Intercontinental Geoinformation Days

igd.mersin.edu.tr

GIS-Based landslide susceptibility mapping using weight of evidence (WoE) and random forest (RF)

Aslan Cihat Basara *[1], **Yasemin Sisman** [2]

[1]Ondokuz Mayis University, Institute of Graduate Studies, Department of Geomatics Engineering, Samsun, Turkey
[2]Ondokuz Mayis University, Faculty of Engineering, Department of Geomatics Engineering, Samsun, Turkey

Keywords
Landslide
Susceptibility Map
Weight of Evidence
Random Forest
Machine Learning
GIS

ABSTRACT

In this study, the Landslide Susceptibility Map of Tokat province was produced. Slope classes, elevation classes, land use classes, geology classes, aspect classes and proximity to fault lines classes were used during the study. The Weight of Evidence method was applied to determine the relationship between the classes of the parameters and the landslide events. Random Forest method was used to determine the weights between parameters. Weighted Overlay operation was applied to the classified and weighted map data using ArcGIS program. As a result of the process, the data were divided into 5 classes and the Landslide Susceptibility Map was produced. When susceptibility classes are examined, it was seen that 92,42% of the old landslide events occurred in high and very high classes.

1. INTRODUCTION

Disasters are events that cause material and moral damages in the society they affect and cause great problems in terms of the consequences they cause in the flow of daily life. The landslides can be defined as the downward movement or sliding of parts such as soil and rocks, under the influence of gravity or external factors such as earthquakes and continuous rains (AFAD 2014).

When the negative effects caused by landslides are carefully examined, it is necessary to first reveal the spatial distribution of existing mass movements and inventory information. Using the available inventory data, landslide susceptibility analysis, risk and hazard values can be determined (Van Westen et al. 2008). Landslide susceptibility analysis, which reveals areas susceptible to possible future landslides, reveals the desire for any landslide to ocur (Guzzetti et al. 2006). Landslide susceptibility maps are of great importance in predicting future landslides and providing land use planning (Basara et al. 2020)

In this study, the Landslide Susceptibility Map of Tokat Province was produced. Location Map given in Figure 1.

Figure 1. Location Map

2. MATERIAL AND METHOD

2.1. Material

There is no standard for the parameters to be used in landslide susceptibility analysis studies. Therefore, the parameters may differ depending on the area to be studied. When the parameters used in the landslide susceptibility analysis were analyzed statistically, the rates in Table 1 were obtained (Tetik Biçer 2017).

* **Corresponding Author**

* (aslancihatbasara@gmail.com) ORCID ID 0000 – 0001 – 6644 – 6097
(ysisman@omu.edu.tr) ORCID ID 0000 – 0002 – 6600 – 0623

Cite this study

Basara A.C & Sisman Y (2021). GIS-Based landslide susceptibility mapping using weight of evidence (WoE) and random forest (RF). 2nd Intercontinental Geoinformation Days (IGD), 72-75, Mersin, Turkey

Table 1. Usage Rates of Parameters

Parameters	Rate (%)	Parameters	Rate (%)
Slope	86,47	Land Use	46,62
Lithology	67,29	Curvature	40,60
Aspect	59,77	Fault Lines	28,57
Elevation	55,64	NDVI	24,06
Drainage Density	50,75	Soil	23,68

In this study, Slope, Aspect, Elevation, Geology, Land Use, Proximity to Fault Lines and Landslide Inventory Map were used. Maps of the material are given in Figure 2-8.

