

LANDSLIDE HAZARD MAPPING USING A RADIAL BASIS FUNCTION NEURAL NETWORK MODEL: A CASE STUDY IN SEMIROM, ISFAHAN, IRAN

H. Yavari¹, P. Pahlavani^{1, *}, B. Bigdeli²

¹ School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran –
(hamidyavari72, pahlavani)@ut.ac.ir

² School of Civil Engineering, Shahrood University of Technology, Shahrood, Iran - bigdeli@shahroodut.ac.ir

KEY WORDS: Landslide, RBF neural network, Logistic Regression

ABSTRACT:

In this paper, Radial Basis Function (RBF) Neural Network and Logistic Regression (LR) models were proposed for hazard prediction of landslides in a part of the Semirom area (Iran) to compare their accuracy and performance. For this purpose, a spatial database of the study area was prepared that consists of 68 landslide locations and 11 influencing information layers including slope, aspect, profile curvature, plan curvature, distance from faults, distance from roads, distance from residential regions, distance from rivers, land use, lithology and rainfall. Landslide hazard maps were prepared for the study area by applying the proposed algorithms. Performance of the models was assessed using the Receiver Operating Characteristic (ROC) curve and area under the ROC curve (AUC). The coefficient of determination (R^2), the root mean square error (RMSE), and the Normal Root Mean Square Error (NRMSE) were calculated for proposed methods. The outcomes showed that the RBF Neural Network has the highest R^2 (0.8224), in comparison to that of the LR model (0.5365). Also, the ROC plots, RMSEs and NRMSEs showed that the proposed RBF Neural Network is much better than the LR model. Consequently, it can be concluded that the RBF Neural Network is the best regression model in this study and it can be considered as a capable method for landslide hazard mapping in landslide-susceptible areas.

1. INTRODUCTION

One of the most sensitive and vital issues in the development projects, such as selecting the highway route and any development of mining and construction, depends on studying the sustainability of the land of the desired project. Most of Iran's area is mountainous and hills. This mainly mountainy topography with active tectonic, various environmental and climatic conditions, tend to provide appropriate situations for frequent landslides. Landslides in most provinces of the country cause life and financial damage every year. In the landslide hazard zonation, the earth's surface is divided into distinct areas of no degree to high degree of landslide risk. The procedure of preparing the zonation map is based on the recognition of natural features and quantitative modeling based on the data of the studied area.

Geospatial information system and geostatistical analyses have been widely applied in analyzing different field of geo-related topics (Delavar, 2004; Pahlavani et al., 2006; Pahlavani et al., 2017; Bahari et al., 2014). In this regard, applying statistical approaches to forecast landslide hazard has generally been seen since the 1990s. The approaches that have been used to generate a landslide hazard zonation map divided into four classes (Youssef et al., 2016): Heuristic, Deterministic, Statistical and combination of the cited approaches.

Several spatial layers impress the process of preparing the landslide hazard map. Some researchers classify the landslides effective layers into two classes of internal and external

factors (Milaghardan et al., 2016). Internal factors do not change over the time, such as slope, elevation, aspect and so on, on the other hand, external factors are dependent on time and not predictable. Some of external factors are earthquake, rainfall and so on.

Pourghasemi et al. (2012) applied a conditional probability model and index of entropy to predict landslide hazard for the Safarood region. The results of the index of entropy illuminated that the height and land use are the most important factors in the landslide hazard zonation. In the conditional probability model, effective layers must be independent, but it can determine the impact of each layer separately. Based on their research, index of entropy has more accuracy than conditional probability model (Pourghasemi et al., 2012).

Catani et al. (2013) used random forest technique to assess the effect of the map scale on the map of landslide hazard zonation and its accuracy (Catani et al., 2013).

Elkadiri et al. (2014) applied logistic regression and artificial neural network to prepare a landslide hazard map. Based on their research, artificial neural network affected by more factors than logistic regression and is more vulnerable (Elkadiri et al., n.d.).

