



38

39 **Introduction**

40 Land degradation is one of the most pressing environmental issues around the globe.
41 Several aspects of this issue have been recognized by the United Nations Convention
42 (Gholami et al. 2019a). Land degradation can be driven by both water and wind, of
43 which the former can have a stronger impact on soil erosion in short time (Gia et al.
44 2018). Spatial mapping of risks of land degradation can provide a basis to support
45 managers and policymakers in risk mitigation and adaptation to aeolian and water
46 erosion.

47 Land degradation driven by aeolian erosion is a known problem (Shi et al. 2004). Dust
48 storms, which are a natural hazard, are associated with soil erosion. This phenomenon
49 has detrimental impacts in the Earth system, e.g., for food security (Boroughani et al.
50 2022), human health (Moridnejad et al., 2015), geochemical conditions (Gholami et al.,
51 2020b), and the Earth's carbon cycle (Gherboudj et al., 2017). Identifying dust sources
52 as potential areas of dust emission is therefore necessary in developing a better
53 understanding of land degradation. Spatial mapping of dust source susceptibility areas
54 (DSSAs) is a crucial step for erosion mitigation and watershed management. Different
55 approaches for identifying DSSAs exist, e.g., using meteorological data (Yang et al.
56 2019), numerical modeling (Péré et al. 2018), and remote sensing (Jafari et al. 2021).
57 Remote sensing can provide worldwide information on aerosol properties (Park et al.
58 2014). The present study uses MODIS satellite images to detect dust aerosols over the
59 Lut Desert.

60 In addition to soil erosion by wind, water-driven soil erosion is a known mechanism for
61 soil degradation. This kind of soil erosion is a known environmental threat and can
62 influence both terrestrial and aquatic systems (Halecki et al. 2018, Sun et al. 2014).
63 Therefore, knowing the spatial distribution of water-induced soil erosion susceptibility
64 areas (SESA) is also necessary. Several numerical models exist for predictions and risk
65 evaluations of water-induced soil erosion (Chicas et al., 2016, Gao et al., 2017, Anache
66 et al., 2018, Gia et al., 2018, Halecki et al., 2018), but none used machine learning to
67 combine different observational data sets for assessing soil erosion. Machine learning
68 has emerged as a subfield of data science and helps to better understand environmental
69 problems (Gholami et al. 2019b). It can integrate data from different sources to create
70 forecasts and discover patterns (Gholami et al. 2020a). In environmental sciences,



71 algorithms such as support vector machine, random forest (RF), artificial neural
72 networks (ANN), and multivariate adaptive regression spline have been applied, e.g.,
73 for groundwater (Lee et al. 2017), gully erosion (Zabihi et al. 2018), sediment
74 contamination (Mirchooli et al. 2019), dust sources (Boroughani et al. 2020), landslides
75 (Youssef and Pourghasemi 2021), floods (Tehrany et al. 2014), and trace elements
76 (Derakhshan-Babaei et al. 2022).

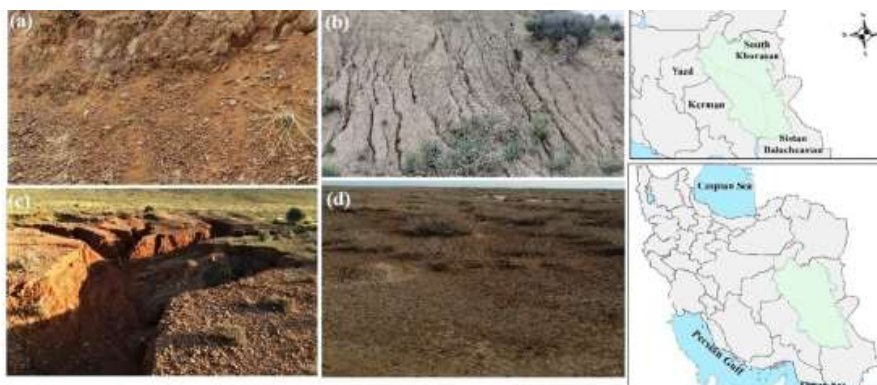
77 The aims of the current study are (1) to assess the spatially resolved contribution of soil
78 erosion by water and wind, and (2) to combine the findings into spatially resolved
79 information on risks for land degradation.

80

81 **2. Data and methods**

82 The focus of this study is on the Lut watershed situated in the east and southeast of Iran
83 covering an area of 206242 km² (28° 10' to 32° 30' N latitude and 55° 45' to 61° 15' E
84 longitude) and is marked in Fig. 1. The region includes the largest desert of the country,
85 the Lut Desert. The region contributes to the increasing dust concentration of southwest
86 Asia (Ebrahimi-khusfi et al. 2021). This area is chosen to develop and test the methods
87 based on regional data on erosion observations with examples shown in Fig. 1a-d. The
88 methods tested in this article could be later transferred to similar assessments in other
89 regions. Choosing this region in Iran is further motivated by the impact of land
90 degradation on the country's economy. It is estimated that land and water degradation
91 cost Iran about US \$12.8 billion per year which is four percent of the total Gross
92 Domestic Product (GDP) (Emadodin et al. 2012). It underlines the impacts of land
93 degradation that goes well beyond impacts on the natural environment.

94



95

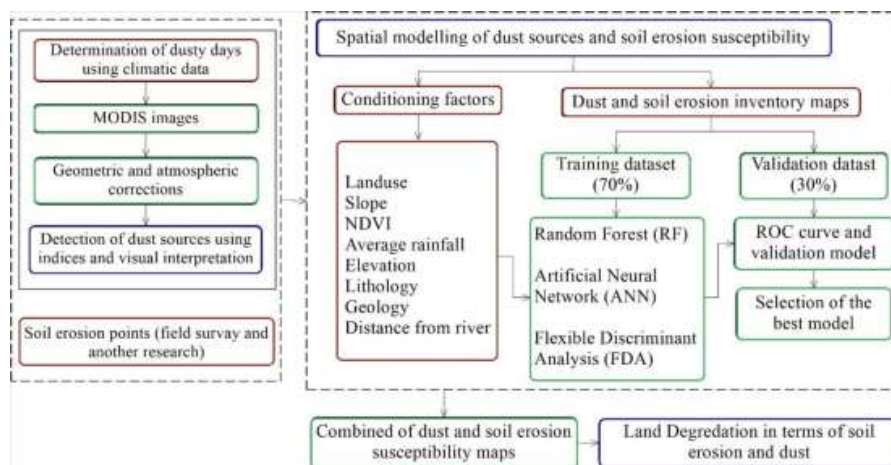


96 Fig.1 Geographical location of the study watershed. Green shading marks the Lut watershed. The Lut
97 Desert is located in the centre of the watershed. Settlements are primarily situated in the northern and
98 south-western parts. Example of soil erosion in the watershed are sheet erosion (a), rill erosion (b),
99 gully erosion (c), and wind erosion (d).

101 2. 1. Land degradation mapping

102 Our land degradation zonation consists of three main processing steps, graphically
103 depicted in Fig. 2. At first, spatial mapping of water erosion is conducted (section
104 2.1.1). In the second step, spatial mapping of dust source susceptibility is carried out
105 with machine learning methods (section 2.1.2). In the last step, the patterns of water
106 erosion and dust source susceptibility are combined to identify risk areas of land
107 degradation (section 2.2.3).

108



109
110 Fig.2 Flowchart of inputs (red boxes), data processing (green boxes), and outputs
111 (blue boxes) in the present study

112

113

114 2.1.1 water erosion map

115 Quantifying the erosion susceptibility of an area requires to determine a spatial
116 distribution of observed water-induced soil erosion that can have different
117 characteristics, e.g., gully erosion, rill erosion, and surface erosion. That information
118 are extracted from data collected during an own field surveys paired with a previous
119 research (Shit et al. 2020). The aim of the field survey for the present study was to
120 identify regions where sheet, rill, and gully erosion took place. This field survey was



121 carried out in accessible parts of the watershed in April 2020. The data set contains the
122 type of water-induced soil erosion along with the geographical location using a Global
123 Positioning System (GPS). A selection of the identified water soil erosions in the study
124 region is shown in Fig. 1.