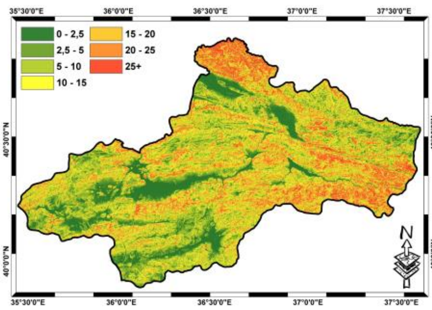

Figure 2. Slope Map

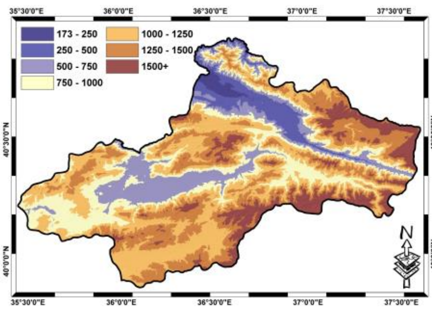

Figure 4. Elevation Map

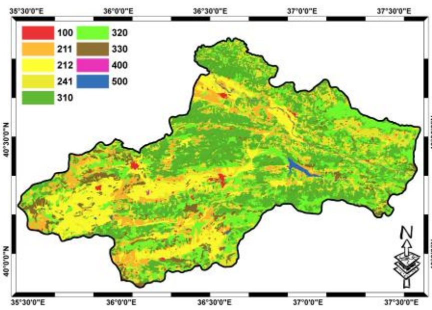

Figure 6. Landuse Map

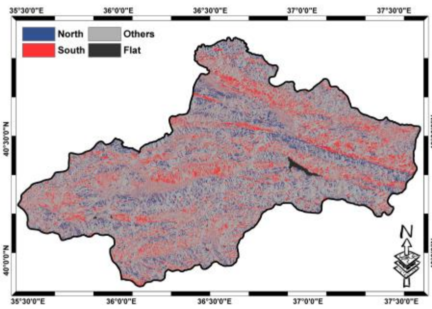

Figure 3. Aspect Map

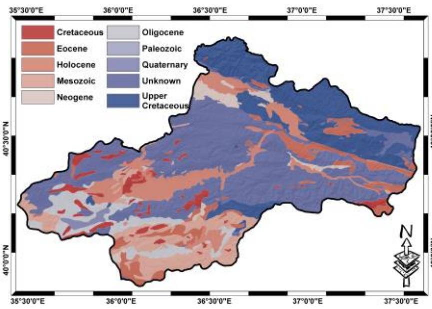

Figure 5. Geology Map

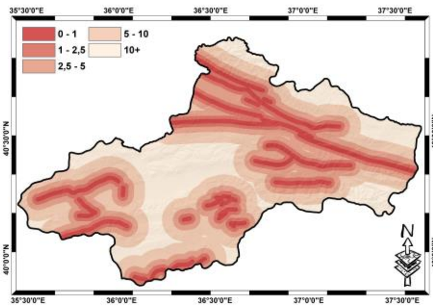

Figure 7. Proximity to Fault Lines Map

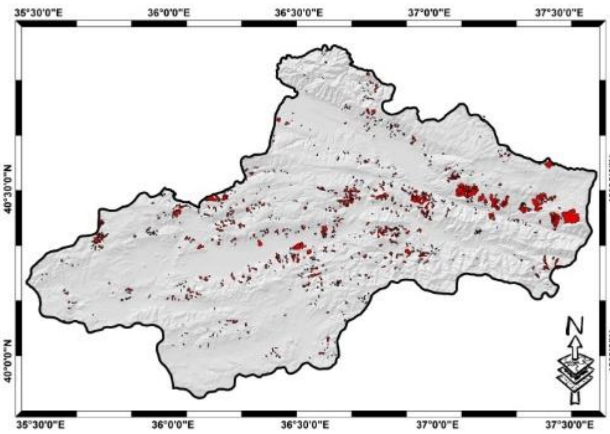

Figure 8. Landslide Inventory Map

2.2. Method

In this study, obtaining the landslide susceptibility map was applied in two stages. In the first part, the Weight of Evidence (WoE) method was applied. In the second part, the weights of the parameters are determined by the Random Forest (RF) algorithm.

2.2.1. Weight of Evidence (WoE)

The Weight of Evidence method has been mathematically expressed by Van Westen et al. (2003) and Regmi et al. (2010). In this study, the weights of the subcategories of the factors affecting the landslide were determined using the equation 1-3 (Regmi et al. 2010; Ozdemir and Altural 2013).

$$W+ = \ln[\,(A1/(A1+A2))\;/\;(A3/(A3+A4))\,] \quad \textbf{(1)}$$
$$W- = \ln[\,(A2/(A1+A2))\;/\;(A4/(A3+A4))\,] \quad \textbf{(2)}$$
$$C = (W+) - (W-) \quad \textbf{(3)}$$

In the equation, A1 refers to the landslide areas in a selected subcategory, A2 refers to the total landslide areas outside the selected category, A3 refers to the areas with no landslides in the selected category, and A4 refers to the total landslide-free areas other than the selected category. While A1 + A2 refers to the total landslide areas, A3 + A4 refers to the total landslide-free areas in the study area. (Regmi et al. 2010).

