Advanced GIS

http://publish.mersin.edu.tr/index.php/agis/index

e-ISSN:2822-7026

Landslide susceptibility mapping of Tokat (Turkey) province using weight of evidence and random forest

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

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

Keywords
Landslide
Susceptibility Map
Weight of Evidence
Random Forest
Machine Learning

ABSTRACT

Landslides are one of the important disasters that have negative effects on people. In this study, the Landslide Susceptibility Map of Tokat (Turkey) 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.

Research Article
Received: 13/09/2021
Revised: 04/10/2021
Accepted: 06/10/2021
Published: 30/09/2021

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 examining the negative effects caused by landslides, it is necessary to know the spatial distribution and inventory information of past landslides. 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 take place (Guzzetti et al., 2006). Landslide susceptibility maps are of great importance in predicting future landslides and providing land use planning (Basara et al., 2020).

Weight of Evidence (WoE) and Random Forest (RF) were used as methods in the study. Slope, Aspect, Elevation, Geology, Land Use, Proximity to fault lines were used as materials. As a result of this study, the landslide susceptibility map divided into 5 sub-sections was produced. The produced map was compared with the previous landslide events in the region. According to this comparison, an accuracy of 92,42% was found.

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

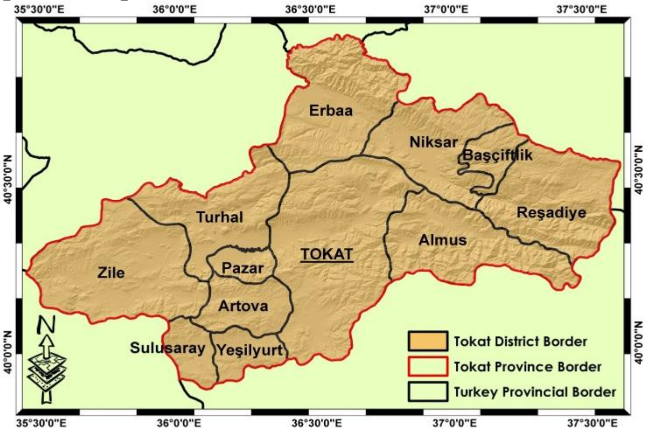

Figure 1. Location Map

* **Corresponding Author**

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

Cite this article

Başara, A. C., & Şişman, Y. (2021). Landslide susceptibility mapping of Tokat (Turkey) province using weight of evidence and random forest. *Advanced GIS,* 1(1), 1-7.

This study is an extended version of the paper presented at the 2nd IGD symposium (Basara & Sisman, 2021).

2. Method

Although there are many landslide susceptibility map applications in the literature (Aleotti & Chowdhury, 1999; Lee & Talib, 2005; Tetik Biçer, 2017), there is no consensus on the methods and parameters used in these applications. There are a lot of landslide susceptibility analysis methods like Frequency Ratio, Analytical Hierarchy Process, Weight of Evidence, Logistic Regression, Fuzzy Logic and Artificial Neural Networks (Basara et al., 2021).

In this study, the Weight of Evidence method which is one of the statistical methods and the Random Forest Algorithm which is one of the machine learning methods, were used together.

2.1. Weight of evidence method (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 & Altural, 2013).

$$W+ = \ln[\ (A1/(A1+A2))\ /\ (A3/(A3+A4))\ ] \quad (1)$$

$$W- = \ln[\ (A2/(A1+A2))\ /\ (A4/(A3+A4))\ ] \quad (2)$$

$$C = (W+) - (W-) \quad (3)$$

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 & Altural, 2013).

2.2. Random forest algorithm (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 & 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).

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

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

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

3. 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).

Table 1. Usage Rates of Parameters

Landslide Parameters	Usage Rate (%)	Landslide Parameters	Usage 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 Groups	23,68

GIS is important for collecting and processing geographic data of objects. Transforming data into geographic information with geographic analysis and viewing geographic data helps to plan activities (Basara et al., 2021).

