10.1016/j.earscirev.2013.12.006

Valentin, C., Poesen, J., Li, Y. 2005. Gully erosion: impacts, factors and control. Catena. 63(2), 132-53. https://doi.org/10.1016/j.catena.2005.06.001

Vanmaercke, M., Poesen, J., Van Mele, B., Demuzere, M., Bruynseels, A., Golosov, V., Bezerra, J. F. R., Bolysov, S., Dvinskih, A., Frankl, A. 2016. How fast do gully headcuts retreat?. Earth-Science Reviews. 154, 336-355. https://doi.org/10.1016/j.earscirev.2016.01.009

Wang, Q., Ni, J., Tenhunen, J. 2005. Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. Global Ecology and Biogeography. 14(4), 379-393. https://doi.org/10.1111/j.1466-822X.2005.00153.x

Wade, A. 1935. The relationship between topography and geology. Aust Surv. 5, 367–371. https://doi.org/10.1080/00050326.1935.10436440

Wang, Q., Li, W., Chen, W., Bal, H. 2015. GIS-based assessment of landslide susceptibility using certainty factor and index of entropy models for the Qianyang County of Baoji city, China. J. Earth Syst. Sci. 124 (7), 1399–1415

Wheeler, D.C. 2014. Geographically Weighted Regression. In: Fischer M., Nijkamp P. (eds) Handbook of Regional Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23430-9_77

Yilmaz, I. 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison (case study: Kat landslides). Computer and Geosciences. 35, 1125-1138. https://doi.org/10.1016/j.cageo.2008.08.007

Yesilnacar, E.K. 2005. The application of computational intelligence to landslide susceptibility mapping in Turkey. University of Melbourne, Department.

Yu, X., Wang, Y., Niu, R., Hu, Y. 2016. A Combination of Geographically Weighted Regression, Particle Swarm Optimization and Support Vector Machine for Landslide Susceptibility Mapping: A Case Study at Wanzhou in the Three Gorges Area, China. Int J Environ Res Public Health. 13, 487. 10.3390/ijerph13050487.

Zakharov, S.A. 1940. Importance of slope aspect and gradient for soil and vegetation distribution in the Great Caucasus. Journal Botanique de l'URSS. 4–5, 387–405. (in Russian)

Zakerinejad, R., Maerker, M. 2014. Prediction of gully erosion susceptibilities using detailed terrain analysis and maximum entropy modeling: a case study in the Mazayejan Plain, Southwest Iran. Geogr. Fis. Din. Quat. 37 (1), 67–76. 10.4461/GFDQ.2014.37.7

Zabihi, M., Mirchooli, F., Motevalli, A., Darvishan, A.K., Pourghasemi, H.R., Zakeri, M.A., Sadighi, F. 2018. Spatial modelling of gully erosion in Mazandaran Province, northern Iran. Catena. 161, 1–13. https://doi.org/10.1016/j.catena.2017.10.010

Zhang, L., Shi, H. 2004. Local modeling of tree growth by geographically weight regression. Forest Science. 50(2), 225-244.

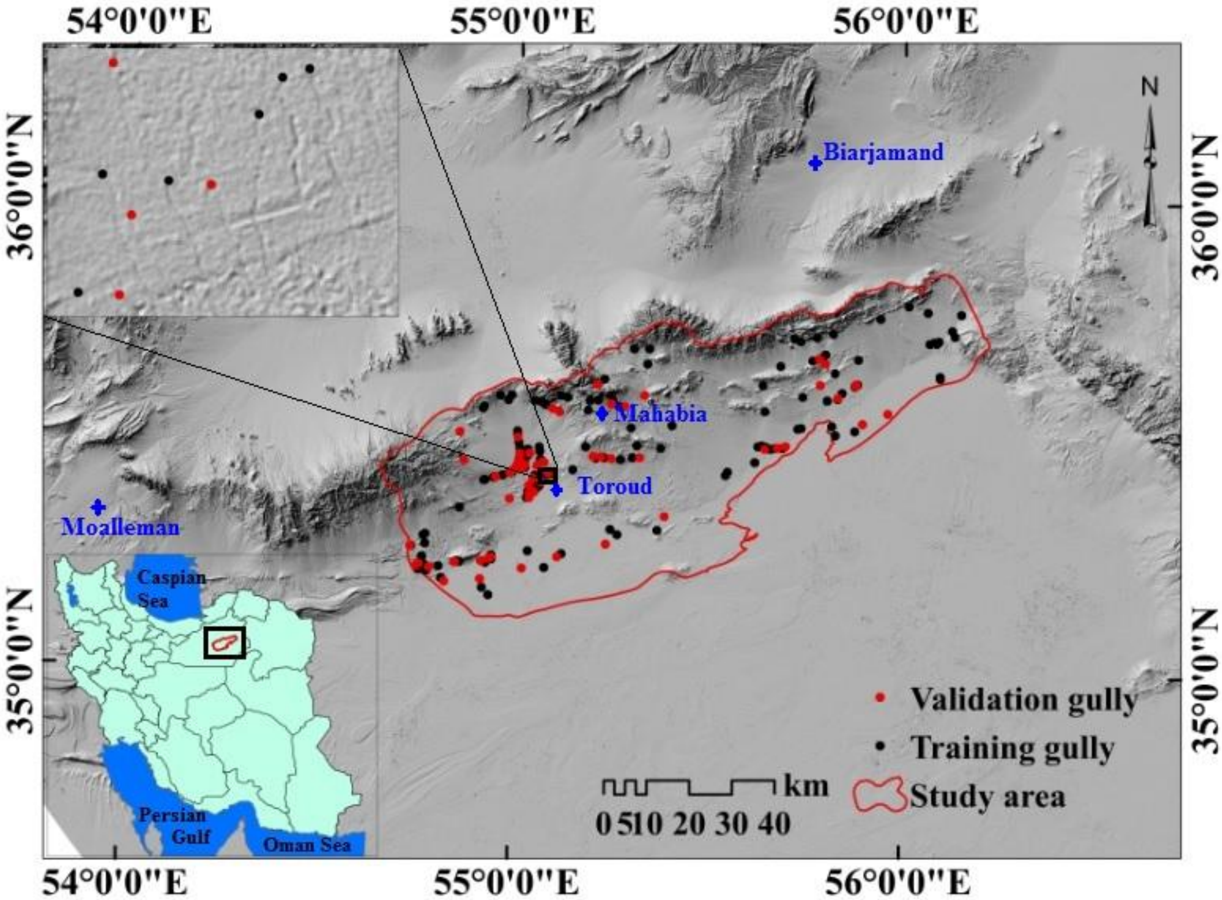


Fig 1. Study area and location of gullies.