The difference between the W + and W- weights is called the contrast of the weights (C). The C value shows the final positional relationship between the landslide event and the forecast variable. A value equal to zero indicates that the subcategory of the factor causing the landslide is not important for the analysis. Positive contrast indicates a positive positional relationship, negative contrast indicates the opposite (Ozdemir and Altural, 2013).

Table 2. Aspect Classes

Attribute	Landslide area	Total area	WoE
North	46,76 km^2	1317,89 km^2	0,0945
South	48,81 km^2	1306,85 km^2	0,1566
Others	232,66 km^2	7349,36 km^2	-0,1310
Flat	0,28 km^2	43,39 km^2	-1,6471

Table 3. Geology Classes

Attribute	Landslide area	Total area	WoE
Cretaceous	6,07 km^2	407,14 km^2	-0,8300
Eocene	38,78 km^2	783,26 km^2	0,4744
Holocene	8,54 km^2	1043,15 km^2	-1,5004
Mesozoic	10,74 km^2	971,05 km^2	-1,1804
Neogene	3,36 km^2	250,12 km^2	-0,9289
Oligocene	3,52 km^2	649,02 km^2	-1,8865
Paleozoic	0,26 km^2	3,43 km^2	0,8992
Quaternary	0,06 km^2	72,22 km^2	-3,7744
Unknown	155,76 km^2	3944,05 km^2	0,3386
Upper Cretaceous	101,68 km^2	1897,02 km^2	0,6777

Table 4. Slope Classes

Attribute	Landslide area	Total area	WoE
0 – 2,5 degree	4,31 km^2	1220,36 km^2	-2,3784
2,5 – 5 degree	17,53 km^2	1169,98 km^2	-0,8736
5 – 10 degree	94,69 km^2	2432,80 km^2	0,2415
10 – 15 degree	104,96 km^2	2119,40 km^2	0,5816
15 – 20 degree	66,70 km^2	1510,13 km^2	0,3752
20 – 25 degree	26,93 km^2	880,08 km^2	-0,0781
25 degree+	13,39 km^2	684,72 km^2	-0,5610

Table 5. Elevation Classes

Attribute	Landslide area	Total area	WoE
173 – 250 m	0,14 km^2	131,32 km^2	-3,5079
250 – 500 m	5,52 km^2	528,47 km^2	-1,2061
500 – 750 m	38,39 km^2	1240,21 km^2	-0,0677
750 – 1000 m	103,66 km^2	1811,46 km^2	0,7679
1000 – 1250 m	116,33 km^2	2936,22 km^2	0,2896
1250 – 1500 m	46,35 km^2	2324,58 km^2	-0,6265
1500 m+	18,16 km^2	1048,20 km^2	-0,7091

Table 6. Land Use Classes

Attribute	Landslide area	Total area	WoE
CORINE.100	4,14 km^2	135,52 km^2	-0,0759
CORINE.211	39,25 km^2	808,52 km^2	0,4526
CORINE.212	12,31 km^2	1337,34 km^2	-1,4041
CORINE.241	134,27 km^2	1919,02 km^2	1,1182
CORINE.310	77,45 km^2	3105,39 km^2	-0,3883
CORINE.320	49,65 km^2	1966,29 km^2	-0,3261
CORINE.330	11,25 km^2	669,71 km^2	-0,7213
CORINE.400	0,00 km^2	2,38 km^2	-11,2991
CORINE.500	0,31 km^2	72,31 km^2	-2,0788

Table 7. Proximity to Fault Lines Classes

Attribute	Landslide area	Total area	WoE
0 - 1 km	29,84 km^2	1094,26 km^2	-0,2117
1 - 2,5 km	52,81 km^2	1563,53 km^2	0,0358
2,5 - 5 km	91,08 km^2	2112,37 km^2	0,3746
5 - 10 km	84,38 km^2	2829,21 km^2	-0,1346
10 km+	70,61 km^2	2421,09 km^2	-0,1573

2.2.2. Random Forest (RF)

Random Forest Method is one of the collective learning algorithms based on using many decision tree models together to solve a specific classification and regression problem (Breiman 2001). The algorithm is based on the principle of combining the estimates made by each of the decision trees that make up the forest and making the final decision for the relevant sample in the process of estimating a sample with an unknown class label (Kuncheva and Whitaker 2003).

