International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

Comparison of GIS based Statistical Methods for Landslide Susceptibility mapping (Case Study Faizabad City, Afghanistan)

Ahmad Qasim Akbar

Abstract— Landslide risk exists in all mountainous regions and every year causes a great life and financial losses. Afghanistan is one of those countries which has mostly suffered from landslide and the phenomena is not only risking the financial asset but it is a major threat for human lives, therefore, a landslide mitigation analysis is a major need for the area. To take a step towards landslide mitigation and prevent the disaster, numbers of GIS based statistical methods exist, but it is still unclear which one has higher accuracy and yet very few studies propose a reliable method, therefore, this study aims to compare the commonly used GIS based bivariate statistical method "Frequency Ratio (FR), Weight of Evidence (WOE)" and multivariate statistical methods "Logistic Regression (LR)" and their combination to achieve higher accuracy for landslide susceptibility map.

This study was implemented in Faizabad city. The high mountains, human activity, seasonal rainfall, and earthquakes are the main reason of the landslide and rock fall which time to time the area is suffering from, the risk is very high in the area but very few studies have been done and there is no data available, Therefore, this study also focuses on data collection, compare the GIS based statistical methodologies and their combination to create high accuracy a landslide susceptibility map for an area where data is limited.

The result was validated using the matrix validation tool and it is proposing that, however, all of the methodologies can achieve an acceptable accuracy but the combination of bivariate and multivariate statistical methods increase the accuracy of the analysis and the combination method overcomes with the demerits of the methodologies alone, or in other hands they become complimentary for each other, the result proves that the combination of Weight of evidence (WOE) and Logistic regression (LR) is achieving higher accuracy and more reliable..

Index Terms— Landslide Susceptibility Mapping, WOE, LR, Combined method, Statistical Methods, Faizabad, Badakhshan

IJSER

1 INTRODUCTION

Generally, a landslide is a downslope movement of soil, rock and organic materials due to factors such as gravity, water, tectonic and human activities, which result in the landform (Highland, L.M., and Bobrowsky, Peter, 2008). Landslide is considered to be the most common geological disaster, which causes life and financial losses, it damages everything which is coming to its path (Bui et al., 2012; Shahabi et al., 2014).

The Faizabad city is located in the northeast part of Afghanistan, and has the population of around 50,000 (Dupree 1977).

The complex geology and tectonic activity of the Hindukush mountain range frequently cause high magnitude earthquake motions in the area (USGS Earthquake Report,2014; USGS Earthquake Report,2015).

Figure 1. April 28,2015 landslide of Faizabad Badakhshan (Guardian Newspapers)

The earthquakes, high altitude, precipitation, human activity, and the combination of several other factor trigger numbers of landslides every year as an example " in April 28, 2015, a landslide killed 52 people was killed (America. Aljazeera, 2015)".

However, the risk of slope failure is high in the area but no action for mitigation has been taken and no data has been collected for this purpose. Therefore, this study aims to collect all the available data for the area and create a high accuracy land-

- *Ahmad Qasim Akbar has a master in Civil engineering and currently Director of GIS and Remote Sensing in Afghanistan Geological Survey, Ministry of Mines and Petroleum of Afghanistan , Email:* qasim1368@live.com
- *Dr. Guangqi Chen is currently Professor faculty of Arts and Science and Department of Civil and Structural Engineering, Graduate School of Engineering, Kyushu University, JAPAN, Email:* chen@civil.kyushu-u.ac.jp
- *Ataullah Yosufzai, Assistant Professor, Department of Geological Engineering and Exploration of Mines, Geology and Mines Faculty, Kabul Polytechnic University/Afghanistan, Currently PhD Student in Department of Geology, Prospecting and Exploration of Mines, Geology and Mines Faculty, Saint Petersburg Mining University/Russia, email:* ata.yosufzai@gmail.com / ata.yosufzai@kpu.edu.af

International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

slide susceptibility map from the available methodologies.

To compare the methodologies 11652 control point equally distributed from the landslide and non-landslide is selected. The result proposes that combination of the bivariate statistical methods “Frequency Ratio, Weight of Evidence” and multivariate statistical method “Logistic Regression” increase the accuracy of the analysis and it is more reliable than the methods alone. Moreover, because of 82.536% of the success rate, the combined method of Weight of evidence (WOE) and Logistic regression (LR) is reliable for the study area.

2 Relevant Data

Data layers are mainly diveded into five types in the study one(Table 1).

Table 1. Shows the used data layers with its type and Source

Data Type	Factors	Type	Source
Topographical Factors	Elevation	Scale	DEM Achieve of USGS (Earth Explorer)
	Slope Angle	Scale	
	Slope Aspect	Categorical	
	Curvature	Scale	
Hydrological Factors	Distance to the River	Scale	Afghan Geological Survey (USGS Open-File Report 2007-1214)
	Precipitation	Scale	Global Precipitation Climatology Centre (GPCC).
	Stream Power index (SPI)	Scale	DEM Achieve of USGS (Earth Explorer)
	Topographical Wetness index (TWI)	Scale	
Geological Factors	Lithology	Categorical	Afghan Geological Survey (USGS Open-File Report 2007-1214)
	Distance to the Faults	Scale	
Landuse	Landcover	Categorical	Landsat 8 Achieve of USGS (Earth Explorer)
Other	Distance to the Roads	Scale	Afghan Geological Survey (USGS Open-File Report 2007-1214)

Mainly, layers such as elevation, slope angle, slope aspect, curvature, steam power index (SPI) and the topographical wetness index (TWI) were extracted in GIS platform from a 30m resolution DEM which was downloaded from USGS Earth Explorer.

A landslide inventory map for the study area was created using the visual interpretation of the high-resolution imagery of GIS base map considering the landslide indicators such as differences in sediments color, the roughness of the structure, and sharp contacts. 202 landslides location were detected and adjusted with Google Earth owing to its 3D view and high resolution.

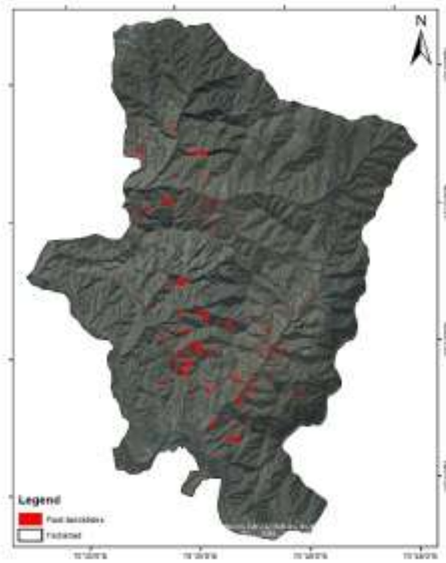


Figure 2. Shows the landslide inventory map which was created using the visual interpretation

2.1 Topographical Factors

Elevation and relief illustrate the potential energy for the mass wasting (Ghimir, M., 2001; Oguchi, T., 1997). The elevation layer of the study area is extracted from DEM and divided into five different classes (Figure 3.a).

