Arabian Journal of Geosciences (2020) 13: 417
https://doi.org/10.1007/s12517-020-05363-1

ORIGINAL PAPER

Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping

Mehdi Vafakhah[1] · Sajad Mohammad Hasani Loor[1] · Hamidreza Pourghasemi[2] · Azadeh Katebikord[1]

Received: 11 August 2018 / Accepted: 28 April 2020 / Published online: 30 May 2020

Abstract
Flood is one of the important destructive natural disasters in the world. Therefore, preparing flood susceptibility map is necessary for flood management and mitigation in a region. This research was planned to compare the performance of frequency ratio (FR), adaptive neuro-fuzzy inference system (ANFIS), and random forest (RF) models for flood susceptibility mapping (FSM) in the Gilan Province, Iran. First, a geospatial database included 220 flood locations and eleven effective flood factors (slope angle, aspect, altitude, distance from rivers, drainage density, lithology, land use, topographic wetness index (TWI), and stream power index (SPI)) were produced. According to flood locations, 30–70% of them were used for training and validation of the models, respectively. Afterward, the mean of Gini reduction was used to determine the priority of effective flood factors. Finally, the receiver operating characteristic (ROC) curve, area under the curve (AUC), was used to evauate and compare the performance of the models. The validation results of the models show that FR, ANFIS, and RF models had 68.6, 63.9, and 71.3% accuracy, respectively. In addition, distance from rivers, altitude, and drainage density was the most important factor for FSM in the study area. The finding of the current research proved a reasonable prediction performance for the models. Therefore, these models can be proposed for preparing FSM in similar climatic and physiographic areas and flood susceptibility maps can be used to manage floodplains in the study area.

Keywords Flooding · Random forest · Adaptive neuro-fuzzy inference system · Frequency ratio

Introduction

According to statistics provided by the United Nations, among the natural disasters, floods and storms have brought the most casualties and losses to human societies. Between 2000 and 2008, nearly 99 million people worldwide have been affected by flooding (Opolot 2013). The flooding trend in recent years suggests that most of Iran's regions are exposed to destructive floods. In addition to financial and psychological damage, floods cause soil erosion and nutrient loss. The continuation of this situation entails irreparable damage to water and soil resources. In the current situation, it is expected that the socio-economic and environmental consequences of floods will be felt more than ever in the context of urbanization, increasing deforestation, and sustainability of rainfall because of climate change in the various regions. Therefore, operation prevention are essential and necessary (Alvarado-Aguilar et al. 2012; Billa et al. 2006; Dang et al. 2011; Huang et al. 2008). One of these measures is to provide a flood map that is necessary in integrated management for sustainable development (Rahmati et al. 2016b).

Lee et al. (2012) found the accuracy of the frequency ratio (FR) model with area under the curve (AUC) of receiver operating characteristics (ROC) of 91.5% for preparing flood susceptibility mapping (FSM) in Busan, South Korea. Tehrany et al. (2013) and Tehrany et al. (2014) provided FSM for Kuala Terengganu County in Malaysia using two sets of decision-

Responsible Editor: Biswajeet Pradhan

✉ Mehdi Vafakhah
vafakhah@modares.ac.ir

Hamidreza Pourghasemi
hr.pourghasemi@shirazu.ac.ir

[1] Department of Watershed Management Engineering, Faculty of Natural Resources, Tarbiat Modares University, Tehran, Iran

[2] Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran

making methods, a combination of FR and logistic regression (ensemble model), support vector machine (SVM), and weight-of-evidence (WOE). They found that the accuracy of decision trees, the ensemble model, SVM, and WOE models were 87%, 90%, 95.67%, and 96.48%, respectively. Youssef et al. (2016b) prepared FSM for Jeddah County in Saudi Arabia using FR and logistic regression (LR). They found that the accuracy of FR and LR models was 89.6% and 91.3%, respectively. Rahmati et al. (2016b) prepared FSM in Golestan Province, Iran, using FR and WOE models. According to the results, these two models had the same and reasonable efficiency for identifying FSM. Khosravi et al. (2016) prepared FSM with four different models: FR, WOE, analytical hierarchy process (AHP), and FR-AHP for Haraz Watershed, Iran. They found that the accuracy of FR, WOE, AHP, and FR-AHP models were 96.55%, 96.95%, 94.99%, and 84.69%, respectively. Rahmati et al. (2016c) evaluated the accurate and reliable performance of AHP technique in potential flood hazard zones identification by comparing with the results of HEC-RAS hydraulic model in some part of the Bashar River downstream of Yasooj city in Iran. Mojaddadi et al. (2017) evaluated the ensemble method of FR and SVM with a radial basis function kernel in FSM for Damansara River catchment in Malaysia. Based on the results, the accuracy of FR-SVM model was 78.9%. Haghizadeh et al. (2017) and Siahkamari et al. (2018) applied the Shannon's entropy, FR, and maximum entropy models for FSM in the Madarsoo Watershed, Golestan Province, Iran. Based on the results, the accuracy of Shannon's entropy, FR, and maximum entropy models was 73.5, 74.3, and 92.6%, respectively. Shafizadeh-Moghadam et al. (2018) compared eight models in FSM for Haraz Watershed, Iran. Based on the results, the highest and lowest accuracy values were reported for boosted regression trees model with 0.975% and generalized linear model with 0.642%, respectively. Darabi et al. (2019) applied the genetic algorithm rule-set production (GARP) and quick unbiased efficient statistical tree (QUEST) models to produce FSM in Sari city, Iran. Based on the results, the accuracy of GARP and QUEST models was 93.5 and 89.2%, respectively. Falah et al. (2019) evaluated an artificial neural network (ANN) model with the the accuracy of 92.0% in FSM for Emam-Ali township in Mashhad city, Iran. Chen et al. (2019a) applied the machine learning-based reduced error pruning trees (REPTree) with bagging (Bag-REPTree) and random subspace (RS-REPTree) ensemble models in FSM for Quannan County, China. Based on the results, the highest accuracy value was reported for RS-REPTree model with 0.907%. Costache and Tien Bui (2019) compared six ensemble models by combining the FR and WOE on the one hand, with ANN, rotation forest (RF), and classification and regression trees (CART) in FSM for Putna River Watershed, Romania. Based on the results, all six ensemble models indicated a high flood prediction performance from 86.8 to 93.9%. Khosravi et al. (2019) compared three mlti-criteria decision-making (MCDM) analysis techniques (VIKOR, TOPSIS, and SAW) along with two machine learning methods (naive bayes trees (NBT) and naive bayes (NB)) in FSM for the Ningdu Watershed, China. Based on the results, all models indicated a high flood prediction capability (> 95%). Mind'je et al. (2019) evaluated the logistic regression model with the accuracy of 79.8% in FSM for Rwanda, Africa.

Therefore, the methods used to investigate and map flood susceptibility in recent years were including FR, multivariate statistical analyzes, WOE, MCDM, LR, and decision tree. In recent years, advanced methods in FSM, data mining have been considered as a useful tool such as ANN (Falah et al. 2019; Costache and Tien Bui 2019), SVM (Tehrany et al. 2014; Mojaddadi et al. 2017), adaptive neuro-fuzzy inference system (ANFIS) with grid partitioning method and metaheuristic optimization algorithms (Razavi Termeh et al. 2018; Tien Bui et al. 2018), and random forest model (RF) (Rahmati and Pourghasemi 2017). ANFIS has showed good results in modeling non-linear processes. ANFIS has adjusted the characteristics of the system according to training data and metadata according to the required accuracy. ANFIS generates FIS by two methods (grid partitioning and subtractive clustering methods) (Vafakhah and Kahneh 2016). RF is a new supervised classification method in modeling process. RF method have been used in various studies such as forest fire, landslide susceptibility mapping (Pachauri et al. 1998; Cevik and Topal 2003; Sidle and Ochiai 2006; Pradhan et al. 2011; Catani et al. 2013; Pourghasemi and Kerle 2016; Youssef et al. 2016a; Chen et al., 2018a, b, c, 2019b, d), ecological studies, and groundwater potential mapping (Rahmati et al. 2015; Naghibi et al. 2016; Chen et al. 2019c). Therefore, new ensemble models should be further investigated.

Given the negative effects of floods, the identification of flood-prone areas is essential. Therefore, due to the aforementioned issues and the lack of information and data of appropriate quality in most watersheds, ANFIS with subtractive clustering method and RF models are used due to the good accuracy in the preparation of FSM and a groundwater potential mapping, as well as landslide susceptibility mapping. In addition, there have been no studies, which compare the ANFIS and RF differences and apply the ANFIS with subtractive clustering method in the preparation of FSM. In the present study, the effectiveness of these two models was investigated in the preparation of the FSM in the Gilan Province, Iran.

Materials and methods

General characteristics of the study area and database

The present study was conducted in the Gilan Province with an area of 14,100 km^2 between the northern latitudes of 36° 34′ 00″ to 38° 27′ 00″ and eastern longitudes of 48° 34′ 00″ to 48° 36′ 00″ at the southern margin of the Caspian Sea. The

Gilan Province is limited to the Caspian Sea from the north, to the Alborz Mountains from the south, to Mazandaran Province from the east, and to Ardabil Province from northwest (Fig. 1). The mean above sea level varies from − 128 m in coastal areas to 2700 m in mountainous areas. The Caspian strip is considered a temperate and humid region, under the influence of the northern Siberian air masses and the western air masses of the Mediterranean and the Atlantic.

