Investigation the sensitivity mapping performance to landslide with using logistic regression method

Farshid Nasrollahi[1] Mohamad reza Javadi[2] Reza Talaie[3]

1-Watershed management educated student, Department of natural resources, Nour Branch. Islamic Azad university, Nour, iran.Member of Nikan.

2- Assistant Professor , Department of Natural, Nour Branch, Islamic Azad university, Nour, iran. (Author and correspondence responsible).

Javadi_desert@gmail.com

3-A member of the agricultural Research Center faculty and natural Resources in Ardabil province.

In this study, in order to landslide sensitivity hazard zoning in the watershed Givichaylogistic regression method was used. at The first, by field visit were identified landslides characterizes in the area (335 cases) and their distribution map distribution was obtained in the area.Then, using GIS techniques in Arc Gis ver9.3 software environment,effecting factors information layers in landslide in the region were prepared. The results showed that, Marniedeposits crushed volcanic lithological units and slopeNeogene clay, 10 to 15 percent in East and Southeast, Land use of rangeland, 100 meter distance from road from the roadand 50 to 150 meters from the canals a significant effect on the landslide in the region. Also, the results of the index Qs, showed that with the amount of 0.755, logistic regression method could offer landslide hazard zoning acceptable in the region.

Key word: watershed areaGivichay, hazard zoning, andslide

Introduction:

Landslides is including mass movements which annually occur in many parts of the world and Iran, and possible loss of life, significant environmental and financial to obtain (4). Landslides are as a form of mass movements that in the detachment of soil and rock material it contains, the slope created (1). Natural features, geological and climatic Iran is such that, in some areas susceptible to this phenomenon, each year significant events, the occurrence, and result in significant damage to resources, such as residential areas, roads, facilities, land agriculture, gardens, and other resources that, the consequences of the economic and social problems will bring the region (4). Therefore, due to the negative effects of this phenomenon, the different aspects of the life of human societies under the influence, necessary and essential to, the factors affecting, and exacerbation of the phenomenon of the high points of of the (مبهم است)چی؟, found in the area of knowledge, so that preventive measures later successfully fulfilled (6). By definition, sensitivity hazard zoning to landslide including the division of the land into separate areas, these areas are classified according to the degree of actual or potential sensitivity-induced landslide on the slope is, to this sensitivity to various points in the region, recognized (4). Different methods are associated with landside hazard zoning, to each according to his concept and structure of several factors used in the landslide, and according to effecting factors weighting, susceptibility hazard map to landslide, to provide the desired area. Extensive studies have been done so far in this connection that the different methods used for landslide hazard zonation, some of which are as follows:

Mousavikhatir et al. (2010), using logistics regression statistics method began the landslide hazard zonation in the Sajjad rood. In this study six factors, distance to roads, slope, lithology, distance from drainage, land use, and the slope, respectively, the most important factors in landslides in the region, he said, and based on the sensitivity point model to landslide, was present for the watershed.

Siddiqi (2011), in a study of the information value statistic method, and logistic regression used for hazard zonation in the watershed area Pahne Kala. He plans landslide hazard zonation, through the frequency percent of landslides in high-risk classes, evaluated. The results showed that, for a zoning map in the area of intelligence value method, according to the assessment index is more appropriate than the logistic regression method.

Abedini and colleagues (2014) in a study to modeling the landslide risk with using logistic regression statistics method in the city of Bijar and stated that, based on ROC5 (9805), prepared sensitivity hazard map has high accuracy and generally about 10.84% of the area surface and high is in high sensitivity risk area and very high sensitivity risk area.

Lee and Min (2001) Sensitivity to slip in Gynkrh area determine with two possibilities method and relative logistic regression. In this modeling of lithology variable and geomorphology (for the shelves, curve slopes), soil characteristics (drainage, composition, texture and thickness), the forest cover and land property was used. However, the success of the results of both methods was satisfactory, but the accuracy of logistic regression method is more than possibilities method.

Nefeslioglu et al (2008) in the study to the landslide zoning began with using logistic regression model and artificial neural networks, and concluded that artificial neural network model compare with logistic regression method has the better result.

Mathew et al. (2009), sensitivity zoning to landslide, part of a region in the Himalayan by using logistic regression stepwise method was performed. Zoning map by curve method ROC, was evaluated. The results showed that landslides are predicted with an accuracy of 91.7% done.

