

OPEN

Assessing and mapping multi-hazard risk susceptibility using a machine learning technique

Hamid Reza Pourghasemi^{1*}, Narges Kariminejad², Mahdis Amiri², Mohsen Edalat³, Mehrdad Zarafshar⁴, Thomas Blaschke⁵ & Artemio Cerdá⁶

The aim of the current study was to suggest a multi-hazard probability assessment in Fars Province, Shiraz City, and its four strategic watersheds. At first, we construct maps depicting the most effective factors on floods (12 factors), forest fires (10 factors), and landslides (10 factors), and used the Boruta algorithm to prioritize the impact of each respective factor on the occurrence of each hazard. Subsequently, flood, landslides, and forest fire susceptibility maps prepared using a Random Forest (RF) model in the R statistical software. Results indicate that 42.83% of the study area are not susceptible to any hazards, while 2.67% of the area is at risk of all three hazards. The results of the multi-hazard map in Shiraz City indicate that 25% of Shiraz city is very susceptible to flooding, while 16% is very susceptible to landslide occurrences. For four strategic watersheds, it is notable that in the Dorodzan Watershed, landslides and floods are the most important hazards; whereas, flood occurrences cover the largest area of the Maharlou Watershed. In contrast, the Tashk-Bakhtegan Watershed is so sensible to floods and landslides, respectively. Finally, in the Ghareaghaj Watershed, forest fire ranks as the strongest hazard, followed by floods. The validation results indicate an AUC of 0.834, 0.939, and 0.943 for the flood, landslide, and forest fire susceptibility maps, respectively. Also, other accuracy measures including, specificity, sensitivity, TSS, CCI, and Gini coefficient confirmed results of the AUC values. These results allow us to forecast the spatial behavior of such multi-hazard events, and researchers and stakeholders alike can apply them to evaluate hazards under various mitigation scenarios.

The Sendai Framework, with its comprehensive vision, recommends more efforts to decrease disaster risk and increase sustainable development. Especially communities who are increasingly susceptible to natural hazards should adhere to these guidelines and plan accordingly. In this regard, the multi-hazard approach is often used in risk reduction projects and studies addressing risks associated with human activities or climate change on a regional and local scale¹. It is obvious that introducing a universal set of multi-hazard assessment techniques is of fundamental importance for reducing disaster risk, and constitutes a valuable asset to share with other stakeholders, including the private sectors, local governments, and other stakeholders.

The use of the term multi-hazard in the current research is related to the objective of risk reduction among natural hazards, including flood, landslides, and forest fires, in a specified spatial distribution in this study^{2,3}. Recently, susceptibility modeling approaches related to single processes have advanced considerably for river floods⁴ and landslides⁵⁻⁷. However, there is still neither a common terminology nor a uniform conceptual approach for analyzing multiple hazards in conjunction. This is not unexpected because multi-hazard analyses are not the sum of single-hazard examinations. The various hazard characteristics and the methods used to analyze them are completely different⁸. A variety of quantification measures and susceptibility descriptions exist, which need to be adapted to enable the comparison of multiple hazards⁹. Also, natural processes have various effects on different elements at risk, and the techniques used to determine vulnerability diverge between hazards³. These topics constitute the main challenges for multi-hazard analyses.

¹Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, 71441-65186, Shiraz, Iran. ²Department of Watershed and Arid Zone Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, 49189-434, Iran. ³Crop Production and Plant Breeding Department, School of Agriculture, Shiraz University, 71441-65186, Shiraz, Iran. ⁴Natural Resources Department, Fars Agricultural and Natural Resources Research and Education Center, AREEO, Shiraz, Fars, Iran. ⁵Department of Geoinformatics-Z-GIS, University of Salzburg, 5020, Salzburg, Austria. ⁶Soil Erosion and Degradation Research Group, Department de Geografia, Universitat de València, València, Spain. *email: hr.pourghasemi@shirazu.ac.ir

The possibility of predicting which areas are susceptible to a specific type of disaster, including landslides or forest fires, is undisputed. The prediction techniques have proven valuable for predicting various characteristics of a natural disaster that has occurred¹⁰. Many researchers recognized that the occurrence of landslides and forest fires is influenced by various aspects that involve human activities and climate conditions^{11,12}. Several methods for spatially modelling landslides and forest fires have been developed^{13,14}.

Moreover, floods affect more than 20,000 human-lives per year¹⁵. In Asia, approximately 90% of all human losses are caused by floods^{16,17}. A flood that occurred in the center of Fars Province (Shiraz City) on March 25, 2019, killed 21 persons, while injuring 164 others and damaging 1186 homes. In terms of economic impact, financial losses were estimated to be about \$ 9,344,615 (<http://www.irna.ir/fars/fa/News/83266320>). In recent years, with the help of GIS and RS technology, the accuracy of flood susceptibility maps has been improved. Techniques include frequency ratio, logistic regression¹⁸, weights-of-evidence¹⁹, fuzzy logic²⁰, artificial neural networks²¹, decision tree²², support vector machines (SVM)²³, and Random forest models²⁴. In this study, the RF model was selected because it is a very fast machine learning method. It produces an accurate classifier with an internal unbiased estimate of generalizability during the forest building processes²⁵. It makes no statistical assumptions, and it is characterized by high prediction performance^{13,26}.

In the present study, the assessment was carried out for Fars Province (133,400 km²), which is strongly affected by floods, landslides, and forest fires. However, the area is also influenced by other climatic hazards (such as gully erosion), which are not systematically recorded at a municipal level. In this research, the first step was to assess the importance of effective factors on flood, landslide, and forest fire occurrence using the Boruta algorithm. Next, the aim was to prepare susceptibility maps for different hazards using the RF data mining algorithm. Then, the three risks were combined in a multi-hazard probability index (MHPI) with respect to their occurrence probability and the range of susceptibility classes. Based on extensive literature review and to the best of our knowledge, no research related to the multi-hazard modeling of floods, landslides, and forest fires exists to date.

Study Area

The study area is in the Fars Province (SE Iran) between 27° 2' to 31° 42' N latitudes and 50° 42' to 55° 36' E longitudes, with an area of approximately 133,400 km² of mainly arid and semi-arid land²⁷. It covers 8.1% of Iran and includes 26 cities (Fig. S1). The Fars Province has three different atmospheric regions: First, the mountainous areas in the north and northwest with moderately cold winters and moderate summers. Second, the central regions with rainy winters and hot, dry summers. And, third, the southern and southeastern regions with cold winters and hot summers²⁸. The geology of the study area is shown in Table S1.