125 We translated the observations of the field survey into maps of non-degraded and
126 degraded areas. These areas were plotted in an inventory map and prepared for further
127 analysis, although not all desert areas are fully covered by the survey.

128

129 **2.1.2 Dust aerosol map**

130 The large desert area to be covered is a motivation for the use of satellite data for
131 estimating dust sources. We used MODIS images from the Terra (morning) and Aqua
132 (afternoon) satellites (Vickery and Eckardt, 2013) to identify dust aerosols. We define
133 dusty days, when the horizontal visibility is less than 2000 m for at least one hour during
134 the day based on available weather stations in Iran (Vickery and Eckardt, 2013;
135 Borouhghani et al., 2021). According to the mentioned condition, more than 500 dusty
136 days were identified during 2010–2021 distributed over the stations in Birjand,
137 Zahedan, Kerman, Bam, Doostabad, Bisheh, Rafsanjan and Mighan. We pair the station
138 observations with satellite data to estimate the spatial extent of the dust aerosol plumes.
139 Due to the overpass of the Terra and Aqua satellites once per day, we acquired 28
140 satellite images from the MODIS sensor that during times when the weather stations
141 had documented dusty conditions in the ten year period. For identifying pixels with dust
142 aerosols in these images, we calculate four different dust indices (BTD2931, BTD3132,
143 NDDI and D) for dust aerosol identification.(Borouhghani et al., 2020, 2021
144 Hahnenberger and Nicoll, 2014).

$$145 \quad B(T, \lambda) = \frac{2hc^2}{\lambda^5 \frac{hc}{(e\lambda kt - 1)}} \quad (1)$$

146 where $B(T, \lambda)$ represents the Planck equation at λ (μm), T is the BT (K), h is the
147 Planck's constant ($6.626 \times 10^{-34} \text{ m}^2\text{kg}^2\text{s}^{-1}$), k is the Boltzmann's constant (1.38×10^{-23})⁵, c
148 is the speed of light ($2.99 \times 10^8 \text{ ms}^{-1}$), and T is the temperature (Hao et al., 2007)
149

$$150 \quad T = \frac{hc}{\lambda k \ln(1 + \frac{2hc^2}{L\lambda^5})} \quad (2)$$

151 Using Planck's equation, the value of the temperature can be derived, where L is the
152 amount of radiance in the images (in $\text{Wm}^{-2}\text{sr}^{-1}\mu\text{m}^{-1}$).
153



154
$$NDDI = (p_{2.13} - p_{0.469}) / (p_{2.13} + p_{0.469})$$

155 (3)

156

157 where $\square_{2.13}$ and $\square_{0.469}$ depict the reflectance value at the top-of-atmosphere at 2.13
158 and 0.469 μm , respectively (Qu et al., 2006)

159

160
$$D = \exp\{-[rr \times a + (BTD - b)]\}$$
 (4)

161 where rr shows the reflectance proportion among wavelengths of 0.54 μm and 0.86 μm
162 and BTD is the difference among the bands 11 and 12 μm ; a and b are constants took
163 during the initial calibration (Eq. 1). (Qu et al., 2006; Miller, 2003; Hao et al., 2007;
164 Boroughani et al., 2020, 2021).

165 We compute false color maps using four combinations of channels (1: NDDI, B4, B3;
166 2: D, BTD_{2931} , NDDI; 3: D, BTD_{3132} , NDDI; and 4: BTD_{2931} , B4, B3) in ENVI
167 software. We choose these four different indices for cross-validating the presence of
168 dust aerosols. With each of these methods we see dust aerosol in different color and
169 quality in the MODIS images on the 28 days. After combining the four methods in the
170 software ENVI, we choose the method that shows the dust plume in the MODIS image
171 more clearly as the best method (Boroughani et al., 2020, 2022). This method is based
172 on a cone of dust diffusion seen in the processed MODIS images, where the apex
173 denotes the dust's source (Lee et al., 2009; Walker et al., 2009). Ultimately, the
174 inventory map of the dust aerosols in the Lut watershed was created.

175

176 **2.2. Identification of key factors controlling for aeolian and water erosion**

177 To develop DSSA and SESA, the identification and selection of appropriate dust
178 sources and soil erosion effective factors is necessary. The main factors affecting DSSA
179 and SESA were selected and constructed based on literature, available data and
180 geographical maps (Torabi et al., 2021; Zabihi et al., 2018; Boroughani et al., 2020;
181 Gholami et al., 2020a). The considered factors in this study included: elevation, land
182 use, slope of terrain, lithology, annual rainfall, distance from rivers, distance from
183 roads, the Topographic Witness Index (TWI), and Normalized Difference Vegetation
184 Index (NDVI). Various sources were used to gather data for these factors, introduced
185 in the following in more detail. All collected data were mapped to a horizontal grid of
186 1km resolution.



187 The shuttle radar topography mission (SRTM) images were used to create the digital
188 elevation model (DEM, , Fig 3c) (Ghorbanzadeh et al., 2018). The lowest and highest
189 elevation of the study area is 124 m in the centre of the desert and 3966 m at the western
190 and eastern margins of the study watershed, respectively (Fig. 3c). Vegetation cover
191 considerably supports soil conservation. Areas with low vegetation cover would be
192 more sensitive to both erosion by water and wind (Arabameri et al., 2019a; Gholami et
193 al. 2019b). Therefore, we use the Normalized Difference Vegetation Index (NDVI) to
194 assess the vegetation cover in the study area from MODIS images following
195 (Arabameri et al., 2019a; Boroughani et al., 2020):

$$196 \quad NDVI = \frac{NIR+R}{NIR-R}$$

197 Where R is the red (0.620-0.670 μm) and NIR is near-infrared bands (0.841-0.876 μm)
198 (Fig. 3d).

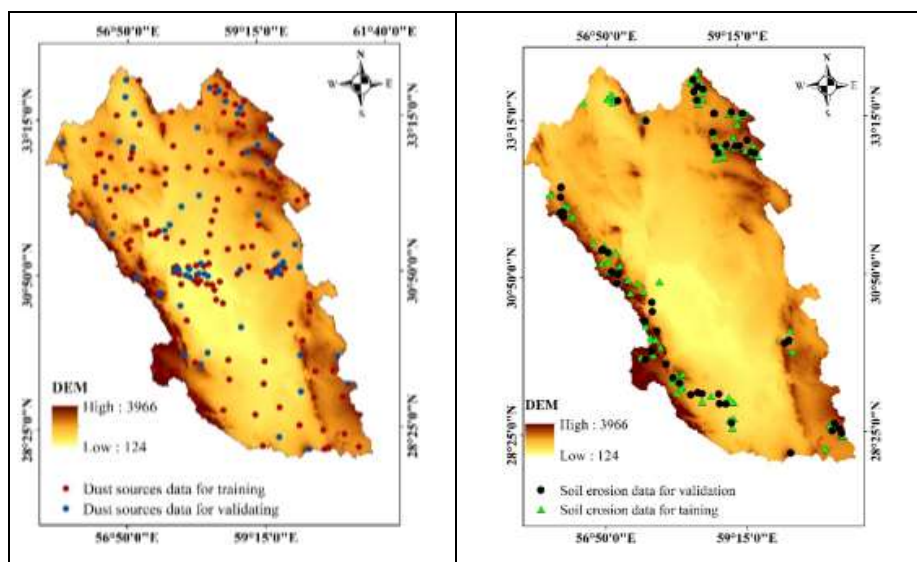
199 Annual rainfall (Fig. 3e) were obtained from Iran Meteorological Organization for the
200 period of 2000-2021. Mean annual rainfall were calculated using 40 different
201 meteorological stations located within or close to the watershed (Fig.3e). The inverse
202 distance weighting (IDW) interpolation method was applied to integrate rainfall over
203 the study area in the ArcGIS environment (Gholami et al., 2020a). Topographic
204 Wetness Index (TWI), which indicates the spatial distribution of areas of potential soil
205 saturation, is an effective factor to indicate water erosion including landslides and also
206 flooding (Arabameri et al., 2019b). TWI which determines the dry and wet zones
207 calculated as (Beven and Kirkby 1979):