Due to the success of the logistic regression method in the different regions zoning the areas with the relative sensitivity of landslides researchers should use this method from many different areas of the world.

The aim of this study was to determine areas with landslide potential different sensitivity with using logistic regression method in the watershed Givi chay in Ardabil province, which, as one of the areas vulnerable to landslide in the province have been identified.

The study area position

The study area was located in the southern province of Ardabil and the western domain of the Talesh Mountain. This area in situations ″56'03˚48, to ″30'42˚48 longitude, and ″42'26˚37, to ″20'55˚37 latitude, and 85 kilometers from the city of Ardabil and is situated in city Kosar. Watershed area is 184719.6 hectares. The study area, the northern part of the basin Gharehsou highlands Trka from the East, the Caspian Sea watershed and Talesh Mountain, and the West and South, the watershed

Ozan is limited. The highest mountains on the border of the eastern basin of the mountain called Ajam with 3016 meters above sea level, the lowest point in its output to 860 meters above sea level is located. The average height of the basin is 1828.5 meters from the sea. Average annual precipitation basin is about 422 mm and annual evaporation rate is 336 mm. The studies regional climate, the cold dry for low regions, to ultra-cold dry for height area varies. In terms of topography, watershed Givi chay has particular topography so that, a lot of ups and downs, high steep, but not flat areas along rivers, is characteristic of the watershed.All the watershed, with the mountainous and hilly with topography is harsh. Land located in the North, Northeast and East of the basin with very complex topography and steep slopes are rough which made agriculture impossible. Most of the agricultural land located in the northwest part of the western basin. The main direction of the slope the southern basin and east-west directions as minor respects, so were the South. Givi chay Permanent River is the most important river and the main drainage area of study. The river, along with Senghor chay and Gzaz chay, hydrographic basins form the main trunk network. In general, the drainage, the length of 3656 kilometers and the number 3494, consisting of permanent and seasonal rivers, streams, stream and rivers dry are responsible for the Givi chay watershed drainage. The main branch of the river, with Heroabad from the heights of the Southeast with a maximum height of 2820 meters above sea level comes and the public along the southeast to the northwest flows and end at a height of 860 m sea in the southwestern part of the field of study abroad and Ozan unloading throughput.In the area, pasture, forest and agricultural lands and orchards that have supplied the needs of people and livestock area. In term of lithology the oldest rock units, Upper Cretaceous carbonate strata of the region in the south and southeast, the basin is formed and apart from damage stone, Oligocene and Neogene that mainly in the middle part of the basin, with a large area extending to other parts of the basin in possession of lava and tuff Cenozoic which is mainly related to the volcanic activity.