Material and Methods

The methodology of the presented study is shown in Fig. 1. The flow chart comprises three main steps, namely 1) data preparation, i.e. obtaining the location of 365 floods, 358 forest fires, and 179 landslides based on intensive fieldwork using a Global Positioning System (GPS) and different province reports; 2) recognizing the most important factors contributing to the occurrence of floods, forest fires, and landslides using the Boruta algorithm; 3) constructing flood, forest fire, and landslide susceptibility maps along with validation processes using the RF model; and, finally, preparing a MHPI in the study area.

Gathering data related to flood, forest fire, and landslide hazards. Detailed terrain mapping was carried out to locate and recognize floods, forest fires, and landslides. Also, the position of these three hazards was recorded using a GPS receiver. All data were used to verify the locations of these hazards mapped during the fieldwork. The susceptibility modeling technique applied in this study requires samples of both hazards and non-hazards to generate susceptibility maps. Of the total recorded hazard events (365 floods (Fig. S2a), 358 forest fires (Fig. S2b), and 179 landslides (Fig. S2c)) that occurred in the Fars Province, 70% were used in the model building process, while the remaining 30% were used in the validation step (Fig. S2). The same number of non-hazard locations was randomly sampled in the study area. Also, the values of all effective factors were extracted for both samples to validate and train datasets for further processing.

Construction of flood, forest fire, and landslide conditioning factors. The main factors influencing the occurrence of forest fires are divided into two groups, namely, biophysical and human factors. The biophysical factors are further divided into atmospheric factors (humidity, rainfall, and temperature), and topographic factors (altitude, slope aspect, slope degree). The human factors include land use, access to the forest, and fuel management processes^{29,30}. Landslides are influenced by a collection of geo-environmental and anthropological factors³¹. To evaluate the importance of the various effective factors controlling hazard locations, 12 (flood), 10 (forest fire), and 10 (landslide) factors were selected. The factors used in this study to predict the occurrence of flood events are altitude, slope angle, aspect, plan curvature, TWI, distance from rivers, distance from roads, drainage density, lithology, rainfall, land use, and soil features. For landslides, the selected influencing factors are altitude, slope aspect, slope degree, plan curvature, profile curvature, distance from rivers, distance from roads, distance from faults, lithology, and land use. Moreover, for the occurrence of forest fires, the selected effective factors are altitude, slope aspect, slope degree, TWI, distance from rivers, distance from roads, drainage density, distance from urban, rainfall, and annual mean temperature. Topographical factors were extracted using a digital elevation model (ASTER-GDEM) with a spatial resolution of 30 m. Data layers were prepared using ArcGIS 10.2.2 with the pixel size of 30 m². The distance from rivers, roads, and urban maps was obtained from the rivers, roads, and urban areas maps, respectively. The lithology map was obtained from the Geological Survey of Iran at a scale of 1:100,000. The land use map of the study area was also obtained from the Natural Resources Office of Fars Province at a scale of 1:100,000 and was updated using Google Earth images. Finally, the soil feature map was prepared

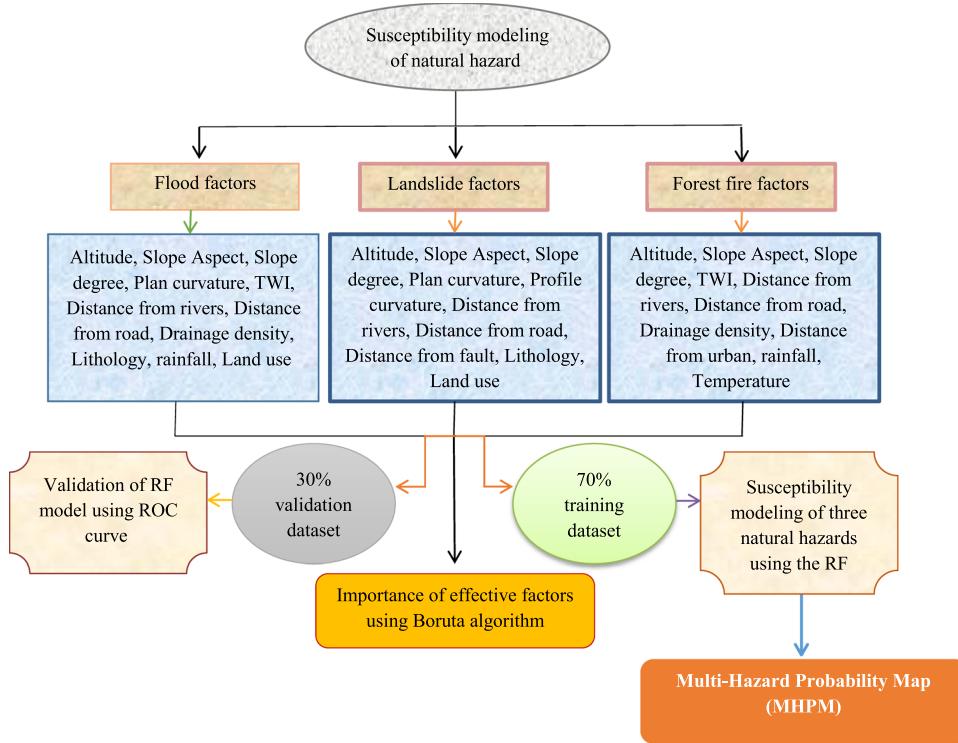


Figure 1. Flowchart of methodology used for multi-hazard spatial modeling in the Fars Province, Iran.

using data of the soil and water research institute of Fars Province, provided at scale of 1:100,000 and detailed in Table 2S. In general, all effective factor maps are shown in Fig. S3 (a-p).

Boruta algorithm. The Boruta algorithm was used to prioritize the selected factors affecting natural hazards. The Boruta algorithm is built on the combined dataset by the Random forest classifier and performed in the R statistical package³². Boruta is based on a similar viewpoint to that which underlies the Random forest classifier. However, by increasing randomness to the system and collecting results from the ensemble of randomized samples, the devious impact of random variation and relations decrease. Here, this extra randomness shall provide us with a clearer view of which properties are really important³³. This algorithm has been successfully applied in predicting gully erosion in Iran³⁴.

Random forest data mining model. Random forest (RF) is a supervised classifier^{35,36} that consists of many decision trees and has low error in contrast to other classification algorithms. In this study, the number of trees, minimum node size, and the number of features were used to split each node³⁵. However, if one of the predictors has a much stronger effect on the predicting function than the other factors, that predictor is going to be the top splitter in all the trees. Consequently, all trees are going to be similarly constructed and, hence, correlated. Averaging predictions from correlated trees may not decrease the variance significantly³⁶.