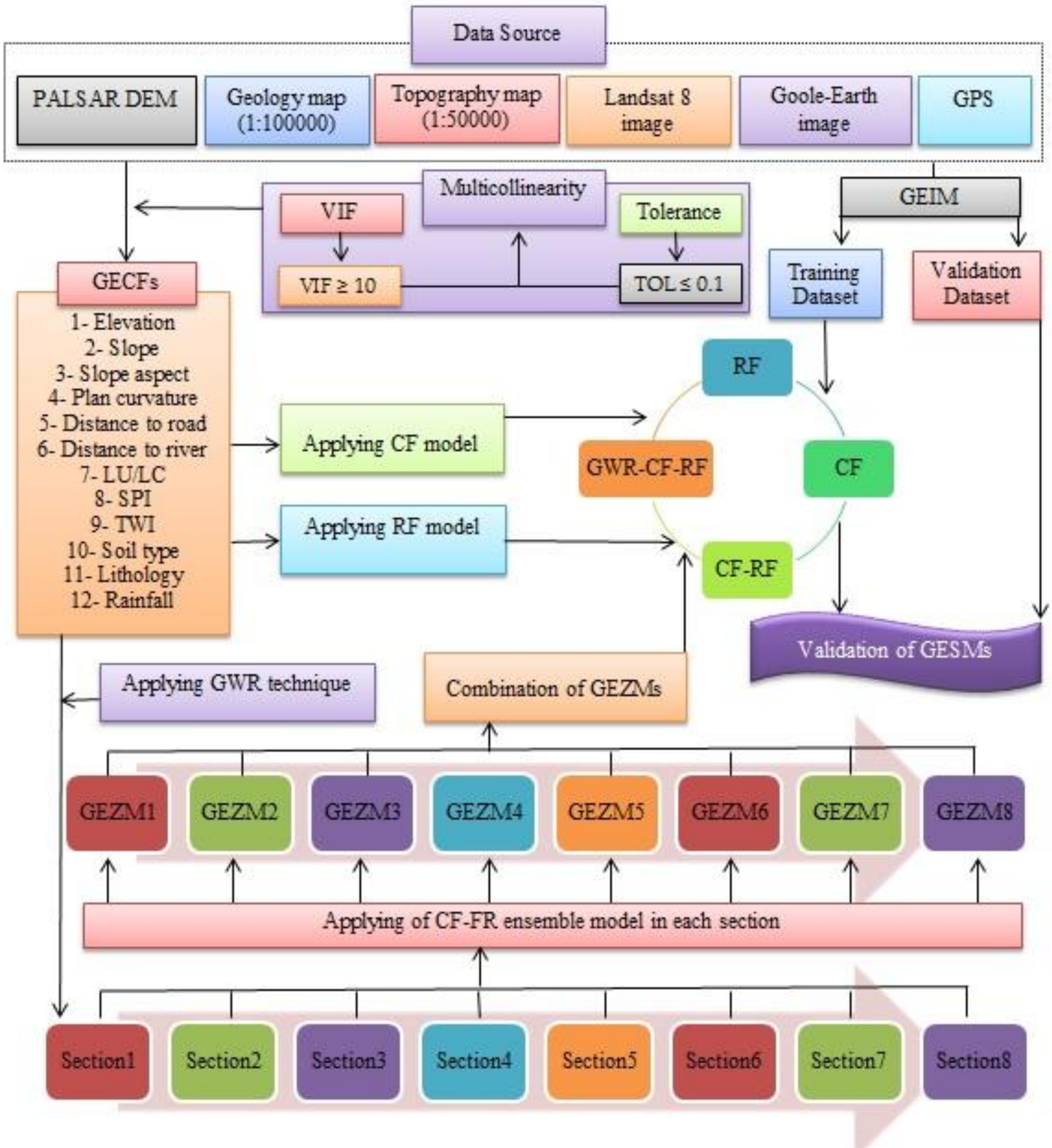


Fig 2. Methodology of research applied in this study.