The slope angle is a primary factor in the dynamics of processes governing land evolution and landslide, it is used as the main triggering factor in a landslide (Bourenane, 2014). Different sediments react differently to the slope angle based on their physical property but in general, as much as the slope angle increases the possibility of slope failure increases. The slope angle layer is extracted from the DEM and divided into seven different classes(Figure 3.b).

Slope aspect indirectly indicates the slope instability based on the influence of the related factors such as exposure to the sunlight, exposure to the wind and soil moisture which specifies the landslide occurrence (Magliulo et al.,2008; Dieu et al.,2011). The slope aspect layer is extracted from the DEM and divided into eight classes (Figure 3.c).

The curvature is extracted from the DEM and represents the morphology for an area and its classified into three classes: class 1, convex; class 2, concave; class 3, planar (straight). Generally, the concave class is considered as a potentially unstable, unlike theconvex class which is more stable for the sliding (Stocking, M.A., 1972) (Figure 3.d)

IJSER

International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

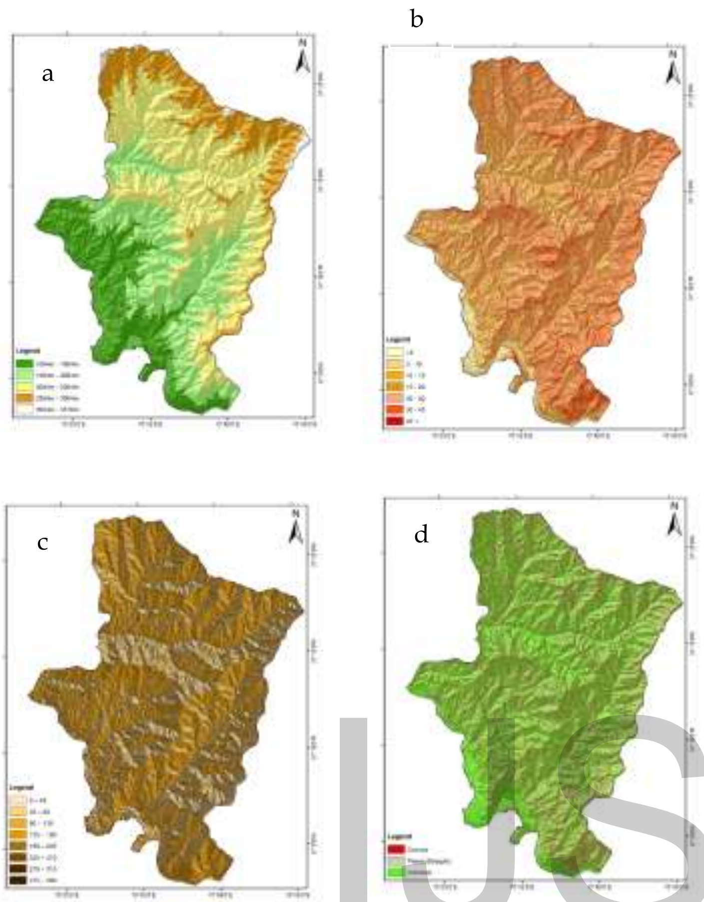


Figure 3. a) Elevation layer b) Slope Angle layer c) Slope Aspect layer d) Curvature layer

2.2 Hydrological Factors

Landslide and Flood has a close relationship because both are related to the intensive perception, surface runoff, which increases the possibility of the landslide in the slopes (Highland, L.M., and Bobrowsky, Peter, 2008)

The river data used for the study area was provided by Afghan Geological Survey (USGS Open-File Report 2007–1214). The area around the river is buffered based on its distance from the river and divided into six different classes(Figure 31.a).

The annual average precipitation rate is about (484.9mm) in the area (NOAA 1964-1983) and it is considered as the primary triggering factor of the landslides (Bourenane, H., 2014)., The precipitation map was downloaded from the Global Precipitation Climatology Centre (GPCC) and divided into five different classes (Figure 31.b)

Stream Power Index (SPI) is the measure of erosive power associated with flowing water based on the assumption that discharge is proportional to the specific catchment area, and it is calculated as below.

$$SPI = {A \cdot \tan\beta}/{b} \quad (1)$$

A is the flow accumulation, (radian) is the slope, and b is the width of a cell through which water flows. Higher SPI value should correspond to a higher likelihood of erosion on the landscape (Wilson and Lorang,2000). The values are divided into three classes (Figure 31.c).

The Topographic Wetness Index (TWI) or Compound Topographic Index (CTI), is a steady-state wetness index. In some areas, TWI has been shown to predict solum depth (Gessler et al., 1995), and it is calculated as below.

$$TWI = \ln\left({A}/{Tan\beta}\right) \quad (2)$$

A is flow accumulation and represents the slope. Higher TWI values represent drainage depressions however lower values represent crests and ridges. The values are classified into three different classes (Figure 31.d).

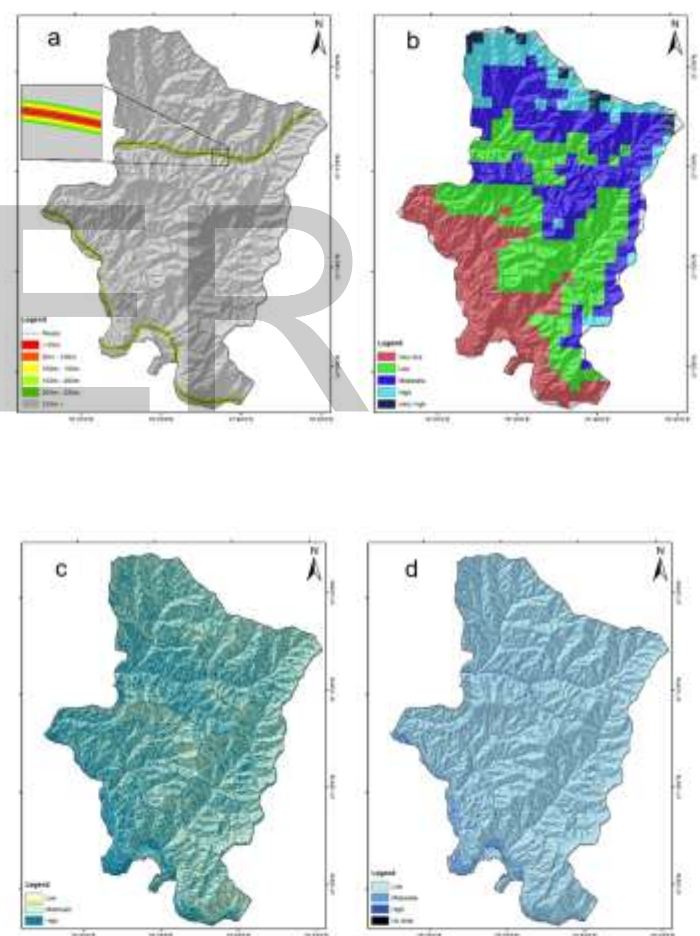


Figure 4. Shows a. Distance to the River b. Precipitation c. Stream Power index (SPI) d. Topographical Wetness index (TWI)