Evaluation of susceptibility maps produced by random forest. The area under the ROC curve (AUC) indicates the capability of a model to properly predict the occurrence or non-occurrence of landslides, forest fires, and floods. The ROC curve represents the trade-off between two rates (the false-positive and true-positive rates on the X and Y axes). The AUC values are interpreted as reflecting the following model accuracies: 0.6–0.7 poor, 0.6–0.7 medium, 0.7–0.8 good, 0.8–0.9 very good, and 0.9–1 excellent^{37,38}. In the current study, different techniques and measures were applied to evaluate the robustness and uncertainty of the RF model for three different hazards, namely, floods, forest fires, and landslides. These accuracy measures are the true positive rate (TPR), false positive rate (FPR), F-measures, fallout, sensitivity, specificity, true skill statistics (TSS), overall accuracy, corrected classified instances (CCI), and the Gini coefficient^{39–41}. All of these indices were calculated based on the four parameters of true negative (TN), false positive (FP), false negative (FN), and true positive (TP).

The TPR (sensitivity) and TNR (specificity) show the probability of correct predictions of the positives and negatives as observed in the reality. The FPR (1–specificity) indicate the probability of incorrect predictions of non-event location as an event. TSS also measure the ability of a predicted value to discriminate between the events and non-events, using all of the elements in the confusion matrix⁴². The CCI considers TN and FN for true- and false-negative predicted events, and TP and FP for true- and false-positive, respectively. The coefficient of variation may often be suggested over the Gini coefficient if a measure of relative precision is selected to evaluate inequality⁴³.

Factors	Mean Importance	Median Importance	Min Importance	Max Importance	Decision
Altitude	6.15	6.08	3.19	8.39	Confirmed
Aspect	12.39	12.41	10.53	14.34	Confirmed
Slope	8.74	8.74	6.77	10.92	Confirmed
Plan curvature	7.43	7.50	5.29	9.14	Confirmed
Distance from roads	9.81	9.84	7.63	11.63	Confirmed
Distance from rivers	14.07	14.24	11.80	16.46	Confirmed
Drainage density	21.21	21.34	19.65	22.90	Confirmed
Rainfall	9.70	9.73	7.98	11.68	Confirmed
TWI	20.97	20.83	18.17	22.96	Confirmed
Lithology	12.07	12.12	10.03	13.55	Confirmed
Land use	33.23	33.38	30.74	35.60	Confirmed
Soil	2.99	2.92	0.85	4.72	Confirmed

Table 1. Considering flood variables importance using by Boruta algorithm.

$$TPR = TP/(TP + FN) \quad (1)$$

$$FPR = FP/(TN + FP) \quad (2)$$

$$F\text{-measure} = 2 * \text{precision} * \text{recall}/(\text{precision} + \text{Recall}) \quad (3)$$

$$\text{Precision} = TP/(TP + FP) \quad (4)$$

$$\text{Recall} = TPR \quad (5)$$

$$\text{Fallout} = FP/(TP + FP) \quad (6)$$

$$\text{Specificity} = TN/(TN + FP) \quad (7)$$

$$\text{Sensitivity} = TP/(TP + FN) \quad (8)$$

$$\text{TSS} = \text{Sensitivity} + \text{Specificity} - 1 \quad (9)$$

$$\text{CCI} = (TN + TP)/(TN + TP + FP + FN) * 100 \quad (10)$$

$$\text{Gini coefficient} = 2 * \text{AUC} - 1 \quad (11)$$

Results

Prioritizing and determining effective factors using the Boruta algorithm. The first aim of using the Boruta algorithm was to select the best conditioning factors for the occurrence of landslides, forest fires, and floods. The resulting rank of features for these three hazards according to their importance is shown in Tables 1–3. According to the mean importance for the occurrence of flood events depicted in Table 1, land use (33.23), drainage density (21.21), and TWI (20.97) are the most important factors, followed by distance from rivers (14.07), aspect (12.39), lithology (12.07), distance from roads (9.81), rainfall (9.70), slope (8.74), plan curvature (7.43), altitude (6.15), and soil (2.99). The highest rank of effective factors for the occurrence of forest fires was assigned to closeness to residential areas (35.36), slope (20.07), aspect (15.03), rainfall (13.48), distance from rivers (9.46), annual mean temperature (8.64), TWI (6.40), and land use (2.71) (Table 2). However, distance from roads (2.24) and altitude (0.08) were found to have no relevance among all considered factors. Additionally, the ranking of effective factors for the occurrence of landslides assigned the highest value to slope (15.95), followed by distance from rivers (12.56), lithology (10.50), land use (8.20), profile curvature (7.08), aspect (6.29), altitude (5.85), and distance from faults (4.82) (Table 3). Distance from roads (1.07) and plan curvature (0.09) were rejected based on this algorithm.

Providing natural hazard susceptibility maps using the RF model. In this study, susceptibility maps were produced for three natural hazards using the RF model (Fig. 2). Based on the flood susceptibility map produced by the RF model, 39.03%, 31.20%, 18.04%, and 11.73% of the total area in Fars Province are considered as having low, moderate, high, and very high flood susceptibility, respectively (Figs. 2a and 3). According to the forest fire susceptibility map, 74.57%, 8.57%, 7.80%, and 9.07% of the total area was classified into low, moderate, high and very high classes of susceptibility, respectively (Figs. 2b and 3). Also, the landslide susceptibility map

Factors	Mean Importance	Median Importance	Min Importance	Max Importance	Decision
Distance from rivers	9.46	9.48	6.66	12.59	Confirmed
Residential areas	35.36	35.29	32.15	39.12	Confirmed
Distance from roads	2.24	2.26	0.37	4.18	Rejected
TWI	6.40	6.46	3.36	8.66	Confirmed
Slope	20.07	20.21	17.38	23.16	Confirmed
Rainfall	13.48	13.42	10.04	16.94	Confirmed
Altitude	0.08	0.18	-1.71	2.12	Rejected
Aspect	15.03	15.11	10.96	17.83	Confirmed
Temperature	8.64	8.64	6.80	11.19	Confirmed
Land use	2.71	2.81	-0.65	4.83	Confirmed

Table 2. Considering forest fire variables importance using by Boruta algorithm.

Factors	Mean Importance	Median Importance	Min Importance	Max Importance	Decision
Plan curvature	0.09	0.03	-2.04	1.73	Rejected
Aspect	6.29	6.21	4.43	8.72	Confirmed
Altitude	5.85	5.86	2.93	8.84	Confirmed
Distance from faults	4.82	4.87	1.67	7.21	Confirmed
Distance from roads	1.07	1.17	-1.93	3.23	Rejected
Distance from rivers	12.56	12.53	10.31	14.22	Confirmed
Profile curvature	7.08	7.12	5.30	9.61	Confirmed
Slope	15.95	15.90	13.51	18.58	Confirmed
Lithology	10.50	10.47	8.76	12.26	Confirmed
Land use	8.20	8.20	6.64	10.90	Confirmed

Table 3. Considering landslides variables importance using by Boruta algorithm.

derived from the RF model implied that the largest part of the study area (49.76%) has low susceptibility to landslide occurrence, 34.08% of the area has medium susceptibility to landslide occurrence, while the high and very high susceptibility classes cover 13.22% and 2.95% of the area, respectively (Figs. 2c and 3). The spatial aggregation of these three susceptibility maps produced by the RF model confirmed that the majority of the study area has a low susceptibility to the occurrence of flood, forest fire, and landslide events.