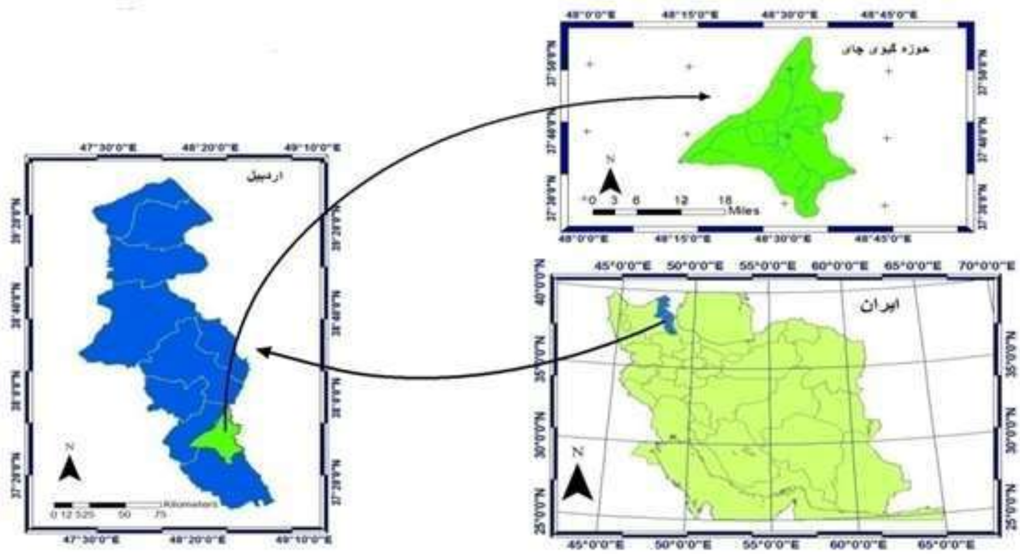

Figure 1: Location of the study area in Iran and Ardabil province

Methodology

In order to do this, first of all basic information about the geological, geomorphological, physiographic and land use were collected. By interpreting aerial photographs, geological map, vegetation and topographic base map, the area was established. All maps using field frequent visits were reviewed. Basic map, by using the software ARC-GIS intro the GIS environment. After digitization, the other physiographic maps are needed including DEM, classes, maps of slope, slope direction were extracted. After this stage with extensive field operations, further geological maps, the relative height difference and distribution of landslide were reviewed.

In this study, the logistic regression method for sensitivity hazard zonation to landslide in the region that the structure of this method are as follows:

An approach that in recent decades, most are used to assess the occurrence of mass movements is statistical methods that logistic regression is one of example it, that with a bit patterns, to explore the relationship between mass movements and Initial data deals. In logistic regression model the probability of landslide is in the range of zero and one, where zero means there is no risk of landslides and one number means that a maximum risk of landslides in the future. Logistic regression model relationship is as follows:

Where:
( ) The risk of landslides
Z= dependent variable

The assumed linear model z relationship (occurrence and lack of occurrence landslides - the risk degree of landslides) can be expressed by the following function:

$$Z=a_0+a_1x_1+...+a_px_p$$

Where
x_p: Independent variables (effecting factors) for p-th variable
a_p : equation coefficient

If all the independent variables and the dependent variables were continuous, there was the linear regression relationship possibility, but because the variables are continuous and are grouped, so the use of logistic regression is necessary. Therefore, linear modelas a logistic model, and wrote by following equation

Where

P = the probability of landslides risk in each pixel on the region domain.

In this study, the pixels size was used 40 meters in 40 meters. So for all the area pixels, (40 × 40 m) and the possibility of landslides as the possibility risk in a numerical range between zero and one, respectively. These values are based on distances method equal to the five categories of risk including: the very low risk (less than 0.2), low risk (0.2 – 0.4), medium (0.4 – 0.6), high (0.6 – 0.8) and high risk (greater than 0.8) was divided.

Conclusion:

In this study, after analyzing aerial photos and previous studies and field visits, a total of 335 cases identified landslide and was registered. Of the total study area, about 84/31 square kilometers, the region is sliding. Compare the number of pixels per unit area lithology, which are landslide, suggests that in the volcanic rock units split, marl and clay sediments of Neogene and alluvial with the origin of Neogene sediments, the percentage of pixels has been more slip.

Directions shelves and slopes, control the amount and direction of the flow of runoff, vegetation density, and temperature and soil moisture are the main factors in the occurrence of landslide (Ayalou and Yamagashy, 2005). In this area, more landslides in the region, in the south-east slopes of the East and the South has happened and in the foothills of the North and the North Eastas well as distribution of the landslide other than the is not significant. Distribution of landslides on slopes with different percentages vary and in term of statistics shows significant difference.

Comparing the slide area ratio in various applications shows that the most landslide in the slopes with the pasture application, occurred. In areas without vegetation and low vegetation, due to the very low thickness of the soil, or the absence of soil, the frequency of landslides is very low.

More than 76 percent of the landslide, which occurred in the vicinity of 50 to 150 meters, can provide the necessary moisture through the soil profile and search channels.

More than 84 percent of the landslide has happened at the optional 100-meter path, perhaps expected the direct impact of roads in landslides as the road and human intervention including agriculture, the increase the degree of instability in the region. It also creates trenches in road path and due to reduced layers resistive force to channel the flow of water and increased permeability can be a factor in landslides.

Region rainfall value was variable and from 330 to 390 mm in year. The effect of rainfall in slopes unstable the climate of the region, the geology structure, topography and permeability of the material depends (4). Therefore, it is possible that, with the increase in rainfall landslide are not exceeded. As well as the type of rainfall, lithology and morphology of the region, the main impact of the landslides in the region. Since the entire basin, the annual precipitation, almost flat and there is no significant change. Therefore, we expect more landslides controlled by other parameters.