2.3 Geological Factors

Sediments or rocks in the study area are divided into three classes: (1) sedimentary such as conglomerate, sandstone, limestone, and gravel stone or as sand, clay, and loess, (2)

IJSER

International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

metamorphic such as marble, gneisses, and quartzite, and (3) igneous such as granite, gabbro and diorite (USGS Open-File Report 2007–1214) (Figure 5.a).

Tectonic activity and geological faults are considered as main triggering factors in slope failure (Paulsen et al. 1998), and the area is suffering from various numbers of normal faults, buried and proven (USGS Open-File Report 2007–1214). The area around the faults are classified into five different classes (Figure 5.b).

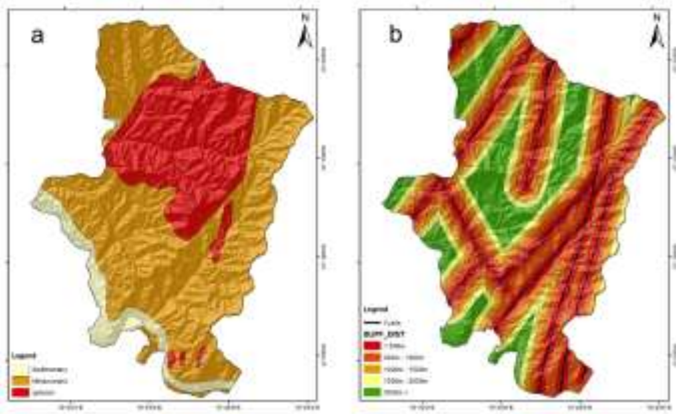


Figure 5. Shows a. Lithology b. Distance to the faults

2.4 Land Cover

The land cover for the study area is extracted from Landsat 8 imagery (USGS Earth Explorer). The band composition and classification was performed using the unsupervised classification in the GIS platform however due to the limitation of the image resolution and permanent snow cover in the area, the image was classified into three classes (Figure 6).

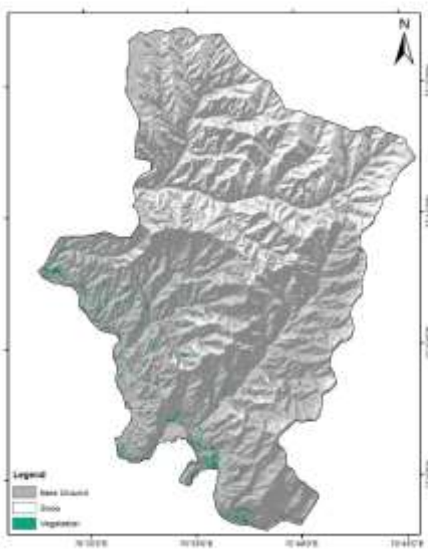


Figure 6. Landcover map of the study area extrated from Landsate 8

2.5 Other

While building a road, normally the extensive excavation, overloading and removing the vegetation happens in slopes which most of the time causes landslides (Highland, L.M., and Bobrowsky, Peter, 2008). The road data is collected from Afghan Geological Survey (USGS Open-File Report 2007–1214) and the area around the roads are divided into five different classes (Figure 7).

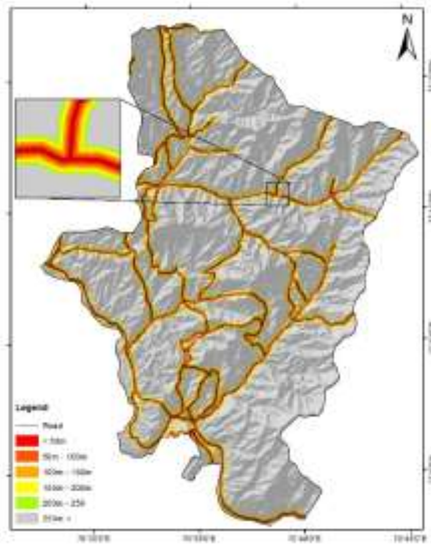


Figure 7. Show the Distance to the Roads layer

IJSER

3 METHODOLOGY

3.1 Frequency Ratio (FR) Method

Frequency Ratio (FR)method is based on the distribution of landslide in each factor's class (Saro and Biswajeet, 2007). It is normally using the ratio of landslide area in a class to the total landslide ratio in the area. To find the relation between landslide occurrence in each factor for each factor's class, a database was developed and FR method was applied (Eq.3)

$$Fr = \ln\left(\frac{Con-prob}{Prior-prob}\right) = ln\left(\frac{\frac{LPC}{TPC}}{\frac{TLP}{TP}}\right) \quad (3)$$

(LPC) is the number of landslide cell in each class. (TPC) is the total number of the cell in the class, (TLP) is the total number of the landslide cell in all the area and (TP) is the total number of the cell.

To calculate the Fr weight for each class of the factors in the database a code was designed in the Matlab environment and the result is given as the result is given as (Table 2) below.

