scientific reports

Check for updates

OPEN

Assessing the susceptibility of schools to flood events in Iran

Saleh Yousefi[1], Hamid Reza Pourghasemi[2]✉, Sayed Naeim Emami[1], Omid Rahmati[3], Shahla Tavangar[4], Soheila Pouyan[2], John P. Tiefenbacher[5], Shahbaz Shamsoddini[1] & Mohammad Nekoeimehr[1]

Catastrophic floods cause deaths, injuries, and property damages in communities around the world. The losses can be worse among those who are more vulnerable to exposure and this can be enhanced by communities' vulnerabilities. People in undeveloped and developing countries, like Iran, are more vulnerable and may be more exposed to flood hazards. In this study we investigate the vulnerabilities of 1622 schools to flood hazard in Chaharmahal and Bakhtiari Province, Iran. We used four machine learning models to produce flood susceptibility maps. The analytic hierarchy process method was enhanced with distance from schools to create a school-focused flood-risk map. The results indicate that 492 rural schools and 147 urban schools are in very high-risk locations. Furthermore, 54% of rural students and 8% of urban students study schools in locations of very high flood risk. The situation should be examined very closely and mitigating actions are urgently needed.

Floods are among the destructive natural hazards. These extreme events can be generated by a number of natural processes or from human activities and catastrophes, including heavy precipitation events, melting snowpack, modified drainage networks, failures of dams, and manipulation of drainage features. Based on recorded data, floods have caused US $700 billion globally and about 7 million deaths since 1900[1]. Floods are about 30% of hazardous events[2,3]. During last few decades, urbanization and increasing in populations have greatly increased exposure of people and properties to floods[4–6]. Some studies indicate that flood frequency and severity may increase as a consequence of global warming and changing climates[7,8].

Floods could be managed and mitigated by soft (nature-based and/or non-structural) and hard (engineered and structural) actions and decisions. The hard actions include dams, diversions, and check dams. Soft actions include land use planning, river restoration, selective siting of buildings, flood prediction modeling, alarm systems, improving public awareness of flood hazards, and education[9,10]. Floods influence soil erosion, enhance natural habitats, support ecological processes, and are important to many aspects of human life.

In communities, children are the most vulnerable to the consequences of flood exposure[11,12]. Schools are settings that concentrate children and need special attention with regard to extreme natural events. Over the last few decades, the frequency of floods has been increasing and loss of life and property has accordingly increased[13–15]. So, it is important to assess susceptibility of schools to flood events to reduce damages and prevent loss of lives. For this purpose, a flood susceptibility and hazard map can be prepared using various techniques or algorithms including statistical and machine learning.

Machine learning (ML) algorithms like logistic regressions[16–18], random forests[19,20], support vector machines[21–24], decision trees[25,26], artificial neural networks[27,28], boosted regression trees[29–31], multivariate adaptive regression splines[29,32], and model-driven architectures[16,33] have been tested for hazard analysis and mapping in literature. The ML approach has been used to evaluate the risk and susceptibility of communities exposed to a number of extreme and hazardous conditions: landslides[34–36], wildfires[37,38], gully erosion processes[39–41], land subsidence[42,43], earthquakes[4,13,44], dust storms[45], and floods[6,7,46]. Flood-hazard vulnerability has been examined by a number of scholars. Ochola et al. [47] studied the susceptibility of schools to floods in the Nyando River basin in Kenya. They analyzed the conditions of 130 schools in the western part of that country and found that 40% were vulnerable to floods. Karmakar et al.[48] conducted a risk-susceptibility analysis of floods in southwestern Ontario, Canada. They evaluated four types of vulnerability—physical, economic, infrastructural, and social—using a

[1]Soil Conservation and Watershed Management Research Department, Chaharmahal and Bakhtiari Agricultural and Natural Resources Research and Education Center, AREEO, Shahrekord, Iran. [2]Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran. [3]Soil Conservation and Watershed Management Research Department, Kurdistan Agricultural and Natural Resources Research and Education Center, AREEO, Sanandaj, Iran. [4]Department of Watershed Management Engineering, College of Natural Resources, Tarbiat Modares University, Tehran, Iran. [5]Department of Geography, Texas State University, San Marcos, TX, USA. ✉email: hr.pourghasemi@shirazu.ac.ir

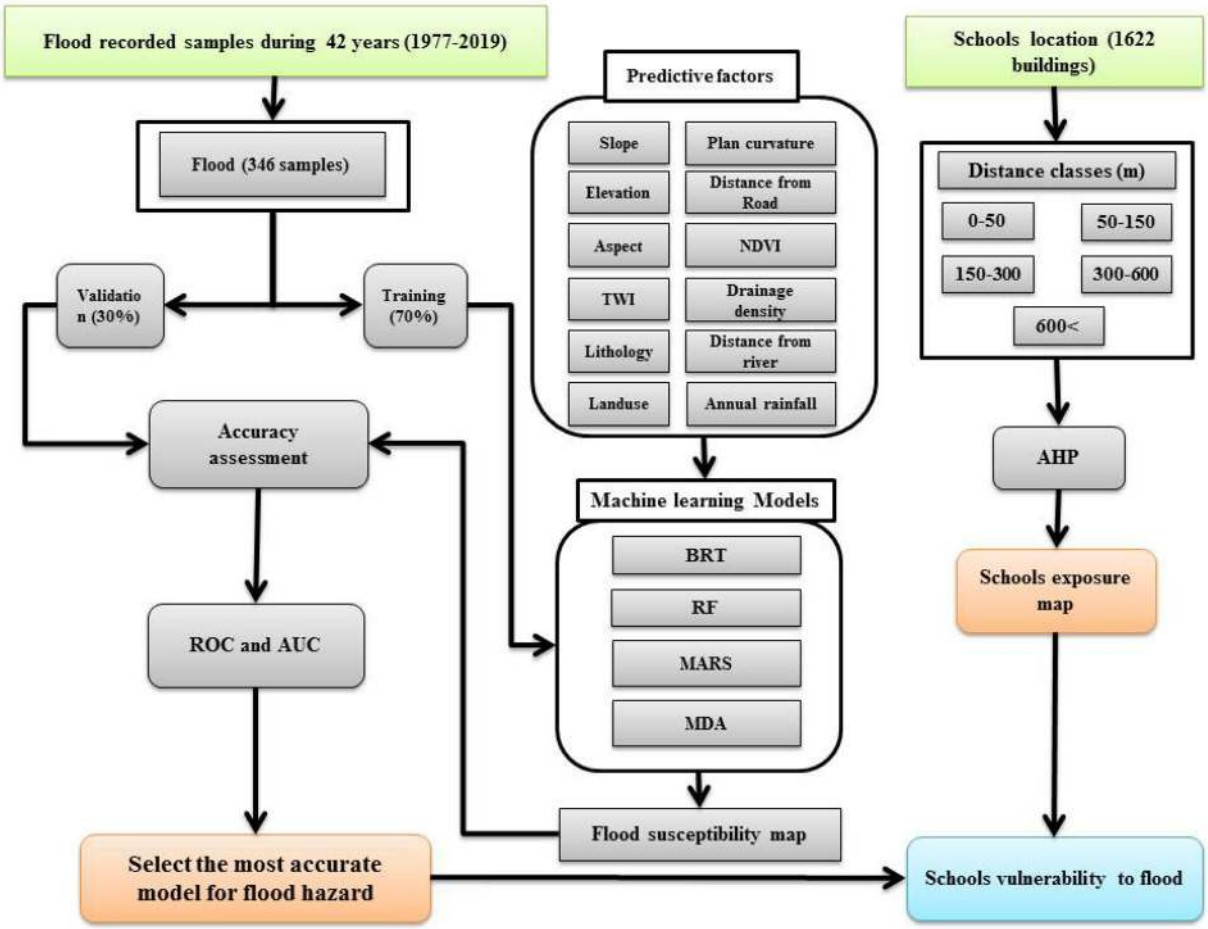

Figure 1. Flowchart of the methodology in present study.

geographic information system (GIS). Balica et al.[49] examined flood susceptibility using parametric and physical models and concluded that parametric modeling has limited accuracy, but provides a simplified view of social indicators of vulnerability. Nabegu[50] studied the vulnerabilities of households to flooding in Kano, Nigeria. They found that houses in the most vulnerable zone were destroyed and 17 people lost their lives during flood events. Eini et al.[51] investigated urban flood susceptibility using ML techniques in Kermanshah, Iran. They prepared flood maps using two ML models—maximum entropy and genetic algorithm—and found that maximum entropy yielded a more accurate flood-susceptibility model. They also determined that infrastructural characteristics had the greatest influence on flood susceptibility. Tascón-González et al.[52] studied social flood-vulnerability in Ponferrada, Spain using analytic hierarchy process (AHP) and found that 34,941 residents were impacted by floods from a dam break, and that 77% of them suffered heavy damages.