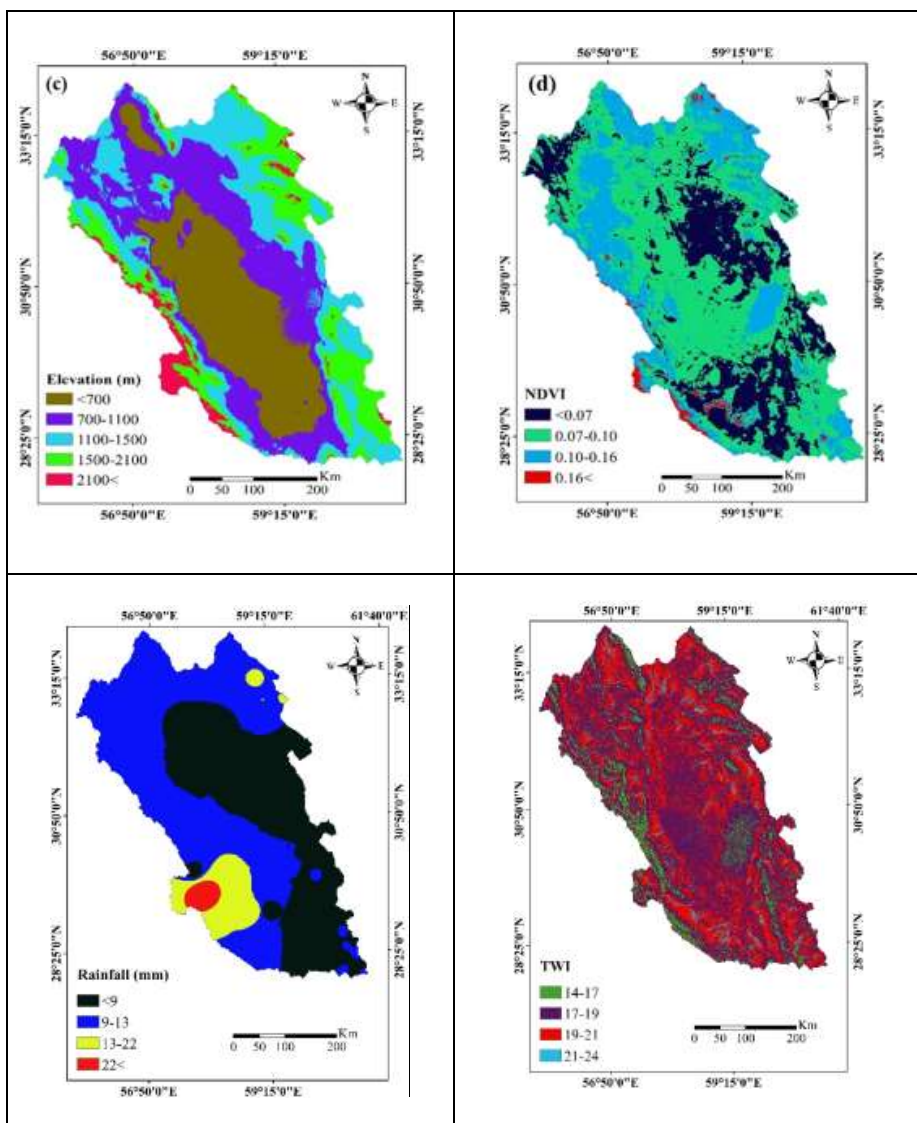
$$208 \quad TWI = \ln\left(\frac{\alpha}{\tan\beta}\right)$$

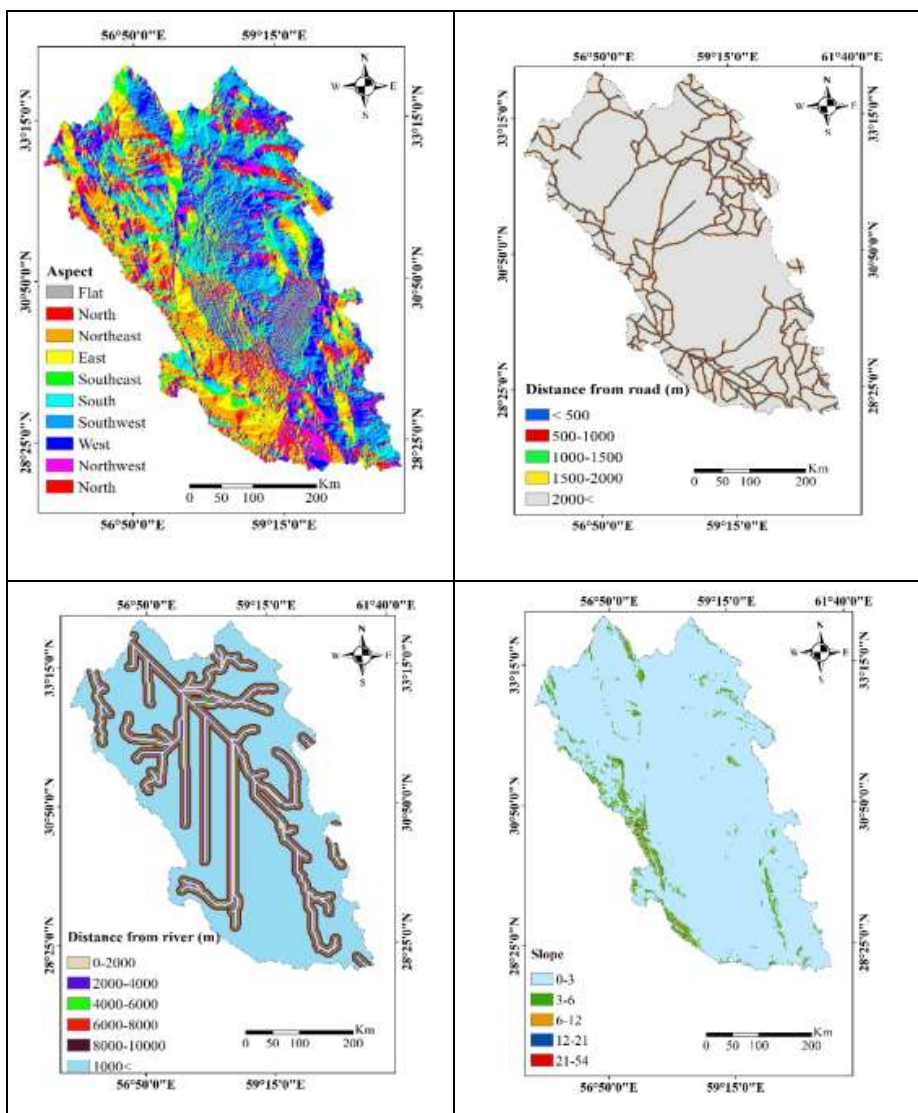
209 where α is the cumulative up-slope area from a point (per unit contour length) and β is
210 the slope angle at that point. This index was calculated in the SAGA-GIS environment
211 and classified into four groups viz. 14-17, 17-19, 17-21, 21-24 (Fig. 3f). The aspect
212 map was also generated using DEM and grouped into ten classes (Fig. 3 g). Distance
213 from road is an indicator of infrastructure development which influences soil erosion
214 and land degradation (Torabi et al., 2021). This factor is shown in five classes in Fig. 3
215 h. Distance from river is one of the most effective factors on water-caused erosion
216 (Amiri et al., 2019) which is classified into six groups (Fig. 3i).