Also, the Alborz Mountain, like a barrier, prevents the outflow of moist air masses and floods to the central parts of Iran, causing significant atmospheric precipitation in northern provinces. In addition to rangeland cover, the Gilan Province is often covered with broad-and-hardwood forests.

After collecting spatial information of flood events using flood database, flood inventory map including 220 flood locations (Fig. 2) was prepared. Also, the geological map of the region at the scale of 1:100,000 from the Geological Survey and Mineral Explorations of Iran (GSI) and topographic map at 1:50,000 scale from the Iran National Cartographic Center were obtained. Land use map was prepared using Landsat 8 OLI images (27 Jun 2016).

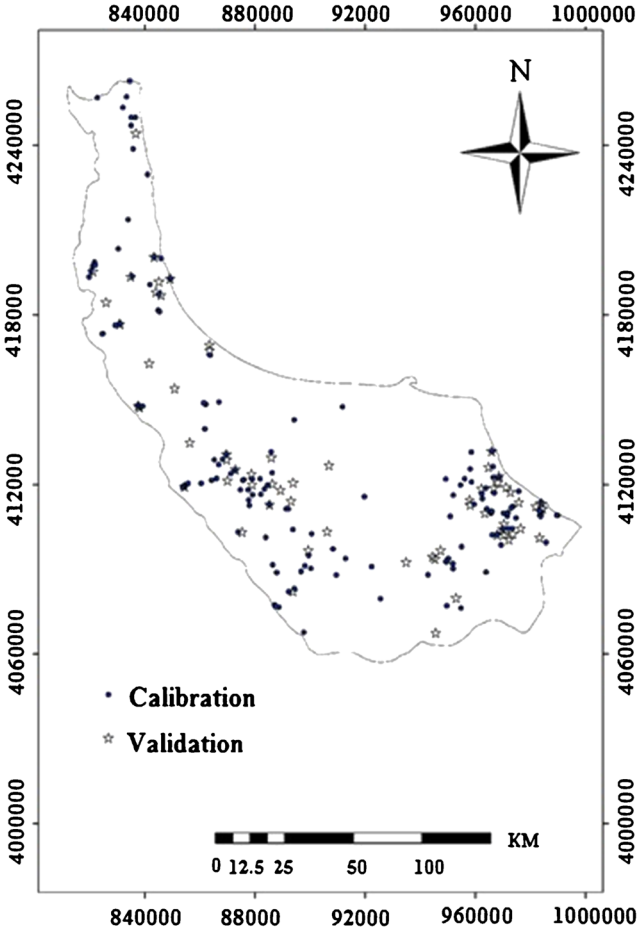

Fig. 2 Map of flood locations in the study area

Methodology

In this study, eleven effective factors including slope angle, aspect, plan curvature, profile curvature, altitude, distance from rivers, drainage density, geology, land use, topographic wetness index (TWI) (Eq. 1) (Beven and Kirkby 1979), and stream power index (SPI) (Eq. 2) (Moore et al. 1991) were used to prepare FSM. Since there are a significant negative correlation between elevation and rainfall in the study area (Fig. 3), we used altitude in this study to avoid collinearity.

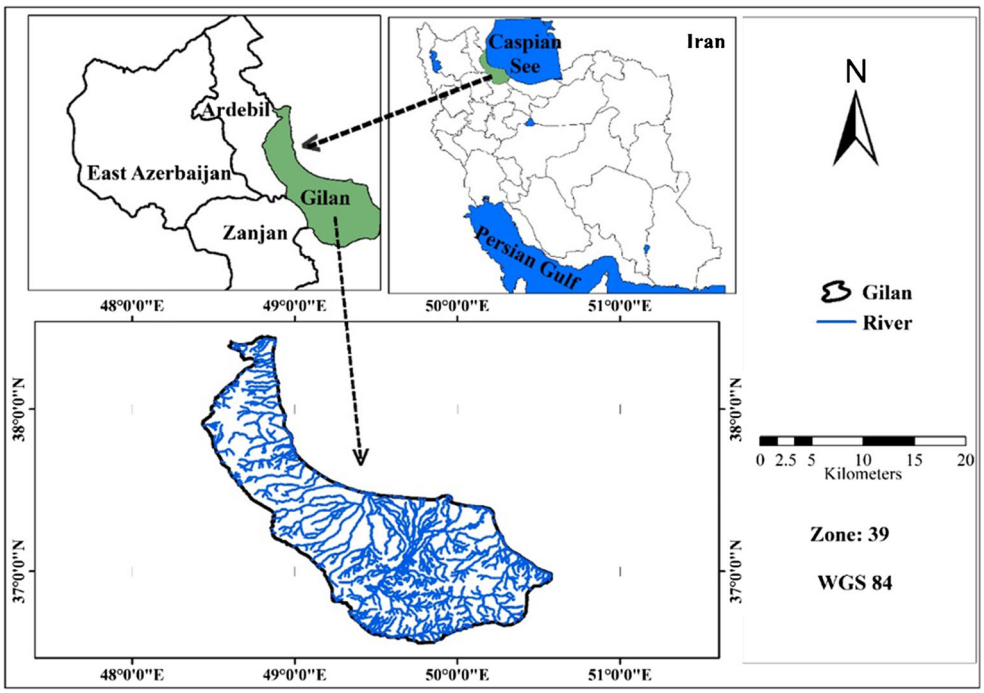

Fig. 1 Location of the study area

Raster grid maps with 30 m × 30 m resolution were then derived for all effective factors. The various stages of the research are presented in Fig. 4.

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

$$SPI = A_s \tan\beta \quad (2)$$

where α is the total upslope catchment area draining downward from a point with a slope angle of β and A_s is the specific area (local upslope area draining through a certain point per unit contour length) of a basin (m^2/m).

Frequency ratio method

After producing and categorizing each of the effective factors in the research, the layers were overlaid with flood inventory map, so that the number of floods in each class could be attributed to different factors. In the next step, the FR coefficients of each class/category were determined through Eq. 3.

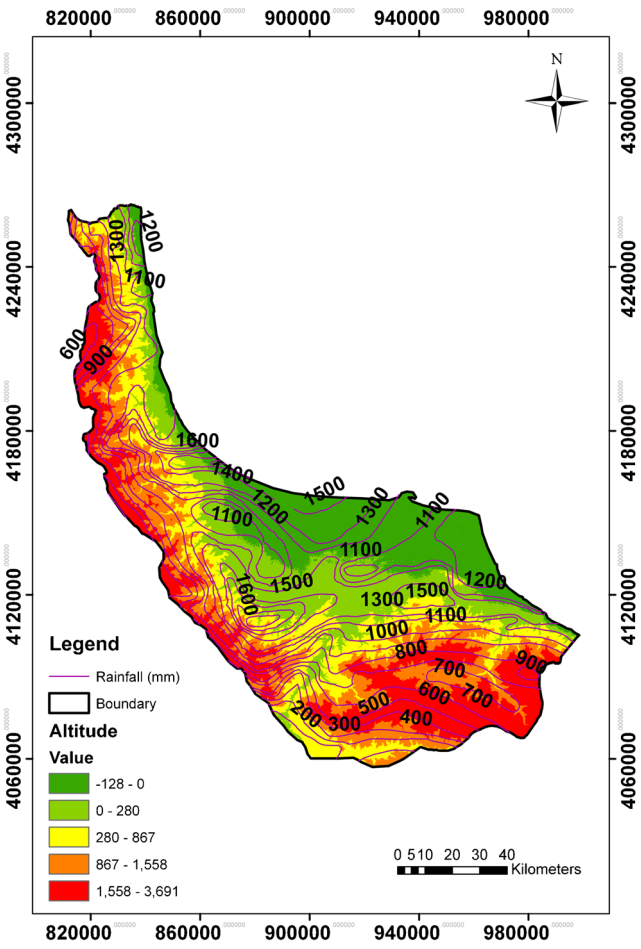

Fig. 3 Altitude map of the study

$$FR = \frac{\text{Percent of flooded pixcels in each class}}{\text{Percent of each factor class pixcels}} \quad (3)$$

Adaptive neuro-fuzzy inference system

There are two methods of grid partitioning and subtractive clustering to generate FIS. In subtractive clustering, the input data is divided into several groups by the size of the impact radius. In this case, the number of linear and non-linear factors has significantly decreased, which facilitates the process of network training. In this research, subtractive clustering method was used to investigate the effect of FIS generation on the performance of ANFIS model. Data standardization (Erdirencelebi and Yalpir 2011; Kisi et al. 2013; Sidle and Ochiai 2006; Vafakhah 2012) was carried out as follows:

$$X_{\text{Standardized}} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (4)$$

where $X_{\text{Standardized}}$ is the standardized value, X_i is the original value, and $X_{\min}$ and $X_{\max}$ are, respectively, the minimum and maximum value.

After data standardization and determining the input variables, the data were divided into two parts: training and testing datasets. Optimal replications and range of influence were determined in the model using hybrid optimization method. The range of influence was changed from 0.4 to 0.8 with steps of 0.01, and the optimal range of influence was determined.