Table 2. The result of the FR weight calculation using the designed Matlab code

Factor	Classes	Total PIX	Landslide PIX	Percent	Fr Weight
Elevation (m)	1064 - 1564	118868	750	17.00295	-0.20376
	1564 - 2046	143828	3065	69.48538	1.013368

International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

	2064 - 2564	207893	596	13.51168	-0.9926
	2564 - 3064	93327	0	0	-0.9926
	3064 - 3576	6314	0	0	-0.9926
Slope (Degree)	< 5	19701	28	0.634777	-1.69428
	5 - 10	43302	107	2.425754	-1.14119
	10 - 15	68228	289	6.551802	-0.60225
	15 - 20	91080	520	11.78871	-0.30373
	20 - 30	198747	1699	38.51734	0.099946
	30 - 45	142099	1717	38.92541	0.445993
	45 <	7073	51	1.1562	-0.07028
Aspect	North - East 1	37338	1041	23.60009	1.282109
	North - East 2	40404	618	14.01043	0.681743
	South - East 1	69656	161	3.649966	-1.20798
	South - East 2	81364	12	0.272047	-3.95984
	South - West 1	93949	39	0.884153	-2.92501
	South - West 2	102967	203	4.602131	-1.36702
	North - West 1	86646	917	20.78894	0.31346
	North - West 2	57906	1420	32.19225	1.153774
Curvature	Concave	114321	978	22.17184	0.100682
	Convex	340604	2613	59.23827	-0.00828
	Planar	115305	820	18.58989	-0.08409
Distance to the River (m)	< 50	545452	4407	99.86404	0.043147
	50 - 100	5701	0	0	-1.49324
	100 - 150	5620	0	0	-1.49324
	150 - 200	5643	0	0	-1.49324
	200 - 250	5560	2	0.045321	-3.06864
	250 <	2301	4	0.090641	-1.49324
Precipitation	Very Low	143651	870	19.71448	-0.23074
	Low	187511	2940	66.62135	0.72048
	Moderate	173843	603	13.66417	-0.78808
	High	60996	0	0	-0.78808
	Very High	12513	0	0	-0.78808
SPI	Low	110480	836	18.95262	-0.02202
	Moderate	345058	2317	52.52777	-0.1415
	High	114692	1258	28.51961	0.349211
TWI	Low	367516	2908	65.92609	0.022621
	Moderate	168745	1338	30.33326	0.02471
	High	33960	165	3.740648	-0.46507
Lithology	Metamorphic Rocks	354444	3713	84.13777	0.30288
	Sedimentary Rocks	43969	1	0.02266	-5.82965
	Igneous rocks	171877	699	15.83956	-0.64329
Distance to the Fault (m)	< 500	114038	790	17.90165	-0.11066
	500 - 1000	155567	1271	28.80127	0.054318
	1000 - 1500	134176	1088	24.65443	0.046779
	1500 - 2000	91368	608	13.77748	-0.15089
	2000 <	75141	656	14.86517	0.120629
LandCover	Bare Ground	422378	3633	82.32495	0.102336
	Snow	138535	778	17.62973	-0.32397
	Vegetation	7434	2	0.045321	-3.36249
Distance to the Road (m)	250 <	440357	2418	54.79266	-0.34305
	<50	28056	497	11.26218	0.828223
	50 - 100	27151	458	10.37843	0.77929
	100 - 150	26098	363	8.225697	0.586379
	150 - 250	24946	352	7.976433	0.600753
	200 - 250	23682	325	7.364605	0.572945

Once the Fr weight is calculated for each factors class, it would be taken into GIS Platform to display each factors susceptibility to the landslide and each factors class has its specific weight as (Table 2). Furthermore, to create the landslide susceptibility map, the calculated Fr weights are summed (Eq.4).

$$LS_{FR} = \sum Fr_1 + Fr_2 + Fr_3 + \cdots + Fr_n \quad (4)$$

LS is the landslide susceptibility index, (Fr) is the Fr weight of each factor's classes. LS is representing the relative hazard of landslide occurrence.
The higher result values, the higher risk of slope failure (Saro and Biswajeet, 2007), therefore, the results were classified using the natural break classification into five different classes "Very Low, Low, Moderate, High, Very High" which represents the level of unstable locations in the maps. This information is applicable to all the methods (Figure 8).

IJSER

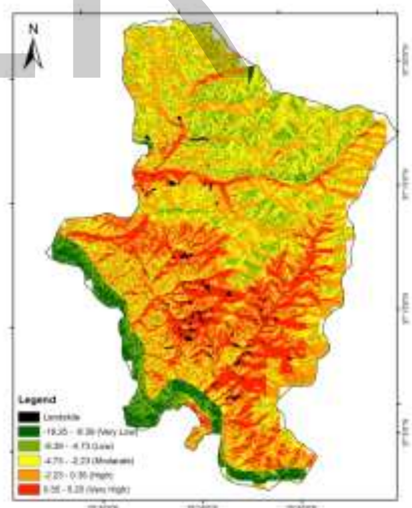


Figure 8. Illustrates the result of the Frequency Ratio Method for the landslide susceptibility mapping of Faizabad City, Afghanistan

3.2 Weight of Evidence (WOE)Method

To evaluate the contribution of each factor towards landslide hazard, the existing landslide distribution, data layer was compared to various thematic data layers separately. (Netra R. Regmi, 2009).

$$W_i^+ = ln\left(\frac{\frac{NPix_1}{Npix_1 + Npix_2}}{\frac{NPix_3}{Npix_3 + Npix_4}}\right) \quad (5)$$

International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

$$W_i^- = ln\left(\frac{\frac{NPix_2}{Npix_1 + Npix_2}}{\frac{NPix_4}{Npix_3 + Npix_4}}\right) \qquad (6)$$

$$(Weight = W_i^+ - W_i^-) \qquad (7)$$

$Npix_1$ is the number of pixels representing the presence of both potential landslide predictive factor and landslides. $Npix_2$ is the number of pixels representing the presence of landslides and absence of potential landslide predictive factor. $Npix_3$ is the number of pixels representing the presence of potential landslide predictive factor and absence of landslides, $Npix_4$ is the number of pixels representing the absence of both potential landslide predictive factor and landslides. Considering the equations above for the weight of evidence and to calculate the weight for each class of the factors a code was designed in the Matlab environment and the result is given as (Table 3) below.