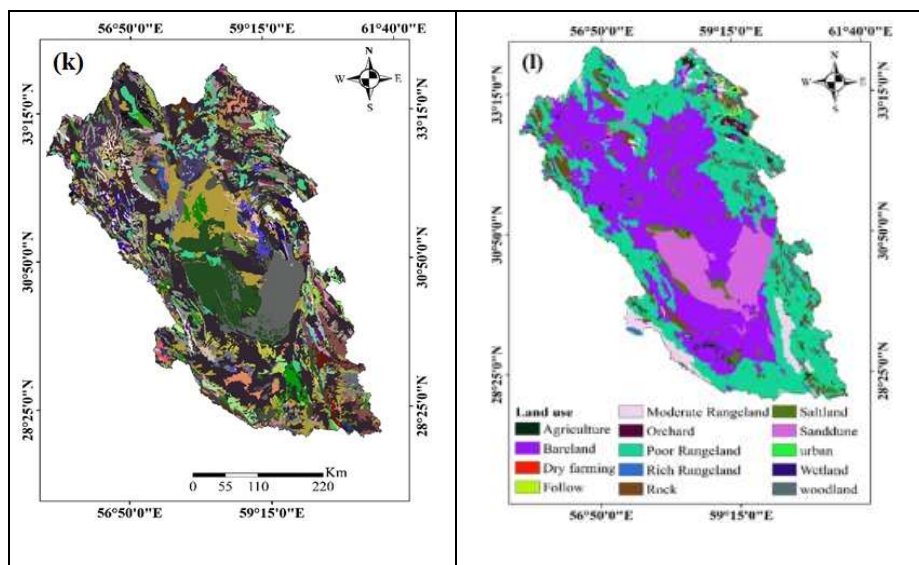


217 The slope map (%) was created using a Digital Elevation Map (DEM, Fig. j) and
218 classified into five groups including 0-3%, 3-6%, 6-12%, 12-21%, and 21-54%. The
219 lithology map indicates eleven different soil classes in the study area (Fig. 3k).
220 Land use and soil maps were obtained from base maps developed by the Iranian Forest,
221 Rangeland, and Watershed Management Organization (<https://frw.ir/>). In the study
222 region, there are fourteen land-use classes including wetlands, rangelands of three states
223 (poor, medium, and rich), dry farming, agricultural lands, urban area, fallow land, rock-
224 covered land, wetland, saltland, woodland, bare surfaces, and sand dunes (Fig. 3m). A
225 large percentage (83%) of the watershed area is covered by bare land, poor rangeland,
226 and sand dunes. All three land use classes are prone to wind erosion due to sparse or no
227 vegetation.









228
229 Fig.3 Location of dust observation points for training and validation (a), water-induced soil erosion
230 points for training and validation (b), and the conditional factors (Elevation (c), NDVI (d), Rainfall (e),
231 TWI (f), Aspect (g), Distance from road (h), Distance from river (i), Slope (j), Lithology (k), Land use
232 (l)) in the watershed.
233

234 2.4. Spatial mapping of DSSA and SESA using machine learning algorithms

235 We combine the two susceptibility maps for DSSA and SESA to create the land
236 degradation hazard map with regards to water- and wind-induced soil erosion. For both
237 types of soil erosion, three machine learning models were constructed and applied. The
238 land degradation susceptibility map was then created by synthesizing the results for
239 both soil erosion types in an ArcGIS 10.5 environment, and the land degradation
240 susceptibility was ultimately evaluated with four classes.

241 A wide range of machine learning algorithms has been applied for spatial mapping of
242 environmental phenomena in the past. The effective factors described in Section 2.2
243 and the inventory maps of water and wind erosion were used as the input of the machine
244 learning algorithms. In the present study, the algorithms of random forest (RF), artificial
245 neural network (ANN), and flexible discriminate analyses (FDA) were used to produce
246 DSSA and SESA maps. We choose three different algorithms to test the dependency of
247 the results on the method as a measure of uncertainty. The three algorithms are
248 described in more detail in the following.
249



250 **2.4.1 Random forest (RF)**

251 Random forest developed by Breiman (2001) is a machine learning algorithm for non-
252 parametric multivariate classification. RF builds multiple trees using a random
253 selection of the training dataset. The data not included are called out-of- bag (OOB)
254 determines the model accuracy using generalization error estimation (Breiman 2001).
255 Diversity among the classification trees increases using resampling the data with
256 replacement and also randomly change of predictors set during tree induction processes
257 (Youssef et al., 2016). Information from numerous decision trees has been combined in
258 the RF algorithm.

259 Generally, it is essential to define two parameters to run the RF model including the
260 number of trees (ntree) and the number of factors prepared from the data shown in Fig.
261 3 (mtry). The former are built while the RF model is running, while the latter is used in
262 the tree-building process. Both the number of trees and factors need to be optimized to
263 minimize the generalization error (Rahmati et al. 2016). The optimisation was done
264 through sensitivity tests.

265

266 **2.4.2 Artificial neural network (ANN)**

267 The artificial neural network (ANN) is a machine learning tool developed by imitating
268 human brain performances and making connections between inputs and outputs
269 (Sakizadeh et al. 2017). The human brain is mimicked in two ways: Firstly, obtaining
270 information and knowledge using a learning process, and secondly, storing knowledge
271 using synaptic weights. Therefore, ANN has been identified as the model that finds the
272 optimal solution for non-linear problems, such as dust source and soil erosion
273 susceptibility, by identifying patterns with conditioning factors (Ghorbanzadeh et al.
274 2019). In an ANN, a neuron is the smallest data processing unit which could make many
275 neural network structures and be used in research for different purposes. The standard
276 structure of ANN consists of three layers, namely, the input layer, hidden layers, and
277 the output layer. The input layer consists of training data and conditioning factors of
278 dust source, the neurons in the hidden layer analyze the complex information contained
279 in the data, and the output layer are the maps of dust source susceptibility. In this
280 structure, the neurons across the same layer are not connected, but they are linked with
281 neurons in the previous and subsequent layers. In ANN, the algorithm determines a
282 weight for each input factor and a transfer function to build results (Kalantar et al.
283 2017).



284

285 **2.4.3 Flexible discriminate analyses (FDA)**

286 The modification of the linear regression model for the application to non-linear
287 problems is the purpose of FDA (Avand et al. 2021). Nonparametric regression models,
288 nonlinear discriminant analysis, and classification methods are combined into one
289 framework. This algorithm is flexible for non-linear classifications because non-linear
290 transformation is used and clusters are soft (Kalantar et al. 2020), here clusters for the
291 relationship between soil erosion and the predictor factors from Fig. 3. In this way,
292 variables in FDA are firstly aligned with the multivariate adaptive regression splines
293 (MARS) and then dimension reduction is performed (Kim and Kim 2021). FDA can
294 overcome the problem of linear discriminant analysis (LDA) and it is minimizing the
295 square average of the residuals (Mosavi et al. 2020), while linear regression is replaced
296 by nonparametric regression in FDA. Therefore, FDA has the potential to apply for
297 non-linear natural problems such as soil erosion, dust, flood, and landslide.

298

299 **2.5. Evaluation of machine learning algorithms**

300 In our DSSA and SESA assessment, 70% of point data are randomly selected for the
301 training dataset and 30% for model validation. The prediction accuracy of the machine
302 learning algorithms is assessed by comparing the DSSA map with the validation dataset
303 of dust sources. These data were extracted from MODIS images and some indicators
304 which was explained in section 2.1.2. The Receiver Operating Characteristic (ROC)
305 curve and the Area Under the Curve (AUC) are applied following past studies that used
306 these to test the prediction skill of a model for the occurrence or non-occurrence of the
307 studied phenomena (Naghibi et al. 2017). The AUC ranges from 0 to 1 in which the
308 models that better perform represent the AUC close to one.