Few have attempted to examine the susceptibility of school locations to floods. A risk assessment of schools in developing countries is very important but has not yet been conducted. This study is the first to investigate the exposure of both urban and rural schools to flood hazards. It has been conducted for the mountainous province of Chaharmahal and Bakhtiari, Iran. The goal is to identify the locations most in need of mitigation to reduce damages and prevent loss of lives. Four ML models were tested and compared for the tasks of mapping flood hazard and assessing schools' exposures.

Materials and methods

Study area. Chaharmahal and Bakhtiari Province is in southwestern Iran in a region dominated by the Zagros Mountains. Having an average elevation of 2153 m above sea level and a range of elevations from 778 to 4203 m, the province is the highest in Iran. The province covers 16,421 km^2 and its population is approximately 947,000. Due to the topographical and climatic conditions of the region, floods occur annually throughout the province.

Methodology. There are five steps to this research: (1) collection and compilation of spatial data; (2) determination of the influence of the independent effective factors on flood probability; (3) production of flood risk maps using four ML algorithms; (4) validation and evaluation of the flood risk maps, and (5) determination of the susceptibility of schools to floods in Chaharmahal and Bakhtiari Province (Fig. 1).

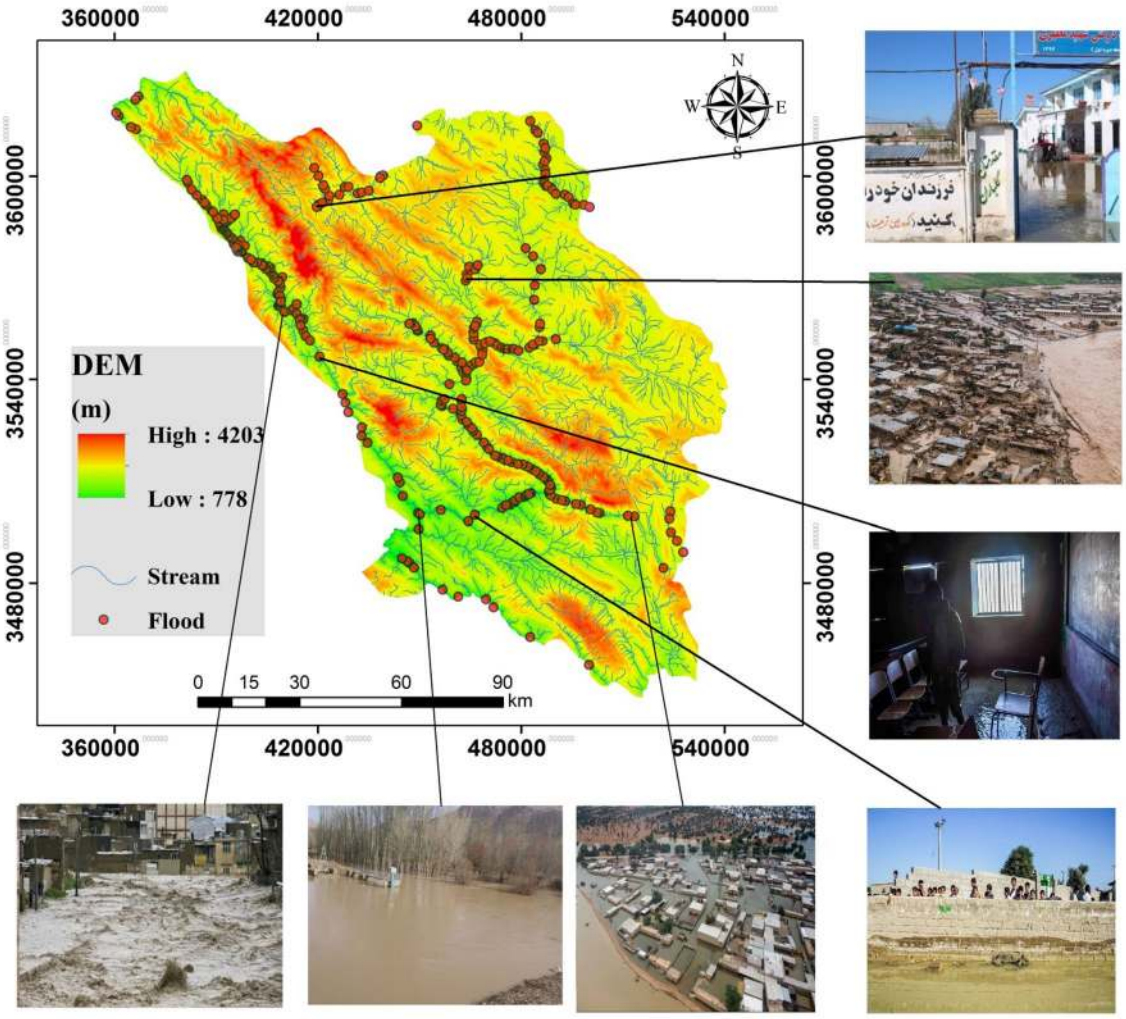

Figure 2. Locations of the floods that occurred between 1977 and 2019 in Chaharmahal and Bakhtiari Province and visual examples of several events.

Collection and compilation of spatial data. To accurately determine flood patterns and frequencies in a region, an accurate and well-distributed sample of flood occurrence must be compiled. Three hundred and forty-six floods that occurred in the province were recorded over a 42-year period (1977–2019) by Iran's Ministry of Energy. The locations of the floods were identified and geo-located using a global position system (GPS) device during extensive field surveys. These points were mapped (Fig. 2). The sample was randomly divided into a modeling set containing 70 percent of the locations and a validation set containing 30% of the sample. As flood occurrence is determined by an interaction of natural and human processes, based on previous studies[15,53–55] 12 of the most important effective factors were identified for use in modeling as input variables. They included elevation, slope, aspect, plan curvature, lithology, drainage density, annual rainfall, topographic wetness index (TWI), normalized difference vegetation index (NDVI), land use type, distance from nearest river, and distance from nearest road. The data were derived from 1:25,000 topographic maps, 1:100,000 geological maps, and OLI Landsat images (from 2018). The 12 data layers were created in ArcGIS 10.4.2 and ENVI 5.3 software. To ensure that the 12 input factors were truly independent of each other (not highly correlated with each other), a multicollinearity test was applied. The Pearson correlation tests showed no significant correlation between the factors, ensuring a more accurate flood risk map (Fig. 3).

Determination of the influence of input factors on flood probability. Some topographic factors can interact to increase the likelihood of flooding. Elevation, aspect, TWI, slope, and plan curvature layers were constructed from 1:25,000 topographic maps (Fig. 4A–E). Vegetation is also integral to hydrological processes. An NDVI layer was extracted from OLI Landsat images from 15 Jun 2017 to indicate vegetation patterns (Fig. 4F). The 1:25,000 topographic maps provided streams and road-network information. These were extracted and used to create raster layers of drainage density, distances from rivers, and distances from roads (Fig. 4G–I). The OLI Landsat images were also used to map land uses (Fig. 4J). Lithological units were extracted from a 1:100,000 geological maps acquired from the Iranian Geology Organization (Fig. 4K). Precipitation is a key factor influencing

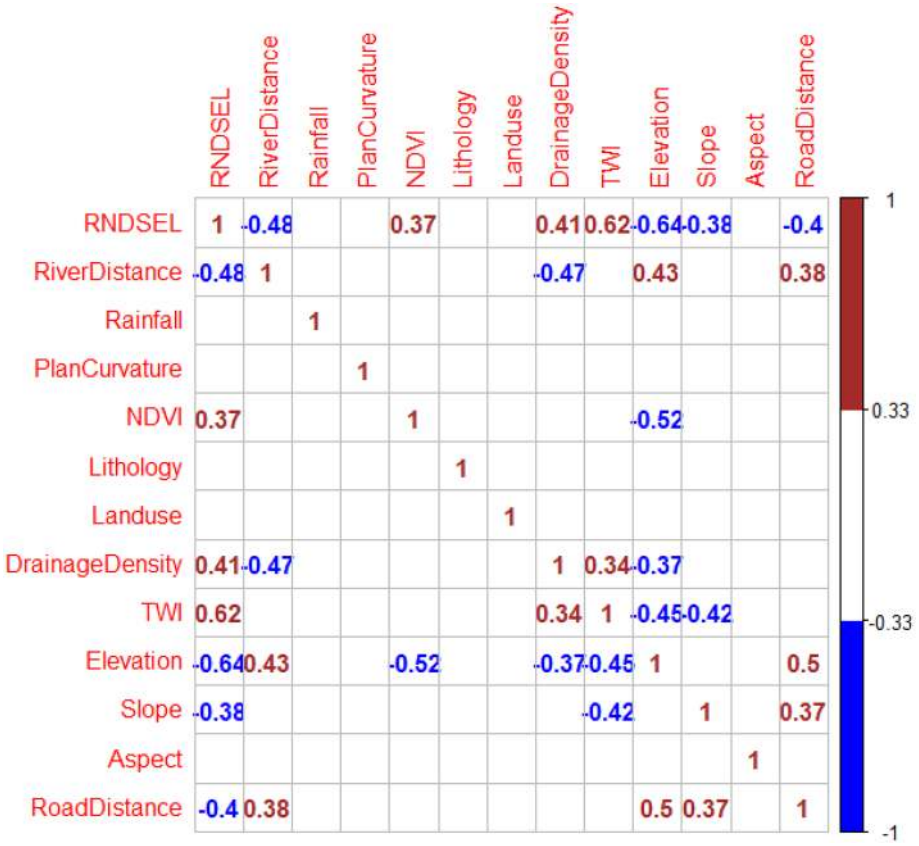

Figure 3. Pearson correlation test among various independent variables (RNDESEL = flood locations).

flood occurrence. Data were gathered from 18 weather stations to determine average annual rainfall from 1982 to 2019 and these data were used to reflect the rainfall factor in flood-risk mapping (Fig. 4L).