Table 3. Shows the Result of the WOE Weight Calculation using the designed Matlab Code

Factor	Classes	Total PIX	Landslide PIX	Percent	Value
Elevation (m)	1064 - 1564	118868	750	17.00295	-0.25298
	1564 - 2046	143828	3065	69.48538	1.928054
	2064 - 2564	207893	596	13.51168	-1.30862
	2564 - 3064	93327	0	0	-1.30862
	3064 - 3576	6314	0	0	-1.30862
Slope (Degree)	< 5	19701	28	0.634777	-1.72964
	5 - 10	43302	107	2.425754	-1.20133
	10 - 15	68228	289	6.551802	-0.66592
	15 - 20	91080	520	11.78871	-0.35475
	20 - 30	198747	1699	38.51734	0.159082
	30 - 45	142099	1717	38.92541	0.658301
	45 <	7073	51	1.1562	-0.07166
Aspect	North - East 1	37338	1041	23.60009	1.505509
	North - East 2	40404	618	14.01043	0.767425
	South - East 1	69656	161	3.649966	-1.3073
	South - East 2	81364	12	0.272047	-4.11996
	South - West 1	93949	39	0.884153	-3.10496
	South - West 2	102967	203	4.602131	-1.52613
	North - West 1	86646	917	20.78894	0.385087
	North - West 2	57906	1420	32.19225	1.454158
Curvature	Concave	114321	978	22.17184	0.128636
	Convex	340604	2613	59.23827	-0.0206
	Planar	115305	820	18.58989	-0.10512
Distance to the River (m)	< 50	545452	4407	99.86404	3.517297
	50 - 100	5701	0	0	-3.08546
	100 - 150	5620	0	0	-3.08546
	150 - 200	5643	0	0	-3.08546
	200 - 250	5560	2	0.045321	-3.08546
	250 <	2301	4	0.090641	-1.50243
Precipitation	Very Low	143651	870	19.71448	-0.2987
	Low	187511	2940	66.62135	1.438015
	Moderate	173843	603	13.66417	-1.00453
	High	60996	0	0	-1.00453
	Very High	12513	0	0	-1.00453
SPI	Low	110480	836	18.95262	-0.02745
	Moderate	345058	2317	52.52777	-0.32826
	High	114692	1258	28.51961	0.464481
TWI	Low	367516	2908	65.92609	0.065489
	Moderate	168745	1338	30.33326	0.035558
	High	33960	165	3.740648	-0.49143
Lithology	Metamorphic Rocks	171877	699	15.83956	-0.83479
	Sedimentary Rocks	354444	3713	84.13777	1.179812
	Igneous rocks	43969	1	0.02266	-5.91805
Distance to the Fault (m)	< 500	114038	790	17.90165	-0.13753
	500 - 1000	155567	1271	28.80127	0.076077
	1000 - 1500	134176	1088	24.65443	0.062111
	1500 - 2000	91368	608	13.77748	-0.17855
	2000 <	75141	656	14.86517	0.141429
LandCover	Bare Ground	422378	3633	82.32495	0.479294
	Snow	138535	778	17.62973	-0.41227
	Vegetation	7434	2	0.045321	-3.38283
Distance to the Road (m)	250 <	440357	2418	54.79266	-1.03824
	<50	28056	497	11.26218	0.907884
	50 - 100	27151	458	10.37843	0.849789
	100 - 150	26098	363	8.225697	0.631911
	150 - 250	24946	352	7.976433	0.645886
	200 - 250	23682	325	7.364605	0.613342

Once the weight of each factor was calculated using the above equation, with the simple summation of all the factors the Landslide Susceptibility indexation map would be extracted

$$LS_{WOE} = Wc_1 + Wc_2 + Wc_3 + \cdots + Wc_n \qquad (8)$$

using (Eq 8).
LS is the landslide susceptibility index, (Wc) is the weightof each factor's classes. LS is representing the relative hazard of landslide occurrence (Figure 9).

International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

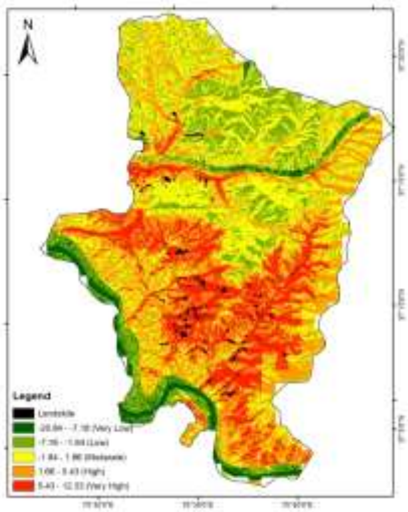


Figure 9. Illustrates the result of the Weight of Evidence Method for the landslide susceptibility mapping of Faizabad City, Afghanistan

Both of the bivariate statistical methodshas some advantages such as, the model can identify the influence of each class within the factor on landslides. Moreover, the method can be used for both scale and categorical factors. In another hand, it cannot identify the possible relationship between the factors (the relation between the slope angle and lithology, which both factors are important in slope failure prediction) and the entire factor which will be used in the analysis should be conditional independent.

IJSER

3.3 Logistic Regression (LR) Method

The principle of logistic regression (LR) rests on the analysis of a problem, in which a result measured with variables such as 0 and 1 or true and false, is determined from one or more independent factors (Menard, 1995). In the case of landslide susceptibility mapping, the goal of LR would be to find the best fitting model to describe the relationship between the presence and absence of landslides in a set of independent parameters such as slope angle, aspect, lithology, etc.
LR does not define susceptibility directly like WOE and FR approaches but an inference can be made using the probability. One of the biggest limitations of this method is that the method cannot be calculated for the categorical data. Therefore, the categorical data in this method was removed from the calculation, which has a significant effect in to the result. Generally, LR involves fitting the dependent variable using an equation below.

$$Y = Logit(p) = ln\left(\frac{p}{1-p}\right) = C_0 + C_1X_1 + C_2X_2 + \cdots + C_nX_n \quad (9)$$

Where p is the probability that the dependent variable (Y) is 1, p1−p is the so-called odds or likelihood ratio, C0is the intercept, and C1,C2……Cn are coefficients, which measure the contribution of independent factors X1,X2……Xnto the variations in Y Considering the above equation, a code was designed in the Matlab platform and the C0is the intercept, and coefficients for each factor was calculated as table below

Independent factors	Coefficients
Constant	-0.482300
TWI	0.014000
SPI	0.059700
Slope	0.011300
Roads	-0.000656
River	0.002600
Curvature	0.002600
Elevation	-0.000652
Fault	0.000003
Precipitation	0.001900

Once we got the constant and the coefficients with the simple summation and multiplication of the independent factors with its coefficients, we can get the result where Y is representing the relative hazard of landslide occurrence (Figure 10).

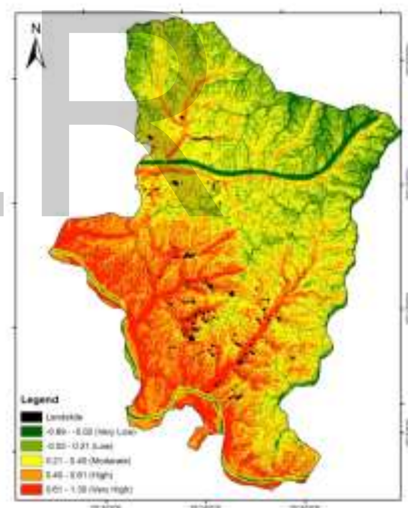


Figure 10. Illustrates the result of the Logistic Regression Method for the landslide susceptibility mapping of Faizabad City, Afghanistan

The problems on identifying the relationship between factors which existed in bivariate statistical method has been solved in model and the output values represents a meaningful probability in a susceptibility map but the model has some weakness such as, the model cannot identify the influence of each class with a factor on landslide and the categorical data can be calculated or if the categorical data has been changed to scale but still too much of it can create immense problems.