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

3.1. Slope

Slope is the main stability parameter that affects shear and normal stresses on the surface. It is more common among researchers that the slope angle is directly proportional to the landslide risk (Karslı et al., 2009; Baeza & Corominas, 2001). Statistical analysis of slopes causing landslides should be made and a decision should be made accordingly (Basara, 2021).

3.2. Aspect

Aspect can be mentioned on the slopes of the same object facing different directions. Aspect is the parameter that shows the direction of the land surface relative to the sun's rays. The direction in which the tangent plane is facing at any point on the surface (Dağ, 2007).

3.3. Elevation

Topographic features vary with altitude. Elevation causes topographical differences in the study area. Altitude controls temperature and vegetation. Landslides, rock and soil properties and other geotechnical parameters are associated with altitude values (Guzzetti et al., 2009).

3.4. Geology

Landslide events are directly related to soil properties such as strength, permeability and hardness (Baeza & Corominas, 2001). Since the geological features will give important information about the landslide sensitivity of the study area, it should be evaluated correctly (Guzzetti et al., 1999).

3.5. Land use

The land use can be the reason of landslide events. Thus, the relationship between the areas like artificial, agricultural, forest, wetlands and water with sparse and dense vegetation and landslides should be evaluated (Basara et al., 2021).

3.6. Proximity to faults

Some landslides can be associated with fault lines areas because of weakness of the material surrounding them. The more buffer zone should be created, taking into account the different proximity for proximity to fault lines. (Wachal & Hudak, 2000). Some inferences can be made as a result of field observations. In this context, it was determined that most of the landslides occurred in regions very close to the faults. (Gökceoglu & Aksoy, 1996).

3.7. Landslide inventory

Landslide inventory is defined as data containing information about the location, type, activity and physical characteristics of landslides in a region. The information about past landslides are obtained as the first step of landslide susceptibility. It is thought that the future landslides may occur under conditions similar to the past landslides. (Varnes, 1984).

For this reason, the Landslide Inventory Map of the study area was created by using the landslide events 1950 - 2021.