Modeling flood risk using four ML algorithms. Boosted regression trees (BRT). The BRT model is a model that combines methods to improve analysis[56]. Since BRT is usually associated with tree-based methods, it is useful for identification of the factors that most impact predictions of an outcome. A benefit of BRT is that it can work even when some data are absent[36]. BRT balances models' performances[37,57] and balances between models' performances[58]. BRT results are conditioned by the number of trees used in the model and the combinations of the trees used. Performance is improved as the number of trees increases[59]. The following features were set for running the BRT model: gbm.x = 2:13, gbm.y = 1, family = "bernoulli", tree.complexity = 5, learning.rate = 0.005, bag.fraction = 0.5. Here, gbm.x = the 12 independent variables and gbm.y = dependent the variable (flood location). The final BRT model had 1850 trees to predict flood locations. Mean total deviance = 1.386, mean residual deviance = 0.059, estimated cross validation deviance = 0.333, and standard error = 0.055.

Mixture discriminant analysis (MDA). The MDA is a supervised classification algorithm based on mixture models. This model is an extension of linear discriminant analysis and is used to estimate density for each class[60]. In general, the MDA model is suitable for modeling multivariate nonlinearity relationships among various parameters within each group. It is also important to determine whether there are underlying sub-classes in each group which can have a positive effect on the factors of the environment or the independent factors[61–64]. The "mda" package[65] was used to run the MDA model.

Random forest (RF). RF is a nonparametric technique based on regression trees[20,40,66]. It is one of the strongest ML models due to the large number of trees that it incorporates[67,68]. RF has several advantages: it is insensitive to noise, it can incorporate most types of data, and it is helps to determine the variables that are most important[14,37,69]. Shahabi et al.[70] indicates that RF is very effective at estimating the relative importance of factors, which aids with decision-making for environmental management. The settings of the RF model were mtry = 4, ntree = 1000, and the estimated out-of-bag (OOB) error rate was 5.27%.

Multivariate adaptive regression splines (MARS). MARS is one of the best regression-based algorithms[13,71]. Its predictions can be made based upon both linear and non-linear relationships between independent factors[72]. This model is very flexible for predicting events based on a set of independent factors. Furthermore, it allows

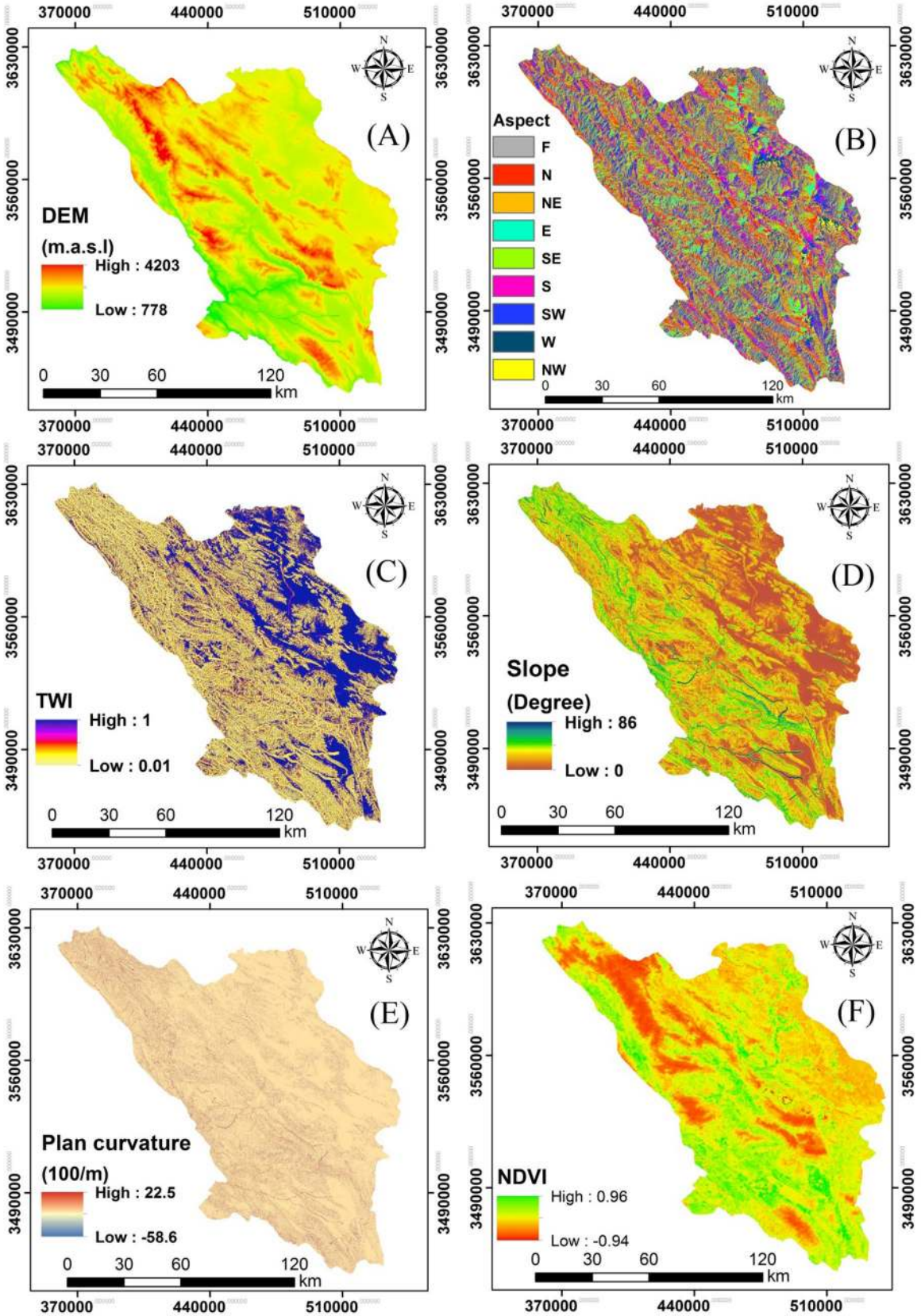

Figure 4. Maps of effective factors in Chaharmahal and Bakhtiari Province: **(A)** elevation, **(B)** aspect, **(C)** TWI, **(D)** slope, **(E)** plan curvature, **(F)** NDVI, **(G)** distance from river, **(H)** drainage density, **(I)** distance from road, **(J)** land use, **(K)** lithology, and **(L)** annual rainfall.