In this study, to investigate the relationship between landslide with fault factor, distance map from fault and develop and landslide map was adjusted. Apparently, with increasing distance from the main fault zone, the number of pixels of a slip, reduced, but an unexpected decline pixels, a slip and slide, than the main fault zone is almost uniform. Therefore, we cannot with certainty, faults, considered the main cause for landslides in the area.

Earthquake is one of the triggers of the slopes in landslides. The effects of earthquakes on unstable slopes, factors such as geology, lithology, Hydrogeology, topography and so on. To simplify the impact of the earthquake on landslide mechanism, shear power, along a non-continuous of surface rupture, but the slope is unstable, it is. To analyze the effects of earthquakes on unstable slopes, the maximum horizontal acceleration caused by the earthquake has been analyzed as a single layer. The results showed that by increasing the maximum horizontal ground acceleration, meaningful changes, and an increase in the amount of pixels involved, will not slip. So the probability of landslides, based on the amount of horizontal acceleration earthquake is unpredictable.

Comparison of the slip pixels ratio and without slide shows that in the lowlands and foothills to mountains high landslide, is higher, this indicates that following landslide and the combination of stone and it means that geological formations morphological conditions because the type of land, it is also a function of the composition of the rock formations of the region.

Height of sea introducing changes and fluctuations of climate. Because of the dispute energy amount of each point that the Earth receives from the sun, it is different. Height changes can have a significant impact on three factors such as temperature, precipitation and humidity make (1). The results showed that high-class, 3000-2800 meters, with the highest weight class over 3000 meters altitude, is the lowest weight. However, the floor height of 2000-1800 meters with a maximum size and floor height of 3000 meters is the lowest area in the region.

The results of the sensitivity analysis of sensitivity zoning map to landslide

After determining effecting factors in landslides and prepare the information layers about the software ARCGIS, to provide the final layer was based on logistic regression method. To run the model was used the Forward Stepwise method. Analyze stepwise by using the maximum likelihood method, conducted. It starts with the arrival of predictor variables. The basic model consists of constant coefficient from variables that are in the next step respectively, the other independent variables are added to the model (Hlmachr and Davis, 2002).

In implementing the model was used the Forward Stepwise method. In this method the variables at less than 5 percent error level identified and entered. In other words, variables that their t and F is more with dependent variables are gradually entered the equation. In this method, at first the independent variable (effecting factor in the landslide) its correlation with the dependent variable is stronger caused R^2 entered to analyze. The second variable is a variable that, after the first variable, the highest increase in R^2 causes. In this way the other variables, the Chi-correlation of each of them with the dependent variable is larger, are entered in the equation.

In the model, 75% of the pixels have landslide, were used and the remaining 25% in order to evaluate it were kept out of the model. Prior to estimate the probability of occurrence of landslides in each pixel, logistic regression equation values was determined and then using logistic regression equations, probability values were calculated. negative values, if significant the factor parent (significance level is less than 0. 05) it means that, with an increase of one point of the independent variable, the value of the logarithm of the odds of landslides, the size of the coefficient reduced that is the landslides risk multiplication sign .

Based on definition function f (z) probability values for all the area pixels calculated. The estimate probability for the area is between 0-0.90. In the model, the coefficients of linear logistic equation was calculated.