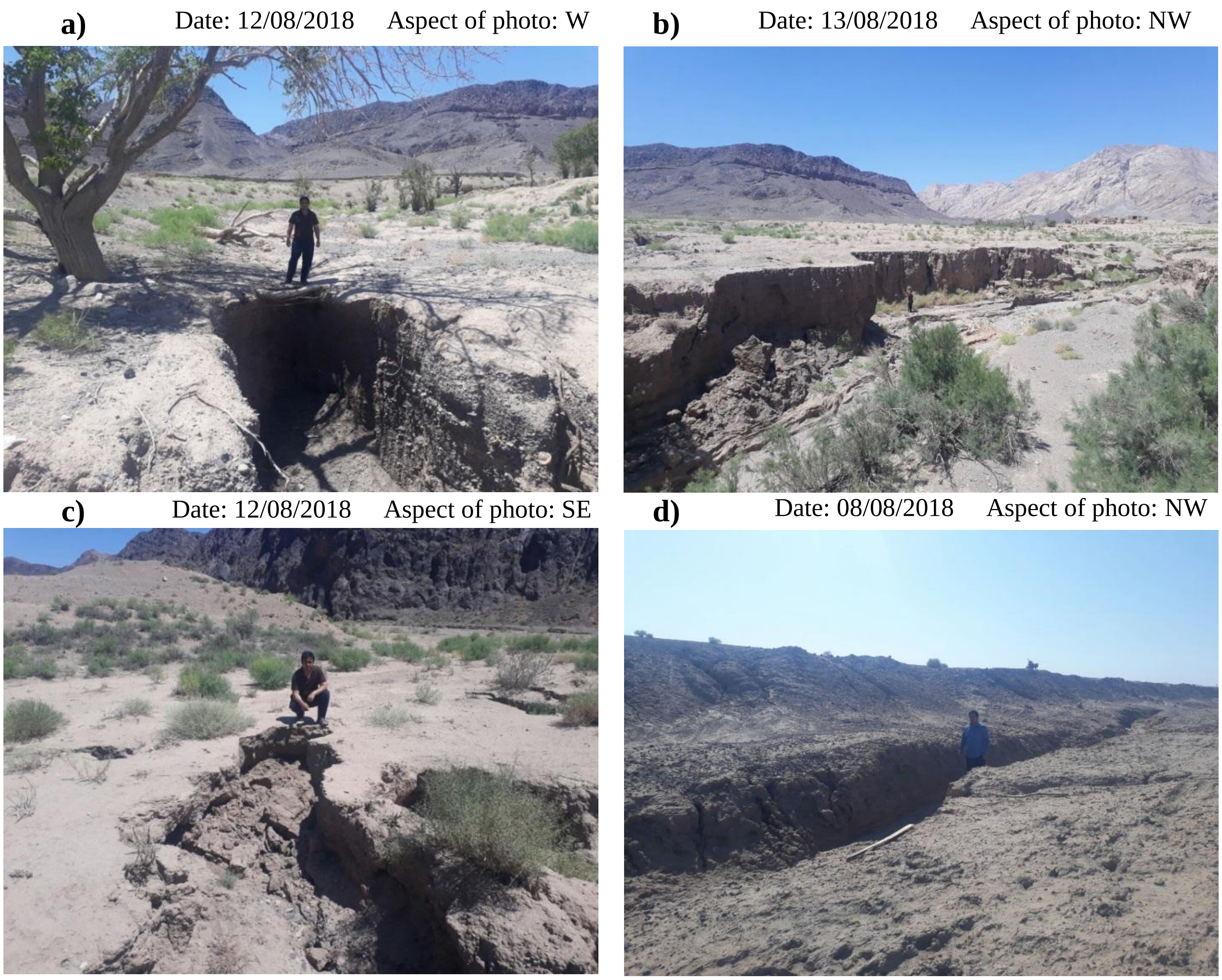


Fig 3. Sample of gullies in the study area and its destructive effect in agricultural lands (a, c), residential area (b), and infrastructure (d).

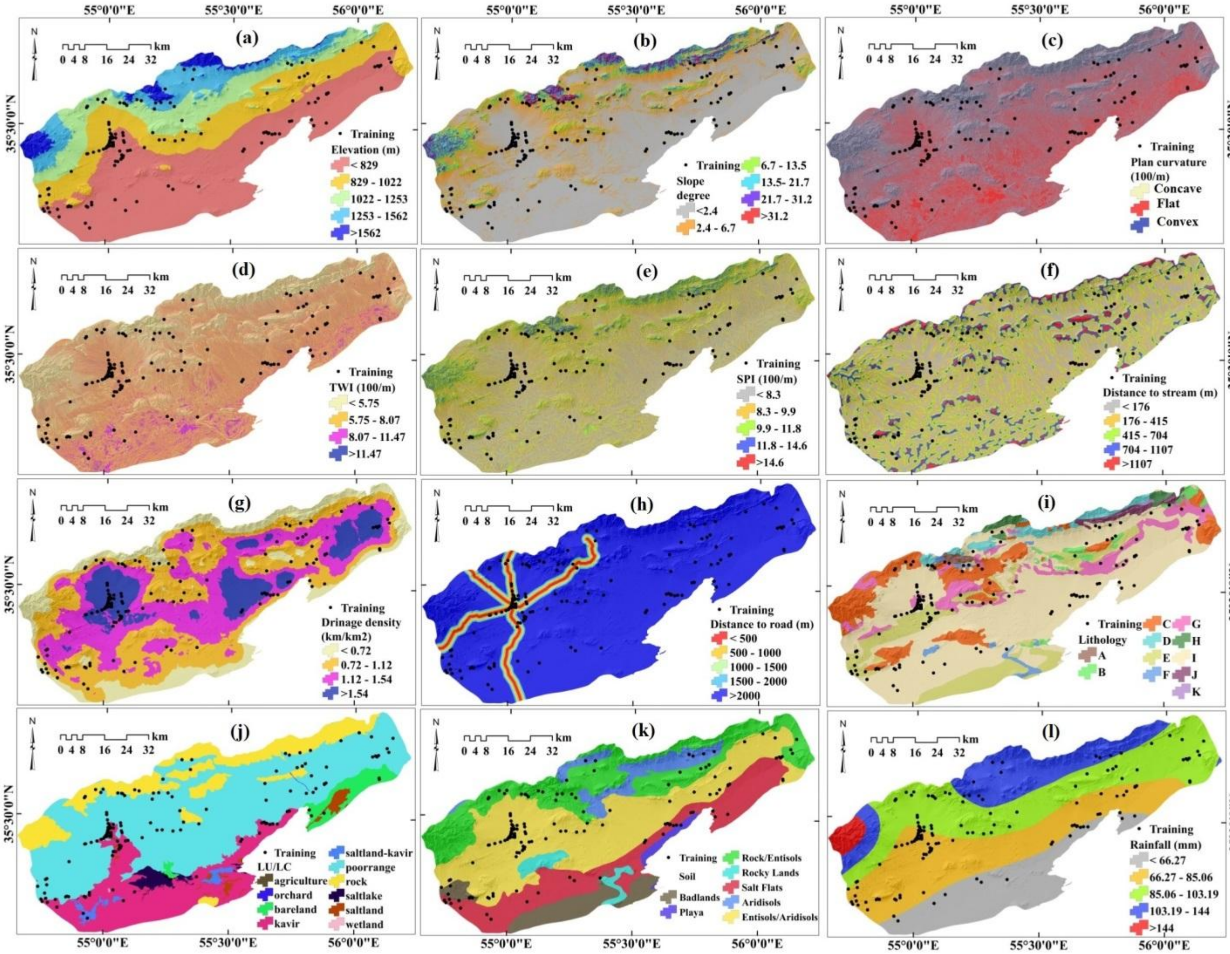


Fig 4. Gully erosion conditioning factors (GECFs). a) elevation, b) slope, c) plan curvature, d) topography wetness index (TWI), e) stream power index (SPI), f) distance to stream, g) drainage density, h) distance to road, i) lithology, j) land use/land cover, k) soil type, and l) rainfall.