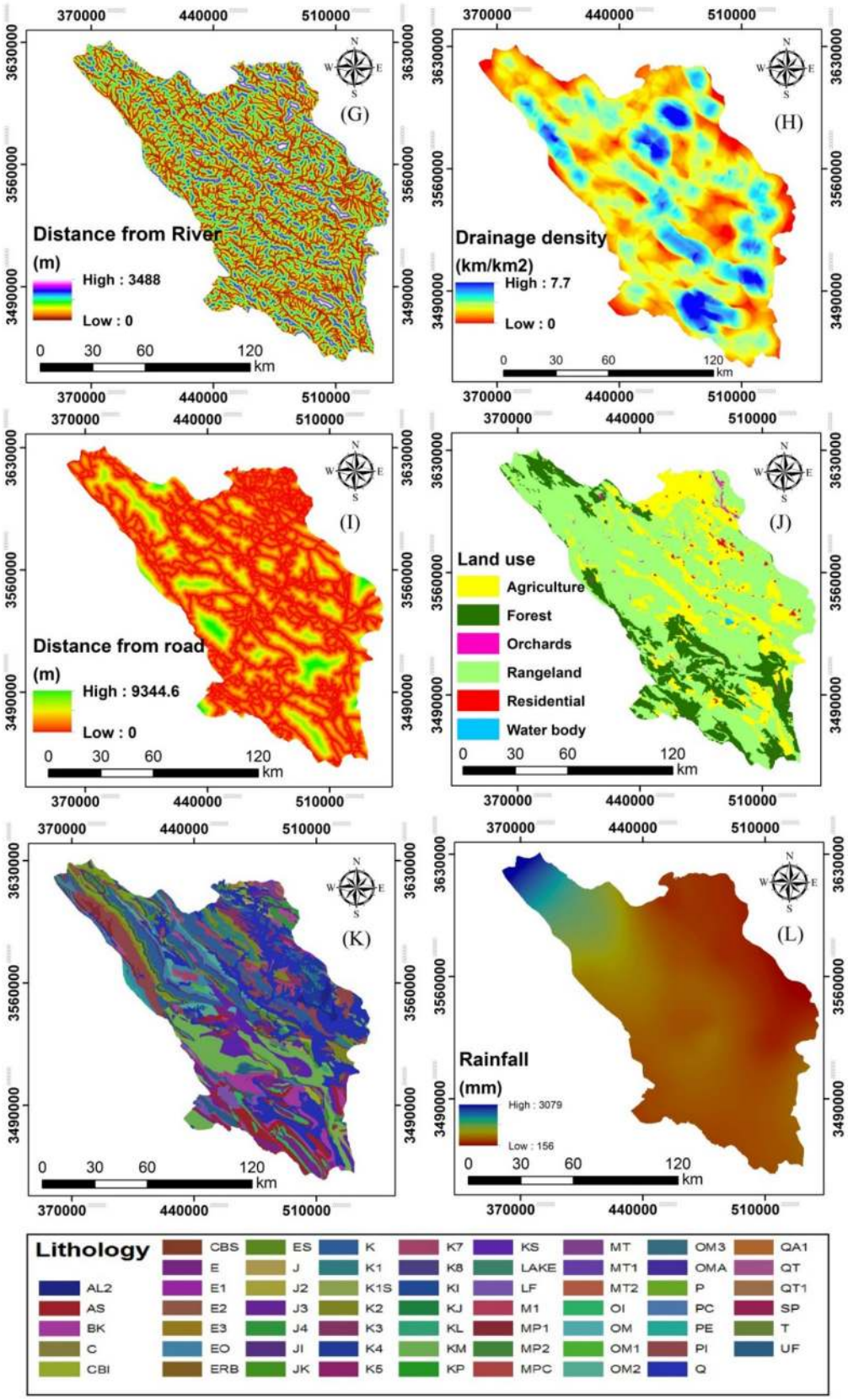

Figure 4. (continued)

County	No. of Schools	No. of Students
Ardal	157	12,195
Brojen	214	22,985
Ben	52	5072
Saman	72	5640
Shahrekord	304	48,688
Farsan	135	16,622
Kohrang	164	10,472
Kiar	151	8610
Lordegan	373	70,990
Total	1622	201,274

Table 1. Distribution of schools and students in 9 counties.

for the determination of the relative importance of the independent variables in the predictions[30,67,71]. MARS determines the relationships between dependent and independent variables and reflects these functions as coefficients so that the impacts of the factors are calculated separately[73]. It defines basic functions by the intervals of the factors[74,75]. MARS has a sensitivity to variable correlations[74] and has been used in many applications to assess geophysical, climatological, environmental, and geomorphological relationships[76–79]. In this method, pruning was "backward" with three penalties. After pruning, generalized R^2 was 0.774, whereas R^2 was 0.824.

R statistical packages used for modelling process. The BRT, MARS, MDA, and RF models were run in R software version R 3.5.3. Each required use of specific packages: "brt"[58], "mda"[61], "MARS"[74], and "randomForest"[80]. R software was used to perform the modeling, analysis, and graphical depictions of the analyses[81,82].

Evaluation of the modeled flood-risk maps. The results of the four ML models were evaluated to identify the most accurate model. The receiver operating characteristic (ROC) curve is a cutoff-independent evaluation approach for determining the goodness-of-fit and predictive performance of models. The area under the ROC curve (AUC) was the analysis of accuracy used[83–85]. The validation data set contained 30% of the flood location sample that was not used for training[38,58,86]. The relative importance of each of the independent factors on the modeled flood predictions were analyzed with least absolute shrinkage and selection operator (LASSO). LASSO is a regression-based method that analyses variable selection and regularization in ML models.

Determination of the proximity of schools to flood zones. The geolocations of 1,622 school buildings that are attended by 201,274 in Chaharmahal and Bakhtiari Province (Table 1) were identified and mapped. Sixty-three percent of schools were in rural areas and 37% in urbanized areas. Thirty-two experts (hydrologist, educational teachers, fluvial geomorphologists, etc.) completed questionnaires about schools' distances to flood zones to reflect the exposure of each school to flood hazard. Consistency ratios (CRs) were calculated to evaluate the consistency of the experts' opinions about school exposures. Arc GIS 10.4.2's Euclidean-distance tool was used to evaluate the proximity of each school to the modeled flood patterns. Using AHP, distances were classified by concentric rings around school buildings (0–50 m, 50–150 m, 150–300 m, 300–600 m and > 600 m) (Fig. 5). Finally, the normalized rates (NR) of the five distance classes were calculated to determine the weight of exposure for each school.

The susceptibility of schools to floods. Based on the natural break algorithm, the flood exposure map was classified into five classes (very low, low, moderate, high and very high) in ArcGIS 10.4.2[15,87]. To generate the final school flood-exposure map, the most accurate flood risk map and school exposure map were fed into the susceptibility equation: Flood susceptibility = Flood risk × School exposure. The susceptibilities of schools in five classes (very low, low, moderate, high and very high) were determined.

Results

Flood risk map. Flood risk maps for Chaharmahal and Bakhtiari Province were produced with BRT, RF, MARS, and MDA algorithms (Fig. 6). The four models generated similar patterns, but they differed in the details of the predictions. The western and southwestern parts of the province are most prone to flood events.

The RF model produced the best flood-risk map (Table 2) by predicting locations that are likely to flood better than the other models. The others, in order of accuracy, were MARS, MDA, and BRT models. But based on AUC analysis, MARS, MDA, and BRT also produced acceptable flood-risk maps (Table 2). The RF model indicates that flood risk in the eastern portion of the study area is much lower than in the central and southern parts of the province.

School-exposure map. Using AHP, the normalized rate (NR) of the five distance classes were determined (Table 3). The exposure map was prepared according to experts' ratings for different school-vulnerability classes

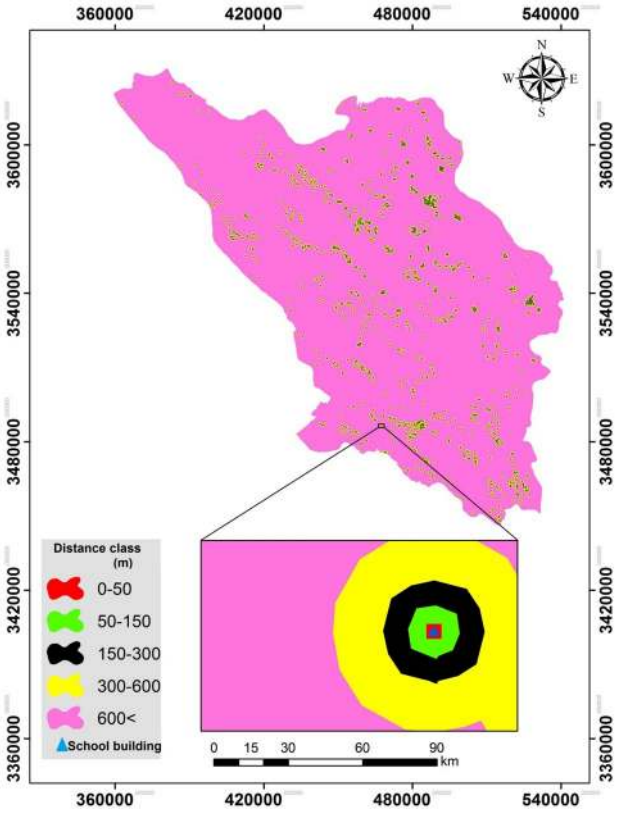

Figure 5. Proximities to schools in Chaharmahal and Bakhtiari Province.

(based on distance from flood hazard zone)[15,88–90] and the AHP results (Fig. 7). A consistency ratio (CR) of 0.08 is an acceptable value.

Susceptibility map. A map of the flood susceptibility of schools (flood risk x exposure of schools) was produced (Fig. 8) using the flood risk modeled using RF. Susceptibility was categorized into five classes based on natural breaks in ArcGIS 10.4.2 (Fig. 9). The results indicate that 69.85% falls into the lowest class of school flood-susceptibility. Only 1.42% of the province has schools that are highly susceptible to flooding and 0.43% has schools in very highly susceptibility circumstances.

Susceptibility of schools to flood. In all, 979 schools serving 123,324 pupils are in conditions of high and very high flood-susceptibility (Table 4). Of these, 492 are rural schools serving 55,395 pupils and 147 urban schools serving 31,245 students in conditions of very high susceptibility (Table 5). Schools of Lordegan County are the most susceptible to floods: 14,299 pupils in 42 urban schools and 36,312 pupils in 196 rural schools are educated in very high flood-susceptible zones.

Discussion

Experts believe that decision makers can reduce losses caused by flood events by implementing mitigation and management actions in watersheds[91–93]. The most important effects of floods are losses of lives, losses of shelter and property, out-migration, disease outbreaks, despair and hopelessness, loss of social capital, and loss of employment. Flood modelling and mapping alone will not reduce hazard and vulnerabilities, but it provides a perspective for mitigation of risk and management of flood hazard in watersheds and in communities. Children are among the most vulnerable in society to hazards and their consequences. As they spend much of their lives in schools, these structures need to be located in places less likely to flood.