Random forest model

To implement this model, firstly, a large number of decision trees were determined on the basis of trial and error. Then, all the trees were combined together for prediction (Cutler et al. 2007). For implementation of RF model, R software and "randomForest" package was used (Naghibi and Pourghasemi 2015; Pourghasemi and Kerle 2016; Rahmati et al. 2016a; Youssef et al. 2016a). When predictor variables and target variables were identified, RF begins with the emergence of a decision tree. This tree does not use all the available data for the tree, and, instead uses the bootstrap sample, which only contains 66% of the original data, which is referred to as the bagging technique (Breiman 2001).

Evaluation of the performance of the models

In this research, 70% and 30% of the flood locations were used for calibration and validation and performance evaluation of the models, respectively. Then, using the receiver operating characteristics (ROC) curve, the accuracy of the flood susceptibility map was determined (Pourghasemi et al. 2012). The area under the ROC curve represents the predicted value

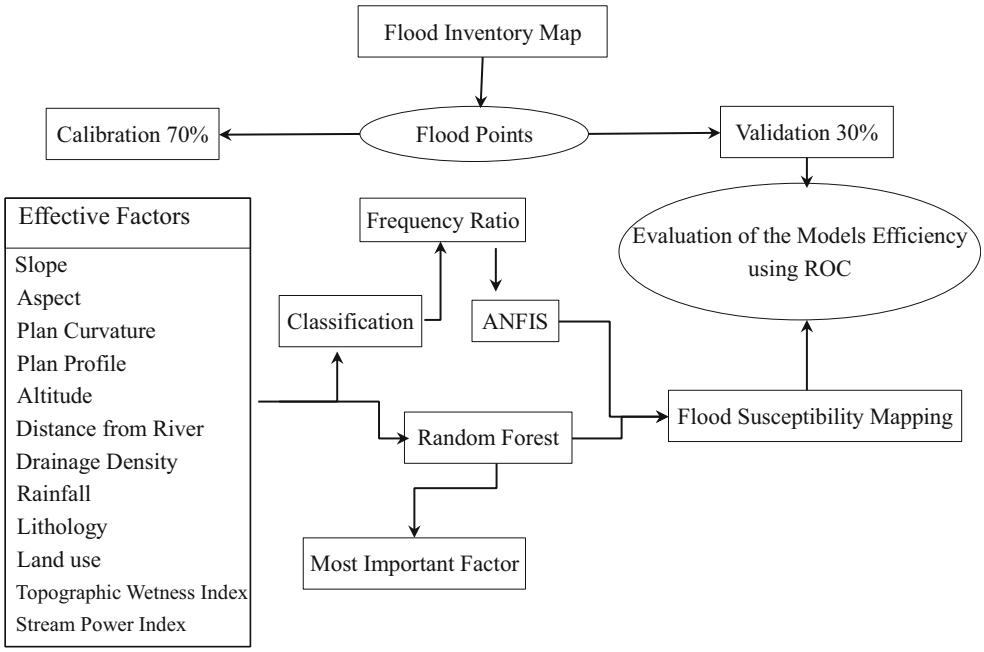

Fig. 4 Flow chart of research method

of the system by describing its ability to accurately estimate events (flood occurrence) and non-occurrence of the event (absence of flood). Using the ROC curve, the accuracy of the model was estimated quantitatively. The most ideal model is the model with the highest the area under the curve (AUC) and AUC values vary from − 1 to 0.5. If the model cannot estimate the flood event better than the probable viewpoint, that is, AUC is less than 0.5, and then the model used does not have usability for prediction. AUC and the estimation are classified as 0.5–0.6 weak, 0.6–0.7 moderate, 0.7–0.8 good, 0.8–0.9 very good, and 0.9–1 excellent (Rahmati et al. 2015).

Results and discussion

The present research aims to provide a FSM map using three models, namely FR, ANFIS, and RF models and to compare their performance together.

Effective flood factors

Slope angle

A digital elevation model (DEM) by 30-m spatial resolution in Arc/GIS environment was used to prepare the slope map of the study area. As can be seen from the Fig. 5, the slope map was divided into five classes by quantile classification method (Tehrany et al. 2015). As shown in Table 1, class 2.36–6.21 degrees has the highest FR, subsequently has the highest value (1.55), which has the most effectiveness on flooding, and class of 76.01–20.72 degrees has the lowest FR (0.53).

Aspect

The aspect due to evapotranspiration and precipitation direction has a great influence on hydrological processes and, as a result, affects weathering and vegetation processes, especially in arid areas (Sidle and Ochiai 2006). The aspect map was prepared in 9 classes of north, northeast, east, southeast, south, southwest, west, northwest, and flat directions (Fig. 5) (Rahmati et al. 2016b). As shown in Table 1, the southeastern direction with a FR of 1.53 has the most effect on flooding and southwest direction with a FR of 0.6 has the least effect on flood.

Plan curvature

The surface curvature (plan curvature) map can be used to describe the divergence and convergence of flows in the basins, trenches, and drainage network (Naghibi and Pourghasemi 2015). The surface curvature map was prepared based on Fig. 5 in three classes including, convex, flat, and concave (Rahmati et al. 2016b). The study of the surface curvature map indicated that flat surfaces with a FR of 1.14 has the greatest effect on flood; in contrast, convex and concave slopes with a FR of 0.91 have the same effect on flood (Table 1).

Profile curvature

The semicircular curvature indicates the ratio of slope gradients in the direction of maximum slope, which was prepared in three classes, namely, concave, flat, and convex (Fig. 5). The study of the profile curvature indicated that the flat surfaces with a FR of 1.59 have the greatest effect and concave slope with a FR of 0.82 has the least effect on flood in the study area (Table 1).

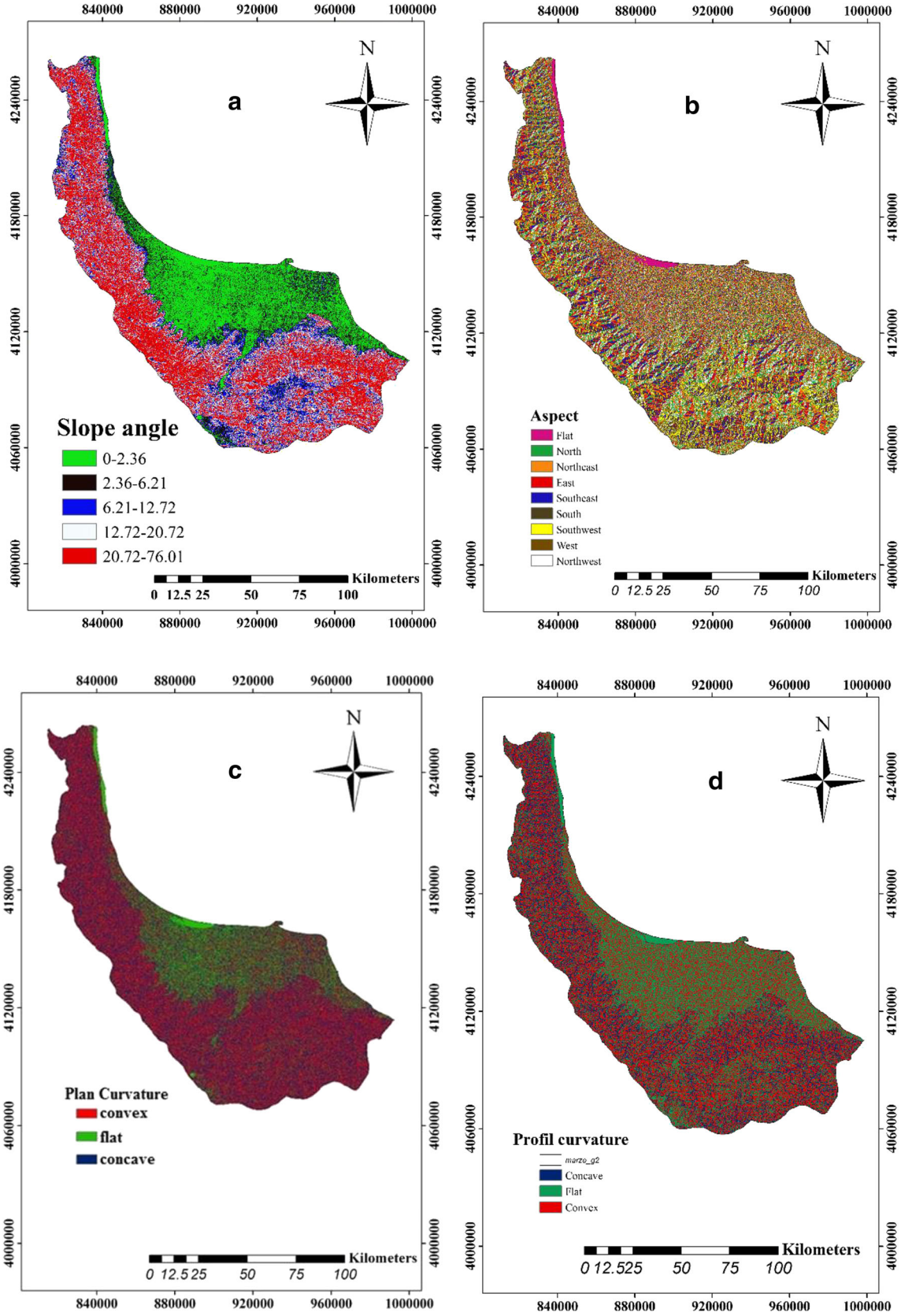

Fig. 5 Maps of the flood-conditioning factors (eleven in all): **a** slope angle, **b** aspect, **c** plan curvature, **d** profile curvature, **e** altitude, **f** distance from rivers, **g** drainage density, **h** land use, **i** geology, **j** topographic wetness index (TWI), and **k** stream power index (SPI)