309

310 **3. Results and Discussion**

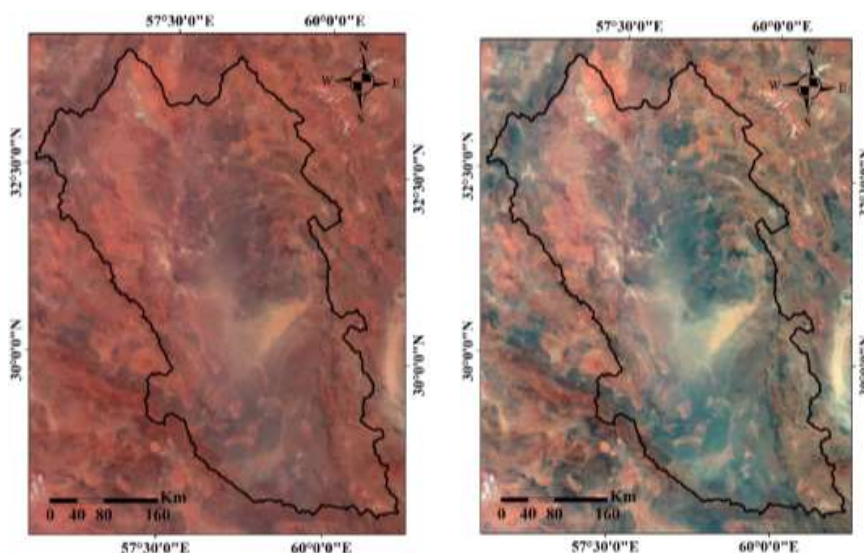
311 **3.1. Spatial distribution of DSSA**

312 **3.1.1. Dust aerosol detection**

313 An illustration of a dust storm seen in MODIS FCC satellite imagery over the Lut
314 watershed on August 7, 2019, is shown in Fig. 4. Following a visual analysis of the
315 images, we determined that the false colour combination (R: BTD2931, G: Band 4, B:



316 Band 3) is the best and applied it to 26 MODIS images of dusty days. As a result, the
317 Lut watershed's dust source locations were identified (Fig. 4).
318



319 Fig.4 The dust storm on 07 August 2019, as seen above is an example of the visual
320 inspection of a dust storm (a) MODIS true colour (Red: Band 5, Green: Band 4, Blue:
321 Band 3), and (b) enhanced MODIS satellite photos, (Red: BTDR2931, Green: Band 4,
322 Blue: Band 3).

323

324 3.1.2 The importance of conditioning factors for DSSA

325 Since multicollinearity among factors has been identified as an obstacle to explaining
326 the results (Roy and Saha 2019), the Variance Inflation Factor (VIF) was calculated to
327 assess the relationships among conditioning factors. This was conducted because
328 multicollinearity among factors will decline the accuracy of the models (Arabameri et
329 al. 2019b). In the present study, VIF values for DSSA mapping range from 1.05 to 1.57
330 which illustrated no collinearity among the eight factors. Therefore, no exclusion was
331 applied and all factors were considered in successor calculations and modeling.

332 The importance and impact of each factor depend on the machine learning algorithms.
333 The result of DSSA mapping using RF showed that NDVI, elevation, land use, and
334 lithology had the greatest degree of effect among conditioning factors. In addition, the
335 impacts of these factors on DSSA have been proved by previous investigations



336 (Gholami et al. 2020a, 2020b). Other factors such as the distance from rivers, lithology,
337 rainfall, and slope were identified as rather weak predictors, respectively. These
338 findings agree with other research (Boroughani and Pourhashemi 2020, Darvand et al.
339 2021).

340 The FDA approach showed that however elevation, NDVI, and land use had the highest
341 effects on dust sources susceptibility, other factors had no impact on DSSA. Similarly,
342 with ANN, elevation, NDVI, and land use were identified as the three most effective
343 factors, and other factors were weaker predictors rather than formers. However these
344 two models of FDA and ANN provide similar results in term of the importance of
345 conditioning factors, FDA could be used rather than ANN because of its higher
346 accuracy which is shown in the next section.

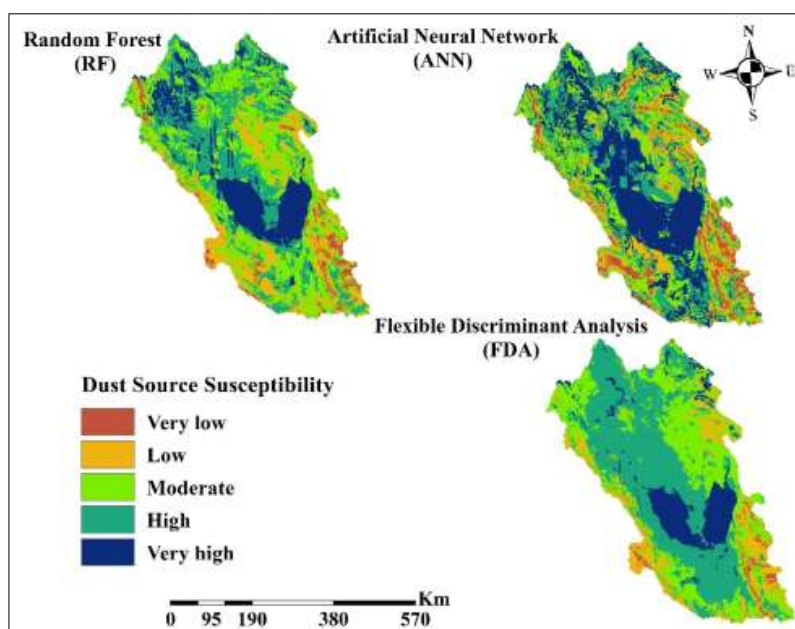
347

348 **3. 1. 3 Spatial distribution of dust source susceptibility**

349 The dust source susceptibility (DSS) maps created by RF, FDA, and ANN are classified
350 into five risk classes (very high, high, moderate, low, and very low) shown in Fig. 5.
351 These classes are set as in earlier studies (Mosavi et al., 2020; Boroughani,
352 Mohammadi, Mirchooli, & Fiedler, 2022). The results of the model evaluation using
353 ROC indicates that the RF model with an accuracy of 75.0% provides the most accurate
354 outputs. FDA and ANN had similar performances with the accuracy of 71.7% and
355 70.7%. In terms of True Skill Statistic (TSS), similar results have been obtained in
356 which RF with accuracy of 45.8% had again the best performance in comparison to
357 FDA (32.4%) and ANN (35.8%). In this way, RF introduces different priorities for the
358 effective factors in comparison with FDA and ANN. RF proposes NDVI, elevation,
359 land use, and lithology as the most important factors, while FDA and ANN suggest
360 elevation, NDVI, and land use as the most influencing factors. The dominance of
361 NDVI, elevation and land use as the most effective factors for DSS is consistent with
362 the understanding of dust source locations that are typically found in topographic
363 depressions with sparse or no vegetation. The DSSA map from RF was selected for
364 further analysis due to the highest accuracy, although the differences to FDA and ANN
365 are in the statistical sense relatively small. According to the DSSA maps, 29% and 17%
366 of the watershed were classified as areas of high and very high DSSA, i.e., almost half
367 of the study area. Only 4% and 16% of the watershed have a very low and low
368 susceptibility to soil erosion through winds, respectively. The spatial extent of high and
369 very high risk areas from RF is smaller than the ones obtained by ANN and FDA. These



370 results are consistent with other research, indicating that RF allows more detailed
371 spatial mapping of dust source susceptibility compared to other machine learning
372 algorithms (Rahmati et al. 2020, Gholami et al. 2019b, Darvand et al. 2021).
373



374
375 Fig. 5 Dust sources susceptibility area (DSSA) based on random forest (RF), artificial neural network
376 (ANN), and flexible discriminate analyses (FDA)
377

378 3.2. Soil erosion susceptibility map

379 3.2.1 Relative influential conditioning factors for SESA

380 For SESA, RF indicates that rainfall, TWI, slope, elevation, land use, and geology are
381 the most important conditioning factors. Distance from roads and rivers was recognized
382 as the least important factors. These findings of the impact of conditioning factors for
383 SESA are similar in other regions (Arabameri et al. 2019a, Hosseinalizadeh et al. 2019).
384 For ANN, TWI, slope, and land use were the most effective factors for prediction which
385 is followed by NDVI, land use, and distance from the river. The results from FDA
386 indicated that the most important conditioning factors are TWI, slope, and elevation,
387 geology, and NDVI.