Pradhan et al. (2015) applied alternating decision tree, two-class kernel logistic regression and support vector machines to study the landslide hazard in the Yihuang area of China. Based on their research, the results of alternating decision tree was

* Corresponding author

much better than the two other approaches. The most effective factor in all three methods was slope layer (Hong et al., 2015).

Pradhan et al. (2016) studied the landslide susceptibility by combination of two models. First, they used shallow landslide stability as a deterministic model and spatial multi-criteria evaluation as a probabilistic model and prepared two landslide susceptibility maps separately. Then these two maps were combined with the frequency ratio method, and concluded that accuracy has increased (Pradhan and kim, 2016).

Haoyuan Hong et al. (2018) applied a new approach that was based on region-partitioning to improve the accuracy of landslide susceptibility mapping. Based on their research, separating the study region into two classes using the proposed method, improved the prediction rate from 0.77 to 0.85 when SVM (Support Vector Machine) was used, and from 0.87 to 0.88 when LR model was used. They also use GWR (Geographically Weighted Regression) to consider the spatial distribution of landslide locations for predicting the landslide. In their study, LR and SVM models performed better than the GWR model, because multicollinearity is significantly stronger in the GWR model than global regression models (Hong et al., 2018).

He et al. (2019) used Naïve Bayes, radial basis function Classifier and RBF Network for landslide hazard zonation in Longhai, located in China. The results of these models were compared and validated using AUC and Friedman and Wilcoxon signed-rank tests. The results of this study revealed that the RBF Classifier model is a reliable method for landslide hazard prediction (He et al., 2019).

Chen et al. (2019) used an innovative ensemble data mining technique involving an Adaptive Neuro-Fuzzy Inference System (ANFIS) optimized by Shuffled Frog Leaping Algorithm (SFLA) and Particle Swarm Optimization (PSO) for landslide susceptibility spatial modeling. They used 18 effective information layers. The produced landslide susceptibility maps were evaluated by the ROC curves. The results showed that based on the training and testing dataset the AUC value was similar. However, the processing time during the training and implementation phase was considerably different. SWARA-ANFIS-PSO appeared to be six times faster with respect to the processing time attained by SWARA-ANFIS-SFLA (Chen et al., 2019).

Mukhiddin Juliev et al. (2019) compared Statistical Index (SI), Frequency Ratio (FR) and Certainty Factor (CF) for landslide susceptibility mapping. Using assessment methods, they concluded that the statistical index method has the best performance in their study (Juliev et al., 2019).

Nsengiyumva et al. (2019) used weights of evidence, logistic regression, frequency ratio and statistical index to produce susceptibility maps for Rwanda area in Africa. Using the ROC curves the produced susceptibility maps were validated. The weights of evidence has the highest AUC value while the statistical index attained the lowest AUC value (Nsengiyumva et al., 2019).

This study aims to implement and compare one regular regression method called LR with one of the Neural Network regressions family methods called RBF Neural Network. Furthermore, to estimate the accuracy and prediction ability of the models, a validation analysis was executed. Section 2 describes the research methods that consist of RBF Neural Network and LR model and evaluation methods. In section 3, the study area is describe. In section 4, the effective

information layers that used in this research is described. Section 5 shows the results of applying the proposed methods. Section 6 explains the results and at last, section 7 explains a conclusion of this study.

2. RESEARCH METHOD

In this research, because of the presence of multiple spatial layers in the Landslides hazard determining, we identify and use independent spatial layers that have correlation coefficients between -0.6 and 0.6. Then to identify and determine the effect of spatial layers on the landslide, RBF Neural Network and Logistic Regression have been used and the accuracy of these two algorithms have been compared.