This study assessed the susceptibility of schools to floods in Chaharmahal and Bakhtiari Province. Four ML algorithms (MARS, MDA, BRT and RF) were used to predict the spatial patterns of floods to determine flood risk. The results of validation of the models' results indicated that RF was the most accurate (AUC = 0.989) of the models. RF uses the most important variables or dividing points within variable subgroups to create a growth tree randomly selected from a set of factors, and thus reduces the importance of each individual regression tree. This shrinks the matching rate, reducing the model error[69]. This method improves the stability and accuracy of the classification, reduces variance, and avoids excessive fitting[67,70]. Finding that the RF model generates an accurate model for prediction and determination of different phenomena is consistent with Taalab et al. (2018), Avand et al., (2019), Hosseinalizadeh et al. (2019), and Gayen et al. (2019). It has been argued that "risk" should be used for flood management as it considers both vulnerability and flood probability simultaneously[25].

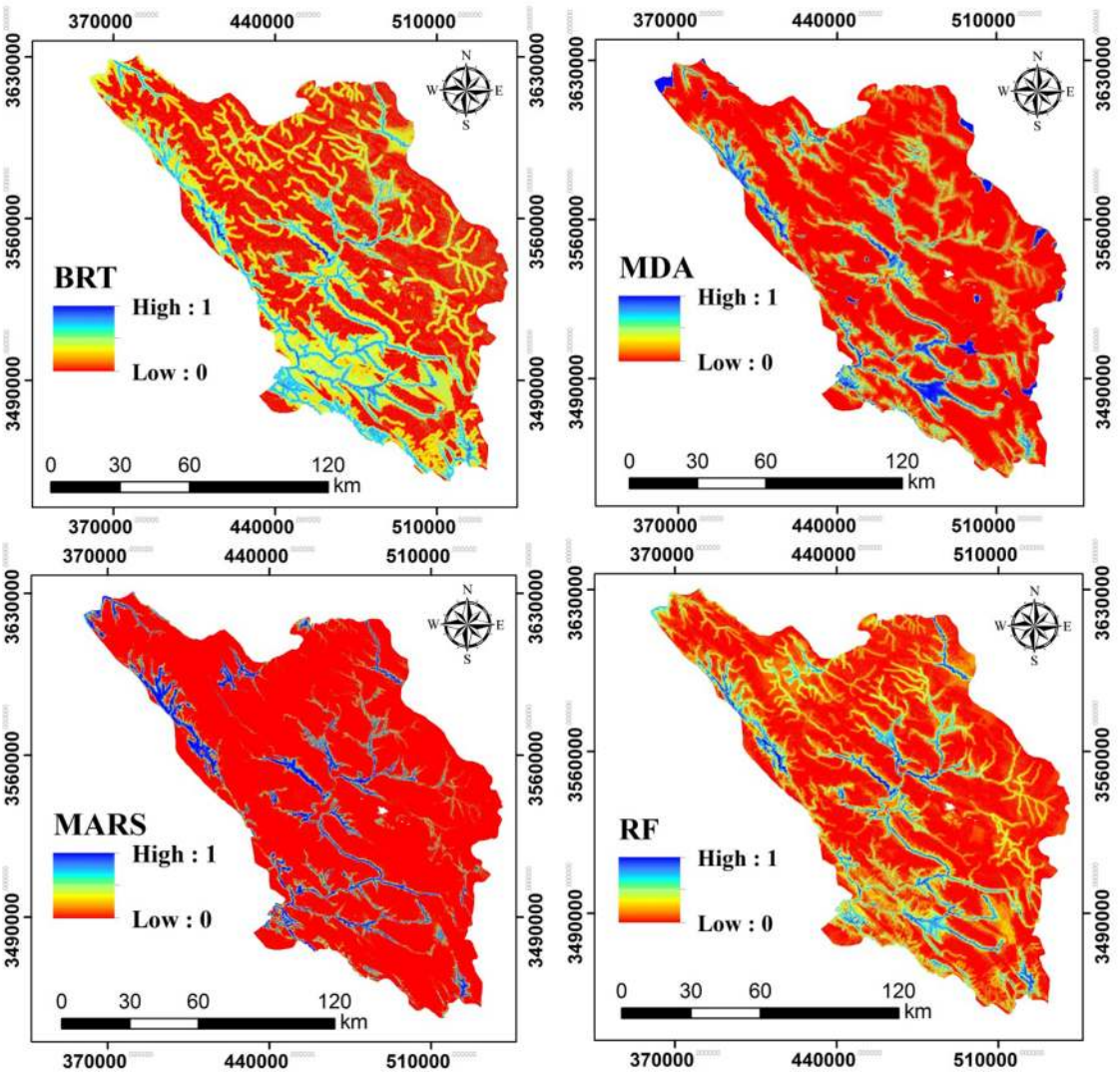

Figure 6. Flood risk maps generated by the four BRT, RF, MARS, and MDA models.

Test result variable(s)	AUC	Standard error	Asymptotic significant	Asymptotic 95% confidence interval	
				Lower bound	Upper bound
RF	0.989	0.006	0.000	0.978	1.000
MDA	0.970	0.010	0.000	0.950	0.990
MARS	0.978	0.010	0.000	0.959	0.997
BRT	0.957	0.013	0.000	0.931	0.983

Table 2. Results of model evaluation using the AUC metric.

The flood risk map allows decision makers to allocate and prioritize places in need of urgent flood mitigation. Based on the flood-risk map produced by the RF model and the school exposure map generated by AHP, the school-susceptibility map was produced (Fig. 10). For vulnerability issues, both quantitative and qualitative datasets were gathered from available reports and through questionnaires and interviews for investigating the different vulnerability dimensions. Social experiences and awareness provide valid information about flood vulnerability[47,96]. There are 1023 and 599 school buildings in the rural and urban parts of the province. Based on the results, 48% (492) of rural schools and 24.5% (147) of urban schools are in conditions of very high flood susceptibility. And 54% (55,395) of students in rural schools and 8% (31,245) in urban schools are in zones of very high school-susceptibility. On the other hand, 76% (297,729) and 2.7% (2733) of children are in very low susceptibility conditions in urban and rural schools, respectively. These results indicate that the rural schools are in more flood-susceptible areas and mitigation of these conditions is urgently needed. This would, most likely,

Distance from school (m)	Exposure class	Rate	NR
0–50	Very high	9	0.345 (10/29)
50–150	High	8	0.276 (8/29)
150–300	Moderate	6	0.207 (6/29)
300–600	Low	4	0.138 (4/29)
600 <	Very low	2	0.069 (2/29)
Total	–	29	1 (29/29)
Consistency ratio = 0.08			

Table 3. Normalized rates of school-exposure classes.

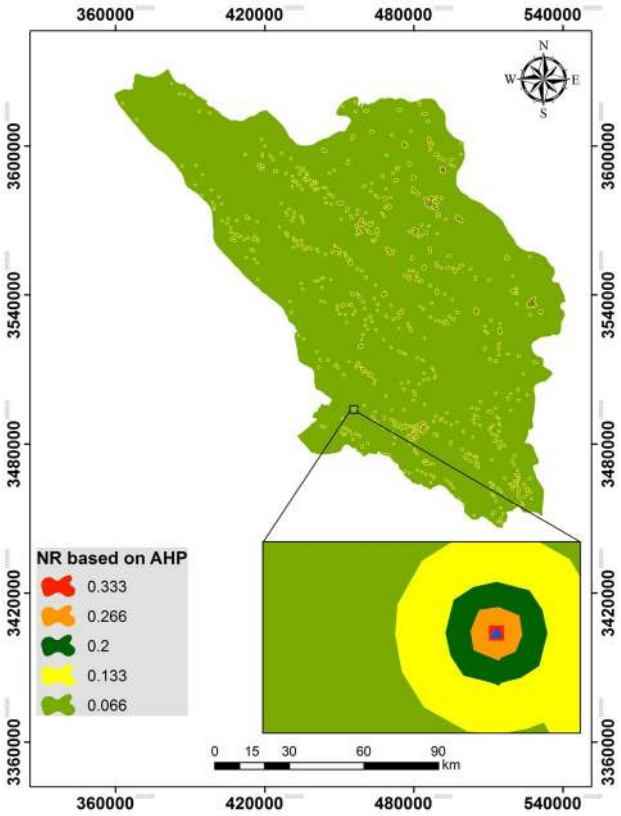

Figure 7. Normalized school-exposure rates based on AHP.

be accomplished by moving schools to less flood-prone locations. Supervision and oversight of the locating, constructing, or reconstructing of schools is usually greater in urban areas than in rural. For this reason, schools and students in urban areas are less susceptible to floods than those in rural areas. The magnitude of challenges for schools and students is expected to grow even further with a population growth, urbanization, and changing climates. Considering the status of schools in terms of flood risk, the current flood defense measures in this province are often unable to cope with additional pressure. As a solution, updated flood risk maps can enhance flood policy and management and can be a rational basis for decision-making.