In order to produce a multi-hazard probability map (Fig. 4), all three hazard susceptibility maps were considered together. The susceptibility classes of the multi-hazard map produced for Fars Province confirmed that 42.83% of the study area is not susceptible to any hazards. Meanwhile, areas of 17.26%, 5.95%, and 14.16% were found to be at risk of floods, landslides, and forest fire, separately. Regarding multi-hazard susceptibility, 0.95% of the study area was found to be at risk of floods and forest fires together. Moreover, the combined risk of forest fires and landslides was detected for 7.28% of the study area. Regarding the combined risk of floods and landslides, 8.87% of the study area was deemed susceptible, while, finally, 2.67% of Fars Province was found to be at risk of all three hazards together (Fig. S4).

Considering natural hazards in Shiraz City (center of Fars Province). The results of our multi-hazard analysis (Fig. 5) presented for Shiraz City indicated that 25% of Shiraz City is very susceptible to flood occurrence, whereas about 1.12% of the study area is susceptible to both floods and forest fires. Furthermore, 18.13% and 3.57% of Shiraz City face the combined risk of landslides and floods, respectively landslides, forest fires, and floods. The flood event that occurred on March 25, 2019, which killed 21 persons, injured 164 others, damaged 1,186 homes, and caused financial losses around \$ 9,344,615, impacted two areas, in particular, namely the Quran Gate and Saadi Zone (<http://www.irna.ir/fars/fa/News/83266320>). One of the most important findings and achievements of this study is that the prepared EMHM could very accurately predict flood events in the areas of Quran Gate and Saadi Zone.

Studying the susceptible watersheds of natural hazards. Dorodzan Watershed is one of the strategic areas of Fars Province and plays a very important role in the agricultural production and self-sufficiency of Iran. It is an area which supplies water resources to Tashk and Bakhtegan lakes and is affected by wind erosion. Figure 6 illustrates that 36.35% and 68.64% of Dorodzan Watershed are covered by the low class of susceptibility to flood and forest fire, respectively. However, regarding landslides, the moderate susceptibility class covers the largest area (42.45%). The Maharloo Watershed, as a second grade watershed of the Ministry of Energy, is the main source of the Kor River. In this watershed (Fig. 6), the moderate class covers the largest area (27.76%), although the classes of low susceptibility to forest fires (88.81%) and landslides (36.48%) covered the greatest area in Maharloo

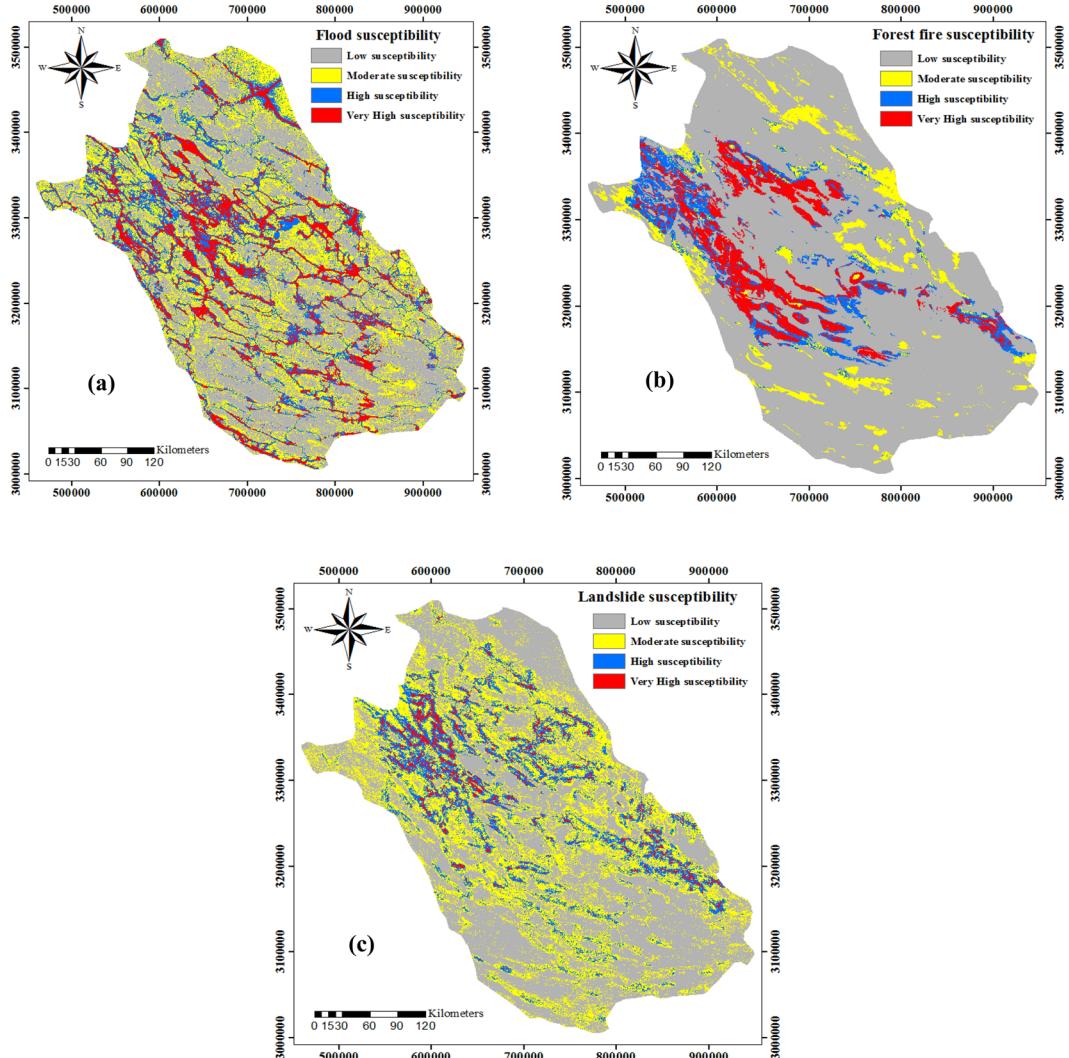


Figure 2. The susceptibility maps of three natural hazards produced using the random forest model.