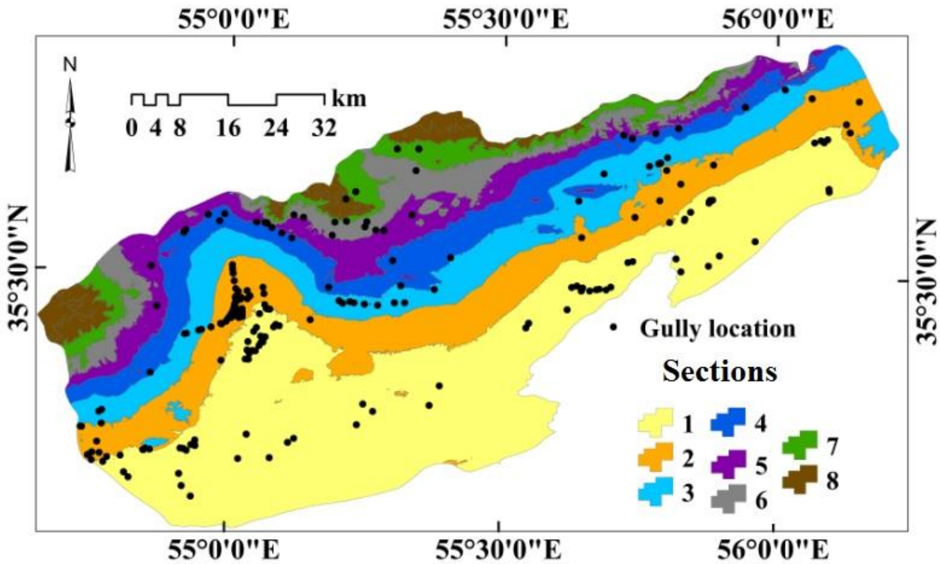


Fig 5. Segmentation of study area using geographically weighted regression (GWR) technique.

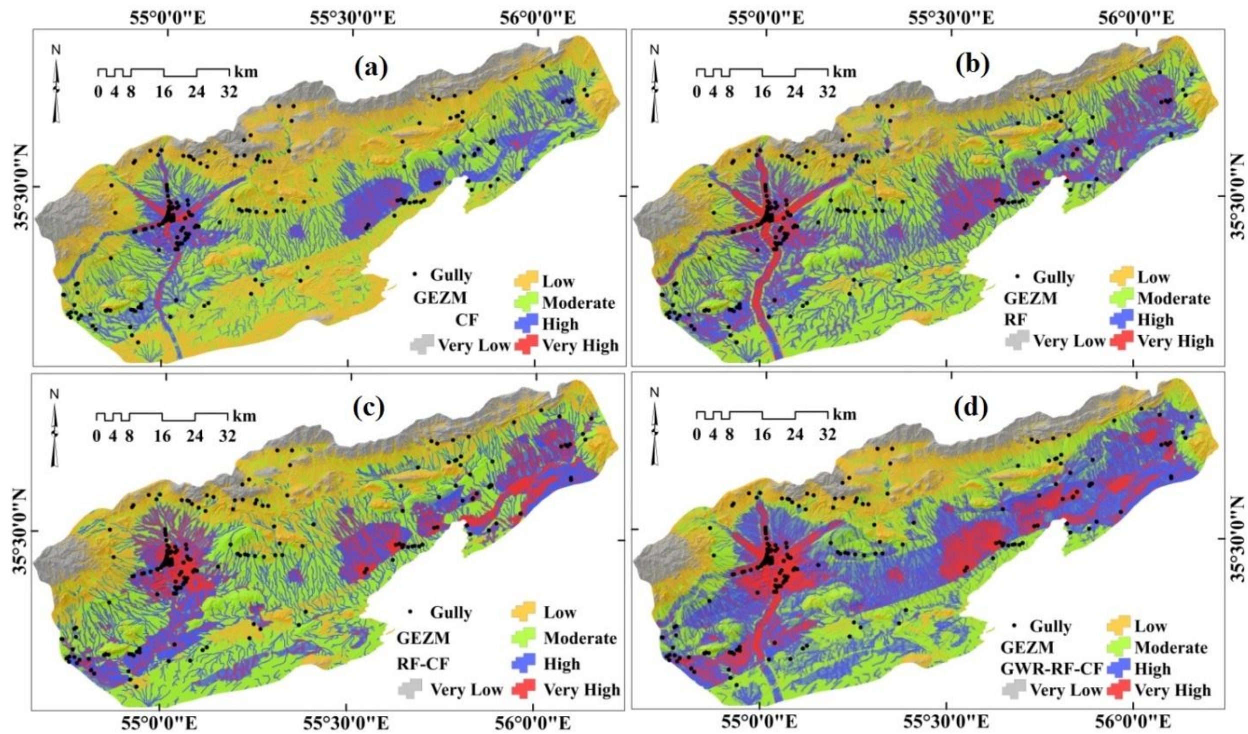


Fig 6. Gully erosion zonation map: (a) certainty factor (CF), (b) random forest (RF), (c) certainty factor - random forest (CF-RF) model, and (d) geographically weighted regression- certainty factor - random forest (GWR-CF-RF) model.