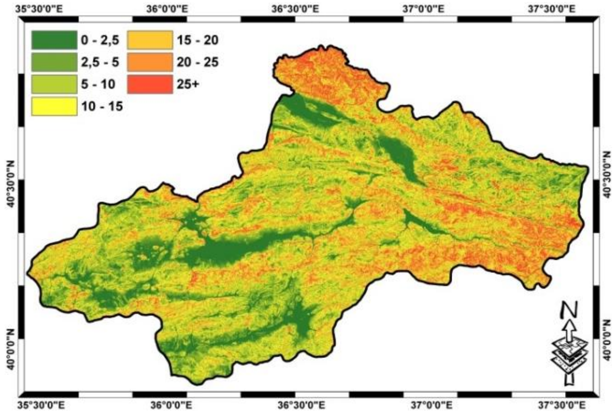

Figure 2. Slope Map

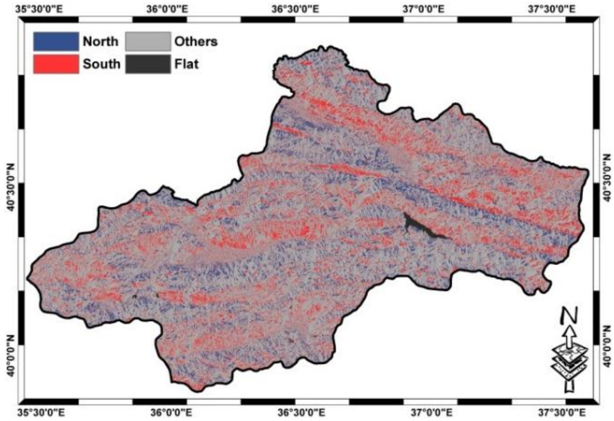

Figure 3. Aspect Map

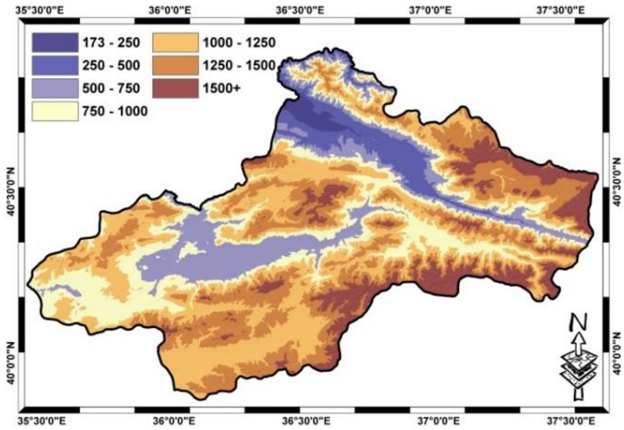

Figure 4. Elevation Map

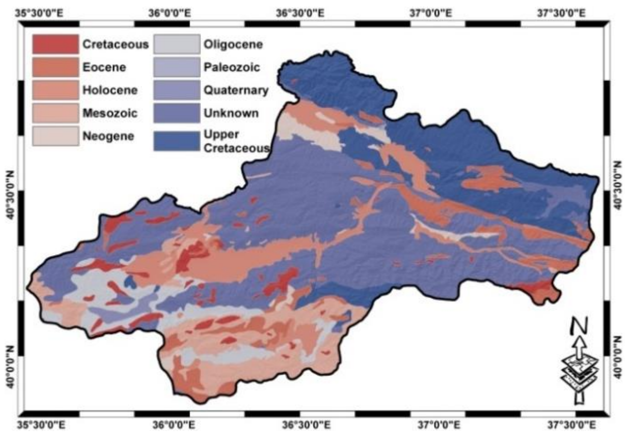

Figure 5. Geology Map

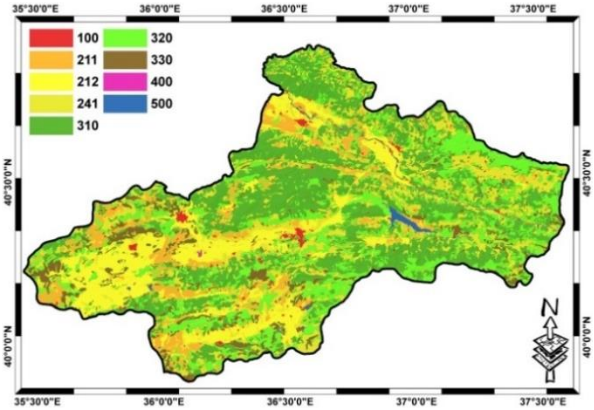

Figure 6. Land Use Map

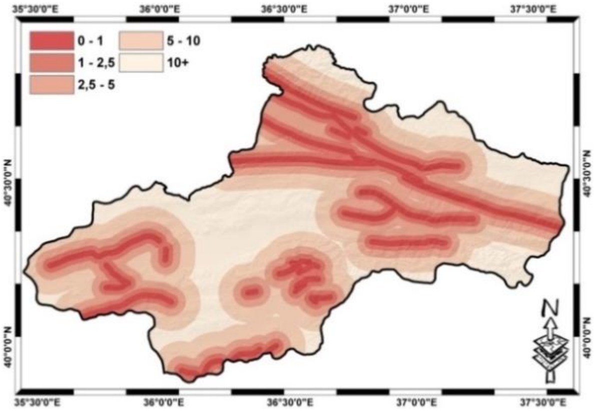

Figure 7. Proximity to Fault Lines Map

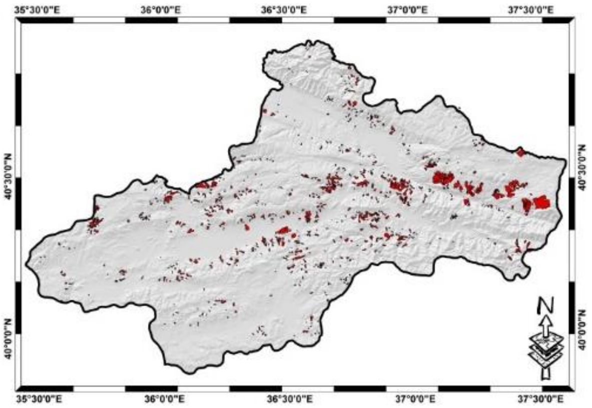

Figure 8. Landslide Inventory Map

4. Results

In this study, the landslide susceptibility map was obtained in two stages. In the first part, the Weight of Evidence (WoE) method was applied. Landslide impact priorities of parameter subclasses were determined. In the second part, the Random Forest (RF) algorithm is applied. Woe data was used in the application of RF. In this way, the priorities of the parameters among each other were determined.

The relationship of the maps with the landslide inventory map was determined using the Weight of Evidence (WoE) Method. The maps were reclassified according to the results of the analysis. The data obtained according to the Weight of Evidence (WoE) method are given in table 2-7.

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

The Random Forest (RF) Algorithm was used to determine the stature of the parameters relative to each other. In the implementation of this process, the data obtained as a result of the Weight of Evidence (WoE) method was used. The data obtained as a result of the Random Forest (RF) Algorithm are given in Table 8.

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 %

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.

5. Discussion

The created landslide susceptibility map was compared with the parameter classes used in the study. Risk values of parameter classes are given in Table 9.

Table 9. Risk Values of Parameter Classes