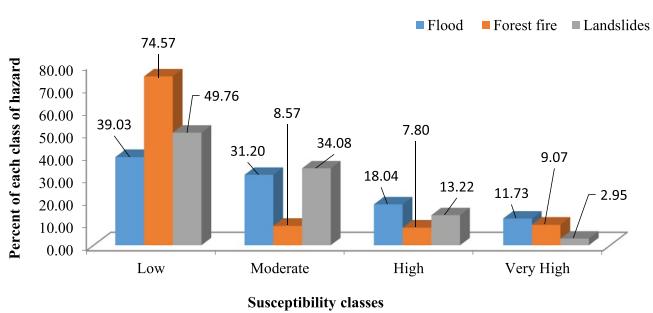


Figure 3. The susceptibility maps of three natural hazards produced using the random forest model.

Watershed. The Ghareaqaj Watershed, which is currently used for drinking and agricultural purposes, is one of the most important rivers in Fars Province. The construction of the Salman Farsi Dam in Qir and Karzin and studies on the construction of the Kavar Dam on this river indicates the importance of the river in the mentioned province. In this watershed (Fig. 6), all three hazards (floods, forest fires, and landslides) pose a low risk (37.68%, 64.71%, and 42.80%). Moreover, the most important source of water supply are the Bakhtegan and Tishak lakes. Based on Fig. 6 (Tashk-Bakhtegan watershed), the low susceptibility class covers the largest area of flood (33.85%) and forest fire (77.59%), while, based on the landslide susceptibility map produced by the RF model, 38.19% of the total area was covered by the moderate class.

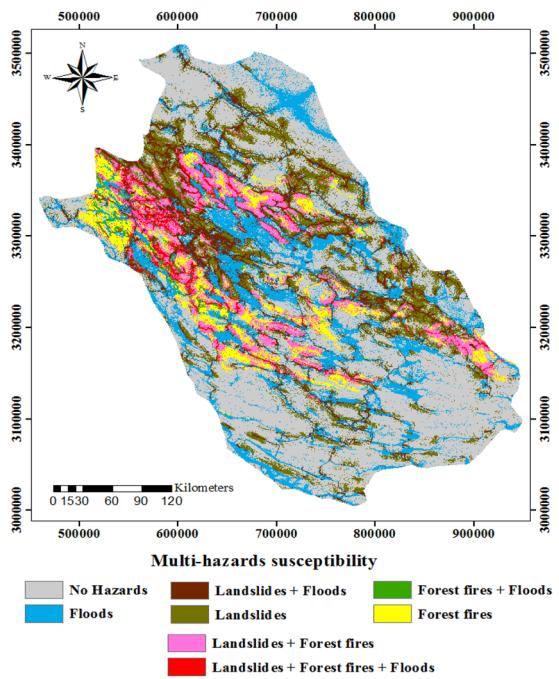


Figure 4. Percentages of susceptibility classes of multi-hazard in Fars province.

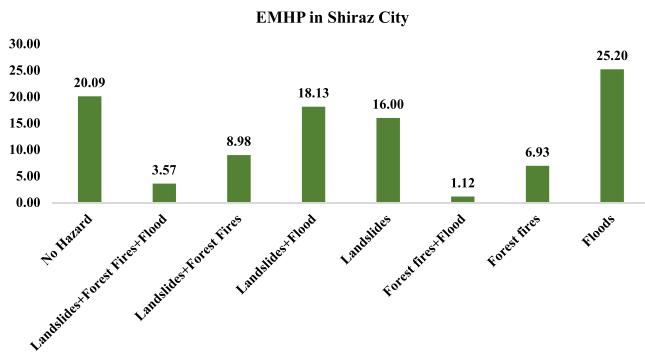


Figure 5. EMHP and percentage of each hazard in Shiraz City.

Studying the validation of natural hazard susceptibility maps. In order to produce natural hazard susceptibility maps, all hazards were divided into two data sets: one for modeling and one for validation. The accuracy of the three maps produced by the RF model was verified using ROC curves (Table S3). The AUC values for the flood, forest fire, and landslide maps were 0.834, 0.943 and 0.939, respectively. Regarding the standard error, floods had the highest value (0.028), followed by forest fires (0.016), and landslides (0.023). Further, the forest fire map had excellent accuracy (0.958), while the model considered the landslide and flood maps as very good. Also, the results of the other measures (Table 4) confirmed the accuracy of the three hazard maps, as, according to Table 4, the F-measure, specificity, and sensitivity of each hazard is more than 0.77. Furthermore, the TSS index is 0.541 for floods, which indicates fair accuracy, whereas its values for landslides and forest fires were 0.889 and 0.850, respectively, indicating an excellent model, based on the findings of Allouche *et al.* (2006)⁴⁴. Also, according to published reports, a Gini coefficient value above 0.6 (60%) indicates a good model in terms of accuracy. On the other hand, when the CCI (overall accuracy) is between 0.6–0.8, it shows that the accuracy of the model is good. So, the RF model is known as an accurate classifier for the three depicted hazards.

Discussion

In this study, the importance of factors controlling landslide, flood, and forest fire locations was analyzed using the Boruta algorithm. The Boruta algorithm provided quantitative results, which is a significant advantage that allows the potential comparison of studies in different regions around the world. As it was already stated, the study area is prone to combinations of landslides, floods, and forest fires. Generally, the development and formation of these natural hazards are controlled by several factors, and the distribution of these hazards cannot be random.

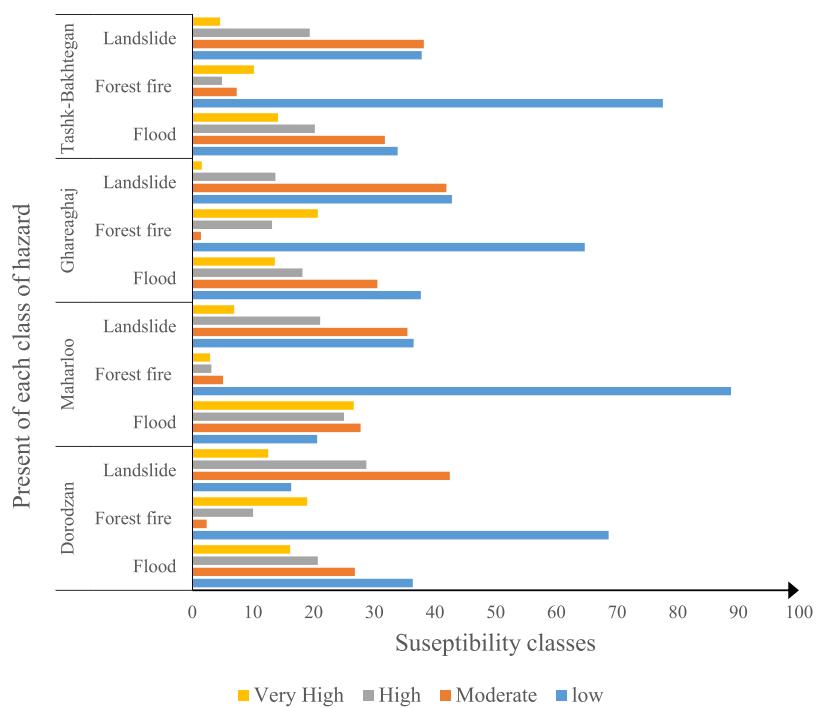


Figure 6. Percentages of susceptibility classes of multi-hazard in Dorodzan, Maharloo, Ghareaghaj, and Tashk-Bakhtegan Watershed.