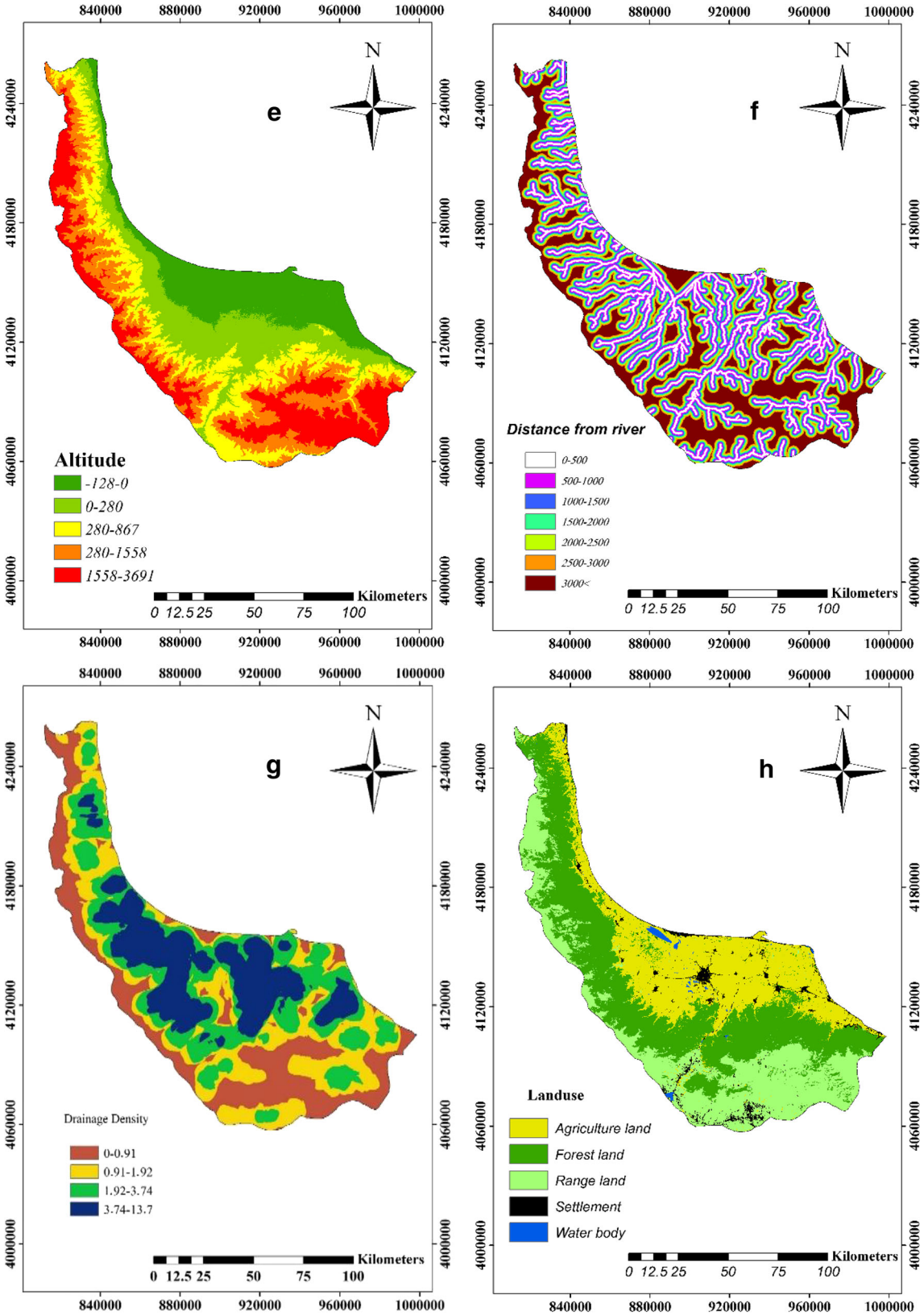

Fig. 5 continued.

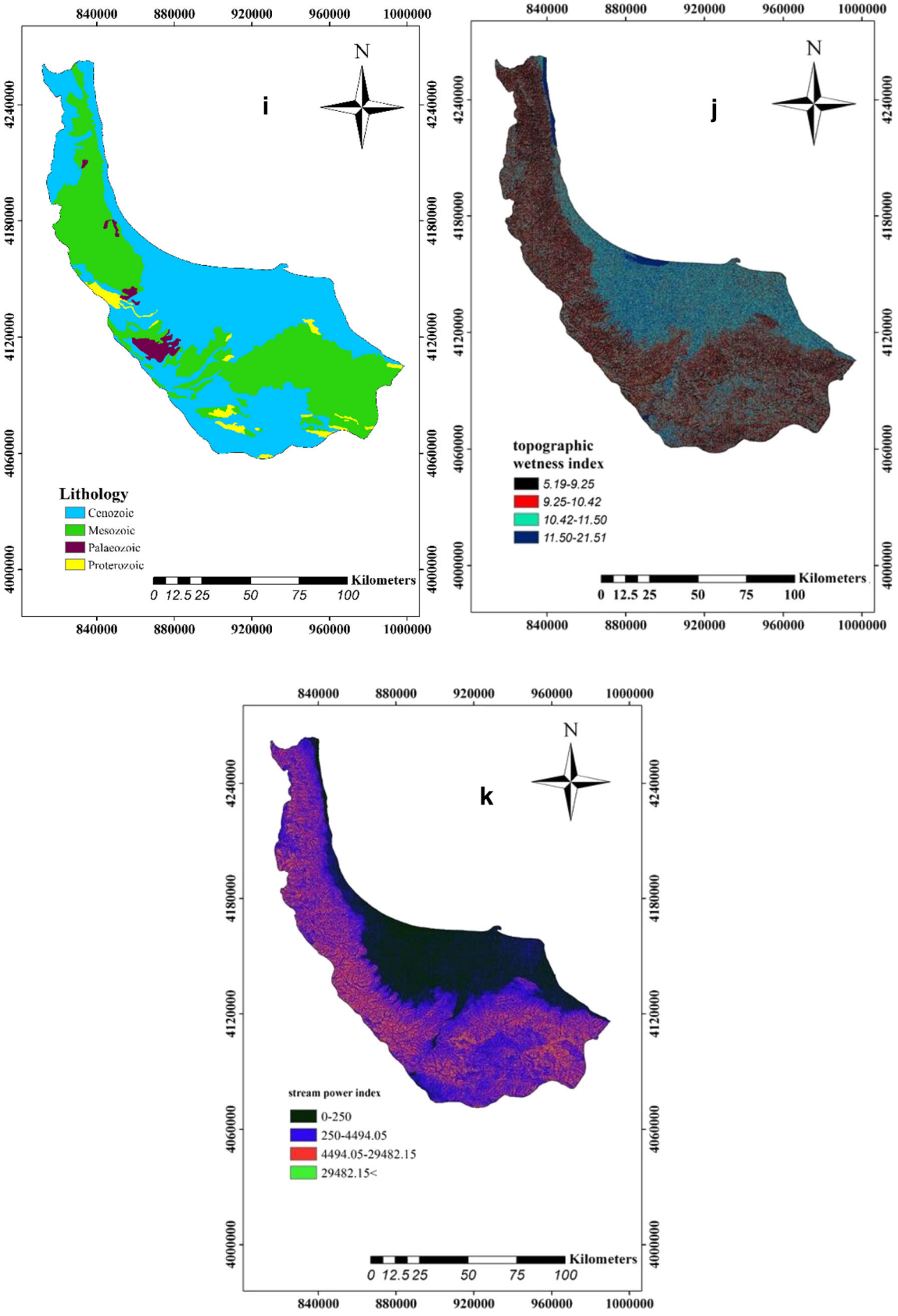

Fig. 5 continued.