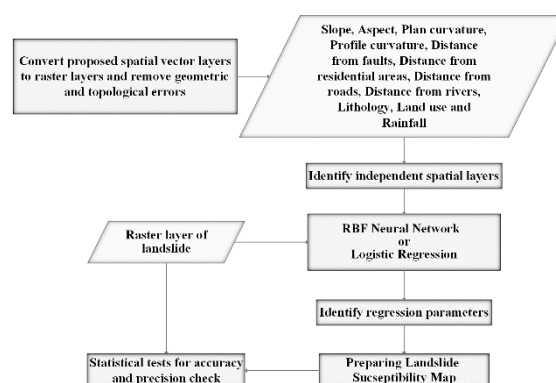


Figure 1. Research procedure

2.1 RBF Neural Network

Radial Basis Function (RBF) Neural Network is a kind of (non)linear neural networks. The RBF Neural Network is a hidden layer neural network that consists of three layers, the input layer, hidden layer, and output layer. The input layer transmits the elements of the input vector to each units in the hidden layer. Each unit in the hidden layer then produces an activation based on the associated RBF. Finally, the output layer calculates a linear combination of the activations of the hidden units.

In the classification case, the output of the model learning of the RBF Neural Network for the input pattern x is as follow (Du and Swamy, 2014):

$$f_i(x) = \sum_{k=1}^m w_{ki} \theta(\|x - \alpha_k\|) \quad (1)$$

, where m is the number of computing units, w_{ki} is the connecting weights between hidden layer and output layer, α_k is the RBF centers, and the function of θ is a Gaussian function. To select the primary hidden unit centers, random selection is used on the training dataset. Furthermore, the primary value of all variance parameter(s) in the network is set to the maximum squared Euclidean distance between any pair of cluster centers.

2.2 Logistic Regression

In logistic regression algorithm (Cox, 1958), the relationship between a dependent and independent variables is calculated using the cumulative logistic distribution as follows:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (2)$$

$$y = \log_e \left(\frac{p}{1-p} \right) \quad (3)$$

$$P = \frac{e^y}{1+e^y} \quad (4)$$

where y is the dependent variable, b_0 is the intercept of the model, n is the number of landslides effective factors, b_i is the weight of each independent factor, x is the independent (landslide effective) factor, and P is the probability of landslide occurrence.

2.3 Evaluation methods

Using the Coefficient of Determination (R^2) is a customary manner to determine the accuracy of regression results which is calculated as follows:

$$R^2 = 1 - \frac{SS_E}{SS_T} \quad (5)$$

$$SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (7)$$

In the above relations, n is the number of observations, y_i is the i^{th} observation, \hat{y}_i is estimated value for i^{th} observation and \bar{y} is the mean of observations.

There are other scientific calculations that can indicate the accuracy of the regression methods. The Root Mean Square Error (RMSE) and Normal Root Mean Square Error (NRMSE) are operational in determining the accuracy.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2} \quad (8)$$

$$NRMSE = RMSE / \sigma_{\hat{y}} \quad (9)$$

3. STUDY AREA LOCATION

The study area location is in Semirom province (Iran), between $30^\circ 44' 59''$ and $31^\circ 7' 30''$ north latitude and $51^\circ 29' 52''$ and $51^\circ 45' 10''$ east longitude. This region is about 60km far from the Semirom city. The area of the study region is about 990 square kilometers and includes the Marbor river basin.