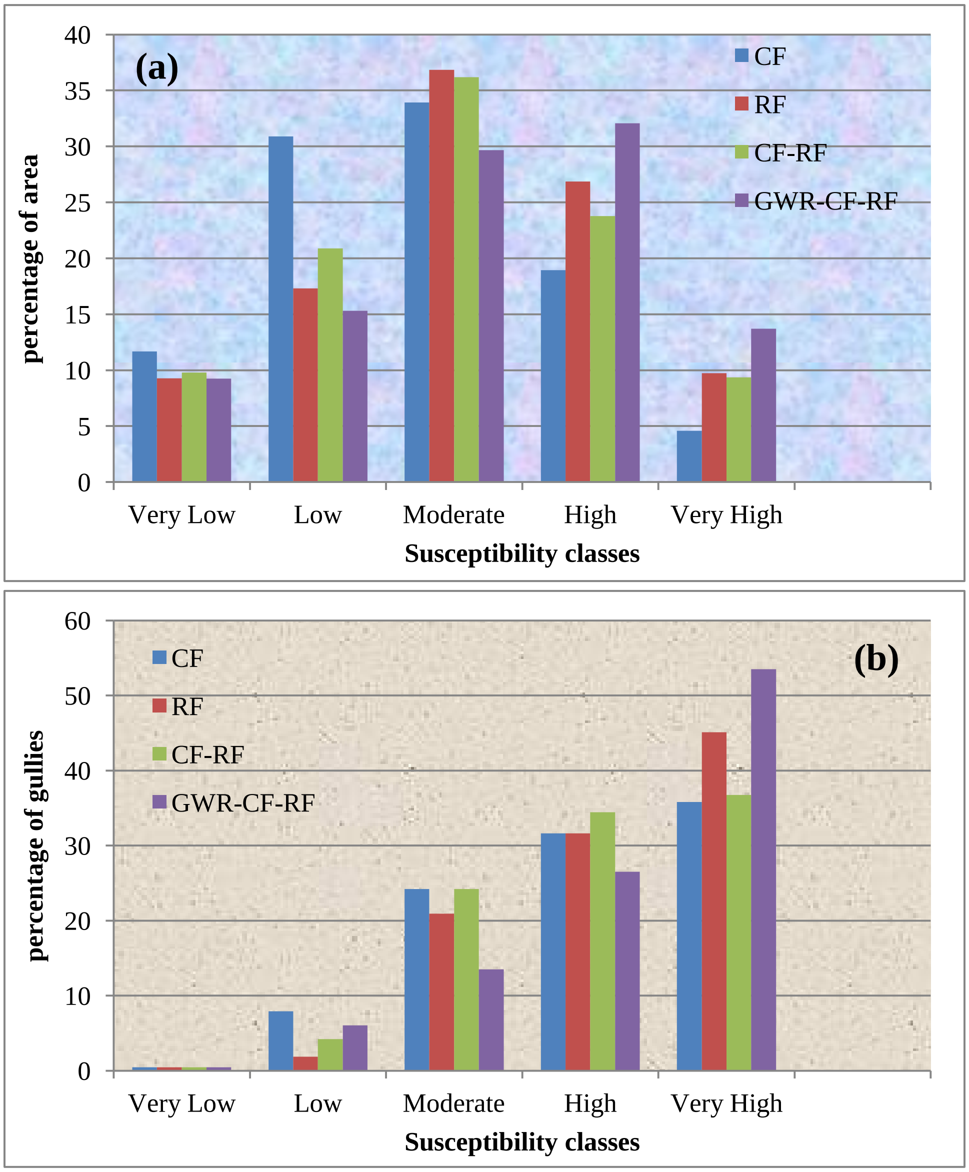


Fig 7. Percentage of area (a) and gullies (b) in each susceptibility classes.

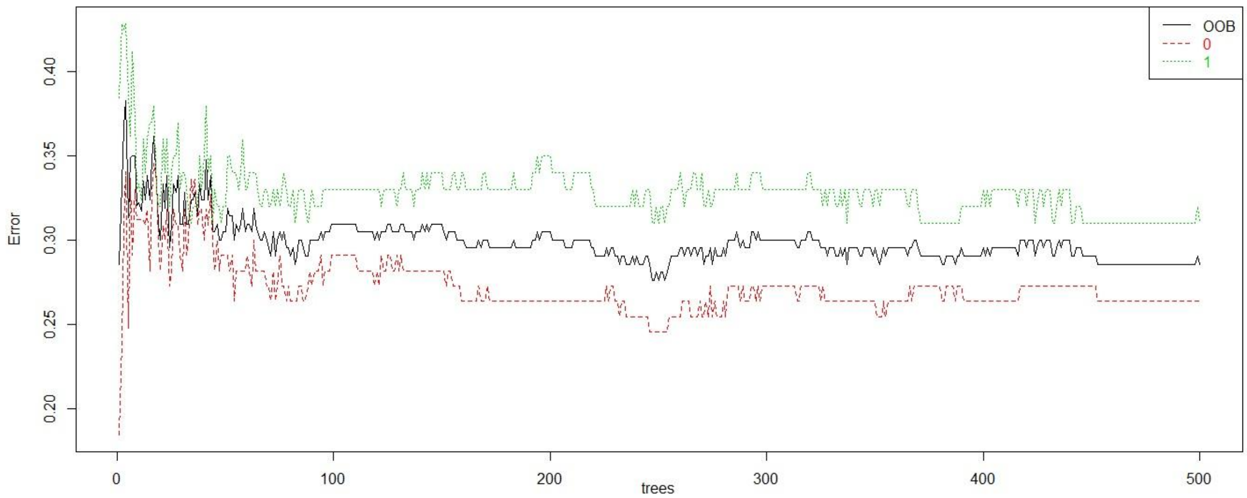


Fig 8. Out-of-bag (OOB) error in random forest (RF) model.

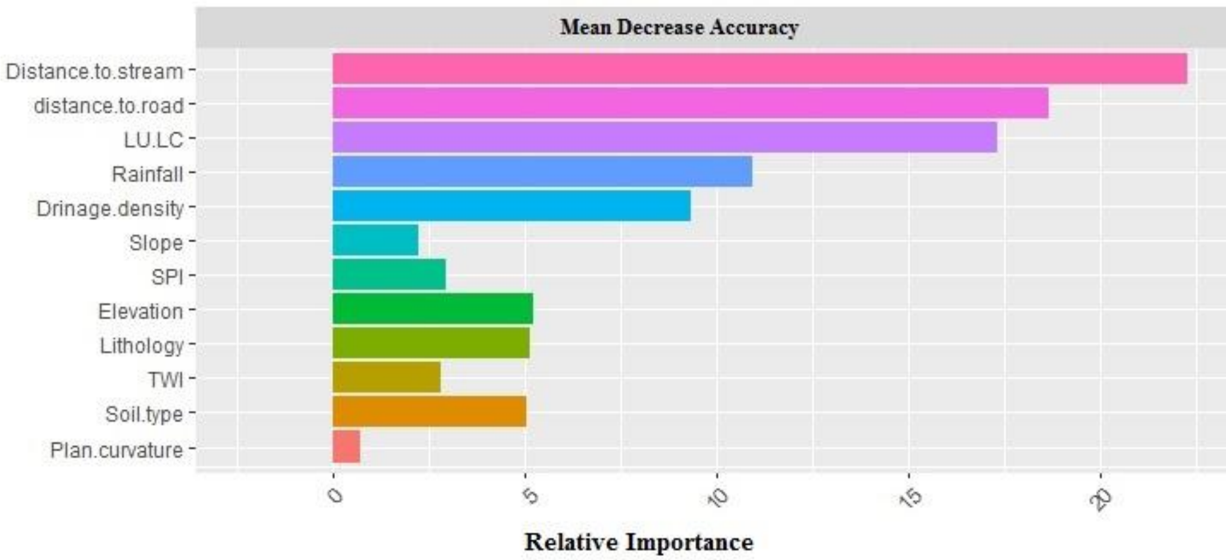


Fig 9. Relative importance of gully-related causal factors (GRCFs) using random forest model.