Table 1 The frequency ratio of previous flood events and flood-risk parameters

Factor	Factor categories	Number of pixel	Pixel percentage	Area (km^2)	Flood number	Flood percentage	Frequency ratio	Standardized values
Slope (degree)	0–2.36	3500063	22.6	3150.06	27	17.53	0.78	0.25
	2.36–6.21	1941432	12.53	1747.29	30	19.48	1.55	1.00
	6.21–12.72	1974087	12.74	1776.68	28	18.18	1.43	0.88
	12.72–20.72	2938889	18.97	2645.00	42	27.27	1.44	0.89
	20.72–76.01	5135219	33.15	4621.70	27	17.53	0.53	0.00
Aspect	Flat	321116	2.07	289.00	2	1.3	0.63	0.03
	North	2281913	14.73	2053.72	19	12.34	0.84	0.25
	Northeast	2229228	14.39	2006.31	24	15.58	1.08	0.49
	East	2054727	13.27	1849.25	17	11.04	0.83	0.24
	Southeast	1908850	12.32	1717.97	30	19.48	1.58	1.00
	South	1786641	11.53	1607.98	24	15.58	1.35	0.77
	Southwest	1518092	9.8	1366.28	9	5.84	0.6	0.00
	West	1534676	9.91	1381.21	12	7.79	0.79	0.20
	Northwest	1854447	11.97	1669.00	17	11.04	0.92	0.33
Profile curvature	Convex	2331987	15.06	2098.79	21	13.64	0.91	0.00
	Flat	5900585	38.09	5310.53	67	43.51	1.14	0.99
	Concave	7257118	46.85	6531.41	66	42.86	0.91	0.04
Plan curvature	Concave	5831819	37.65	2098.79	54	35.06	0.93	0.14
	Flat	2781229	17.96	5310.53	44	28.57	1.59	1.00
	Convex	6876642	44.39	6531.41	56	36.36	0.82	0.00
Altitude (m)	< 0	3069862	19.82	2762.88	23	14.94	0.75	0.22
	0–280	3123853	20.17	2811.47	61	39.61	1.96	1.00
	280–867	3117052	20.12	2805.35	34	22.08	1.1	0.44
	867–1558	3114014	20.1	2802.61	23	14.94	0.74	0.21
	> 1558	3064909	19.79	2758.42	13	8.44	0.43	0.00
Distance from rivers (m)	0–500	2541781	16.41	2287.60	62	40.26	2.45	1.00
	500–1000	2321717	14.99	2089.55	31	20.13	1.34	0.47
	1000–1500	2101608	13.57	1891.45	9	5.84	0.43	0.04
	1500–2000	1877242	12.12	1689.52	17	11.04	0.91	0.27
	2000–1500	1614702	10.42	1453.23	13	8.44	0.81	0.22
	2500–3000	1355669	8.75	1220.10	9	5.84	0.67	0.15
	> 3000	3676971	23.74	3309.27	13	8.44	0.36	0.00
Drainage density (km/km^2)	>1.36	3757335	24.26	13940.7	34	22.08	0.91	0.00
	1.36–2.11	4005297	25.86	3604.77	37	24.03	0.93	0.08
	2.11–13.34	3866761	24.96	3480.08	39	25.32	1.01	0.44
	13.34–17.49	3860297	24.92	3474.27	44	28.57	1.15	1.00
Land use	Agriculture	4690060	30.28	4221.05	50	32.47	1.07	0.40
	Forest	6061261	39.13	5455.13	55	35.71	0.91	0.34
	Range	4195704	27.09	3776.13	37	24.03	0.89	0.33
	Settlement	454782	2.94	409.30	12	7.79	2.65	1.00
	Water body	87883	0.57	79.09	0	0.00	0.00	0.00
Geology	Cenozoic	8656838	55.89	7791.15	76	49.35	0.88	0.26
	Mesozoic	6126054	39.55	5513.45	70	45.45	1.15	0.44
	Paleozoic	307501	1.99	276.75	6	3.90	1.96	1.00
	Proterozoic	399297	2.58	359.37	2	1.30	0.50	0.00
Topographic wetness index	< 9.25	3639404	23.50	3275.46	20	12.99	0.55	0.00
	9.25–10.42	3795007	24.50	3415.51	44	28.57	1.17	0.75
	10.42–11.50	3957464	25.55	3561.72	34	22.08	0.86	0.38
	11.50–21.51	4097815	26.46	3688.03	56	36.36	1.37	1.00
Stream power index	0–250	4277995	27.62	3850.20	33	21.43	0.78	0.00
	250–4494.05	7200935	46.49	6480.84	82	53.25	1.15	0.58
	4494.05–29482.15	3514071	22.69	3162.66	32	20.78	0.92	0.22
	> 29482.15	496689	3.21	447.02	7	4.55	1.42	1.00

Altitude

Altitude can be considered as one of the effective factors in flood studies (Tehrany et al. 2015). It is almost impossible to absorb flood in highlands. Altitude class map were prepared using the quantile method into 5 categories less than 0, 0–280, 280–687, 687–1558, and more than 1558 m (Tehrany et al., 2013, 2014, 2015. As shown in Table 1, the class 0–280 m with a FR of 1.96 has the most effect and the class of more than 1558 m with a FR of 0.43 has the least effect on flood in the study area (Fig. 5).

Distance from rivers

Distance from rivers plays a significant role in the velocity and extent of flood (Glenn et al. 2012). Map of distance from river was prepared and classified into seven classes (Fig. 5) (Darabi et al. 2019). The distance of less than 500 m with a FR of 2.45 has a significant effect on flood in the study area. Also, the class of more than 3000 m has a FR of 0.36 (Table 1).

Drainage density

Drainage density map was prepared using stream lines in km/km^2 and then classified into four different classes using quantile classification method (Rahmati et al. 2016b) (Fig. 5). As shown in Table 1, class of 13.34–17.49 with a FR of 1.15 has the highest effect and the class less than 1.36 has the least effect on flood in the study area.

Land use

Land use directly or indirectly affects some of the components of hydrological processes such as infiltration, evapotranspiration, and runoff generation (Rahmati et al. 2016b). In the post-processing part of results, after preparing error matrix, accuracy of classification results was done based on overall accuracy, kappa coefficient, producers accuracy, and user accuracy. As shown in Table 1, residential areas with a FR of 2.56 have a great influence on flood in the study area (Fig. 5).

Geology

Petrology can be considered as an important factor in hydrology and sedimentation field of drainage. A petrologic map was prepared based on geological eras in four classes of Cenozoic, Mesozoic, Paleozoic, and Proterozoic (Fig. 5). As shown in Table 1, Paleozoic rocks have the most impact and the Cenozoic rocks have the least impact on flood in the study area.

Topographic wetness index

TWI indicates the amount of flow accumulation in each area from a drainage basin and the trend of water to downward slope by the force of gravity. TWI map was prepared in four different classes using quantile classification (Fig. 5). As shown in Table 1, 11.5–21.52 class with a FR of 1.37 has the greatest impact, and the class of less than 9.25 with a the FR of 0.55 has the least impact on flood in the study area.

Stream power index

SWI shows the flow power in terms of erosion. SWI map was prepared in four different classes using quantile classification (Fig. 5). As shown in Table 1, the class that is more than 29,482.15 with a FR of 1.42 has the greatest impact and a class of 0–250 with a FR of 0.78 has the least impact on flooding.

Flood map results using frequency ratio model

After weighing each effective factor in flood, the weights were finally summed up in the ArcGIS environment and the final FSM was prepared and classified into four categories of low, moderate, high, and very high susceptibility (Fig. 6) using quantile method (Rahmati et al., 2015; Zabihi et al. 2015). Based on the results obtained (Table 2), 19.95% of the area measuring 2780.83km^2 were placed with a high sensitivity to flood.

FR was used to determine the correlation of between flood points and effective factors. The results of the significant relationship between flood locations and effective factors using FR were presented in Table 1. The ratio one shows a moderate relationship between the flood points and the effective factors (Pradhan et al. 2011). If the value of the ratio is greater than one, there is high correlation, and if less than one, there is low correlation between flood points and effective factors (Lee

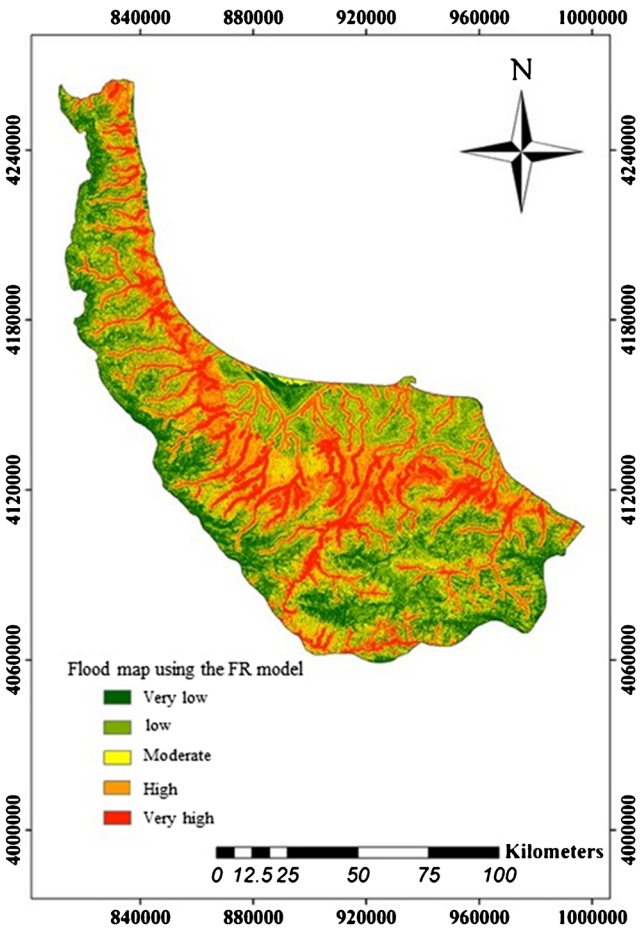

Fig. 6 Flood susceptibility map using FR model

Table 2 The number and percentage of floods in FSM using FR model

Flood susceptibility categories	Area (km^2)	Flood number	Flood percentage	Frequency ratio
Low	3590.13	15	9.74	0.38
Moderate	4944.51	30	19.48	0.55
High	3811.51	48	31.17	1.14
Very high	1594.55	61	39.61	3.46

et al. 2012). For altitude, zero to 280 class with a FR of 1.96 and the class greater than 1558 with a FR of 0.43 had the highest and lowest effect on flooding in the study area, respectively. This is due to the fact that floods are often formed in low altitude areas, while flood formation is impossible in high altitude areas (Rahmati et al. 2016b). Investigating the relationship between distance from rivers and flood events showed that distance of less than 500 m with a FR of 2.45 had a significant effect on flooding in the study area. Also, the class distance between 500–1000 m had a FR of more than one, which indicated the wide coverage of alluvial flood in the Gilan Province (Tehrany et al. 2013). Also, the class of more than 3000 m from rivers with a FR of 0.34 had the lowest effect on flooding in the study area.