Susceptibility to floods in the study area should dictate that the 43% of pupils and 39% of schools in very high zones be relocated to safer places. Students are not spared from floods; they suffer losses, too. Damage to school buildings may make them unsafe to the point where they may need to be demolished and rebuilt. The traumas of disasters have been substantiated to impact students' psyches. Shahrekord (home of the province capital) and Brojen counties have 29 and 28 schools, respectively, in zones of very low flood-susceptibility, making them the sub-regions with the safest schools in Chaharmahal and Bakhtiari Province. Historically, villages and cities have been built in flood zones. Moving the buildings and properties to safer locations requires too much money, and is usually socially unacceptable. An alternative is to rebuild schools and public places with stronger, more flood-resistant materials. The cities of Shahrekord and Brojen are the largest and most important in the province, and it is known that governors focus their efforts in these communities to achieve more satisfaction among the

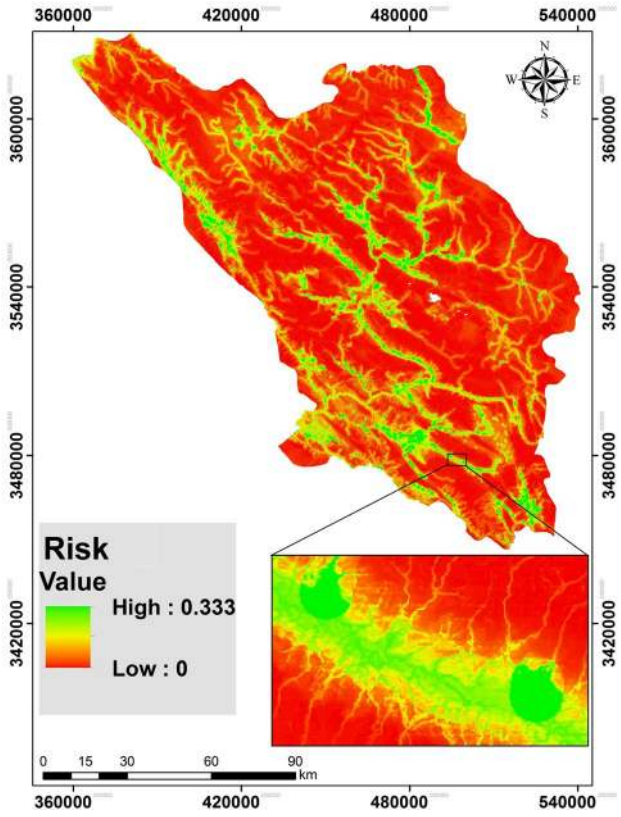

Figure 8. The susceptibility map of schools to flood hazard in study area.

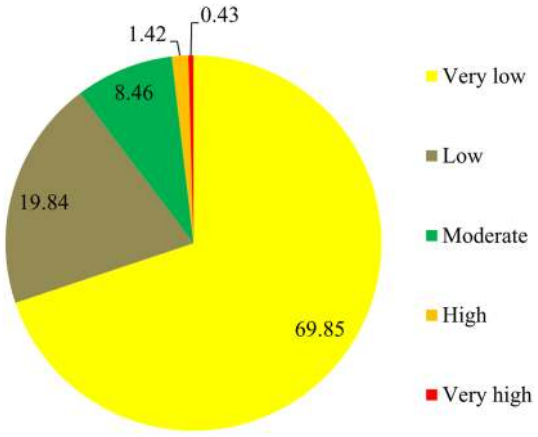

Figure 9. Percentages of the province covered in the five flood-susceptibility classes for schools.

Risk class	No. of schools	No. of school class room	No. of students	Average of students per class
Very low	83	617	10,462	17
Low	186	1188	22,684	19
Moderate	374	2341	44,804	19
High	340	1990	36,684	18
Very high	639	3486	86,640	25
Total	1622	9622	201,274	21

Table 4. Flood susceptibilities of schools by the accurate model.

County	School type	Very low		Low		Moderate		High		Very high	
		No. of schools	No. of students	No. of schools	No. of students	No. of schools	No. of students	No. of schools	No. of students	No. of schools	No. of students
Ardal	Urban	–	–	–	–	–	–	2	80	18	2239
	Rural	1	4	10	274	37	3646	38	2734	51	3218
Brojen	Urban	18	2513	43	7001	36	4156	36	4556	3	202
	Rural	10	696	22	1483	20	1314	19	969	5	94
Ben	Urban	–	–	4	439	8	915	4	354	4	499
	Rural	7	1097	8	781	11	572	5	394	1	40
Saman	Urban	–	–	1	165	15	2217	1	166	–	–
	Rural	–	–	–	–	5	452	2	118	47	2522
Shahrekord	Urban	25	4511	33	7561	59	10,721	42	8711	52	9704
	Rural	4	634	10	878	21	1596	28	2144	27	2101
Farsan	Urban	2	128	6	778	31	4104	23	3265	19	3481
	Rural	1	21	6	32	5	89	12	1658	31	3191
Kohrang	Urban	3	418	–	–	1	178	7	610	1	4
	Rural	7	185	22	1054	31	2264	26	1070	66	4688
Kiar	Urban	1	159	2	185	22	1996	14	1562	8	818
	Rural	2	11	9	199	7	63	20	409	66	3207
Lordegan	Urban	–	–	–	–	5	1222	5	396	42	14,299
	Rural	2	118	10	1854	60	9299	53	7490	196	36,312

Table 5. Distribution of schools and students by flood-susceptibility class and county.

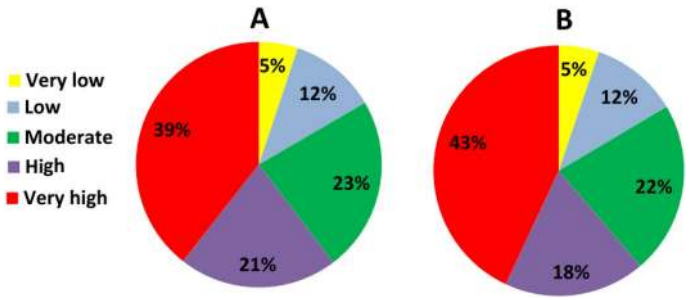

Figure 10. Distribution of schools **(A)** and pupils **(B)** of the Chaharmal and Bakhtiari Province in the five flood-susceptibility classes.

residents. Still, there are 9704 pupils in 52 schools in Shahrekord City located in very highly susceptible zones and they remain in serious danger from flooding.

This study described the root causes of flood risk related to schools and provided insight into flood-risk management. Students and children are the future of any country and growing them in a safe environment is essential for any government. This study showed that many of the schools of Chaharmahal and Bakhtiari Province are in worrisome locations, and this concern is even more acute in rural areas. For this reason, it is recommended that safety managers examine the locations of school buildings in their jurisdiction to identify those that are in the most precarious locations. The susceptibility of all future school sites should be carefully considered before they are constructed. All schools should be located in places that are as risk-free as possible. Schools that are at high or very high levels of flood-susceptibility should be relocated to safer places at the earliest possible opportunity, before the next flood disaster occurs. In addition, flood control measures can help to reduce flood risk when building new schools is impractical.

Conclusion

The susceptibility of schools to floods in the Chaharmahal and Bakhtiari Province, Iran was assessed. Thirty-nine percent of schools are in zones of very high flood susceptibility and urgent action is need by decision makers. Additionally, the susceptibility of rural schools to floods is greater than it is for the schools in urban areas. A total of 86,640 pupils attend schools in locations of very high flood susceptibility in the province. In addition to relocating schools in dangerous places, decision makers should enhance public knowledge and awareness of the threats faced by schools and by children. Results of studies like this one can help raise public awareness, which is an effective soft measure to reduce unavoidable negative impacts of floods. Reducing deaths, damages, and

disruptions caused by floods could be facilitated in by education at all levels of society, in schools and publicly. Drills and simulations should be held in schools and rural areas to build preparedness for flood events. Assessment of the susceptibility of schools to flood risks in a mountainous region of Iran is but one part of a management to reduce the likelihood that extreme flood events will turn into tragic disasters.