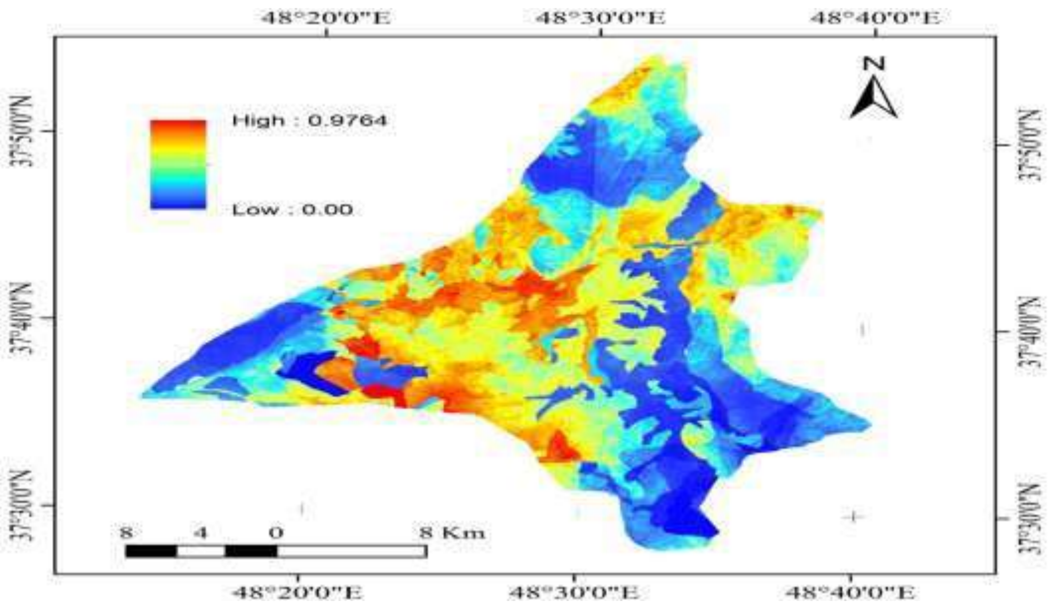

Figure 2: Map point landslide risk possibility with using logistic regression method

Evaluation logistic regression model accuracy

This model was carried out in two phases. In the first stage of model implementation, there is no independent variable in the model. The output of this stage is called the block (0). In the second stage the independent variables, respectively entered to the model and the results in the form of block 1 is shown. In the first stage (zero block) the success of the model in predicting the occurrence or non-occurrence of landslides showed that 100% of pixels which have landslides were correctly predicted while none of pixels that had not landslide correctly predicted.

Also in this stage of the analysis the model successfully in the predicting was 50 percent. Table (1), the success of the model in predicting the occurrence or non-occurrence of landslides in each pixel after enter the independent variables shows. In tenth step input of independent variables in the model 72.4% of the total pixels correctly predicted that compared with 50% zero model and has a great improvement. The percentage of pixels have a landslide which correctly predicted by the model, the model is called model sensitivity (true positive).

The likelihood power or the model resolution based on the percentage of correctly classified pixels which are non-slip were shown (true negative). This model is based on area under the curve ROC (AUC) is about 72.8% of the pixels that have slip and non-slip has grouped properly.

Table 1: The success of the logistic regression model in the pixels grouped that have slip and non-slip based on cross assessment method

step	Views	Forecasts		
		landslides		correct Percent
		Non-occurrence	occurrence	

10 step	landslides	Non-occurrence	9840	5580	63.8
		occurrence	2821	12599	81.7
	total Percent				72.8

The total number of pixels in the region are 495,991 pixels, of this number20, 675 pixels with slip (% 4.2) and 475,316 pixels, is non-slip (95.8%).

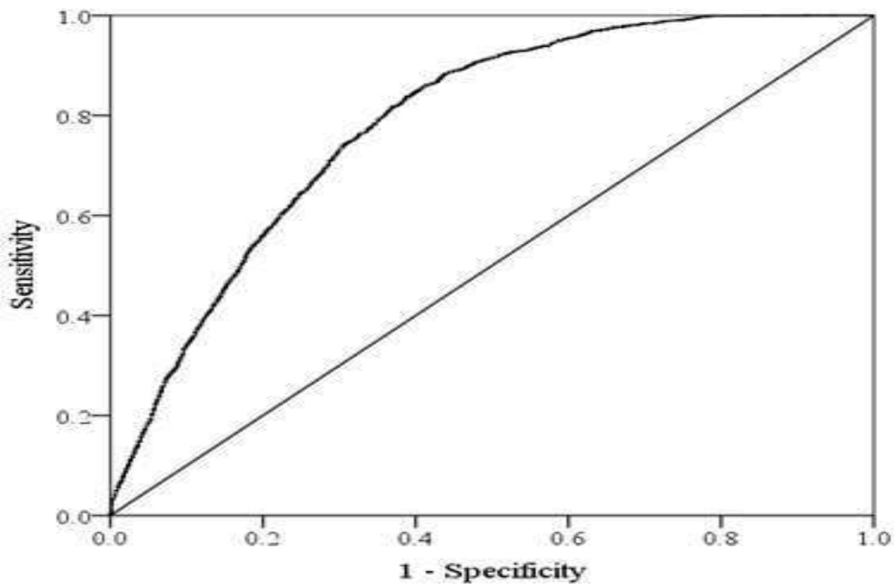

Figure 3: The area of under the curve ROCC (with the 0.782) for the logistic regression model