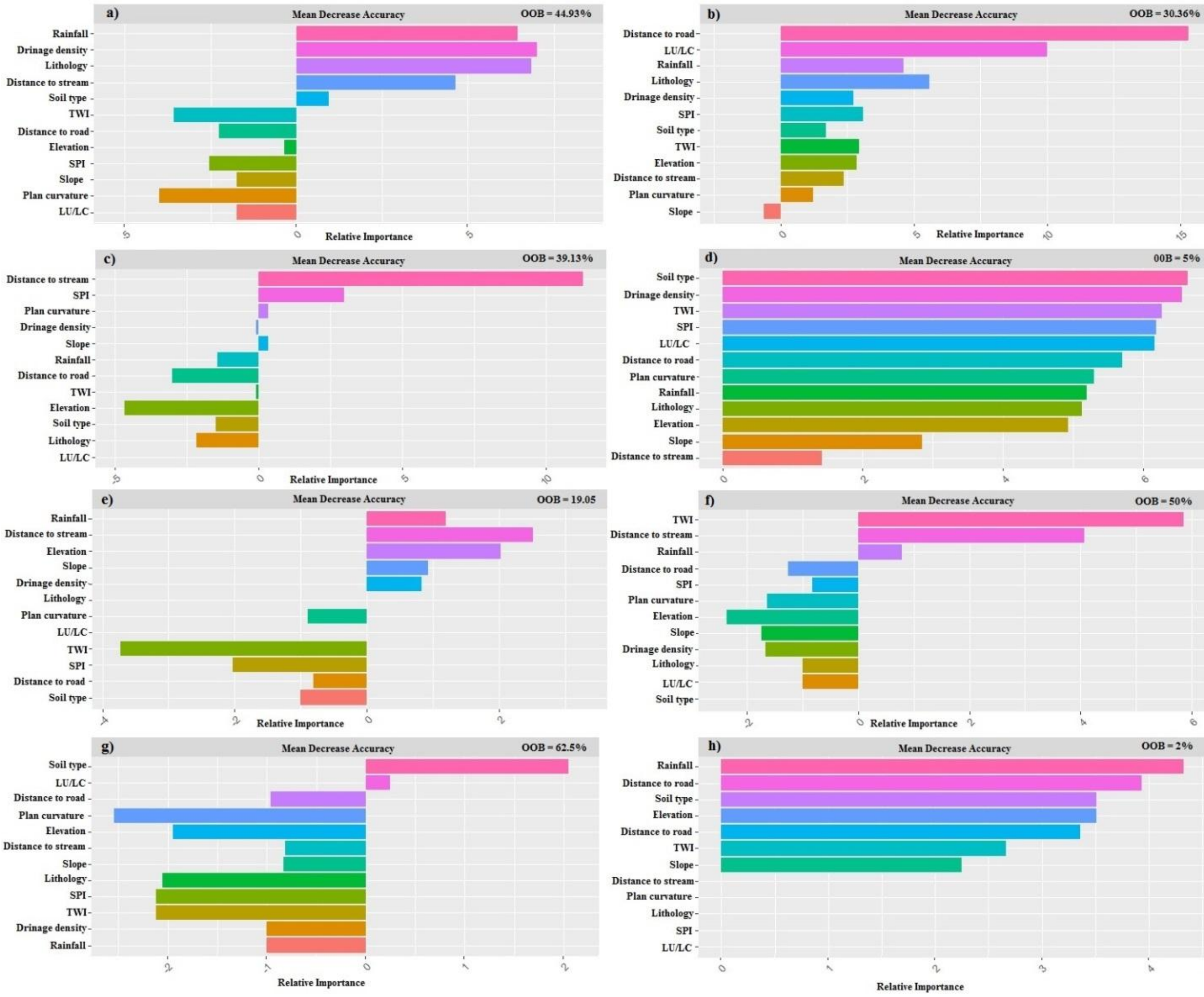


Fig 10. Importance of gully-related causal factors (GRCFs) in sections using random forest model.

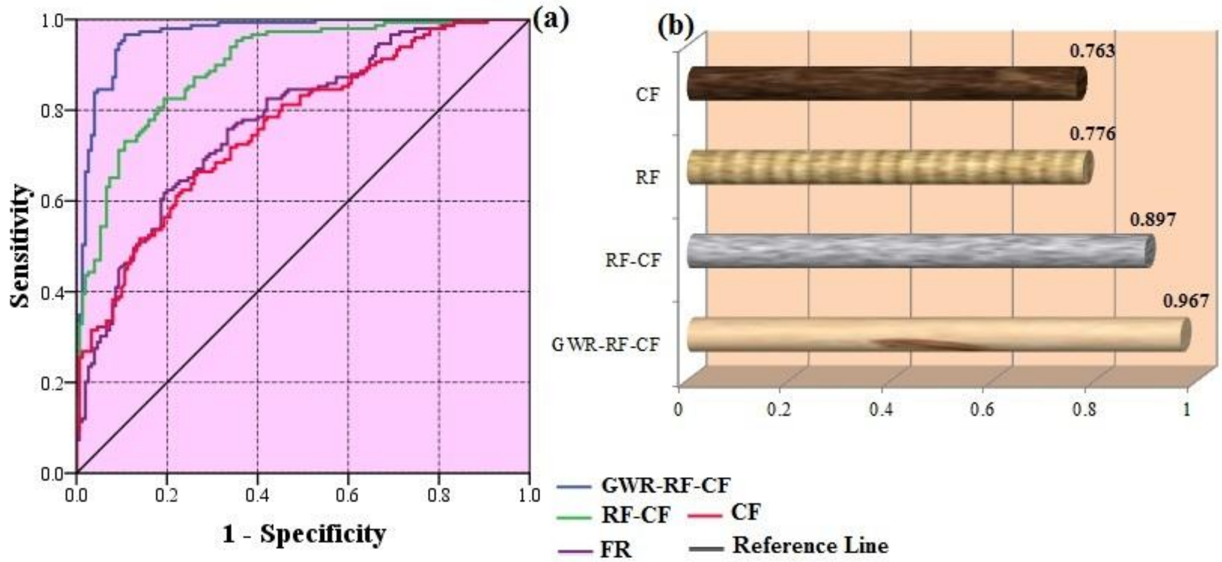


Fig 11. Validation of models: (a) Receiver Operating Characteristic (ROC) curve, and (b) area under curve values.