The general formula of the Random Forest algorithm is defined as in Equation 4. Since the algorithm produces K number of decision trees, the predicted value (P) is given by the average of the predicted values (T) in all trees (Costa et al. 2020). Generalization error in Random Forest algorithm is defined as in Equation 6. The "x and y" values here are the landslide conditioning factors showing the x-y space and the probability above mg and are defined as in Equation 5-6. The "I" values here measure the extent to which the average number of votes in random vectors exceeds the average vote for any other output for correct output (Masetic et al. 2016).

$$P = \frac{1}{K}\sum_{k=1}^{K} T \quad \textbf{(4)}$$

$$\mathrm{GE} = P_{x,y}\ (\ \mathrm{mg}\ (\mathrm{x,y}) < 0\ ) \quad \textbf{(5)}$$

$$\mathrm{mg}(\mathrm{x,y}) = \mathrm{av_k I}(\mathrm{h_k}(\mathrm{x}) = \mathrm{y}) - \mathrm{max_{j \neq y} av_k I}(\mathrm{h_k}(\mathrm{x}) = \mathrm{j}) \quad \textbf{(6)}$$

Table 8. Random Forest Data

Parameters	Variable İmportance	Standard Deviation	Weight
Land Use	66,909	0,261	27 %
Aspect	10,073	0,036	4 %
Slope	40,407	0,172	16 %
Proximity to Faults	35,912	0,243	14 %
Geology	42,681	0,392	17 %
Elevation	52,508	0,215	21 %

3. RESULTS

The parameters to be used in the study were mapped with the help of ARCGIS. The relationship of the maps with the landslide inventory map was determined using the Weight of Evidence (WoE) Method. Maps were reclassified according to the analysis result. The Random Forest (RF) Algorithm was used to determine the importance of the parameters relative to each other. Finally, the Landslide Susceptibility Map was produced by processing the data with Weighted Overlay analysis. The map produced was reclassified 5 as very low, low, medium, high and very high. Landslide susceptibility map is given in Figure 9.

4. DISCUSSION AND CONCLUSION

The areas and rates of the landslide susceptibility classes are given in Table 9.

Table 9. Landslide Susceptibility Classes

	Landslide area (km^2)	Total area (km^2)	Landslide incident (%)	Total area (%)
Very Low	0,00	187,79	0,00%	1,88%
Low	0,82	878,77	0,25%	8,78%
Medium	24,13	2160,46	7,33%	21,58%
High	125,91	4905,13	38,25%	48,99%
Very High	178,35	1880,26	54,18%	18,78%