3.4 Combination of methods

The bivariate analysis is a quantitative method that applies bivariate data and then makes comparisons in order to find

International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

any significant relationships. Meanwhile, multivariate analysis is a method that simultaneously observes and analyzes two or more variables of interest (Nguyen 2014) and to overcome with the strength and weakness of the bivariate and multiversity statistical methods, the combination is used and in the process the above mentioned problems will be solved or in another hand, they can become complimentary for each other. As an example, the multivariate has the limitation of using the categorical data but in bivariate methods, such a problem does not exist. Therefore, in this study, we have used the Combination technique to increase the accuracy of the analysis.

3.4.1 Combination of Frequency ratio method and Logistic regression.

Combination of the methods follows the same path as the main approach follows. In the combination of frequency ratio method, firstly the distribution of landslide in each factor's class will be calculated. It is normally using the ratio of landslide in a class to the total landslide ratio in the area. Once the Fr weight index for each factor's class is calculated then sample points will be taken from the data for the logistic regression method. The combined method is following the same role as the logistic regression method.

$$Y_{FR} = C_0 + C_1 Fr_1 + C_2 Fr_2 + \cdots + C_n Fr_n \qquad (11)$$

C_0 is the constant, and C_1, C_2……C_n are coefficients, which measure the contribution of independent factors Fr_1, Fr_2……Fr_n to the variations in Y. To get the constant and the coefficient for the equation, the designed code which was used to calculate the logistic regression approach is used (Figure 11).

IJSER

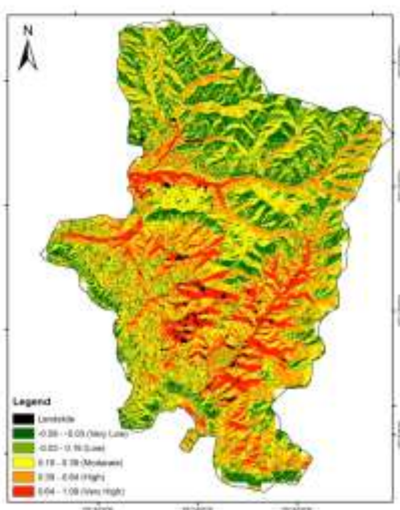


Figure 11. Illustrates the Combination of Frequency Ratio and Logistic Regression Methods for the landslide susceptibility mapping of Faizabad City, Afghanistan.

3.4.2 Combination of Weight of evidence and Logistic regression

Similarly, the combine method of weight of evidence and logistic regression method follows the same role. First, the weight will be calculated from each factors class, same as the weight of evidence method (Eq. 34,35,36). Once the weight is calculated then the result of the weight of evidence method will be used in the logistic regression method to find the best fit model to separate the landslide from non-landslide in a set

$$Y_{WOE} = C_0 + C_1 \text{Weight}_1 + C_2 \text{Weight}_2 + \cdots + C_n \text{Weight}_n \qquad (11)$$

of independent parameters.
C_0 is the intercept, and C_1, C_2……C_n are coefficients, which measure the contribution of independent factors Weight_1, Weight_2……Weight_n to the variations in Y (Figure 12).

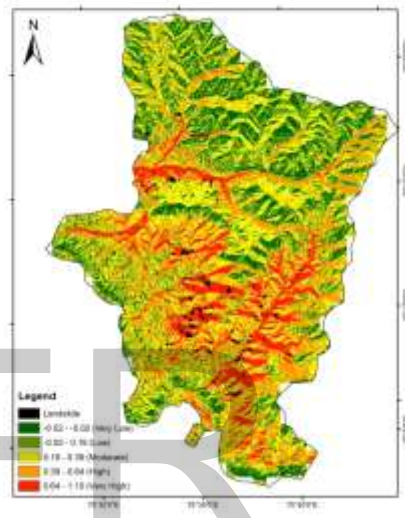


Figure 12. Illustrates the Combination of Weight of Evidence and Logistic Regression Methods for the landslide susceptibility mapping of Faizabad City, Afghanistan.

The higher result values, the higher risk of slope failure (Saro and Biswajeet, 2007), therefore, the results were classified into five different classes "Very Low, Low, Moderate, High, Very High" which represents the level of unstable locations in the maps. Moreover, visually looking at the maps, it shows that all past occurred landslides were located in the two "high and very high risk" classes which indicate the high accuracy of the analysis.

4 Result and Discussion

To define the accuracy of the result, the matrix validation method is used, this method not only gives the success rate of a methodology but it gives the false alarm rate and miss alarm rate. To operate the matrix validation method a contingency table (Table 4) was built from the control point as below.

Table 4. Contingency Table Used to Validate the Result

	Predicted		
	Total	Landslide	No landslide
Actual	Landslide	Success(A)	Miss-alarm(B)
	No landslide	False-alarm (C)	Success(D)

International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

Considering the contagious table, three indexes (1) success rate (Eq.13), (2) miss alarm rate (Eq.14)and (3) false alarm rate (Eq.15) can be evaluated for efficiency.

$$Success\ Rate = \frac{A+B}{A+B+C+D} \qquad (13)$$

$$Miss\ Alarm\ Rate = \frac{B}{A+B} \qquad (14)$$

$$Fale\ Alarm\ Rate = \frac{C}{A+C} \qquad (15)$$

Success rate shows the percentage of the points that are correctly classified, miss alarm rate shows the percentage of the points that are landslide occurrence but predicted as a non-landslide that is an important rate for landslide hazard mapping. The higher the miss alarm rate values, the higher number of landslide points are predicted as non-landslide. On the other hand, false – alarm rate shows the percentage of the non-landslide points incorrectly classified as a landslide. The higher the false alarm rate, the higher false information in the landslide hazard prevention.

To validate the landslide susceptibility map and find the suitable critical boundary for the very high-risk zone 11652 control point equally distributed from the landslide and non-landslide area is taken to define the accuracy only success rate is shown (Figure 13).

Figure 13. Illustrates the validation result of the methods.

The result (Figure 13) shows that all the methodologies used in this study is giving an acceptable accuracy and the combination of the bivariate statistical methods "Frequency Ratio and Weight of Evidence" and multivariate statistical method "Logistic Regression" increases the accuracy of the analysis and they are more reliable than the methods alone and the combination method of weight of evidence (WOE) and logistic regression (LR) is more reliable because of its higher success rate.