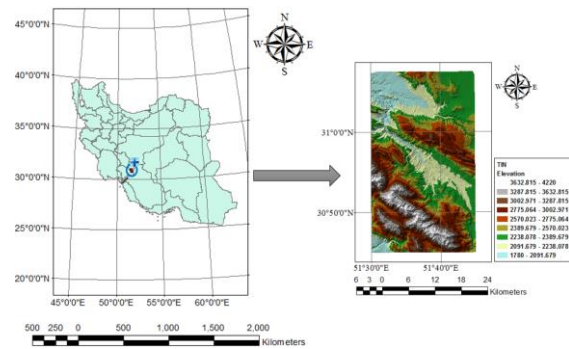


Figure 2. Study area

4. EFFECTIVE INFORMATION LAYERS

In this research we prepare a spatial database of the study area that consists of 68 landslide locations and 11 influencing information layers. These layers including slope, aspect, profile curvature, plan curvature, land use, lithology and rainfall were prepared using the 1: 25000 maps of the Iran's surveying organization, the digital elevation model with 30-meter spatial resolution, the 1:100,000 geological map of the Iran's geological organization and the precipitation data of the Iran's meteorological organization for the desired region. Also, the landslides map of the region has been prepared from the 1:100,000 geological map. Using Euclidean distance on related layers, the maps of distance from faults, distance from roads, distance from residential regions and distance from rivers were prepared. In order to determine the rainfall, the average rainfall data in the last 10 years was used in 19 stations around the study area. Using collected data to specify the amount of rainfall in every location in the study area is not possible. For this purpose, rainfall rates for each point was interpolated using the collected data and position of each station and general kriging interpolation with an exponential semivariogram model and a resolution of 30-meter.

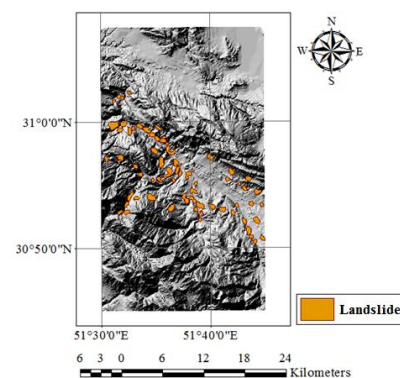
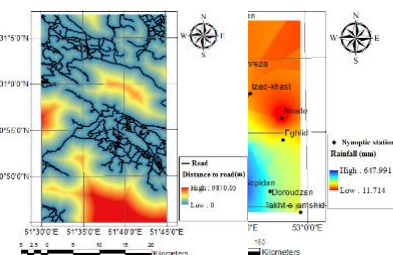


Figure 3. Landslide inventory map



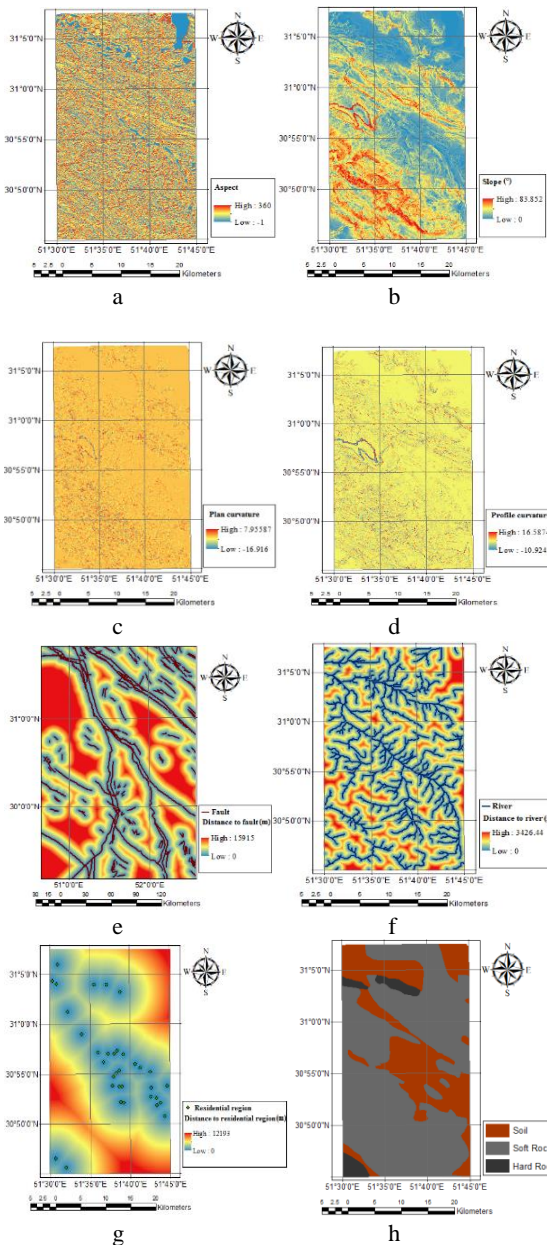


Figure 4. Effective information layers. . a Aspect, b Slope c Plan curvature, d profile curvature, e Faults and distance from faults, f Rivers and distance from rivers, g Residential regions and distance from residential regions, h Lithology