388 A large area of the watershed is land with typically little rain and vegetation cover such
389 that bare soil is the main physical attribute in the watershed. This kind of surface is



390 known to be prone to water-induced soil erosion, when rain events occur. The erosion
391 can be particularly pronounced over slopes. This understanding is consistent with all
392 algorithms pointing to a major role of TWI and slope for SESA.

393 Some environmental factors (rainfall, TWI, slope, elevation, and geology) influence
394 SESA more than DSSA. Land use as a human-induced conditioning factor, however,
395 affects both SESA and DSSA, which underlines the importance of land-use planning
396 and management.

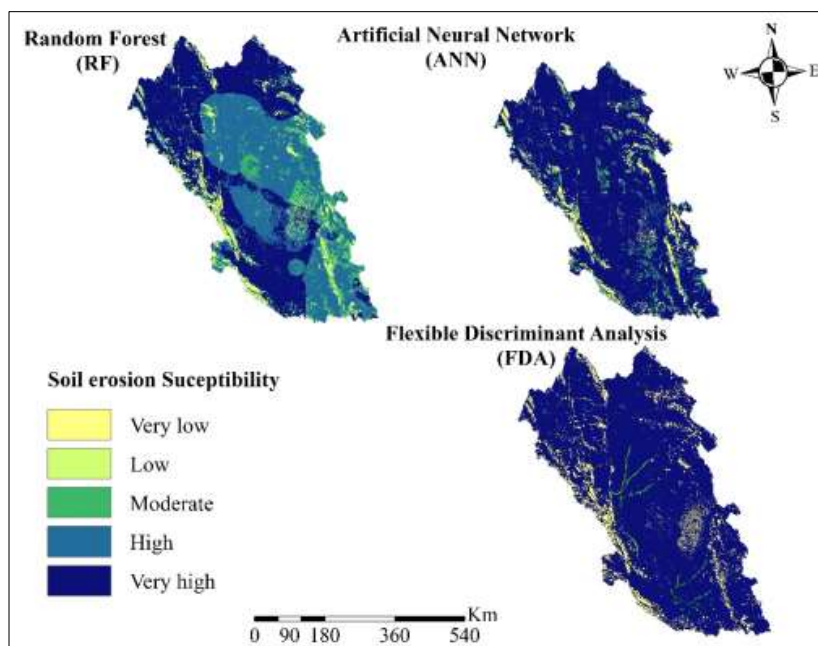
397

398 **3.2.2. Spatial modeling of SESA**

399 Fig. 6 shows the SESA predictions from the three machine learning algorithms,
400 classified by the soil erosion risk in the ArcGIS environment. Validation of the three
401 machine learning algorithms highlights that RF was again the most reliable algorithm
402 amongst the three, indicated by the best prediction rate. Based on ROC, RF yields a
403 94% accuracy for SESA (Fig. 6c). The ROC coefficient of ANN and FDA were slightly
404 lower, but still high with an accuracy of 91% and 89%, respectively. In the case of the
405 TSS index, better performance was obtained again for RF (89%) rather than ANN
406 (78%) and FDA (78%).

407 The majority of the land in the watershed (81%) has a high and very high risk for water-
408 induced soil erosion by RF. This is slightly lower than for ANN and FDA which
409 classified 85% and 89% of the watershed as high and very high susceptible areas. The
410 high and very high susceptible areas for water-driven soil erosion are mostly located in
411 the north and south-west parts of the watershed. The high and very high susceptible
412 areas have socio-economic implications, particularly because most settlements and
413 cities of the watershed are located in the same regions. This can mean that human
414 activity is a contributing factor for the water-induced soil erosion.

415



416
417 Fig. 6 soil erosion susceptibility areas map (GESM) using random forest (RF), artificial neural network
418 (ANN), and flexible discriminate analyses (FDA)
419

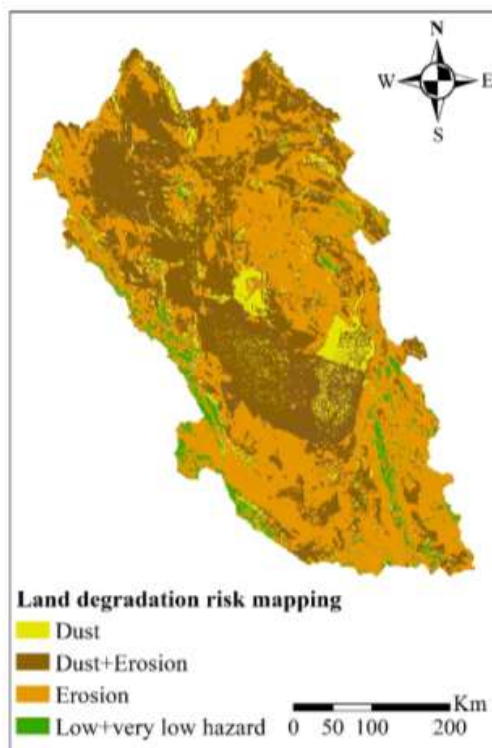
420 3.3. Land degradation susceptibility

421 The majority of the study watershed is susceptible to a substantial risk for land
422 degradation. The spatial distribution of land degradation susceptibility, shown in Fig.
423 7, indicates that only 4% of the land area has low to very low risks of land degradation.
424 Areas susceptible to both soil erosion by water and winds together constitute 43% of
425 the total area. Approximately 45% and 8% of the study area are at risk to soil erosion
426 by water and wind, respectively. Taken together, it means that the majority of the Lut
427 watershed falls under the category of land degradation risks. The watershed accounts
428 for 12.5% of the total land of Iran. The findings of the present study is therefore
429 consistent with a report that indicated water erosion as an environmental hazard in Iran
430 (Bui et al. 2019).

431 The areas that falls under the category of both kind of land degradation might be most
432 vulnerable concerning local self-sufficiency for food security and sustainability of
433 human activities. For instance, dust storms drive water loss through failure of
434 agricultural crops in Iran (Boroughani et al. 2022). Moreover, the adverse impacts of



435 water-induced soil erosion is known from numerous other regions (Lal and
436 Moldenhauer 2008, Gao et al. 2015, Standardi et al. 2018).



437
438 Fig. 7 Land degradation susceptibility map in terms of soil erosion and dust sources areas
439

440 **Conclusion**

441 Investigation of soil erosion through water along with wind-driven soil erosion from
442 dust sources have received little attention in past studies, despite their importance for
443 land degradation with associated social, economic, and environmental impacts. The
444 present study used several different data sets, conducted a field survey and paired the
445 data with three different machine learning algorithms to construct spatial maps for areas
446 of risk for land degradation for the Lut watershed in Iran. Three machine learning
447 algorithms were successfully applied to create land susceptibility maps describing dust
448 aerosol occurrence considering methodological uncertainty. In addition, these models
449 were used to identify the areas prone to soil erosion by surface water runoff. These
450 obtained maps were synthesized to generate a single map for risks of land degradation.



451 The results of the present study show that the random forest algorithm outperformed
452 the other two machine learning approaches for both dust sources and soil erosion
453 susceptibility mapping with an accuracy of 75% and 94%, respectively.

454 As expected, the vegetation cover, terrain elevation, land use, and geology were
455 important prerequisites for dust-emission occurrence in the watershed, while rainfall,
456 Topographical Wetness Index (TWI), terrain slope, terrain elevation, land use, and
457 geology were identified as the most influential factors for water-induced soil erosion.