Parameters	Classes	Class Risk (%)	Class Area (%)
Slope	0 – 2,5 degree	12,49	12,18
	2,5 – 5 degree	45,14	11,66
	5 – 10 degree	74,53	24,31
	10 – 15 degree	84,92	21,14
	15 – 20 degree	87,10	15,08
	20 – 25 degree	82,15	8,81
	25 degree+	66,69	6,82
Elevation	173 – 250 m	0,40	1,30
	250 – 500 m	32,79	5,27
	500 – 750 m	53,79	12,37
	750 – 1000 m	75,92	18,08
	1000 – 1250 m	78,13	29,30
	1250 – 1500 m	67,97	23,21
	1500 m+	66,74	10,48
Aspect	North	68,45	13,18
	South	68,56	13,06
	Others	67,90	73,33
	Flat	0,80	0,43
Land Use	CORINE.100	41,43	1,35
	CORINE.211	76,66	8,08
	CORINE.212	5,45	13,35

Table 9 is continued.

	CORINE.241	93,81	19,15
	CORINE.310	81,41	31,01
	CORINE.320	70,90	19,63
	CORINE.330	47,25	6,69
	CORINE.400	0,00	0,02
	CORINE.500	0,59	0,72
Geology	Cretaceous	39,83	4,07
	Eocene	84,59	7,82
	Holocene	19,31	10,41
	Mesozoic	38,86	9,73
	Neogene	39,08	2,49
	Oligocene	18,85	6,50
	Paleozoic	100,00	0,03
	Quaternary	0,15	0,72
	Unknown	88,46	39,37
	Upper Cretaceous	88,51	18,84
Proximity to Fault Lines	0 - 1 km	60,49	10,91
	1 - 2,5 km	64,09	15,61
	2,5 - 5 km	72,25	21,09
	5 - 10 km	67,79	28,26
	10 km+	69,50	24,13

When the data in the table are examined, it has been determined that slope, elevation, land use and geology parameters are important for the study area. Parameter subclasses were evident in the creation of different risk groups.

It has been determined that the aspect parameter is not important for the study area as there is no distinctiveness in the subgroups. In future studies, the parameter can be made meaningful by examining it in different classes.

The fact that the risk ratios in the subgroups were very close to each other showed that the parameter of proximity to the fault lines was not important for the study area. In future studies, the parameter can be made meaningful by examining it with different proximity classes.