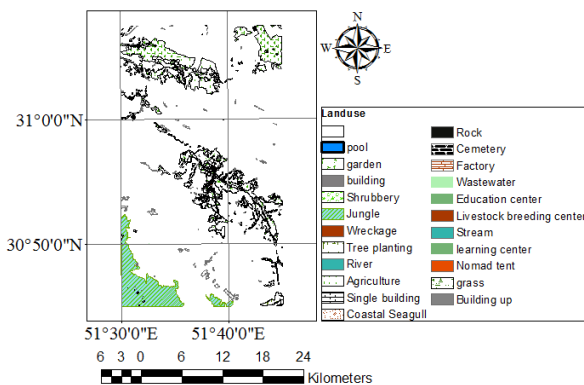


Figure 5. Effective information layers: Land use

Figure 6. Effective information layers: a Precipitation, b Roads and distance and roads

In this study, lithology layer is classified into three clusters of soils, soft rock and hard rock (Kamranzad et al., 2016).

5. EXPERIMENTAL RESULTS

In the first step of research procedure, the correlation coefficients between the information layers were calculated. To calculate the correlation coefficient, following relationships were used:

$$Cov(X,Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n} \quad (8)$$

$$r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad (9)$$

In the above relationships, X and Y are two data sets, \bar{X} and \bar{Y} are means of these two data sets, n is the number of each data set, COV is the covariance between two data sets, σ_X and σ_Y are the standard deviations and r is the correlation coefficient. In this study, layers that their correlation coefficients was between -0.6 and 0.6 were used.

number	Layer's name
1	Distance to residential places (m)
2	Distance to rivers (m)
3	Distance to roads (m)
4	Lithology
5	Plan curvature
6	Profile curvature
7	Aspect
8	Slope (degree)
9	Distance to faults (m)
10	Land use
11	Rainfall (mm)

Table1. Information layers

After applying the proposed methods, the following results were obtained that described in the table (2).

Regression Algorithm	RMSE	NRMSE	R^2
RBF Neural Network	0.2809	0.4774	0.8224
Logistic Regression	1.2505	1.4405	0.5365

Table 2: RMSE, NRMSE, R^2 values for different regression algorithms

After applying the proposed methods, the regression coefficients, the estimated values of the observations, and the residuals were specified. Using the estimated values of the

landslide for sample points, landslide hazard zonation map for each approach was prepared using Kriging interpolation method with exponential semivariogram model and resolution of 30-meter. The landslide hazard maps are categorized into five categories based on the natural break method: very low risk areas, low risk areas, moderate risk areas, high risk areas, and very high risk areas.

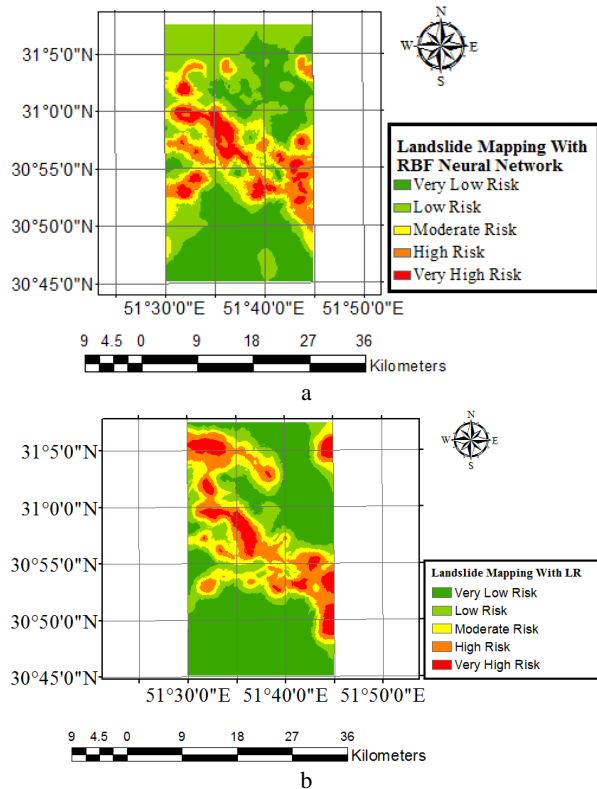


Figure 7. a Landslide hazard map with RBF Neural Network, b Landslide hazard map with LR