In the case of land use, the analysis of the FR to residential areas was 2.56, which showed a strong relationship with flood and residential land use in the study area. The main reason for this is that residential areas are often covered with impenetrable surfaces such as asphalt causing flooding. This finding is consistent with those of other studies (Lee et al. 2012; Rahmati et al. 2016b; Tehrany et al. 2014, 2015) and suggests that other land uses should not be changed to residential land use in the study area and watershed areas. The FRs obtained for forest, pasture, and farming land uses were 0.91, 0.89, and 0.71, respectively. The lower FR of pasture to the forest can be attributed to their high height, which is consistent with Rahmati et al. (2016b) results. Due to the presence of tree cover, forest land use is able to stop flood and reduces runoff generation and has a much lower FR than agricultural land use. As a result, it has a lower impact on flooding in the study area than the agricultural and residential land uses. For geology factor, while the Paleozoic formation forms less than 2% of the study area, they have the greatest effect on flooding. The stones of Mesozoic and Paleozoic formations with a FR of more than one have the most effect on flooding and the stones of Cenozoic and Proterozoic areas with a FR of less than one had the least effect on flooding. This result is consistent with Khosravi et al. (2016) result. The integration of the geology and river maps of the study area showed that due to the small area of Paleozoic rocks, many rivers flow through it, and the protozoan rocks are mostly found in mountainous areas with low flood occurrence. The FR of drainage density factor for different classes of less than 1.36, 1.36–2.11, 2.11–34.13, and 34.13 to 49.17 km^2 were 0.91, 0.93, 1.01, and 1.15, respectively. With increasing drainage density, the FR increases, which has been emphasized in various studies (Cevik and Topal 2003; Nagarajan et al. 2000; Pachauri et al. 1998). Once drainage density increases, penetration decreases and surface runoff increases. As a result, with increasing drainage density, the risk of flooding increases in the lower lands. The results showed that the FRs for the south, southeast, and northeastern directions were more than one and therefore, they had the most impact on flooding in the study area. For the other six directions, the FR was less than one and the direction to the southwest with a FR of 0.6 had the least impact on flooding. These results are different with the results of Rahmati et al. (2016b). The reason for this mismatch can be climatic, geomorphic, and geological differences in two study areas. Slope angel factor can be used as another indicator to prove the negative relationship between high altitude and flooding. The lowest FR was 0.53 for class of 20.72–76.01. The highest FR was obtained for class of 2.36–6.21. As can be seen from the Table 1, flooding occurred in areas with high slope, but in lower areas, it occurred more frequently. The study of plan curvature indicated that the highest FR was equal to 1.14 for smooth surfaces with the most impact on flooding. The reason for this can be proved by the natural properties of flooding, where flooding occurs more often in flat areas. Overall, according to Lee et al. (2012), Tehrany et al. (2014), (2015), and Rahmati et al. (2016b), the most susceptible areas to flooding were the areas with lowest elevation, minimum slope angle, flat area, and close to rivers. The lowest FR is related to convex ranges.

Also, the study of the profile curvature factor, such as plan curvature, showed that it had the most FR relative to the smooth surfaces, and the convex and concave ranges with the same FR had the least impact on flooding. A survey of TWI indicated that the highest FR was equal to 1.37 for the last class (11.52–21.5) and the lowest FR was equal to 0.55 for the first class (less than 25.9). The more the TWI values, the higher the flood volume; this is due to the conditions of soil saturation. The more saturated the soil, the greater the flooding.

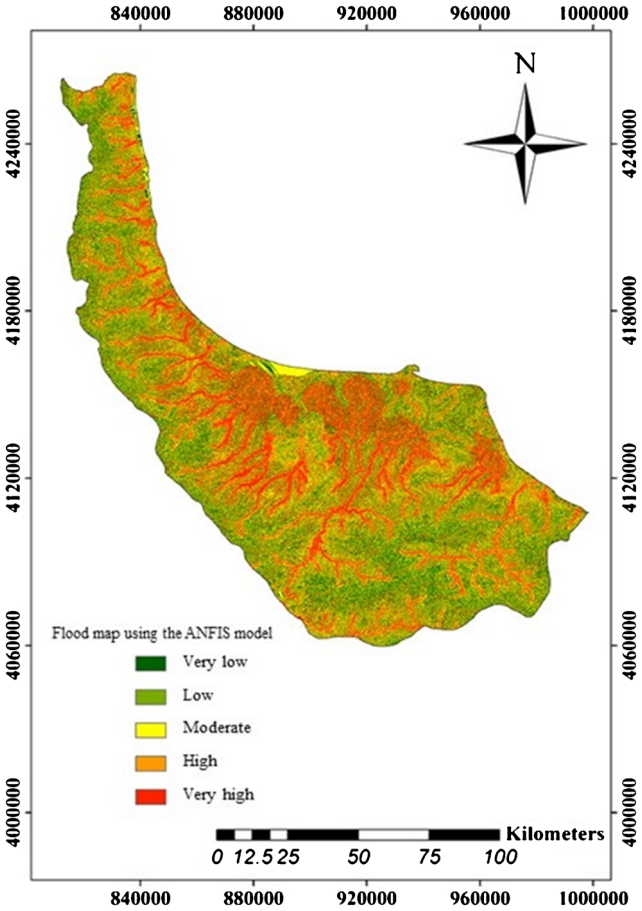

Fig. 7 Flood susceptibility map using ANFIS model

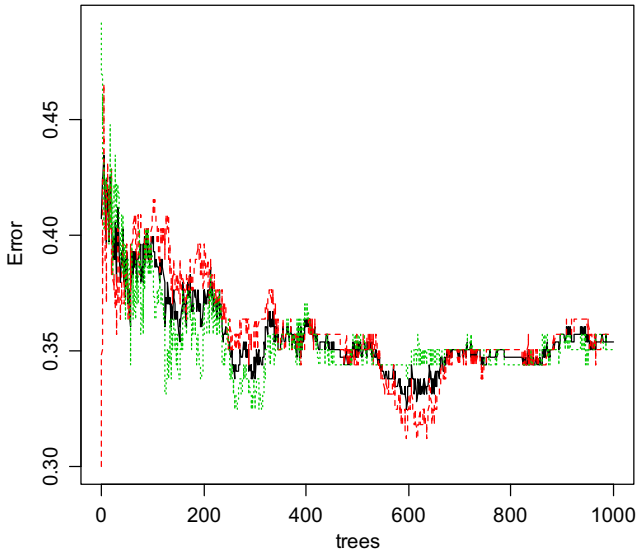

Fig. 8 The overall error rate for RF model (OOB (black line); no flood (red line); flood (green line)

In the case of the SPI, the class higher than 15.29482 with a FR of 42.1 and the class 0.250 with a FR of 0.78 had the highest and lowest impact on flooding, respectively. SPI shows the flow power in terms of erosion (Tehrany et al. 2015). As a result, the higher the flow power, the more flooding.

ANFIS

Data standardization results

FR were standardized for classes of each effective factor using Eq. (2). This equation indicates the standardization between two values of zero to one, which is presented in the last column of the Table 1.

Subtractive clustering results

After normalizing the weights obtained from the FR model, the map of each factor affecting flood, such as slope angle, aspect, altitude, distance from rivers, drainage density, geology, land use, plan curvature, profile curvature, TWI, and SPI in ArcGIS environment were prepared based on standardized weights.

In the subtractive clustering, with a change in the range of influence from 0.4 to 0.8, with steps of 0.01 in different repetitions, considering the lowest error in the testing stage, the optimal range of influence was determined 0.52 by trial and error.

Flood map results using ANFIS model

After determining the optimal range of influence and repetition, the ANFIS model was performed and then all the maps of the effective factors were normalized, and entered into Matlab 2014 software pixel-to-pixel. The ANFIS model was implemented and the weight of each pixel was obtained and finally a flood map was provided (Fig. 7), classified based on quantile method. By studying Table 3, it is obvious that the moderate class has the largest area percentage (34.45%),

Table 3 The number and percentage of floods in FSM using ANFIS model

Flood susceptibility categories	Area (km^2)	Flood number	Flood percentage	Frequency ratio
Low	3648.98	11	7.14	0.27
Moderate	4802.12	31	20.13	0.58
High	3704.21	43	27.92	1.05
Very high	1785.42	69	44.81	3.50

Table 4 Error matrix of RF model

	Observed	Predicted		Error
		Without Flood	Flood	
Without flood	154	97	57	0.37
Flood	154	54	100	0.35

followed by high (26.57%), low (26.17%), and very high (12.81%) classes.