Table 2: Results of the logistic regression model test

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	37823.740[a]	0.148	0.197
2	36252.129[a]	0.190	0.253
3	35046.496[a]	0.221	0.295
4	34453.765[a]	0.236	0.315
5	34321.539[a]	0.239	0.319
6	34021.660[a]	0.247	0.329
7	33936.260[a]	0.249	0.332
8	33859.101[a]	0.251	0.334
9	33762.234[a]	0.253	0.337
10	33736.701[a]	0.254	0.338
11	33728.011[a]	0.254	0.338

Hasmr and Lmshv test with significant quantities of less than 0.05 indicating very good performance of the model compared to zero mood. So running the model with input independent variables (effecting factors in region landslide) better than the performed default model in the zero block (Table 3).

Table3: the result of Hasmr and Lmshv test

Step	Significant level	Degrees of fredom	Chi-square
1	1.000	5	0.000
2	0.000	8	131.241
3	0.000	8	105.519
4	0.000	8	746.414
5	0.000	8	847.720
6	0.000	8	342.685
7	0.000	8	395.345
8	0.000	8	358.453
9	0.000	8	287.675
10	0.000	8	261.746
11	0.000	8	352.208

The results of sensitivity map ranking to the landslide

Practical use from sensitivity map to the landslide, need to classify it to the different degrees of sensitivity. To this end, the predicted probability for the occurrence of landslides share in 5rank, very low, low, medium, high and very high (Figure 3).

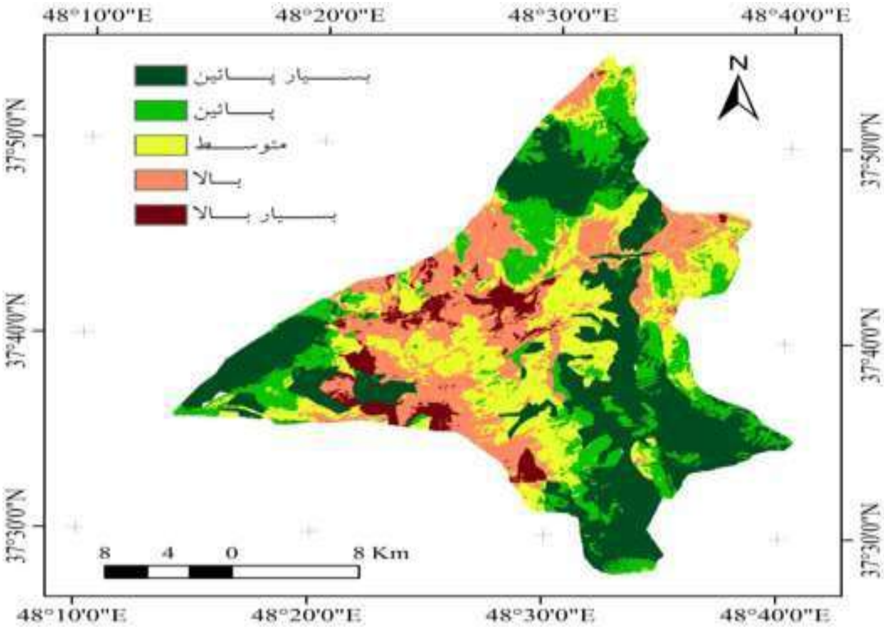

Figure 4-12: sensitivity ranking map to the landslide according logistic regression model

In this study, the condensation relative index (DR) in order to compare the danger zone (classes each agent) and index of quality total (QS) used to determine the appropriateness of the using methods. In other words, the index of quality total and accuracy, appropriate model in accordance with the study area and the compression ratio index, the zone resolution index or risk categories, each method of zoning models, is introduced.

On a map zoning method, category slide density Dr = 1equivalent to the slide density mean in total of region and density ratio equivalent to 2, has slide density equivalent to twice the density of the slip region. Therefore, separation between risk categories by density ratio index lines is more favorable and the map prepared by the method has accuracy and more favorable.

The total amount of quality (Qs), indicating the right or the utility performance of the method for predicting the risk of landslides in the region. Usually the value of this index for different method across zero and seven. Although in theory there is no limit to it. In the method evaluation the amount of quality total (Qs) is more, method has accuracy or more utility in separating.