6. DISCUSSION

The results of the present study disclosed that the RBF neural network had significant prediction performance than the LR model. Also the amount of RMSE and NRMSE show the performance of the RBF neural network in the interpolation on the data of this study. For further accuracy assessment, the ROC curve and the area under the ROC curve were used. The AUC denotes the predictive power of the system based on its ability to estimate the true occurrence. In this research, the ROC curve is plotted using the primary and predicted data.

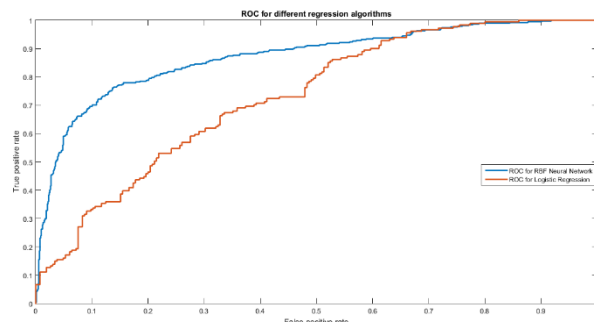


Figure 8. ROC curve to examine the accuracy of the methods

Regression Algorithm	RBF Neural Network	LR
AUC	0.8994	0.7207

Table 3. Comparison of the Area Under the ROC Curve in the various interpolation methods

As shown in Table (2) and (3), the RBF Neural Network in this study has far better results than the Logistic Regression algorithm. Table (4) compares the percentage of different parts of the zonation maps obtained by different methods.

Regression Algorithm	RBF Neural Network	LR
Area percentage		
Very Low Risk	31.13	48.29
Low Risk	33.51	16.96
Moderate Risk	16.96	15.98
High Risk	10.87	12.04
Very High Risk	7.53	6.72

Table 4: the percentage of different parts of the landslide hazard maps

7. CONCLUSION

Determining landslide-susceptible areas is an important phase in land use planning. An appropriate land use plan can help managers to decrease financial injuries and loss of life resulted by landslides. The current study evaluated and compared the performance of RBF neural network and Logistic regression models for their use in landslide hazard mapping. In the present research, in the first step, we tried to determine the effective factors in the landslide occurrence for Semirom of Isfahan province in Iran. Using the effective information layers and proposed methods, the landslide hazard maps were modeled. The results verifies that the RBF Neural Network is the best regression model in this study and it can be considered as a reliable method for landslide hazard mapping in landslide-susceptible areas. To increase the accuracy of landslide hazard mapping, it is suggested that more information layers such as altitude, soil thickness, groundwater depth, volcanic map, water flow strength index, slope length and topographic humidity index be used. This information was not used in this research due to lack of this information.

REFERENCES

Bahari, R.A., Abbaspour, R.A., Pahlavani, P., 2014, November. Prediction of PM2.5 concentrations using temperature inversion effects based on an artificial neural network. In The ISPRS international conference of Geospatial information research (Vol. 15, p. 17).

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. <https://doi.org/10.5194/nhess-13-2815-2013>

Chen, W., Panahi, M., Tsangaratos, P., Shahabi, H., Ilia, I., Panahi, S., Li, S., Jaafari, A., Ahmad, B.B., 2019. Applying population-based evolutionary algorithms and a neuro-fuzzy