6. Conclusion

The areas and rates of the landslide susceptibility classes are given in Table 10.

Table 10. 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

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.

Method and parameter selection for landslide susceptibility analysis is a step that needs attention. In the result of working, it was determined that the sub-classification step of the parameter is as important as the parameter selection in landslide susceptibility analysis studies.

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.

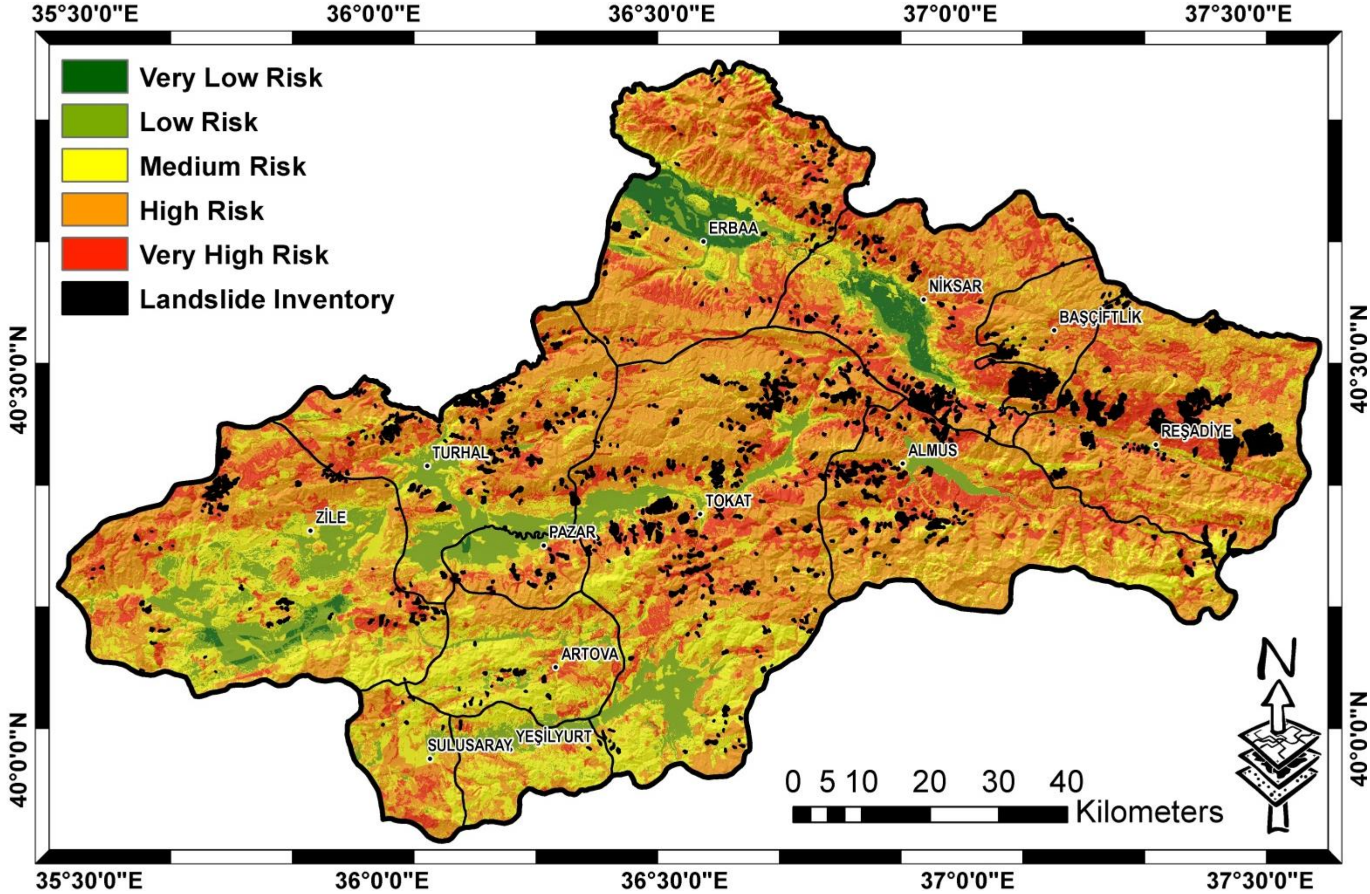

Figure 9. Landslide Susceptibility Map

Acknowledgement

It is an extended version of the paper titled "GIS-Based landslide susceptibility mapping using weight of evidence (WoE) and random forest (RF)" presented at the 2nd IGD Symposium.

Author Contributions

1st Author: Conceptualization, Methodology, Software, Data Curation, Writing-Original Draft Preparation, Validation, Visualization
2nd Author: Investigation, Reviewing and Editing

Statement of Conflicts of Interest

The authors declare no conflicts of interest.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

References

Aleotti, P., & Chowdhury, R. (1999). Landslide hazard assessment: summary review and new perspectives. *Bulletin of Engineering Geology and the Environment*, 58 (1), 21-44.

AFAD. (2014). *Annotated glossary of disaster management terms*. Disaster and Emergency Management Presidency Earthquake Department.

Baeza, C. & Corominas, J. (2001). Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. Earth Surface Processes and Landforms: *The Journal of the British Geomorphological Research Group,* 26(12), 1251-1263.

Basara, A. C., Tabar, M. E., & Sisman, Y. (2020). GIS-Based landslide susceptibility mapping using frequency ratio and AHP methods. *Intercontinental Geoinformation Days (IGD),* 223-226, Mersin, Turkey.

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.

Basara, A. C., Tabar, M. E. & Sisman, Y. (2021). Landslide Susceptibility Mapping of Samsun (Turkey).