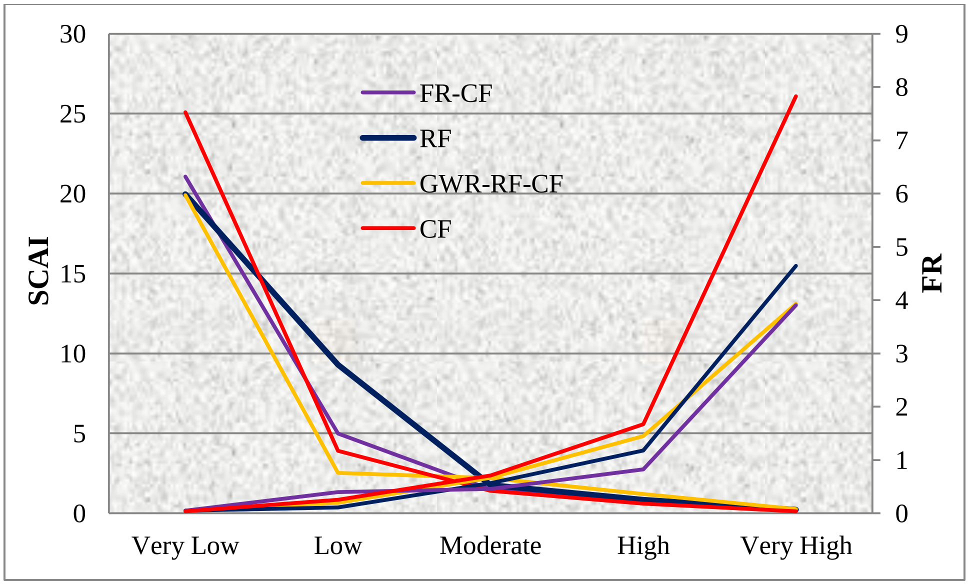


Fig 12. Values of frequency ratio and seed cell area index for different susceptibility classes.

Table 1. Lithology of study area

Group	Code	Lithology	Geological age
1	Qs,d	Unconsolidated wind-blown sand deposit including sand dunes	Quaternary
	Qcf	Clay flat	
	Qsf	Salt flat	
	Qal	Stream channel, braided channel and flood plain deposits	
	Qft1	High level piedmont fan and valley terrace deposits	
	Qft2	Low level piedment fan and valley terrace deposits	
	Qsl	salt lake	
2	E1m	Marl, gypsiferous marl and limestone	Early Eocene
3	Ea.bvt	Andesitic to basaltic volcanic tuff	Eocene
	Ed.avs	Dacitic to Andesitic volcano sediment	
	Ek	Well bedded green tuff and tuffaceous shale (KARAJ FM)	
	Edav	Dacitic to Andesitic volcanic	
	Ed.avb	Dacitic to Andesitic volcanobreccia	
	Eavb	Andesitic volcanobreccia	
	E2s	Sandstone, marl and limestone	
4	Egr	Granite	Pre-Cambrian
	pCmt1	Medium - grade, regional metamorphic rocks (Amphibolite Facies)	
	pCgn	Gneiss, granite gneiss and locally including migmatite	
	pCmt2	Low - grade, regional metamorphic rocks (Green Schist Facies)	
5	mb	Marble	Triassic
6	Jph	Phyllite, slate and meta-sandstone (Hamadan Phyllites)	Jurassic
	Jdav	Jurassic dacite to andesite lava flows	
7	PlQc	Fluvial conglomerate, Piedmont conglomerate and sandstone.	Pliocene-Quaternary
8	K	Cretaceous rocks in general	Cretaceous
	K1-2lm	Albian - Cenomanian marl and argillaceous limestone	
9	TRJs	Dark grey shale and sandstone (SHEMSHAK FM.)	Triassic-Jurassic
10	Murm	Light - red to brown marl and gypsiferous marl with sandstone intercalations	Miocene
	Mur	Red marl, gypsiferous marl, sandstone and conglomerate (Upper red Fm.)	
11	E1c	Pale-red, polygenic conglomerate and sandstone	Paleocene-Eocene

Table 2. Overview of factors used in gully erosion susceptibility mapping

Factor	Range		Classes	Method	Reference
	min	max			
Elevation	683	2297	1. (< 829m), 2. (829m-1022m), 3. (1022m-1253m), 4. (1253m-1562m), 5 (>1562m)	Natural break	Arabameri et al. (2018c)
Slope	0.00	69.22	1. (< 2.4°), 2. (2.4° - 6.7°), 3. (6.7° - 13.5°), 4. (13.5° - 21.7°), 5. (>21.7°	Natural break	Arabameri et al. (2018b)
Plan curvature	-8.37	10.87	1. Concave (< -0.05), 2. Flat (-0.05 – 0.05), 3. Convex (> 0.5)	Natural break	Arabameri et al. (2018a)
TWI	1.37	22.48	1. (<5.75), 2. (5.75-8.07), 3. (8.07-11.47), 4. (>11.47)	Natural break	Arabameri et al. (2018a)
SPI	6.27	24.16	1. (<8.3), 2. (8.3-9.9), 3. (9.9-11.08), 4. (11.08-14.6), 5. (>14.6)	Natural break	Arabameri et al. (2018b)
Distance to stream	0	3208.3	1. (<176m), 2. (176m-415m), 3. (415m-704m), 4. (704m-1107m), 5. (>1107m)	Natural break	Arabameri et al. (2018c)
Drainage density	0.031	2.45	1. (<0.72 km/km2), 2. (0.72 -1.12 km/km2), 3. (1.12-1.54 km/km2), 4. (>1.54 km/km2)		Arabameri et al. (2018a)
Distance to road	0	53381.2	1. (<500m), 2. (500m-1000m), 3. (1000m-1500m), 4. (1500m-2000m), 5. (>2000m)	Natural break	Arabameri et al. (2018c)
Lithology	-	-	Group1, Group 2, Group 3, Group 4, Group 5, Group 6, Group 7, Group 8, Group 9, Group 10 Group 11	Lithological units	-
LU/LC	-	-	1. (Agriculture), 2. (Orchard), 3. (Bare land), 4. (Kavir), 5. (Salt land-kavir), 6. (Poor range), 7. (Rock), 8. (Salt lake), 9. (Salt land) 10. (Wetland)	Supervised classification	-
Soil type	-	-	1. (Bad Lands), 2. (Playa), 3. (Rock Outcrops / Entisols), 4. (Rocky Lands), 5. (Salt Flats), 6. (Aridisols), 7. (Entisols / Aridisols)	Supervised	-
Rainfall	43.40	208.7	1. (<66.27mm), 2. (66.27 - 85.06mm), 3. (85.06 -103.19mm), 4. (103.19 – 144mm), 5. (> 144mm)	Natural break	Arabameri et al. (2018b)