- system for modeling landslide susceptibility. *CATENA* 172, 212–231. <https://doi.org/10.1016/j.catena.2018.08.025>
- Cox, D.R., 1958. The Regression Analysis of Binary Sequences. *J. R. Stat. Soc. Ser. B Methodol.* 20, 215–242.
- Delavar, M.R., Samadzadegan, F., Pahlavani, P., 2004. A GIS-Assisted optimal urban route finding approach based on genetic algorithms. *International archives of photogrammetry remote sensing and spatial information sciences*, 35(Part 2), pp.305-308.
- Du, K.-L., Swamy, M.N.S., 2014. Radial Basis Function Networks, in: Du, K.-L., Swamy, M.N.S. (Eds.), *Neural Networks and Statistical Learning*. Springer London, London, pp. 299–335. https://doi.org/10.1007/978-1-4471-5571-3_10
- Elkadiri, R., Sultan, M., Youssef, A.M., Elbayoumi, T., Chase, R., Bulkhi, A.B., Al-Katheeri, M.M., n.d. A Remote Sensing-Based Approach for Debris-Flow Susceptibility Assessment Using Artificial Neural Networks and Logistic Regression Modeling. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 7, 4818–4835.
- He, Q., Shahabi, H., Shirzadi, A., Li, S., Chen, W., Wang, N., Chai, H., Bian, H., Ma, J., Chen, Y., Wang, X., Chapi, K., Ahmad, B.B., 2019. Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. *Sci. Total Environ.* 663, 1–15. <https://doi.org/10.1016/j.scitotenv.2019.01.329>
- Hong, H., Pradhan, B., Sameen, M.I., Kalantar, B., Zhu, A., Chen, W., 2018. Improving the accuracy of landslide susceptibility model using a novel region-partitioning approach. *Landslides* 15, 753–772. <https://doi.org/10.1007/s10346-017-0906-8>
- Hong, H., Pradhan, B., Xu, C., Tien Bui, D., 2015. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *CATENA* 133, 266–281. <https://doi.org/10.1016/j.catena.2015.05.019>
- Juliev, M., Mergili, M., Mondal, I., Nurtaev, B., Pulatov, A., Hübl, J., 2019. Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan. *Sci. Total Environ.* 653, 801–814. <https://doi.org/10.1016/j.scitotenv.2018.10.431>
- Kamranzad, F., Mohasel Afshar, E., Mojarab, M., Memarian, H., 2016. Landslide Hazard Zonation in Tehran Province Using Data-Driven and AHP Methods. *J. Geosci.* 25, 101–114. <https://doi.org/10.22071/gsj.2015.41372>
- Milaghardan, A.H., Delavar, M., Chehregan, A., 2016. Uncertainty in landslide occurrence prediction using Dempster–Shafer theory. *Model. Earth Syst. Environ.* 2. <https://doi.org/10.1007/s40808-016-0240-5>
- Nsengiyumva, J.B., Luo, G., Amanambu, A.C., Mind’je, R., Habiyaremye, G., Karamage, F., Ochege, F.U., Mupenzi, C., 2019. Comparing probabilistic and statistical methods in landslide susceptibility modeling in Rwanda/Centre-Eastern Africa. *Sci. Total Environ.* 659, 1457–1472. <https://doi.org/10.1016/j.scitotenv.2018.12.248>
- Pahlavani, P., Samadzadegan, F., Delavar, M.R., 2006, September. A GIS-based approach for urban multi-criteria quasi optimized route guidance by considering unspecified site satisfaction. In *International Conference on Geographic Information Science* (pp. 287-303). Springer, Berlin, Heidelberg.
- Pahlavani, P., Sheikhan, H., Bigdeli, B., 2017. Assessment of an air pollution monitoring network to generate urban air pollution maps using Shannon information index, fuzzy overlay, and Dempster-Shafer theory, A case study: Tehran, Iran. *Atmospheric environment*, 167, 254-269.
- Pourghasemi, H.R., Mohammady, M., Pradhan, B., 2012. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *CATENA* 97, 71–84. <https://doi.org/10.1016/j.catena.2012.05.005>
- Pradhan, A., kim, Y.-T., 2016. Evaluation of a combined spatial multi-criteria evaluation model and deterministic model for landslide susceptibility mapping. *Catena* 140, 125–139. <https://doi.org/10.1016/j.catena.2016.01.022>
- Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M., 2016. 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.