Hazards	TN	FP	FN	TP	TPR	FPR	F-measures	Fallout	Specificity	Sensitivity	TSS	CCI	Gini
Flood	84	25	25	84	0.771	0.229	0.771	0.229	0.771	0.771	0.541	77.06	0.668
Landslides	48	6	6	48	0.889	0.111	0.889	0.111	0.889	0.889	0.778	88.89	0.878
Forest fire	91	16	16	91	0.850	0.150	0.850	0.150	0.850	0.850	0.701	85.05	0.886

Table 4. Different robustness measures for validation of the built model of each hazard.

The most statistically significant relationship among factors, based on the Boruta algorithm, was found between flood location and land use, and land use presented as the most important factor influencing flood hazards among all considered variables. Wheater and Evans (2009)⁴⁵ implied that land use affects the hydrology that determines water resources leading to flood hazards. It is increasingly identified that the management of water and land are strongly linked. Generally, steeper slopes are more vulnerable to massive erosion, including landslides. The steepness of slopes is reported as a factor of primary importance that promotes high runoff velocity, which results in this type of erosion. Regarding slope and aspect, forest fires predominantly occur in the steep slopes of the southern areas, as vegetation is typically dry. The effects of slope and aspect on fire behavior in the occurrence of forest fires have been reported by Adab *et al.* (2013)⁴⁶. According to Pourghasemi (2016)⁴⁷, topographic data (i.e. slope and aspect) are the most important factors for forest fire assessment. Contrary to the above results, Bui *et al.* (2017)⁴⁸ found that NDVI (Normalized Difference Vegetation Index) had the strongest impact on the occurrence of forest fires. However, Hong *et al.* (2017)⁴⁹ and Gigovic *et al.* (2019)⁵⁰ respectively demonstrated that slope has a significant positive effect on the occurrence of forest fire events. In relation to flood effective factors, the research carried out by Liu *et al.* (2005)⁵¹ confirmed that the urbanization scenario has a strong influence on heightening flood volume. For instance, afforestation has a positive impact, while deforestation has a negative impact on the occurrence of floods.

Besides determining variable importance, the RF model was used to prepare susceptibility maps for landslides, floods, and forest fires, first separately (Fig. 2), and then jointly in the form of a multi-hazard map (Fig. 4). The susceptibility maps for floods, landslides, and forest fires revealed that most of the study area is characterized by low susceptibility to each hazard when analyzed separately (Fig. 3). The multi-hazard probability map modeled by RF revealed that the most parts of the study area are not susceptible to any hazards, whereas few areas are at risk of all three hazards together (Fig. 3). Floods are recognized as the most dangerous hazard in the study area, followed by landslides and forest fires (Fig. S4). Further, effective flood risk reduction requires more analysis of this individual hazard and its interaction with the other hazards. Additionally, the validation of the RF models determined an excellent accuracy of the forest fire and landslide susceptibility maps (Table S3). Pourghasemi *et al.* (2019)⁵² produced a susceptibility map for three hazards (i.e. landslides, floods, and earthquakes) using the ensemble model named SWARA-ANFIS-GWO. They showed that 17.14% of the area is affected by no hazards, whereas most parts were susceptible to landslide and flood hazards together (33.70%). They also indicated accuracies of

84% and 80% for flood and landslide maps, respectively. Skilodimou *et al.* (2019)⁵³ applied the analytical hierarchy process (AHP) to produce separate maps for landslide, flood, and earthquake hazards and combined them into a single multi-hazard map. They showed that 80% of the landslide occurrences and all the recorded flood events fall within the boundaries of the moderate, low and very low susceptibility classes.

There are several advantages that make the RF model suitable for the approach in the present study. First, it is a simple, fast algorithm that makes no statistical assumptions and is characterized by a high prediction performance^{54,55}. It produces an internally unbiased evaluation of generalizability with an accurate classifier during the forest building processes⁵⁶ and provides better consistency of results and robustness of forecasts⁵⁶. The RF can precisely handle heterogeneous inputs of different nature and scalability from different sources^{55,57}. Another important benefit of the RF model is that there are significant criteria that indicate the importance of each predictor variable^{55,58}. However, it has some sources of uncertainty that are frequently unacknowledged or even unrecognized.

One source of uncertainty in the modeling process is related to the gathered data. It is important to consider non-linear correlations among dependent and independent variables; this problem can be solved by machine learning techniques. One of the advantages of machine learning techniques in comparison to traditional methods (bivariate and multivariate statistical methods) is that the ML algorithms can deal with noises in the data and are also accurate in the presence of uncertain data and limited measurement errors. Quality of data is also important. In the current study, different extensive field surveys were conducted to collect suitable data for all three hazards; however, according to the accuracy of the flood susceptibility map (Tables S3 and 4), there appears to be greater uncertainty compared to the landslide and forest fire hazards, because the selection of flood locations is so difficult compared to other hazards. Another uncertainty source is the accuracy of the built model. For solving this problem, different techniques were applied, and the results are presented in Table 4. According to Table 4, the achieved results of the AUC values confirmed the accuracy of the built model for the three examined hazards, namely floods, landslides, and forest fires. Also, dividing the entire dataset into two sets for training (70%) and validation (30%) can be effective in decreasing uncertainties in a model's performance. Another uncertainty source is limitations of the learned model that the ML techniques such as the RF isn't faced to this problem, meanwhile this algorithm for removing this uncertainty, used from error rates (Table 4) and out-of-bag indicator. Results of the out-of-bag values for forest fires, landslides, and floods were 3.55%, 15.6%, and 22.27%, respectively.