458 Based on the land degradation map, almost the entire study region is at risk. A large
459 fraction of 43% of the area is prone to both high wind-driven plus water-driven soil
460 erosion. In addition to these areas, another 45% and 8% of the area have a risk for water-
461 driven and wind-driven soil erosion, respectively. These results can potentially be
462 useful for managers and policy makers to identify local hotspots for land degradation
463 to implement mitigation and adaptation measures in this watershed. Future studies
464 could work on improving the spatial resolution and coverage of the risk assessment for
465 providing more information on risks for land degradation. It requires more
466 measurements for soil erosion by water and winds to train the machine learning models.

467

468 **Acknowledgement**

469 SF acknowledges funding from the German Research Foundation (DFG) for SFB
470 1502/1–2022 (Project: 450058266).

471

472 **Conflict of Interest**

473 The authors declare that there is no conflict of interests regarding the publication of
474 this article.

475

476 **References**

477 Amiri M, Pourghasemi HR, Ghanbarian GA, Afzali SF (2019) Assessment of the importance
478 of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping
479 using three machine learning algorithms. *Geoderma* 340:55–69.
480 <https://doi.org/10.1016/j.geoderma.2018.12.042>

481 Anache JAA, Flanagan DC, Srivastava A, Wendland EC (2018) Land use and climate change
482 impacts on runoff and soil erosion at the hillslope scale in the Brazilian Cerrado. *Science of the*
483 *Total Environment* 622–623:140–151. <https://doi.org/10.1016/j.scitotenv.2017.11.257>

484 Arabameri A, Chen W, Loche M, et al (2019a) Comparison of machine learning models for
485 gully erosion susceptibility mapping. *Geoscience Frontiers*.



- 486 <https://doi.org/10.1016/j.gsf.2019.11.009>
- 487 Arabameri A, Pradhan B, Rezaei K (2019b) Gully erosion zonation mapping using integrated
488 geographically weighted regression with certainty factor and random forest models in GIS.
489 *Journal of Environmental Management* 232:928–942.
490 <https://doi.org/10.1016/j.jenvman.2018.11.110>
- 491 Avand M, Moradi HR, Lasbooyee MR (2021) Spatial prediction of future flood risk: An
492 approach to the effects of climate change. *Geosciences (Switzerland)* 11:1–20.
493 <https://doi.org/10.3390/geosciences11010025>
- 494 Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin
495 hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin
496 versant. *Hydrological sciences journal* 24:43–69
- 497 Boroughani M, Mohammadi M, Mirchooli F, Fiedler S (2022) Assessment of the impact of
498 dust aerosols on crop and water loss in the Great Salt Desert in Iran. *Computers and Electronics
499 in Agriculture* 192:106605
- 500 Boroughani, M., Pourhashemi, S., Gholami, H., & Kaskaoutis, D. G. 2021. Predicting of dust
501 storm source by combining remote sensing, statistic-based predictive models and game theory
502 in the Sistan watershed, southwestern Asia. *Journal of Arid Land*, 13(11), 1103-1121.
503
- 504 Boroughani M, Pourhashemi S (2020) Susceptibility Zoning of Dust Source Areas by Data
505 Mining Methods over Khorasan Razavi Province. *Quarterly journal of Environmental Erosion
506 Research* 9:1–22
- 507 Boroughani M, Pourhashemi S, Hashemi H, et al (2020) Application of remote sensing
508 techniques and machine learning algorithms in dust source detection and dust source
509 susceptibility mapping. *Ecological Informatics* 56:101059.
510 <https://doi.org/10.1016/j.ecoinf.2020.101059>
- 511 Breiman L (2001) Random forests. *Machine Learning* 45:5–32.
512 <https://doi.org/10.1023/A:1010933404324>
- 513 Bui DT, Shirzadi A, Shahabi H, et al (2019) A novel ensemble artificial intelligence approach
514 for gully erosion mapping in a semi-arid watershed (Iran). *Sensors (Switzerland)* 19:.
515 <https://doi.org/10.3390/s19112444>
- 516 Chicas SD, Omine K, Ford JB (2016) Identifying erosion hotspots and assessing communities
517 'perspectives on the drivers , underlying causes and impacts of soil erosion in Toledo 's Rio
518 Grande Watershed : Belize. *Applied Geography* 68:57–67.
519 <https://doi.org/10.1016/j.apgeog.2015.11.010>
- 520 Darvand S, Khosravi H, Keshtkar H, et al (2021) Comparison of machine learning models to
521 prioritize susceptible areas to dust production. *Journal of Range and Watershed Management*



- 522 74:53–68
- 523 Derakhshan-Babaei F, Mirchooli F, Mohammadi M, et al (2022) Tracking the origin of trace
524 metals in a watershed by identifying fingerprints of soils, landscape and river sediments.
525 *Science of The Total Environment* 155583
- 526 Ebrahimi-khusfi Z, Taghizadeh-mehrjardi R, Mirakbari M (2021) Evaluation of machine
527 learning models for predicting the temporal variations of dust storm index in arid regions of
528 Iran. *Atmospheric Pollution Research* 12:134–147. <https://doi.org/10.1016/j.apr.2020.08.029>
- 529 Emadodin I, Narita D, Rudolf H (2012) Soil degradation and agricultural sustainability : an
530 overview from Iran. *Environment, Development and Sustainability* 14:611–625.
531 <https://doi.org/10.1007/s10668-012-9351-y>
- 532 Gao L, Bowker MA, Xu M, et al (2017) Biological soil crusts decrease erodibility by modifying
533 inherent soil properties on the Loess Plateau, China. *Soil Biology and Biochemistry* 105:49–
534 58. <https://doi.org/10.1016/j.soilbio.2016.11.009>
- 535 Gao X, Xie Y, Liu G, et al (2015) Effects of soil erosion on soybean yield as estimated by
536 simulating gradually eroded soil profiles. *Soil and Tillage Research* 145:126–134
- 537 Garosi Y, Sheklabadi M, Conoscenti C, et al (2019) Assessing the performance of GIS- based
538 machine learning models with different accuracy measures for determining susceptibility to
539 gully erosion. *Science of the Total Environment* 664:1117–1132.
540 <https://doi.org/10.1016/j.scitotenv.2019.02.093>
- 541 Gholami H, Kordestani MD, Li J, et al (2019a) Diverse sources of aeolian sediment revealed
542 in an arid landscape in southeastern Iran using a modified Bayesian un-mixing model. *Aeolian*
543 *Research* 41:100547
- 544 Gholami H, Mohamadifar A, Sorooshian A, Jansen JD (2020a) Machine-learning algorithms
545 for predicting land susceptibility to dust emissions : The case of the Jazmurian Basin , Iran.
546 *Atmospheric Pollution Research* 11:1303–1315. <https://doi.org/10.1016/j.apr.2020.05.009>
- 547 Gholami H, Mohammadifar A, Collins AL (2019b) Spatial mapping of the provenance of storm
548 dust: Application of data mining and ensemble modelling Hamid. *Atmospheric Research*
549 104716. <https://doi.org/10.1016/j.atmosres.2019.104716>
- 550 Gholami H, Mohammadifar A, Pourghasemi HR, Collins AL (2020b) A new integrated data
551 mining model to map spatial variation in the susceptibility of land to act as a source of aeolian
552 dust. *Environmental Science and Pollution Research* 27:42022–42039
- 553 Ghorbanzadeh O, Kamran KV, Blaschke T, et al (2019) Spatial Prediction of Wildfire
554 Susceptibility Using Field Survey GPS Data and Machine Learning Approaches. *fire* 2:1–23
- 555 Gia T, Degener J, Kappas M (2018) Integrated universal soil loss equation (USLE) and
556 Geographical Information System (GIS) for soil erosion estimation in A Sap basin : Central
557 Vietnam. *International Soil and Water Conservation Research* 6:99–110.
558 <https://doi.org/10.1016/j.iswcr.2018.01.001>