province using frequency ratio and AHP methods. *Turkish Journal of Geographic Information Systems*, 3(1), 24-30.

Basara, A. C. (2021). P*roduction of landslide susceptibility maps by statistical methods and investigation of spatial susceptibility (Publication no. 679574)* [Master Thesis, Ondokuz Mayıs University]. YÖK National Thesis Center.

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.

Dağ, S. (2007). *Landslide susceptibility analysis of Çayeli region (Rize) by statistical methods* (Publication no. 212109) [Doctoral Thesis, Karadeniz Technical University]. YÖK National Thesis Center.

Gökceoglu, C., & Aksoy, H. (1996). Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. *Engineering Geology*, 44(1-4), 147-161.

Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi- scale study, Central Italy. *Geomorphology*, 31(1-4), 181-216.

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

Guzzetti, F., Ardizzone, F., Cardinali, M., Rossi, M., & Valigi, D. (2009). Landslide volumes and landslide mobilization rates in Umbria, central Italy. *Earth and Planetary Science Letters*, 279(3-4), 222-229.

Karslı, F., Atasoy, M., Yalcin, A., Reis, S., Demir, O., & Gokceoglu, C. (2009). Effects of land-use changes on landslides in a landslide-prone area (Ardesen, Rize, NE Turkey). *Environmental monitoring and Assessment,* 156(1), 241-255.

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.

Lee , S., & Talib, J. A. (2005). Probabilistic landslide susceptibility and factor effect analysis. *Environmental Geology*, 47:7, 982-990.

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

Varnes, D. J. (1984), Landslide hazard zonation: a review of principles and practice, Commission of Landslides of the IAEG, UNESCO. *Natural Hazards*, No. 3, 61 pp.

Wachal, D. J., & Hudak, P. F. (2000). Mapping landslide susceptibility in Travis County, Texas, USA. *GeoJournal,* 51 (3), 245-253.

Tetik Biçer, Ç. (2017). *A semi-quantitative evaluation of landslide risk mapping* (Publication no. 465288) [Doctoral Thesis, Hacettepe University]. YÖK National Thesis Center.