Table 4: Results of evaluate the accuracy and efficiency from logistic regression model

Zoning model	Risk category sensitivity on the map	Ai area of risk floor in map)km^2(	Si Landside area total in each floor(km^2)	DR (density ratio)	Area percent	S Area ratio of each risk floor to regain total area	**Q_s**
logistic regression	Very low	239	2.32	0.21	0.009	0.31	0.755
	Low	87	0.96	2.6	0.011	0.11	
	Average	253	15.2	1.42	0.06	0.33	
	High	136	12.28	2.14	0.09	0.17	
	Very high	46	1.4	0.71	0.03	0.06	
	total	681	32.16				

The results of the data in the above table indicate that in both methods, the compression ratio (Dr) which in some sources it is mentioned the name of Landslide Index. While relatively good resolution, in accordance with the risk zone, and increased. This resolution represents an average accuracy of the danger zone separation in the procedure which corroborates research Shirani 2007 and 1995. Also, as shown in the above table, the amount of total quality Qs, which implies the comparison and evaluation methods, in relation to each other in logistic regression method the value maximum (0.755) is obtained, the utility is good.

The results of this study with the results Ashqly Farahani, 2000, and Haqshenas, 1995, great Mousavi Khatir et al. (2010) the logistic regression method as an appropriate and efficient method, in most circumstances based on quality index as mentioned, logistic regression method has a good performance in the study area. Due to the compression ratio mainly, the main causes of the instability of the slopes, in area Kosar (Givi) can be associated with loose materials, the amount and direction of the slope of the land (vegetation) outlined. Other factors such as access roads negative effect on the range and scope as well as leaching, and leaching of the heel, by the river, affecting instability have been diagnosed. It is worth noting that, several other factors can also influence the process of a landslide in the range studied, among which is the creation of trenches.

Reference

[1] Amin Zadeh, M. 1997, Assess slope stability and control in the watershed of Jiroft, Master Thesis Engineering Geology, Tarbiat Modarres University. 145 pages.

[2] Hassanzadeh Nofoty, M, 1999, Landslide hazard zonation Shlmanrvd basin, MA thesis Tehran University.

[3] Haqshenas, A, 1994 landslide hazard zonation and its relationship with sediment production in the watershed Taleghan MA thesis, Faculty of Science, Tarbiat Modarres University

[4] Shariat Jafarie, M, 1995, landslide (principles of natural slope stability). Commercial Press, 218 pages.

[5] Komak Panah, A, Hafezi moqadasie, N, 1993, Landslide hazard zonation methods. Proceedings of the Workshop on Evaluation of landslide mitigation strategies in the country. International Institute of Seismology and Earthquake Engineering. From page 390 to 414.

[6] Memarian, H, 2005, Geological Engineering, Tehran University Pub .210 page.

[7] Mousavi Khateir, Z, Kavyan, A, Soleimani, K, 2010, Landslide Susceptibility Mapping Sajaroud watershed using a logistic regression model, Journal of Science and Technology of Agriculture and Natural Resources, Soil and Water Sciences (14).

[8] Siddiqi, M, 2012, landslide hazard zonation using logistic regression, and the value of intelligence in the catchment area of incoming client Sari, MSc thesis, Islamic Azad University's .115 light.

[9] Abedini, M, Ghasemian, M, Shirzadi, AS, 2014. Landslide risk modeling using logistic regression model Bijar watershed. Journal of Geography and Development, No. 37, pages 85 to 105.

[10] Lee,S., K. Min. 2001. Stastistical analysis of landslide susceptibility at Yongin, Korea. Environmental geology 40, pp. 1095-1113.

[11] Ayalew L., H. Yamagishi, 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65:15.

[12] Mathew, J., V. K. Jha, G. S. Rawat. 2009. Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiveroperating characteristic curve method. Landslides 6, pp.17–26.

[13] Ohlmacher, C.G., C. J. Davis.(2003). Using multiple regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Eng Geol 69:331–343.

[14] Nefeslioglu, H.A., Gokceoglu, C., Sonmez, H(2008). An assessment on the use of logisticregression and artificial neural networks withdifferent sampling strategies for the preparationof landslide susceptibility maps. Engineering

[15]Geology 97