Nowadays, the necessity of using machine learning techniques is increasingly emphasized in the susceptibility modeling of geomorphological features and processes³⁷. A universal framework describing which factors to compare is required. This general framework can be semi-quantitative, qualitative, or quantitative³. It should be suitable for both single hazard and multi-hazard assessments, because multi-hazard evaluation plays the main role in reducing disaster risk and provides crucial information for sharing with the other stakeholders, such as local governments and private sectors⁵⁵. Considering multi hazards jointly and applying the same technique to analyze them can give us a comprehensive view of the changes occurring in the environment. Further, a synthesized multi-hazard probability map supports planners in sustainable development and adaptive management because this map provides homogenized information about different environmental hazards for a specific area⁶⁴. It means that the potential use of hazard evaluation becomes obvious when considering all hazards together, on the basis of which plans and projects can be implemented considering this comprehensive view of a region⁵⁹. From this point of view, a multi-hazard probability map can be used for integrated and comprehensive watershed management and land use planning and, consequently, for the sustainable development of a region.

Conclusion

A better understanding of the factors controlling flood, forest fire, and landslide occurrence is crucial to the sustainable development of regions prone to these three hazards, such as the Fars Province. In this study, 365 floods, 358 forest fires, and 179 landslides were mapped for an area of 133,400 km². The Boruta algorithm enabled us to analyze the impact of effective factors on the occurrence of three different natural hazards. According to the Boruta algorithm, the most important factor controlling flood occurrence in the study area was land use, followed by drainage density, and TWI. Among the different factors controlling forest fire occurrence, residential areas ranked highest, followed by slope, and aspect. Moreover, the highest rank of conditioning factors regarding landslide occurrence was found to be slope, followed by distance from rivers, and lithology. The RF model was also applied to prepare a susceptibility map of flood, landslide, and forest fire locations. The multi-hazard probability map produced for floods, forest fires, and landslides in Fars Province revealed that the majority of the land is not prone to any hazards. Total areas of 17.26%, 5.95%, and 14.16% were found to be at risk of floods, landslides, and forest fire, separately. However, 2.67% of Fars Province was found to be at risk of all three hazards together. Based on the AUC values, the best accuracy was determined for the forest fire susceptibility map, followed by the maps produced for landslides, and floods. Further, the multi-hazard probability map prepared in this study can be used for integrated and comprehensive watershed management and land use planning and, consequently, for sustainable development in the study region.

Received: 4 June 2019; Accepted: 10 February 2020;

Published online: 21 February 2020

References

1. Nations, U. UNISDR terminology on disaster risk reduction. *United Nations Office for Disaster Risk Reduction, Report* (2009).
2. Duncan, M., Edwards, S., Kilburn, C., Twigg, J., & Crowley, K. An interrelated hazards approach to anticipating evolving risk. In: GFDRR (Ed.), *The Making of a Riskier Future: How Our Decisions Are Shaping Future Disaster Risk*. Global Facility for Disaster Reduction and Recovery, Washington, USA, 114–121 (2016).
3. Kappes, M. S., Keiler, M., von Elverfeldt, K. & Glade, T. 2012. Challenges of analyzing multi-hazard risk: a review. *Nat. Hazards*. **64**(2), 1925–1958 (2012).

4. Hunter, N. M., Bates, P. D., Horritt, M. S. & Wilson, M. D. Simple spatially-distributed models for predicting flood inundation: a review. *Geomorphology* **90**(3–4), 208–225 (2007).
5. Dai, F. C., Lee, C. F. & Ngai, Y. Y. Landslide risk assessment and management: an overview. *Eng. Geol.* **64**(1), 65–87 (2002).
6. Crozier, M. J., & Glade, T. Landslide hazard and risk: issues, concepts and approach. *Landslide Hazard and Risk* 1–40 (2005).
7. Fell, R., Ho, K., Lacasse, S. & Leroi, E. A framework for landslide risk assessment. In: Proceedings of the International Conference on Landslide Risk Management, Vancouver, Canada, 3–25 (2005).
8. Carpignano, A., Golia, E., Di Mauro, C., Bouchon, S. & Nordvik, J. P. A methodological approach for the definition of multi-risk maps at regional level: first application. *J. Risk Res.* **12**(3–4), 513–534 (2009).
9. Marzocchi, W., Garcia-Aristizabal, A., Gasparini, P., Mastellone, M. L. & Di Ruocco, A. Basic principles of multi-risk assessment: a case study in Italy. *Nat. Hazards* **62**(2), 551–573 (2012).
10. Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A. & Chakraborty, B. A review on application of data mining techniques to combat natural disasters. *Ain Shams Eng. J.* **9**(3), 365–378 (2018).
11. Chen, W., Pourghasemi, H. R. & Zhao, Z. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto Int.* **32**(4), 367–385 (2017).
12. Hong, H. *et al.* Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China. *Sci. Total Environ.* **630**, 1044–1056 (2017).
13. Naghibi, S. A., Ahmadi, K. & Daneshi, A. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resour. Manage.* **31**, 2761–2775 (2017).
14. Kamalakkannan, J., Chakrabortty, A., Bothra, G., Pare, P. & Kumar, C.P. Forest fire prediction to prevent environmental hazards using data mining approach. In *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology* 615–622 (2019). Springer, Singapore.
15. Sarhadi, A., Soltani, S. & Modarres, R. Probabilistic flood inundation mapping of ungauged rivers: Linking GIS techniques and frequency analysis. *J. Hydrol.* **458**, 68–86 (2012).
16. Dutta, D. & Herath, S. Trend of floods in Asia and flood risk management with integrated river basin approach. In *Proceedings of the 2nd international conference of Asia-Pacific hydrology and water resources Association, Singapore* (1), 55–63 (2004).
17. Khosravi, K. *et al.* A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total Environ.* **627**, 744–755 (2018).
18. Nandi, A., Mandal, A., Wilson, M. & Smith, D. Flood hazard mapping in Jamaica using principal component analysis and logistic regression. *Environ. Earth Sci.* **75**, 465 (2016).
19. Tehrany, M. S., Pradhan, B. & Jebur, M. N. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J. Hydrol.* **512**, 332–343 (2014).
20. Pulvirenti, L., Pierdicca, N., Chini, M. & Guerriero, L. An algorithm for operational flood mapping from Synthetic Aperture Radar (SAR) data using fuzzy logic. *Nat. Hazards Earth Sys. Sci.* **11**, 529–540 (2011).
21. Kia, M. B. *et al.* An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environ. Earth Sci.* **67**(1), 251–264 (2012).
22. Tehrany, M. S., Pradhan, B. & Jebur, M. N. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *J. Hydrol.* **504**, 69–79 (2013).
23. Tehrany, M. S., Pradhan, B. & Jebur, M. N. Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. *Stoch Environ. Res. Risk Assess.* **29**, 1149–1165 (2015).
24. Wang, Z. *et al.* Flood hazard risk assessment model based on random forest. *J. Hydrol.* **527**, 1130–1141 (2015).
25. Hosseinalizadeh, M. *et al.* 7 Tiefenbacher, J.P. Gully headcut susceptibility modeling using functional trees, naïve Bayes tree, and random forest models. *Geoderma* **342**, 1–11 (2019).
26. Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H. R. & Feizizadeh, B. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology* **298**, 118–137 (2017).
27. Nafarzadegan, A. R. *et al.* Drought area monitoring during the past three decades in Fars province, Iran. *Quaternary Int.* **250**, 27–36 (2012).
28. Zare, M. *et al.* Spatiotemporal clustering of cutaneous leishmaniasis in Fars province, Iran. *Asian Pacific J. Tropical Biomed.* **7**(10), 862–869 (2017).
29. Countryman, C. M. The fire environment concept. USDA forest service, pacific southwest forest and range experiment station. *General Technical Report PSW 7*, 12 (1972).
30. Dickson, B. G. *et al.* Mapping the probability of large fire occurrence in northern Arizona, USA. *Landscape Ecol.* **21**(5), 747–761 (2006).
31. Moosavi, V. & Niazi, Y. Development of hybrid wavelet packet-statistical models (WP-SM) for landslide susceptibility mapping. *Landslides* **13**(1), 97–114 (2016).
32. Liaw, A. & Wiener, M. Classification and regression by random forest. *R news* **2**(3), 18–22 (2002).
33. Kursa, M. B. & Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Softw.* **36**(11), 1–13 (2010).
34. Amiri, M., Pourghasemi, H. R., Ghanbarian, G. A. & Afzali, S. F. Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. *Geoderma* **340**, 55–69 (2019).
35. Farnaaz, N. & Jabbar, M. A. Random forest modeling for network intrusion detection system. *Procedia Comput. Sci.* **89**, 213–217 (2016).
36. Partopour, B., Paffenroth, R. C. & Dixon, A. G. Random forests for mapping and analysis of microkinetics models. *Comput. & Chem. Eng.* **115**, 286–294 (2018).
37. Pourghasemi, H. R., Yousefi, S., Kornejady, A. & Cerdà, A. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Sci. Total Environ.* **609**, 764–775 (2017).
38. Zabihi, M. *et al.* Spatial modelling of gully erosion in Mazandaran Province, northern Iran. *Catena* **161**, 1–13 (2018).
39. Somodi, I., Lepesi, N. & Botta-Dukat, Z. Prevalence dependence in model goodness measures with special emphasis on true skill statistics. *Ecol. & Evolution* **7**, 863–872 (2017).
40. Fielding, A. H. & Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **24**(1), 38–49 (1997).
41. Rossi, M. & Reichenbach, P. LAND-SE: a software for statistically based landslide susceptibility zonation, version 1.0. *Geoscientific Model. Development* **9**, 3533–3543 (2016).
42. Rahmati, O. *et al.* PMT: New analytical framework for automated evaluation of geo-environmental modelling approaches. *Sci. Total Environ.* **664**, 296–311 (2019).
43. Liu, H., Jia, Y., Niu, C. and Gan, Y., 2019. Spatial Pattern Analysis of Regional Water Use Profile Based on the Gini Coefficient and Location Quotient. *JAWRA Journal of the American Water Resources Association*.
44. Allouche, O., Tsoar, A. & Kadmon, R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**, 1223–1232 (2006).
45. Wheater, H. & Evans, E. Land use, water management and future flood risk. *Land use Policy* **26**, S251–S264 (2009).
46. Adab, H., Kanniah, K. D. & Solaimani, K. Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. *Nat. Hazards* **65**(3), 1723–1743 (2013).