- 559 Halecki W, Kruk E, Ryczek M (2018) Land Use Policy Loss of topsoil and soil erosion by
560 water in agricultural areas : A multi- criteria approach for various land use scenarios in the
561 Western Carpathians using a SWAT model. Land Use Policy 73:363–372.
562 <https://doi.org/10.1016/j.landusepol.2018.01.041>
- 563 Hosseinalizadeh M, Kariminejad N, Rahmati O, et al (2019) How can statistical and artificial
564 intelligence approaches predict piping erosion susceptibility? Science of the Total Environment
565 646:1554–1566. <https://doi.org/10.1016/j.scitotenv.2018.07.396>
- 566 Hahnenberger, M., Nicoll, K., 2014. Geomorphic and land cover identification of dust sources
567 in the eastern Great Basin of Utah, U.S.A. Geomorphology 204 (2), 657–672.
568 <https://doi.org/10.1016/j.geomorph.2013.09.013>.
- 569 Jafari M, Mesbahzadeh T, Masoudi R, et al (2021) Dust storm surveying and detection using
570 remote sensing data, wind tracing, and atmospheric thermodynamic conditions (case study:
571 Isfahan Province, Iran). Air Quality, Atmosphere & Health 1–11
- 572 Kalantar B, Pradhan B, Naghibi SA, et al (2017) Assessment of the effects of training data
573 selection on the landslide susceptibility mapping: a comparison between support vector
574 machine (SVM), logistic regression (LR) and artificial neural networks (ANN). Geomatics,
575 Natural Hazards and Risk 5705:1–21. <https://doi.org/10.1080/19475705.2017.1407368>
- 576 Kalantar B, Ueda N, Saeidi V, et al (2020) Landslide susceptibility mapping: Machine and
577 ensemble learning based on remote sensing big data. Remote Sensing 12:1–23.
578 <https://doi.org/10.3390/rs12111737>
- 579 Kim JW, Kim HG (2021) Landslide susceptibility analysis by type of cultural heritage site
580 using ensemble model: Case study of the Chungcheong Region of South Korea. Sensors and
581 Materials 33:3819–3833. <https://doi.org/10.18494/SAM.2021.3593>
- 582 Lal R, Moldenhauer WC (2008) Effects of soil erosion on crop productivity. Effects of soil
583 erosion on crop productivity 5:303–367. <https://doi.org/10.1080/07352688709382244>
- 584 Lee S, Hong S-M, Jung H-S (2017) GIS-based groundwater potential mapping using artificial
585 neural network and support vector machine models: the case of Boryeong city in Korea.
586 Geocarto International 6049:1–15. <https://doi.org/10.1080/10106049.2017.1303091>
- 587 Lee, J. A., Gill, T. E., Mulligan, K. R., Acosta, M. D., Perez, A. E. 2009. Land use/land cover
588 and point sources of the 15 December 2003 dust storm in southwestern North
589 America. Geomorphology, 105(1-2), 18-27
- 590 Mirchooli F, Motevalli A, Pourghasemi HR, et al (2019) How do data-mining models consider
591 arsenic contamination in sediments and variables importance? Environmental Monitoring and
592 Assessment 191:. <https://doi.org/10.1007/s10661-019-7979-x>
- 593 Mosavi A, Golshan M, Janizadeh S, et al (2020) Ensemble models of GLM, FDA, MARS, and
594 RF for flood and erosion susceptibility mapping: a priority assessment of sub-basins. Geocarto
595 International. <https://doi.org/10.1080/10106049.2020.1829101>



- 596 Naghibi SA, Ahmadi K, Daneshi A (2017) Application of Support Vector Machine, Random
597 Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential
598 Mapping. *Water Resources Management* 31:2761–2775. <https://doi.org/10.1007/s11269-017->
599 1660-3
- 600 Park S, Kim J, Lee J, et al (2014) Combined dust detection algorithm by using MODIS infrared
601 channels over East Asia. *Remote Sensing of Environment* 141:24–39.
602 <https://doi.org/10.1016/j.rse.2013.09.019>
- 603 Péré J-C, Rivellini L, Crumeyrolle S, et al (2018) Simulation of African dust properties and
604 radiative effects during the 2015 SHADOW campaign in Senegal. *Atmospheric Research*
605 199:14–28
- 606 Rahmati O, Mohammadi F, Saeid S, et al (2020) Identifying sources of dust aerosol using a
607 new framework based on remote sensing and modelling. *Science of the Total Environment*
608 737:139508. <https://doi.org/10.1016/j.scitotenv.2020.139508>
- 609 Rahmati O, Pourghasemi HR, Melesse AM (2016) Application of GIS-based data driven
610 random forest and maximum entropy models for groundwater potential mapping: A case study
611 at Mehran Region, Iran. *Catena* 137:360–372. <https://doi.org/10.1016/j.catena.2015.10.010>
- 612 Roy J, Saha S (2019) GIS-based Gully Erosion Susceptibility Evaluation Using Frequency
613 Ratio , Cosine Amplitude and Logistic Regression Ensembled with fuzzy logic in Hinglo River
614 Basin , India. *Remote Sensing Applications: Society and Environment* 15:100247.
615 <https://doi.org/10.1016/j.rsase.2019.100247>
- 616 Sakizadeh M, Mirzaei R, Ghorbani H (2017) Support vector machine and artificial neural
617 network to model soil pollution : a case study in Semnan Province , Iran. *Neural Computing*
618 and Applications 28:3229–3238. <https://doi.org/10.1007/s00521-016-2231-x>
- 619 Shi P, Yan P, Yuan Y, Nearing MA (2004) Wind erosion research in China: Past, present and
620 future. *Progress in Physical Geography* 28:366–386.
621 <https://doi.org/10.1191/0309133304pp416ra>
- 622 Shit PK, Pourghasemi H reza, Bhunia GS (2020) Gully Erosion Studies from India and
623 Surrounding Regions
- 624 Standardi G, Panagos P, Montanarella L, et al (2018) Cost of agricultural productivity loss due
625 to soil erosion in the European Union : From direct cost evaluation approaches to the use of
626 macroeconomic models. *Land Degradation & Development* 29:471–484.
627 <https://doi.org/10.1002/ldr.2879>
- 628 Sun W, Shao Q, Liu J, Zhai J (2014) Assessing the effects of land use and topography on soil
629 erosion on the Loess Plateau in China. *Catena* 121:151–163.
630 <https://doi.org/10.1016/j.catena.2014.05.009>
- 631 Tehrany MS, Pradhan B, Jebur MN (2014) Flood susceptibility mapping using a novel
632 ensemble weights-of-evidence and support vector machine models in GIS. *Journal of*



- 633 Hydrology 512:332–343. <https://doi.org/10.1016/j.jhydrol.2014.03.008>
- 634 Yang M, Zhu X, Pan H, et al (2019) Changes of the relationship between spring sand dust
635 frequency and large-scale atmospheric circulation. Atmospheric Research 226:102–109.
636 <https://doi.org/10.1016/j.atmosres.2019.04.004>
- 637 Youssef AM, Pourghasemi HR (2021) Landslide susceptibility mapping using machine
638 learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi
639 Arabia. Geoscience Frontiers 12:639–655
- 640 Zabihi M, Mirchooli F, Motevalli A, et al (2018) Spatial modelling of gully erosion in
641 Mazandaran Province, northern Iran. Catena 161:1–13.
642 <https://doi.org/10.1016/j.catena.2017.10.010>
- 643 Zerihun M, Mohammedyasin MS, Sewnet D, et al (2018) Assessment of soil erosion using
644 RUSLE, GIS and remote sensing in NW Ethiopia. Geoderma Regional 12:83–90.
645 <https://doi.org/10.1016/j.geodrs.2018.01.002>
- 646 Vickery, K., Eckardt, F. 2013. Dust emission controls on the lower Kuiseb River valley, central
647 Namib. Aeolian Res. 10, 125–133. <https://doi.org/10.1016/j.aeolia.2013.02.006>.
- 648 Walker, A.L., Liu, M., Miller, S.D., Richardson, K.A., Westphal, D.L., 2009. Development of
649 a dust source database for mesoscale forecasting in Southwest Asia. J. Geophys. Res. 114 (18),
650 1–24. <https://doi.org/10.1029/2008JD011541>.