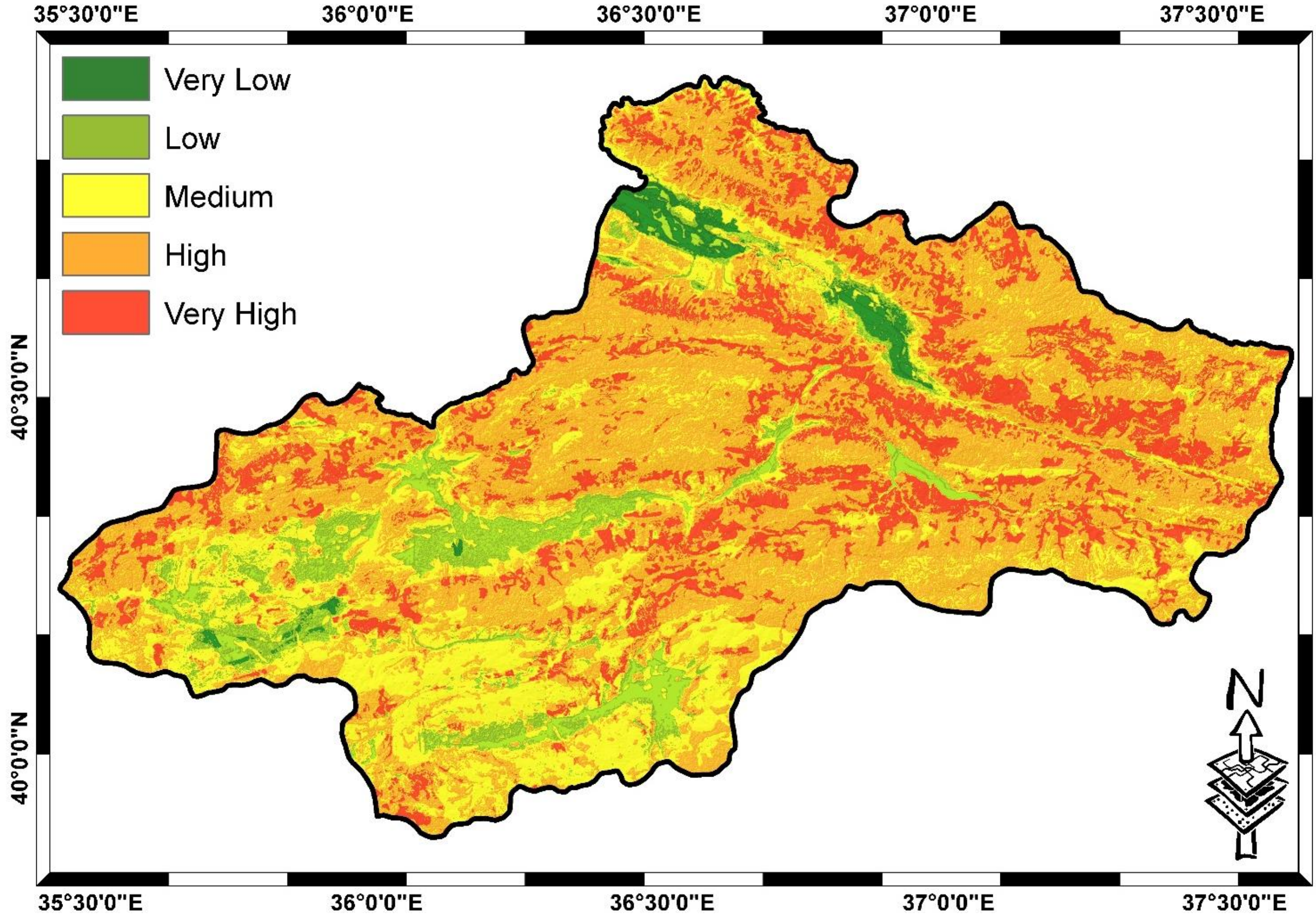

Figure 9. Landslide Susceptibility Map

When susceptibility classes are examined it was seen that 92,42% of the old landslide events occurred in high and very high class, 7,33% occurred in middle class and 0,25% occurred in low and very low class.

In the spatially analysis of landslide events, it was seen that the sensitivity classes are examined spatially, high-risk areas constitute 67,77% of all areas, medium-risk areas constitute 21,58% of all areas and low-risk areas constitute 10,65% of all areas.

As a result, it is possible to say the following. Susceptibility mapping is very important to prevent material and moral losses that may occur due to disasters.

REFERENCES

AFAD 2014. Açıklamalı Afet Yönetimi Terimleri Sözlüğü. TC Başbakanlık Afet ve Acil Durum Yönetimi Başkanlığı Deprem Dairesi Başkanlığı, Ankara.

Basara A C, Tabar M E & Sisman Y (2020). GIS-Based Landslide Susceptibility Mapping Using Frequency Ratio and AHP Methods in Samsun Province. Intercontinental Geoinformation Days (IGD), 223-226, Mersin, Turkey.

Breiman L (2001). Random forests. Machine learning, 45(1), 5-32.

Costa I S L, Serafim I C C D O, Tavares F M & Polo H J D O (2020). Uranium anomalies detection through Random Forest regression. Exploration Geophysics, 51(5), 555-569.

Guzzetti F, Reichenbach P, Ardizzone F, Cardinali M & Galli M (2006). Estimating the quality of landslide susceptibility models. Geomorphology, 81:1-2, 166-184.

Kuncheva L I & Whitaker C J (2003). Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Machine learning, 51(2), 181-207.

Masetic Z & Subasi A (2016). Congestive heart failure detection using random forest classifier. Comput. Methods Programs Biomed. 2016, 130, 54–64.

Ozdemir A & Altural T (2013). A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. Journal of Asian Earth Sciences, 64, 180-197.

Regmi N R, Giardino J R & Vitek J D (2010). Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. Geomorphology, 115(1-2), 172-187.

Van Westen C, Rengers N & Soeters R (2003). Use of geomorphological information in indirect landslide susceptibility assessment. Natural Hazards, 30:3, 399-419.

Van Westen C J, Castellanos E & Kuriakose S L (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. Engineering Geology, 102:3-4, 112-131.

Tetik Biçer Ç (2017). Heyelan Risk Haritalaması Üzerine Yarı Sayısal Bir Değerlendirme.