Random forest model results

RF model is another decision tree technique in this study. As mentioned, for the implementation of this technique, the "randomForest" package was used in the R software. First, based on the error rate diagram, the random variables were determined using the caret packet, the optimal number of random variables for splitting each node was determined, and then with the help of this number of random variable, the error rate diagram was drawn in terms of the number of trees, and decision was made regarding the number of optimal trees. Accordingly, the final RF model with 3 random variables (mtry) was implemented for each node and 1000 trees (ntree).

RF algorithm is based on a handful of decision trees and is currently one of the best learning algorithms. The OOB (Out-Of-Bag) was used to evaluate the performance of the model. As shown in Fig. 8, OOB is a function of trees and is reduced when more trees are added to the RF algorithm.

The performance results of the RF algorithm are presented in Table 4 and Fig. 8. OOB and mean squared error (MSE) values were calculated as 0.36 and 0.29, respectively. Considering OOB and MSE values, flood modeling of the study area can be carried out well. As can be seen from the Table 4, out of 154 flood pixels, 54 pixels were mistakenly predicted without flood, 100 pixels were correctly predicted, and among 154 pixels without flood, 97 pixels were properly predicted without flood, and 57 pixels were mistakenly predicted as flood.

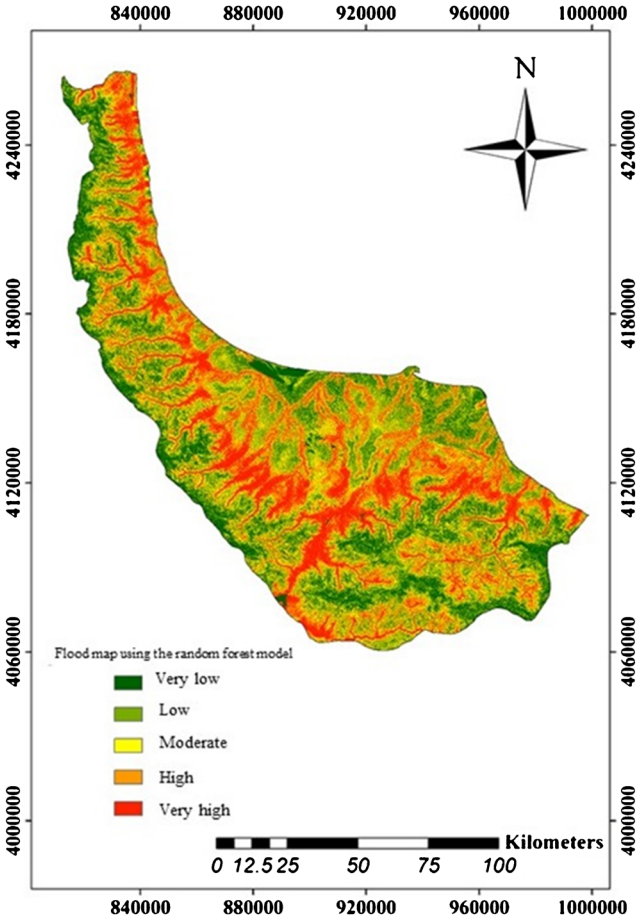

Fig. 9 Flood susceptibility map using RF model

Flood susceptibility map results using RF model

After preparing the effective factor maps, the flood inventory map containing 154 flood points and 154 non-flood points were discontinued and the value of each of the effective factors was obtained for each point. Then, the data were entered into the R software, using the RF package, using the quantile classification method; it was divided into four classes of low, moderate, high, and very high sensitivity. By studying Table 5, it is obvious that the moderate class has the largest area percentage (25.47%), followed by high (25.36%), very high (25.11%), and low (24.06%) classes (Fig. 9).

Table 5 The number and percentage of floods in FSM using RF model

Flood susceptibility categories	Area (km^2)	Flood number	Flood percentage	Frequency ratio
Low	3354.12	0	0.00	0.00
Moderate	3550.27	2	1.30	0.05
High	3535.27	8	5.19	0.20
Very high	3501.15	144	93.51	3.72

Table 6 Accuracy mean decrease and mean decrease Gini of effective flood factors in RF model for the preparation of FSM in the Gilan Province

Factor	Flood	Without flood	Accuracy mean decrease	Mean decrease Gini
Topographic wetness index	4.41	0.37	3.61	10.51
Aspect	− 4.54	6.30	1.35	9.72
Altitude (m)	17.88	4.78	17.79	15.42
Land use	10.35	− 4.52	5.52	2.45
Lithology	6.97	− 1.62	4.58	1.94
Plan curvature	− 3.72	1.75	− 1.37	7.66
Profile curvature	1.98	3.06	3.62	9.03
Drainage density (km/km^2)	1.57	6.57	5.93	11.12
Distance from river (m)	17.76	20.40	26.64	18.75
Stream power index	− 4.52	1.35	− 1.20	1.15
Slope	4.39	− 2.13	1.93	10.64

Prioritizing effective factors in flood

The mean of Gini reduction was used to determine the priority of effective flood factors (Table 6).

As can be seen from the Table 6, distance from rivers, altitude, and drainage density is the most important factor for flood and the least important factor is geology. The results of Tehrany et al. (2015) found that altitude and slope angle factors were the most important factors affecting on flooding. Therefore, the importance of variables in the flood map was influenced by the method used in the research and the characteristics of the study area. In other words, different geological conditions, topography, and weather in a region can change the priority of the factors affecting the flood mapping.

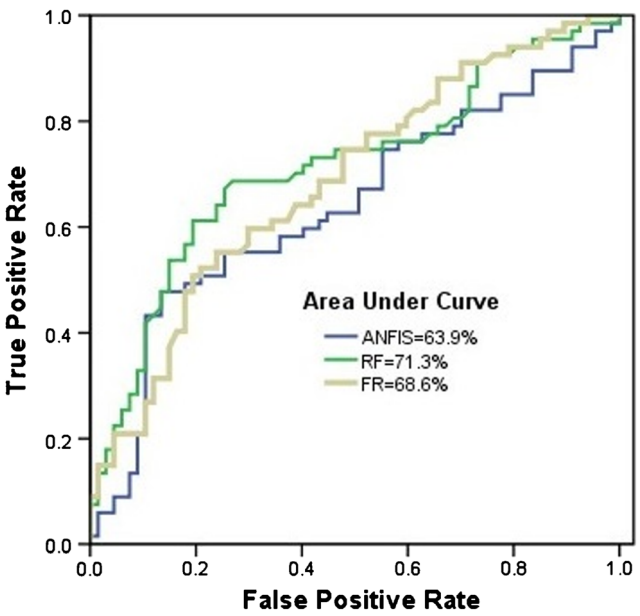

Fig. 10 ROC curve for three models

Evaluation of the model efficiency

One of the most methods for determining accuracy and evaluation of models is the ROC curve. In this study, 30% of the floods were used for validation stage. The results of evaluation of FR, ANFIS, and RF using ROC curve (Fig. 10) showed that the obtained AUC for models were 68.6, 63.9, and 71.3%, respectively.

Rahmati and Pourghasemi (2017) compared the evidential belief function (EBF), RF, and boosted regression trees (BRT) models for mapping flood susceptibility in the Galikesh region, Iran. They found that the EBF and BRT models with the AUC values of78.67% and 78.22%, respectively, were superior to the RF model with an AUC value of 73.33%. Although the EBF and BRT models were outperformed than the RF model, There are several advantages for the RF model over other models, as it can work with data lost in calibration and validation data, and because of its hybrid design, it can estimate even when some inputs are lost. In addition, in the modeling process, the importance of the factors are identified which can be useful (Ball et al. 2009). Another benefit of RF model is able to model non-linear relationships between the flood occurrence and related conditioning factors. Also, the RF model do not need to check statistical assumptions (e.g., outliers and data normalization). As shown in Fig. 10, the RF model with an AUC value of 71.3% had a good performance and the FR and ANFIS models with AUC values of 68.6% and 63.9%, respectively, had a moderate performance in providing FSM. This study shows that the performance of FR and ANFIS models is similar, whereas Razavi Termeh et al. 2018) indicated that a combination of ANFIS with metaheuristic optimization was superior to the FR model. In addition, Tien Bui et al. (2018) proposed three new hybrid artificial intelligence optimization models, namely, ANFIS with cultural (ANFIS-CA), bees (ANFIS-BA), and invasive weed optimization (ANFIS-IWO) algorithms for FSM in the Haraz Watershed, Iran. They found that the AUC values of

ANFIS-CA, ANFIS-BA, and ANFIS-IWO were 92.1%, 93.9%, and 94.4%, respectively.

The current research like other studies has some limitation. Most importantly, non-flood locations were randomly selected in the study area, the locations are considered as non-flood locations that did not have flood occurrence (Chen et al. 2019a). Thus, the new methods must be applied to select reliable non-flood locations. The generated flood map can be used in management and enforcement work to prevent and reduce floods in the future.