However, for a hazard mapping the used classification is not enough because its associated with the higher value of miss alarm rate and the higher the false alarm rate, the more we are miss predicting the landslide. Moreover, the classification has to be with an acceptable success rate, miss alarm rate, and false alarm rate based. Therefore, designed tool was used to calculated the three rate for all of the value of the threshold which will allow the user to define a suitable critical boundaries for the susceptibility map (Table 5).

Table 5. Shows the Matrix Validation Result of (WOE) and (LR) Combined Methods

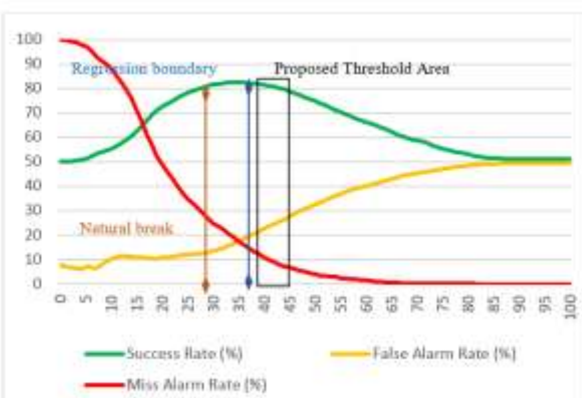
Percentage	Threshold	Success Rate (%)	False Alarm Rate (%)	Miss Alarm Rate (%)
0	1.092	50.008	0	99.983
1	1.076	50.051	20	99.864
2	1.06	50.212	6.897	99.542
3	1.044	50.466	6.349	98.999
4	1.028	50.857	6.087	98.168
5	1.012	51.213	7.186	97.371
6	0.996	52.061	6.762	95.556
7	0.98	53.155	6.542	93.215
8	0.964	53.944	7.957	91.366
9	0.948	54.529	9.174	89.924
10	0.932	55.377	10.076	87.888
11	0.916	56.446	10.825	85.327
12	0.9	57.549	11.438	82.663
13	0.884	58.999	11.249	79.389
14	0.868	60.534	11.139	75.912
15	0.852	62.697	10.914	71.06
16	0.836	64.665	10.758	66.65
17	0.82	67.082	10.448	61.323
18	0.804	69.033	10.437	56.913
19	0.788	71.103	10.357	52.282
20	0.772	72.46	10.735	48.94
21	0.756	73.834	10.842	45.734
22	0.74	74.979	11.219	42.816
23	0.724	76.2	11.532	39.746
24	0.708	77.294	11.8	36.98
25	0.692	78.304	11.863	34.589

IJSER

International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

26	0.676	78.94	12.106	32.875
27	0.66	79.72	12.306	30.857
28	0.644	80.475	12.736	28.634
29	0.628	81.213	13.023	26.582
30	0.612	81.747	13.526	24.733
31	0.596	81.934	14.15	23.528
32	0.58	82.12	14.811	22.239
33	0.564	82.349	15.615	20.611
34	0.548	82.536	16.374	19.084
35	0.532	82.434	17.714	17.337
36	0.516	82.349	18.562	16.2
37	0.5	82.332	19.524	14.623
38	0.484	82.197	20.528	13.181
39	0.469	81.79	21.675	12.095
40	0.453	81.493	22.646	10.941
41	0.437	81.128	23.65	9.805
42	0.421	80.653	24.699	8.77
43	0.405	80.271	25.538	7.854
44	0.389	79.61	26.548	7.26
45	0.373	78.753	27.709	6.751
46	0.357	77.998	28.717	6.226
47	0.341	77.235	29.741	5.547
48	0.325	76.438	30.738	4.936
49	0.309	75.717	31.629	4.292
50	0.293	74.919	32.499	3.885
51	0.277	74.071	33.396	3.444

Based on the (**Table 6**) result, figure below can be ploted.



Based on the result the threshold used in the natural break classification of GIS (**Figure 14**a) not only it is not displaying the high but has 80.475% of success rate, 28.634% of miss alarm rate and about 12.736% of false alarm rate. However, the 0.5 the regression critical boundary is giving about 82.332% of success rate, 14.623% of miss alarm rate and about 19.524% which is the closest acurecy to the highest success rate but it is just one critical boundary it does not provide the information for the other classes of landslide susceptibility mapping, therefore, using the result obtained from the matrix validation method the threshold can be selected based on the study propose in an area. If the study area is in a city with a high population then the miss alarm rate has to lower, on the other hand, for a rural area the higher miss alarm rate is acceptable unless there is no financial threat. Based on the result the (0.453) is usedas a threshold which has 81.493% of success rate, 10.941 % of miss alarm rate and about 22.646% of false alarm rate for the study area (**Figure 14**b) and the rest of the classification was selected using the result of designed tool.

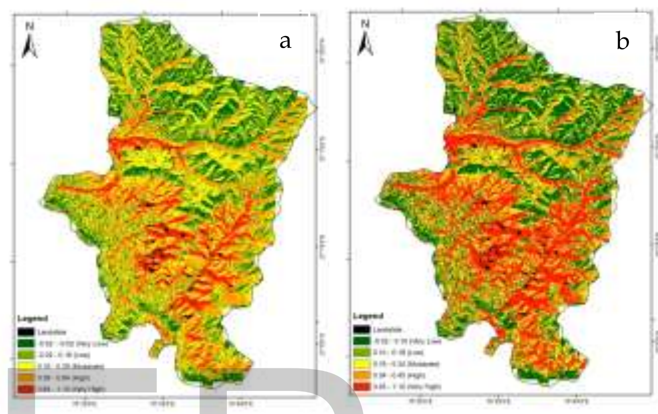


Figure 14. shows a) Illustrates the Natural Break Classification of Combined Frequency Ratio and Logistic Regression Methods for the Landslide Susceptibility Mapping of Faizabad City Afghanistan.
b) Illustrates Modified Map of Combined Frequency Ratio and Logistic Regression Methods for the Landslide Susceptibility Mapping of Faizabad City Afghanistan.

5 Conclusion

Every year thousand of people die, Injure and lose their property because of the unexpected landslide and the phenomena not only affect people but also affects the economy, damage buildings, lifeline and it will damage everything that comes to its path. To take a step toward the hazard mitigation and to prevent such a catastrophe, identification of the potentially hazardous locations are needed.