Received: 17 July 2020; Accepted: 14 October 2020
Published online: 22 October 2020

References

1. Costache, R. & Tien Bui, D. Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. *Sci. Total Environ.* **691**, 1098–1118 (2019).
2. Maantay, J. & Maroko, A. Mapping urban risk: Flood hazards, race, & environmental justice in New York. *Appl. Geogr.* **29**, 111–124 (2009).
3. Mirzaee, S. *et al.* Effects of hydrological events on morphological evolution of a fluvial system. *J. Hydrol.* **563**, 33–42 (2018).
4. 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).
5. Hajdukiewicz, H., Wyżga, B., Mikuś, P., Zawiejska, J. & Radecki-Pawlik, A. Impact of a large flood on mountain river habitats, channel morphology, and valley infrastructure. *Geomorphology* **272**, 55–67 (2016).
6. Rahmati, O. & Pourghasemi, H. R. Identification of critical flood prone areas in data-scarce and ungauged regions: A comparison of three data mining models. *Water Resour. Manag.* **31**, 1473–1487 (2017).
7. Hayri Kesikoglu, M., Haluk Atasever, U., Dadaser-Celik, F. & Ozkan, C. Performance of ANN, SVM and MLH techniques for land use/cover change detection at Sultan Marshes wetland, Turkey. *Water Sci. Technol.* **80**, 466–477 (2019).
8. Overeem, I., Kettner, a. J. & Syvitski, J. P. M. Impacts of humans on river fluxes and morphology. *Treatise Geomorphol.* **9**, 828–842 (2013).
9. Barredo, J. I. Major flood disasters in Europe: 1950–2005. *Nat. Haz.* **42**, 125–148 (2007).
10. Tien Bui, D. *et al.* A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Sci. Total Environ.* **701**, 134413 (2020).
11. Ruiz-Villanueva, V. *et al.* Impacts of a large flood along a mountain river basin: The importance of channel widening and estimating the large wood budget in the upper Emme River (Switzerland). *Earth Surf. Dyn.* **6**, 1115–1137 (2018).
12. Dong, Q., Wang, X., Ai, X. & Zhang, Y. Study on flood classification based on project pursuit and particle swarm optimization algorithm. *J. China Hydrol.* **4**, 2 (2007).
13. Pourghasemi, H. R., Gayen, A., Edalat, M., Zarafshar, M. & Tiefenbacher, J. P. Is multi-hazard mapping effective in assessing natural hazards and integrated watershed management?. *Geosci. Front.* https://doi.org/10.1016/j.gsf.2019.10.008 (2019).
14. Bui, D. T. *et al.* Novel hybrid evolutionary algorithms for spatial prediction of floods. *Sci. Rep.* **8**, 1–14 (2018).
15. Rahmati, O. *et al.* Multi-hazard exposure mapping using machine learning techniques: A case study from Iran. *Remote Sens.* **11**, 1943 (2019).
16. Marjanović, M., Kovačević, M., Bajat, B. & Voženílek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **123**, 225–234 (2011).
17. Jebur, M. N., Pradhan, B. & Tehrany, M. S. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. *Remote Sens. Environ.* **152**, 150–165 (2014).
18. Umar, Z., Pradhan, B., Ahmad, A., Jebur, M. N. & Tehrany, M. S. Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. *CATENA* **118**, 124–135 (2014).
19. Nabiollahi, K., Eskandari, S., Taghizadeh-Mehrjardi, R., Kerry, R. & Triantafilis, J. Assessing soil organic carbon stocks under land-use change scenarios using random forest models. *Carbon Manag.* **10**, 63–77 (2019).
20. Thanh Noi, P. & Kappas, M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors* **18**, 18 (2018).
21. Remesan, R., Bray, M., Shamim, M. A. & Han, D. Rainfall-runoff modelling using a wavelet-based hybrid SVM scheme. in *IAHS-AISH Publication* **331**, 41–50 (IAHS Press, 2009).
22. Rahimian Boogar, A., Salehi, H., Pourghasemi, H. R. & Blaschke, T. Predicting habitat suitability and conserving *Juniperus* spp. habitat using SVM and maximum entropy machine learning techniques. *Water* (*Switzerland*) **11**, 2049 (2019).
23. Pozdnoukhov, A., Purves, R. S. & Kanevski, M. Applying machine learning methods to avalanche forecasting. *Ann. Glaciol.* **49**, 107–113 (2008).
24. Pham, B. T. & Prakash, I. Evaluation and comparison of LogitBoost ensemble, Fisher's linear discriminant analysis, logistic regression and support vector machines methods for landslide susceptibility mapping. *Geocarto Int.* **34**, 316–333 (2019).
25. Mosavi, A., Ozturk, P. & Chau, K. W. Flood prediction using machine learning models: Literature review. *Water* (*Switzerland*) **10**, 1536 (2018).
26. Gokceoglu, C., Nefeslioglu, H. A., Sezer, E., Bozkir, A. S. & Duman, T. Y. Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. *Math. Probl. Eng.* **2010** (2010).
27. Sari, P. A. *et al.* Developing a hybrid adoptive neuro-fuzzy inference system in predicting safety of factors of slopes subjected to surface eco-protection techniques. *Eng. Comput.* 1–8, https://doi.org/10.1007/s00366-019-00768-3 (2019).
28. Tabari, H., Abghari, H. & Hosseinzadeh Talaee, P. Temporal trends and spatial characteristics of drought and rainfall in arid and semiarid regions of Iran. *Hydrol. Process.* **26**, 3351–3361 (2012).
29. Dodangeh, E. *et al.* Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Sci. Total Environ.* **705**, 135983 (2020).
30. Zabihi, M., Pourghasemi, H. R., Motevalli, A. & Zakeri, M. A. Gully erosion modeling using GIS-based data mining techniques in Northern Iran: A comparison between boosted regression tree and multivariate adaptive regression spline. in *Advances in Natural and Technological Hazards Research* Vol. 48 1–26 (Springer, 2019).
31. Pourtaghi, Z. S., Pourghasemi, H. R., Aretano, R. & Semeraro, T. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecol. Indic.* **64**, 72–84 (2016).
32. Vorpahl, P., Elsenbeer, H., Märker, M. & Schröder, B. How can statistical models help to determine driving factors of landslides?. *Ecol. Model.* **239**, 27–39 (2012).
33. Perry, G. L. W. & Dickson, M. E. Using machine learning to predict geomorphic disturbance: The effects of sample size, sample prevalence, and sampling strategy. *J. Geophys. Res. Earth Surf.* **123**, 2954–2970 (2018).
34. Pourghasemi, H. R., Pradhan, B., Gokceoglu, C., Mohammadi, M. & Moradi, H. R. Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. *Arab. J. Geosci.* **6**, 2351–2365 (2013).
35. Shi, Y. & Jin, F. Landslide stability analysis based on generalized information entropy. in *Proceedings—2009 International Conference on Environmental Science and Information Application Technology, ESIAT 2009*, Vol. 2 83–85 (IEEE, 2009).