Conclusions

The results of this study showed that by integrating geographic information system (GIS) and factors influencing flood conditions and the historical data of recorded flood, the interaction between factors influencing watershed floods can be studied. In general, GIS is one of the most powerful tools for analyzing and displaying spatial data in the management of watersheds. For this purpose, eleven factors were used as effective factors for FSM. Among the effective factors on flood, the physical characteristics of the basin are very important due to the high stability and low variability. These properties directly affect the hydrological regime and indirectly affect the climate of the area. The prioritization of effective factors using the mean decrease Gini showed that distance from rivers, altitude, and drainage density had the most effect, respectively, and SPI and geology had the least effect on flooding in the study area. Meanwhile, FSM with RF and ANFIS models and its accuracy evaluation using ROC revealed a reasonable accuracy of the RF model (71.3%) compared with ANFIS model (63.9%) in the study area. Therefore, RF model is suitable for identifying areas with flood susceptibility. Finally, use of RF model is proposed for FSM, especially in developing countries.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Alvarado-Aguilar D, Jiménez JA, Nicholls RJ (2012) Flood hazard and damage assessment in the Ebro Delta (NW Mediterranean) to relative sea level rise. Nat Hazards 62:1301–1321

Ball RL, Tissot P, Zimmer B, Sterba-Boatwright B (2009) Comparison of random forest, artificial neural network, and multi-linear regression: a water temperature prediction case. In: Seventh Conference on Artificial Intelligence and its Applications to the Environmental Sciences. New Orleans, LA

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. Hydrol Sci J 24:43–69

Billa L, Shattri M, Rodzi Mahmud A, Halim Ghazali A (2006) Comprehensive planning and the role of SDSS in flood disaster management in Malaysia. Disaster Prevent Manag: Int J 15:233–240

Breiman L (2001) Random forests. Mach Learn 45:5–32

Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Nat Hazards Earth Syst Sci 13:2815–2831

Cevik E, Topal T (2003) GIS-based landslide susceptibility mapping for a problematic segment of the natural gas pipeline, Hendek (Turkey). Environ Geol 44:949–962

Chen W, Shahabi H, Zhang S, Khosravi K, Shirzadi A, Chapi K, Pham BT, Zhang T, Zhang L, Chai H, Ma J, Chen Y, Wang X, Li R, Ahmad BB (2018a) Landslide susceptibility modeling based on GIS and novel bagging-based kernel logistic. Regression Appl Sci 8:2540

Chen W, Shahabi H, Shirzadi A, Hong H, Akgun A, Tian Y, Liu J, Zhu AX, Li S (2018b) Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. Bull Eng Geol Environ 1-23

Chen W, Zhang S, Li R, Shahabi H (2018c) Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. Sci Total Environ 644:1006–1018

Chen W, Hong H, Li S, Shahabi H, Wang Y, Wang X, Ahmad BB (2019a) Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. J Hydrol 575:864–873

Chen W, Panahi M, Tsangaratos P, Shahabi H, Ilia I, Panahi S, Li S, Jaafari A, Ahmad BB (2019b) Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. Catena 172:212–231

Chen W, Pradhan B, Li S, Shahabi H, Rizeei HM, Hou E, Wang S (2019c) Novel hybrid integration approach of bagging-based fisher's linear discriminant function for groundwater potential analysis. Nat Resources Res 1-20

Chen W, Zhao X, Shahabi H, Shirzadi A, Khosravi K, Chai H, Zhang S, Zhang L, Ma J, Chen Y, Wang X (2019d) Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. Geocarto Int 1-25

Costache R, Tien Bui D (2019) 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

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. Ecology 88: 2783–2792

Dang NM, Babel MS, Luong HT (2011) Evaluation of food risk parameters in the day river flood diversion area, Red River delta. Vietnam Nat Hazards 56:169–194

Darabi H, Choubin B, Rahmati O, Haghighi AT, Pradhan B, Kløve B (2019) Urban flood risk mapping using the GARP and QUEST models: a comparative study of machine learning techniques. J Hydrol 569:142–154

Erdirencelebi D, Yalpir S (2011) Adaptive network fuzzy inference system modeling for the input selection and prediction of anaerobic digestion effluent quality. Appl Math Model 35:3821–3832

Falah F, Rahmati O, Rostami M, Ahmadisharaf E, Daliakopoulos IN, Pourghasemi HR (2019) Artificial neural networks for flood susceptibility mapping in data-scarce urban areas. In: Spatial modeling in GIS and R for earth and environmental sciences. Elsevier, pp 323-336

Glenn EP, Morino K, Nagler PL, Murray RS, Pearlstein S, Hultine KR (2012) Roles of saltcedar (Tamarix spp.) and capillary rise in

salinizing a non-flooding terrace on a flow-regulated desert river. J Arid Environ 79:56–65

Haghizadeh A, Siahkamari S, Haghiabi AH, Rahmati O (2017) Forecasting flood-prone areas using Shannon's entropy model. J Earth Syst Sci 126:39

Huang X, Tan H, Zhou J, Yang T, Benjamin A, Wen SW, Li S, Liu A, Li X, Fen S, Li X (2008) Flood hazard in Hunan province of China: an economic loss analysis. Nat Hazards 47:65–73

Khosravi K, Nohani E, Maroufinia E, Pourghasemi HR (2016) A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Nat Hazards 83:947–987

Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, Dou J, Ly HB, Gróf G, Ho HL, Hong H (2019) A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. J Hydrol 573:311–323

Kisi O, Shiri J, Tombul M (2013) Modeling rainfall-runoff process using soft computing techniques. Comput Geosci 51:108–117

Lee M-J, Kang J-E, Jeon S (2012) Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS. In: Geoscience and remote sensing symposium (IGARSS), 2012 IEEE International. IEEE, pp 895-898

Mind'je R, Li L, Amanambu AC, Nahayo L, Nsengiyumva JB, Gasirabo A, Mindje M (2019) Flood susceptibility modeling and hazard perception in Rwanda. Int J Disaster Risk Reduct 38:101211

Mojaddadi H, Pradhan B, Nampak H, Ahmad N, Ghazali AHB (2017) Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS Geomatics. Nat Hazards Risk 8:1080–1102

Moore ID, Grayson R, Ladson A (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol Process 5:3–30

Nagarajan R, Roy A, Kumar RV, Mukherjee A, Khire M (2000) Landslide hazard susceptibility mapping based on terrain and climatic factors for tropical monsoon regions. Bull Eng Geol Environ 58:275–287

Naghibi SA, Pourghasemi HR (2015) A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. Water Resour Manag 29:5217–5236

Naghibi SA, Pourghasemi HR, Dixon B (2016) GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ Monit Assess 188:44

Opolot E (2013) Application of remote sensing and geographical information systems in flood management: a review. Res J Appl Sci Eng Technol 6:1884–1894

Pachauri A, Gupta P, Chander R (1998) Landslide zoning in a part of the Garhwal Himalayas. Environ Geol 36:325–334

Pourghasemi HR, Kerle N (2016) Random forests and evidential belief function-based landslide susceptibility assessment in western Mazandaran province, Iran. Environ Earth Sci 75:185

Pourghasemi HR, Pradhan B, Gokceoglu C (2012) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed. Iran Natural Hazards 63:965–996

Pradhan B, Mansor S, Pirasteh S, Buchroithner MF (2011) Landslide hazard and risk analyses at a landslide prone catchment area using statistical based geospatial model. Int J Remote Sens 32:4075–4087

Rahmati O, Pourghasemi HR (2017) 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

Rahmati O, Pourghasemi HR, Melesse AM (2016a) Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. Catena 137:360–372

Rahmati O, Pourghasemi HR, Zeinivand H (2016b) Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan province, Iran. Geocarto Int 31:42–70

Rahmati O, Zeinivand H, Besharat M (2016c) Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis. Geomatics Nat Hazards Risk 7:1000–1017

Rahmati O, Samani AN, Mahdavi M, Pourghasemi HR, Zeinivand H (2015) Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. Arab J Geosci 8:7059–7071

Razavi Termeh SV, Kornejady A, Pourghasemi HR, Keesstra S (2018) Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. Sci Total Environ 615:438–451

Shafizadeh-Moghadam H, Valavi R, Shahabi H, Chapi K, Shirzadi A (2018) Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. J Environ Manag 217:1–11

Siahkamari S, Haghizadeh A, Zeinivand H, Tahmasebipour N, Rahmati O (2018) Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. Geocarto Int 33:927–941

Sidle RC, Ochiai H (2006) Landslides: processes, prediction, and land use Water Resources Monograph Series

Tehrany MS, Pradhan B, Jebur MN (2013) 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

Tehrany MS, Pradhan B, Jebur MN (2014) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J Hydrol 512:332–343

Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena 125:91–101

Tien Bui D, Khosravi K, Li S, Shahabi H, Panahi M, Singh V, Chapi K, Shirzadi A, Panahi S, Chen W, Bin Ahmad B (2018) New hybrids of anfis with several optimization algorithms for flood susceptibility modeling. Water 10:1210

Vafakhah M (2012) Application of artificial neural networks and adaptive neuro-fuzzy inference system models to short-term streamflow forecasting. Can J Civ Eng 39:402–414

Vafakhah M, Kahneh E (2016) A comparative assessment of adaptive neuro-fuzzy inference system, artificial neural network and regression for modelling stage-discharge relationship. Int J Hydrol Sci Technol 6:143–159

Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2016a) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin. Asir Region, Saudi Arabia Landslides 13:839–856

Youssef AM, Pradhan B, Sefry SA (2016b) Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. Environ Earth Sci 75:12

Zabihi M, Pourghasemi HR, Behzadfar M (2015) Groundwater potential mapping using Shannon's entropy and random forest models in the Bojnourd Plain. Iranian Journal of Eco-hydrology 2(2):221–232