47. Pourghasemi, H. R. GIS-based forest fire susceptibility mapping in Iran: a comparison between evidential belief function and binary logistic regression models. *Scandinavian J. Forest Res.* **31**(1), 80–98 (2016).
48. Tien Bui, D. T. *et al.* A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agricultural and Forest Meteorology* **233**, 32–44 (2017).
49. Hong, H., Pourghasemi, H. R. & Pourtaghi, Z. S. Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* **259**, 105–118 (2016).
50. Gigovic, L., Pourghasemi, H. R., Drobnjak, S. & Bai, S. Testing a new ensemble model based on SVM and random forest in forest fire susceptibility assessment and its mapping in Serbia's Tara National Park. *Forests* **10**(408), 1–20 (2019).
51. Liu, Y. B., De Smedt, F., Hoffmann, L. & Pfister, L. Assessing land use impacts on flood processes in complex terrain by using GIS and modeling approach. *Environ. Model. & Assess.* **9**(4), 227–235 (2005).
52. Pourghasemi, H. R., Gayen, A., Panahi, M., Rezaie, F. & Blaschke, T. Multi-hazard probability assessment and mapping in Iran. *Sci. Total Environ.* **692**, 556–571 (2019).
53. Skilodimou, H. D., Bathrellos, G. D., Chousianitis, K., Youssef, A. M. & Pradhan, B. Multi-hazard assessment modeling via multi-criteria analysis and GIS: a case study. *Environ Earth Sci.* **78**(2), 47 (2019).
54. Kariminejad, N., Hosseinalizadeh, M., Pourghasemi, H. R., Bernatek-Jakiel, A. & Alinejad, M. GIS-based susceptibility assessment of the occurrence of gully headcuts and pipe collapses in a semi-arid environment: Golestan Province, NE Iran. *Land Degradation & Development*. <https://doi.org/10.1002/ldr.3397> (2019).
55. Golkarian, A., Naghibi, S. A., Kalantar, B. & Pradhan, B. Groundwater potential mapping using C5. 0, random forest, and multivariate adaptive regression spline models in GIS. *Environ. Monit. Assess.* **190**(3), 149 (2018).
56. Wang, B. *et al.* Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. *Ecol. Indicators* **88**, 425–438 (2018).
57. Chan, J. C. W. & Paelinckx, D. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ectope mapping using airborne hyperspectral imagery. *Remote Sen. Environ.* **112**(6), 2999–3011 (2008).
58. Gounaris, D., Chorianopoulos, I., Symeonakis, E. & Koukoulas, S. A random forest-cellular automata modelling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales. *Sci. Total Environ.* **646**, 320–335 (2019).
59. Tiepolo, M., Bacci, M. & Bracco, S. Multihazard risk assessment for planning with climate in the Dosso Region, Niger. *Climate* **6**(3), 67 (2018).

Acknowledgements

The study was supported by the College of Agriculture, Shiraz University (Grant No. 96GRD1M271143) and by the Austrian Science Fund FWF through the GIScience Doctoral College (DK W 1237-N23) at the University of Salzburg.

Author contributions

H.R.P., N.K. and M.A., M.E., M.Z., A.C. and T.B. designed experiments, run models, analyzed results, wrote and reviewed manuscript. T.B. critically discussed the results and helped with the writing. All authors reviewed the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-60191-3>.

Correspondence and requests for materials should be addressed to H.R.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020