36. Youssef, A. M., Pourghasemi, H. R., El-Haddad, B. A. & Dhahry, B. K. Landslide susceptibility maps using different probabilistic and bivariate statistical models and comparison of their performance at Wadi Itwad Basin, Asir Region, Saudi Arabia. *Bull. Eng. Geol. Environ.* **75**, 63–87 (2016).
37. Rodrigues, M. & De la Riva, J. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environ. Model. Softw.* **57**, 192–201 (2014).
38. Pourtaghi, Z. S., Pourghasemi, H. R. & Rossi, M. Forest fire susceptibility mapping in the Minudasht forests, Golestan province, Iran. *Environ. Earth Sci.* **73**, 1515–1533 (2015).
39. Nhu, V. H. *et al.* GIS-Based gully erosion susceptibility mapping: A comparison of computational ensemble data mining models. *Appl. Sci.* **10**, 2039 (2020).
40. Arabameri, A., Pradhan, B. & Rezaei, K. Spatial prediction of gully erosion using ALOS PALSAR data and ensemble bivariate and data mining models. *Geosci. J.* **23**, 669–686 (2019).
41. Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H. R. & Feizizadeh, B. Evaluating the influence of geo-environmental factors on gully erosion in a semi-arid region of Iran: An integrated framework. *Sci. Total Environ.* **579**, 913–927 (2017).
42. Abdollahi, S., Pourghasemi, H. R., Ghanbarian, G. A. & Safaeian, R. Prioritization of effective factors in the occurrence of land subsidence and its susceptibility mapping using an SVM model and their different kernel functions. *Bull. Eng. Geol. Environ.* **78**, 4017–4034 (2019).
43. Mohammady, M., Pourghasemi, H. R. & Amiri, M. Assessment of land subsidence susceptibility in Semnan plain (Iran): A comparison of support vector machine and weights of evidence data mining algorithms. *Nat. Haz.* **99**, 951–971 (2019).
44. Yariyan, P., Avand, M., Soltani, F. & Ghorbanzadeh, O. *SS Symmetry Earthquake Vulnerability Mapping Using Different*. (2020).
45. Theilen-Willige, B. & Wenzel, H. Remote sensing and GIS contribution to a natural hazard database in western Saudi Arabia. *Geosciences (Switzerland)* **9**, 8–15 (2019).
46. Pyayt, A. L., Mokhov, I. I., Lang, B., Krzhizhanovskaya, V. V. & Meijer, R. J. Machine learning methods for environmental monitoring and flood protection. *World Acad. Sci. Eng. Technol.* **78**, 118–123 (2011).
47. Ochola, S. O., Eitel, B. & Olago, D. O. Vulnerability of schools to floods in Nyando River catchment, Kenya. *Disasters* **34**, 732–754 (2010).
48. Karmakar, S., Simonovic, S. P., Peck, A. & Black, J. An information system for risk-vulnerability assessment to flood. *J. Geogr. Inf. Syst.* **2**, 129 (2010).
49. Balica, S. F., Popescu, I., Beevers, L. & Wright, N. G. Parametric and physically based modelling techniques for flood risk and vulnerability assessment: A comparison. *Environ. Model. Softw.* **41**, 84–92 (2013).
50. Nabegu, A. B. Analysis of vulnerability to flood disaster in Kano State, Nigeria. *Greener J. Phys. Sci.* **4**, 22–29 (2014).
51. Eini, M., Kaboli, H. S., Rashidian, M. & Hedayat, H. Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts. *Int. J. Disaster Risk Reduct.* 101687 (2020).
52. Tascón-González, L., Ferrer-Julià, M., Ruiz, M. & García-Meléndez, E. Social vulnerability assessment for flood risk analysis. *Water* **12**, 558 (2020).
53. Yousefi, S. *et al.* A machine learning framework for multi-hazards modeling and mapping in a mountainous area. *Sci. Rep.* **10**, 1–14 (2020).
54. Al-Abadi, A. M. Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: A comparative study. *Arab. J. Geosci.* **11**, 218 (2018).
55. Sachdeva, S., Bhatia, T. & Verma, A. K. Flood susceptibility mapping using GIS-based support vector machine and particle swarm optimization: A case study in Uttarakhand (India). in *8th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2017* 1–7 (IEEE, 2017). https://doi.org/10.1109/ICCCNT.2017.8204182.
56. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
57. Friedman, J., Hastie, T. & Tibshirani, R. Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *Ann. Stat.* **28**, 337–407 (2000).
58. Achour, Y. & Pourghasemi, H. R. How do machine learning techniques help in increasing accuracy of landslide susceptibility maps?. *Geosci. Front.* https://doi.org/10.1016/j.gsf.2019.10.001 (2019).
59. Reineking, B. & Schröder, B. Constrain to perform: Regularization of habitat models. *Ecol. Modell.* **193**, 675–690 (2006).
60. Fraley, C. & Raftery, A. E. Model-based clustering, discriminant analysis, and density estimation. *J. Am Stat. Assoc.* **97**, 611–631 (2002).
61. Hosseinalizadeh, M. *et al.* How can statistical and artificial intelligence approaches predict piping erosion susceptibility?. *Sci. Total Environ.* **646**, 1554–1566 (2019).
62. Pourghasemi, H. R., Yousefi, S., Sadhasivam, N. & Eskandari, S. Assessing, mapping, and optimizing the locations of sediment control check dams construction. *Sci. Total Environ.* 139954 (2020).
63. Rausch, J. R. & Kelley, K. A comparison of linear and mixture models for discriminant analysis under nonnormality. *Behav. Res. Methods.* **41**(1), 85–98 (2009).
64. Li, X. & Wang, Y. Applying various algorithms for species distribution modelling. *Integr. Zool.* **8**, 124–135. https://doi.org/10.1111/1749-4877.12000 (2013).
65. Hastie, T., Tibshiran, R., Leisch, F., Hornik, K., Ripley, B.D. Mixture and flexible discriminant analysis. (2017). https://cran.r-project.org/web/packages/mda/mda.pdf.
66. Hawryło, P., Bednarz, B., Wężyk, P. & Szostak, M. Estimating defoliation of Scots pine stands using machine learning methods and vegetation indices of Sentinel-2. *Eur. J. Remote Sens.* **51** (2018).
67. Gayen, A., Pourghasemi, H. R., Saha, S., Keesstra, S. & Bai, S. Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Sci. Total Environ.* **668**, 124–138 (2019).
68. Kim, J. C., Lee, S., Jung, H. S. & Lee, S. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto Int.* **33**, 1000–1015 (2018).
69. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
70. Shahabi, H. *et al.* A semi-automated object-based gully networks detection using different machine learning models: A case study of Bowen catchment, Queensland, Australia. *Sensors (Switzerland)* **19**, 4893 (2019).
71. Adnan, R. M. *et al.* Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *J. Hydrol.* 124371 (2019). https://doi.org/10.1016/J.JHYDROL.2019.124371
72. Gu, C. & Wahba, G. Discussion: Multivariate adaptive regression splines. *Ann. Stat.* **19**, 115–123 (1991).
73. Busto Serrano, N., Suárez Sánchez, A., Sánchez Lasheras, F., Iglesias-Rodríguez, F. J. & Fidalgo Valverde, G. Identification of gender differences in the factors influencing shoulders, neck and upper limb MSD by means of multivariate adaptive regression splines (MARS). *Appl. Ergon.* **82**, 102981 (2020).
74. Deichmann, J., Eshghi, A., Haughton, D., Sayek, S. & Teebagy, N. Application of multiple adaptive regression splines (mars) in direct response modeling. *J. Interact. Mark.* **16**, 15–27 (2002).
75. Lazarus, E. D. & Constantine, J. A. Generic theory for channel sinuosity. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 8447–8452 (2013).
76. Corte-Real, J., Zhang, X. & Wang, X. Downscaling GCM information to regional scales: a non-parametric multivariate regression approach. *Clim. Dyn.* **11**, 413–424 (1995).
77. Hjort, J. & Luoto, M. Statistical methods for geomorphic distribution modeling. *Treatise Geomorphol.* **2**, 59–73 (2013).

78. Abdulelah Al-Sudani, Z., Salih, S. Q., sharafati, A. & Yaseen, Z. M. Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. *J. Hydrol.* **573**, 1–12 (2019).
79. Tien Bui, D. *et al.* Land subsidence susceptibility mapping in South Korea using machine learning algorithms. *Sensors (Switzerland)* **18**, 2464 (2018).
80. Hosseinalizadeh, M. *et al.* Gully headcut susceptibility modeling using functional trees, naïve Bayes tree, and random forest models. *Geoderma* **342**, 1–11 (2019).
81. Taylor, R. A. J., Digby, P. G. N. & Kempton, R. A. Multivariate analysis of ecological communities. *J. Anim. Ecol.* **56** (1987).
82. Pertille, R. H., Sachet, M. R., Guerrezi, M. T. & Citadin, I. An R package to quantify different chilling and heat models for temperate fruit trees. *Comput. Electron. Agric.* **167**, 105067 (2019).
83. Fielding, A. H. & Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **24**, 38–49 (1997).
84. Merow, C., Smith, M. J. & Silander, J. A. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography (Cop.)* **36**, 1058–1069 (2013).
85. 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).
86. Rahmati, O. *et al.* GIS-based site selection for check dams in watersheds: Considering geomorphometric and topo-hydrological factors. *Sustain.* **11**, 5639 (2019).
87. Yousefi, S. *et al.* A novel GIS-based ensemble technique for rangeland downward trend mapping as an ecological indicator change. *Ecol. Indic.* **117**, 106591 (2020).
88. Krois, J. & Schulte, A. GIS-based multi-criteria evaluation to identify potential sites for soil and water conservation techniques in the Ronquillo watershed, northern Peru. *Appl. Geogr.* **51**, 131–142 (2014).
89. Yilmaz, B. *Application of GIS-Based Fuzzy Logic and Analytical Hierarchy Process (AHP) to Snow Avalanche Susceptibility Mapping,* North San Juan, Colorado. (2016).
90. Pourghasemi, H. R., Beheshtirad, M. & Pradhan, B. A comparative assessment of prediction capabilities of modified analytical hierarchy process (M-AHP) and Mamdani fuzzy logic models using Netcad-GIS for forest fire susceptibility mapping. *Geomat. Nat. Haz. Risk* **7**, 861–885 (2016).
91. Dean, D. J. & Schmidt, J. C. The geomorphic effectiveness of a large flood on the Rio Grande in the Big Bend region: Insights on geomorphic controls and post-flood geomorphic response. *Geomorphology* **201**, 183–198 (2013).
92. Borga, M., Boscolo, P., Zanon, F. & Sangati, M. Hydrometeorological analysis of the 29 August 2003 flash flood in the Eastern Italian Alps. *J. Hydrometeorol.* **8**, 1049–1067 (2007).
93. Hong, H. *et al.* Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Sci. Total Environ.* **621**, 1124–1141 (2018).
94. Taalab, K., Cheng, T. & Zhang, Y. Mapping landslide susceptibility and types using random forest. *Big Earth Data* **2**, 159–178 (2018).
95. Avand, M. *et al.* A comparative assessment of random forest and k-nearest neighbor classifiers for gully erosion susceptibility mapping. *Water (Switzerland)* **11**, 2076 (2019).
96. Santos, M., Aguiar, M., Oliveira, A., Magalhães, L. & Pereira, F. Vulnerability to mass movements' hazards. Contribution of sociology to increasing community resilience. in *Advances in Natural Hazards and Hydrological Risks: Meeting the Challenge* 105–108 (Springer, 2020).

Acknowledgements

This work was supported by College of Agriculture, Shiraz University (Grant No. 98GRC1M271143).

Author contributions

S.Y., H.R.P., S.N.E., O.R., S.T., S.P., J.P.T., S.S., and M.N. designed experiments, run models, analyzed results, wrote and reviewed manuscript. All authors reviewed the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

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.