Table 3. Multicollinearity analysis among independent variables

GRCFs	TOL	VIF	GRCFs	TOL	VIF
Lithology	0.771	1.298	Drainage density	0.468	2.136
LU/LC	0.620	1.613	distance to road	0.704	1.421
Soil type	0.545	1.836	Distance to stream	0.671	1.490
TWI	0.225	4.024	Plan curvature	0.850	1.176
SPI	0.242	4.127	Rainfall	0.299	3.528
Elevation	0.246	4.204	Slope	0.294	3.399

Table 4. Spatial relation between gully-related causal factors and gullies using CF model

Factors	Classes	Number of pixels in domain		Number of gullies		CF
		No.	%	No.	%	
Elevation (m)	< 829	2949953	46.10	92	61.33	0.25
	829-1022	1532604	23.95	32	21.33	-0.12
	1022-1253	1007146	15.74	17	11.33	-0.39
	1253-1562	680354	10.63	9	6.00	-0.77
	>1562	229233	3.58	0	0	0
Slope (°)	< 2.4	4169415	65.16	108	72.00	0.10
	2.4 - 6.7	1287734	20.13	29	19.33	-0.04
	6.7 - 13.5	416673	6.51	11	7.33	0.13
	13.5 - 21.7	254777	3.98	1	0.67	-0.83
	>21.7	183533	2.87	1	0.67	-0.77
Plan curvature (100/m)	concave	2023109	31.61	51	34.00	0.07
	flat	2314183	36.16	62	41.33	0.13
	convex	2061997	32.22	37	24.67	-0.31
TWI (100/m)	<5.75	1827861	28.57	28	18.67	-0.53
	5.75-8.07	2897322	45.28	66	44.00	-0.03
	8.07-11.47	1293381	20.21	27	18.00	-0.12
	>11.47	379680	5.93	29	19.33	0.69
SPI (100/m)	<8.3	1914085	29.92	49	32.67	0.08
	8.3-9.9	2166242	33.86	44	29.33	-0.15
	9.9-11.08	1439789	22.50	24	16.00	-0.41
	11.08-14.6	661893	10.34	8	5.33	-0.94
	>14.6	216235	3.38	25	16.67	0.80
Distance to stream (m)	<176	1500338	23.45	85	56.67	0.59
	176-415	1129109	17.64	22	14.67	-0.07
	415-704	1000724	15.64	17	11.33	-0.11
	704-1107	684925	10.70	8	5.33	-0.21
	>1107	2084245	32.57	18	12.00	-0.26
Drainage density (km/km^2)	<0.72	1079031	16.86	10	6.67	-0.53
	0.72-1.12	2350393	36.73	55	36.67	0.00
	1.12-1.54	2058537	32.17	37	24.67	-0.30
	>1.54	911380	14.24	48	32.00	0.55
Distance to road (m)	<500	179336	2.80	30	20.00	0.86
	500-1000	171453	2.68	7	4.67	0.43
	1000-1500	166580	2.60	2	1.33	-0.95
	1500-2000	162592	2.54	3	2.00	-0.27
	>2000	5719380	89.37	108	72.00	-0.24
Lithology	Group1	40974	0.64	1	0.67	0.04
	Group 2	117752	1.84	1	0.67	-0.64
	Group 3	860303	13.45	13	8.67	-0.36
	Group 4	201318	3.15	2	1.33	-0.58
	Group 5	817439	12.78	31	20.67	0.62
	Group 6	82525	1.29	1	0.67	-0.48
	Group 7	558785	8.74	25	16.67	0.91
	Group 8	84393	1.32	1	0.67	-0.49
	Group 9	3532744	55.24	75	50.00	-0.09
	Group 10	92648	1.45	0	0.00	-1.00

Landuse / landcover	Group 11	6424	0.10	0	0.00	-1.00
	Agriculture	2602	0.04	0	0.00	-1.00
	Orchard	1735	0.03	0	0.00	-1.00
	Bare land	247940	3.87	17	11.33	0.66
	Kavir	1245296	19.46	56	37.33	0.48
	Salt land-kavir	76751	1.20	1	0.67	-0.80
	Poor range	3594153	56.16	74	49.33	-0.14
	Rock	1038247	16.22	0	0.00	-1.00
	Salt lake	104556	1.63	0	0.00	-1.00
	Salt land	87059	1.36	2	1.33	-0.02
	Wetland2	1002	0.02	0	0.00	-1.00
	Bad Lands	475443	7.43	7	4.67	-0.59
Soil type	Playa	79142	1.24	1	0.67	-0.86
	Rock Outcrops / Entisols	1547437	24.18	27	18.00	-0.34
	Rocky Lands	160310	2.51	0	0.00	-1.00
	Salt Flats	1201003	18.77	27	18.00	-0.04
	Aridisols	436939	6.83	7	4.67	-0.46
	Entisols / Aridisols	2499067	39.05	81	54.00	0.28
Rainfall (mm)	<66.27	1170801	18.30	12	8.00	-0.56
	66.27 - 85.06	1919813	30.00	74	49.33	0.64
	85.06 -103.19	2103806	32.88	52	34.67	0.05
	103.19 - 144	1077669	16.84	12	8.00	-0.52
	> 144	127252	1.99	0	0.00	-1.00

Table 5. Confusion matrix from RF model (0 = non-gully or negative, 1 = gully or positive)

observation	predicted	
	0	1
0	81	29
1	31	69