Numbers of statistical methods have been proposed however very few proposed one specificreliable method, therefore, this study was performed in search of one reliable method for landslide susceptibility mapping. Therefore, the bivariate statistical methods “Frequency Ratio and Weight of Evidence”, the multivariate statistical method “Logistic Regression” and their combination were used to create a landslide susceptibility map with a higher accuracy for the study area and the accuracy of the obtained result was examined using the matrix validation method. From results, it can be concluded that however all of the used methodologies give an acceptable result but the combination of bivariate statistical methods and multivariate statistical method gives a higher accuracy for analysis and they are complimentary for each other. Moreover, the combination of Weight of evidence (WOE) and Logistic Regression(LR) methods are more reliable.

IJSER

International Journal of Scientific & Engineering Research Volume 12, Issue 11, November-2021
ISSN 2229-5518

Furthermore, the accuracy of landslide susceptibility map not only depends on the quality of the raw data and the methodologies used but defining a suitable a classification to separate the landslide from the non-landslide is also important for a landslide susceptibility mapping and as concusion the new designed tool to increase the classification accurecy of landslide susceptibility map is proved to be usefull.

Acknowledgment

The authors wish to thank Mr. Thomas Mack, Program Chief of the U.S. Geological Survey's Afghanistan Project, for the support which made it possible to publish this article.

References

[1] Bourenane H., BouhadadY., Guettouche M.S.& Braham M. (2014) GIS-Based Landslide Susceptibility zonation using bivariate statistical and expert approaches in the city of Constantine (northeast Algeria). Bull Eng Geol Environ Doi 10.1007/s10064 – 014 – 0616.

[2] Bui, D.T., Pradhan, B., Lorfman, O., Revhaug, I., and Dick, O.B, 2012, Landslide susceptibility assessment in the Hoa Binh province of Vietnam: A comparison of the Levenberg–Marquardt and Bayesian regularized neural networks. Geomorphology, 171–172, 12–29. doi:10.1016/j.geomorph.2012.04.023

[3] Dieu TB, Owe L, Inge R, Oystein D(2011) Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression nat hazard 59:1413 – 14444.doi:10.1007/s11069 – 011 – 9844 – 2

[4] Dupree, Nancy Hatch (1977): An Historical Guide to Afghanistan. 1st Edition: 1970. 2nd Edition. Revised and Enlarged. Afghan Tourist Organization. (Chap. 28).

[5] Evert Hoek 2006, "Practical Rock Engineering" / Dr. Evert Hoek, Consulting Engineer Inc. 102 – 3200 Capilano Crescent North Vancouver, British Columbia Canada V7R 4H7

[6] G.F. Bonham-Carter Geographic Information Systems for Geoscientists, Modeling with GIS Pregame on Press, Oxford (1994) p. 398

[7] Gessler, P.E.; Moore, I.D., McKenzie N.J., Ryan P.J. (1995) International Journal of GIS 9, p. 421-432Netra

[8] Ghimir, M., 2001. Geo-Hydrological hazard and risk zonation of Banganga watershed using GIS and remote sensing. Journal of Nepal geological Society 23, 99-110

[9] Greenway, D.R., 1987. Vegetation and slope stability. In: Anderson, M.G., Richards, K.S.(Eds.), Slope stability. Wiley, New York, pp.187-230.

[10] Highland, L.M., and Bobrowsky, Peter, 2008, The landslide handbook—A guide to understanding landslides: Reston, Virginia, U.S. Geological Survey Circular 1325, 33 p.

[11] H. Vijith . G. Madhu, 2008, Estimating potential landslide sites of an upland sub-watershed in Western Ghat's of Kerala (India) through frequency ratio and GIS. Environ Geol (2008) 55:1397–1405, DOI 10.1007/s00254-007-1090-2

[12] Magliulo P, Di Lisio A, Russo F, Felano A(2008) Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: a case study in southern Italy. Nat Hazards 47:411 – 435.doi:10.1007/s11069 – 008 – 9230 – x

[13] Oguchi, T., 1997. Drainage density and relative relief in humid steep mountains with frequent slope failure. Earth Surface Processes and Landforms 22, 107-120

[14] Pradhan B, Lee S(2010a) Landslide Susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environ Modelling Softw 25:747-759

[15] S. Lee, J.A. Talib Probabilistic landslide susceptibility and factor effect analysis Environmental Geology, 47 (2005), pp. 982-990

[16] Saro Lee. Biswajeet Pradhan 2007 Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. Landslides (2007) 4:33–41, DOI 10.1007/s10346-006-0047-y

[17] Shabanzadeh Karim, 2011, Zoning Landslide by Use of Frequency Ratio Method (Case Study: Deylaman Region). Middle-East Journal of Scientific Research 9 (5): 578-583, 2011, ISSN 1990-9233, IDOSI Publications, 2011

[18] Shahabi, H., Hhezri, S., Ahmhad, B.B., and Hashim, M., 2014, Landslide susceptibility mapping at central Zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio, and logistic regression models. Catena, 115, 55–70.

[19] Stocking, M.A., 1972. Relief analysis and soil erosion in Rhodesia using multivariate techniques. Zeitchrift für geomorphologie 16, 432-443

[20] Styczen, M.E., Morgan, R.P.C., 1995. Engineering properties of vegetation.In: Morgan, R.P.C., Rickson, R.J.(Eds.), Slope Stabilization and Erosion Control: a bioengineering Approach. E&FN Spon, London, pp.5-58.

[21] USGS Earthquake Hazards Program: Earthquake Report: HINDU KUSH REGION, AFGHANISTAN"115 Killed & 538 injured; 7,630 homes,12 schools, 17 mosques, 20 office buildings have been damaged in 9 provinces by #AfghanistanEarthquake". president.gov.af. 27 October 2015.

[22] USGS Open-File Report 2007–1214, Afghanistan Geological Survey Open-File Report, 2007-1214, USGS Afghanistan Project Product No. 157

[23] Wilson, J.P., and Lorang M.S. 2000. Chapter 6, Spatial Models of Soil Erosion and GIS. Spatial Models and GIS: New Potential and New Models, 83-86.

[24] Yalcin, 2011. Volume 85, A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey, 274–287

[25] https://now.mmedia.me/lb/en/archive/545658-at-least-350-dead-as-landslide-hits-afghan-villages Published: 2/05/2014 08:38 PM

[26] http://america.aljazeera.com/articles/2015/4/28/afghan-official-at-least-52-dead-in-landslide-in-northeast.html April 28, 2015 7:45AM ET

[27] http://www.noaa.gov 1964-1983 report

[28] GPCC "http://gpcc.dwd.de"

[29] USGS Earth explorer Archive "https://earthexplorer.usgs.gov/"

IJSER