Research Article

Lutfiye Kusak*, Fatma Bunyan Unel, Aydın Alptekin, Mehmet Ozgur Celik, and Murat Yakar

Apriori association rule and *K*-means clustering algorithms for interpretation of pre-event landslide areas and landslide inventory mapping

https://doi.org/10.1515/geo-2020-0299
received April 22, 2021; accepted September 13, 2021

Abstract: In this paper, an inventory of the landslide that occurred in Karahacılı at the end of 2019 was created and the pre-landslide conditions of the region were evaluated with traditional statistical and spatial data mining methods. The current orthophoto of the region was created by unmanned aerial vehicle (UAV). In this way, the landslide areas in the region were easily determined. According to this, it was determined that the areas affected by the landslides had an average slide of 26.56 m horizontally. The relationships among the topographic, hydrographic, and vegetative factors of the region were revealed using the Apriori algorithm. It was determined that the areas with low vegetation in the study area with 55% confidence were of a Strong Slope feature from the Apriori algorithm. In addition, the cluster distributions formed by these factors were determined by *K*-means. Among the five clusters created with *K*-means, it was determined that the study area was 38% in the southeast, had a Strong Slope, Low Vegetation, Non-Stream Line, and a slope less than 140 m. *K*-means results of the study were made with performance metrics. Average accuracy, recall, specificity, precision, and F-1 score were found as 0.77, 0.69, 0.84, and 0.73 respectively.

Keywords: Apriori algorithm, *K*-means algorithm, landslide inventory, landslide, UAV

1 Introduction

Landslides accounted for 5.2% of all the natural hazards that occurred in the world between 1998 and 2017, during which approximately 4.8 million people were affected and 18,414 people died [1,2]. According to inventory records, between 23,286 landslides occurred in Turkey between 1950 and 2019 [3]. Moreover, a total of 1,343 people died in 389 landslides that occurred between 1929 and 2019 [4].

Landslides occur as a result of preliminary factors including geological, topographic [5,6], environmental and triggering factors such as earthquakes, precipitation and people [7]. In addition to causing destruction in urban and rural settlements, landslide movements such as flowing, sliding, falling and toppling [8–11], block agricultural lands, forest areas [12], roads that provide rural and urban connections [13,14], communication systems [15], railways, drainage and irrigation channels [16], resulting in financial and moral losses.

Landslides are frequent in Turkey, and often occurred in the same areas and turn into natural disasters due to geographical structure [17], climatic characteristics, incorrect land usage, presence of active earthquake zones, high mountainous areas, and the high slope height across the country [18].

Landslide inventories and maps are produced in order to record landslide zones, understand their distribution, take the necessary precautions for such zones, effectively tackle the losses caused by landslides and understand the environmental variables that cause landslide formation [17,19–24].

The stages of preparing landslide inventory maps vary according to the type of map to be prepared. The methods used in preparing inventory maps also vary according to the purpose of the paper, the size of the study area [25], and the availability of the region. In general, the methods consist of investigation of the field where the event occurs [15,24,26], evaluating aerial

* **Corresponding author: Lutfiye Kusak,** Department of Geomatics Engineering, Faculty of Engineering, Mersin University, Mersin, Turkey, e-mail: lutfiyekusak@mersin.edu.tr
Fatma Bunyan Unel, Mehmet Ozgur Celik, Murat Yakar: Department of Geomatics Engineering, Faculty of Engineering, Mersin University, Mersin, Turkey
Aydın Alptekin: Geological Engineering Department, Mersin University, Mersin, Turkey

photographs [15,25,27,28], remote sensing [29–35] using modern tools such as Light Detection And Ranging (LiDAR) [36,37], and Unmanned Aerial Vehicle (UAV) [38–41].

In recent years, UAVs have been used in disaster-related studies due to the fact that they support reproducible measurement studies, are low cost and of high resolution. In addition, UAVs offer serious advantages including the ability to obtain digital elevation map (DEM) data [42–46], and rapid data acquisition after the landslide event [12,47].

In addition to them, there are studies in which data mining methods have been used for the interpretation of data sets in landslide studies. Data mining, in which methods such as clustering, classification, association analysis are used, is preferred in making predictions for the future as well as establishing meaningful relationships in data sets [48–51]. Spatial data mining is the process of discovering/obtaining interesting, previously unknown but potentially useful patterns from large spatial data sets using traditional data mining methods [52]. There are studies that have used data mining to identify landslide areas [53], improve the quality of inventory maps [54], susceptibility [55], hazard and risk analysis, and to identify and evaluate areas that have been or will be exposed to landslides [56–58].

K-means, one of the clustering algorithms, used in susceptibility [59], landslide detection [60–63], preparing inventory maps [64].

The Apriori algorithm, one of the association rule, is used in detecting landslide movements [65], defining the relations between landslide precursor factors in the landslide inventory [56,66,67], preparing susceptibility maps [68], as well as revealing the landslide deformation [66] in non-linear models.

In addition to all these tools and methods, geological, topographic [69], infrastructure, and buildings maps, land use maps [70] different thematic maps such as slope and aspect maps obtained from DEM data and geomorphological variables [13,71], Google Earth data [72], historical sources, technical reports, newspapers, the internet, social media data [73] and meteorological data are among the important data sources [17,70] as an ancillary data used in the preparation of landslide inventories.

The aim of this paper is to create an inventory map of the landslide, which occurred in January 2019 in Karahacılı District, and to evaluate the area before the landslide event using traditional statistics and spatial data mining. This paper consisted of three basic stages: field studies, evaluation and the analysis of existing data. First, for the field investigation of the area, a flight was carried out with a UAV, on 28 June 2019 and then an orthophoto of the area was created. Although it had been about six months since the landslide event in the study area, traces of the landslide were still visible. The sliding surfaces for the roads, creek, and three agricultural lands, were determined and manually drawn by visual interpretation and using an orthomosaic map prepared in 2018 photogrammetric methods and the orthophoto of UAV produced as a result of the field investigation. Then, in order to evaluate the situation prior to the landslide, topographic factors such as aspect, slope and hydrological factor maps such as streams were created by using the existing DEM data of 5 m resolution prepared previously. The Normalized Difference Vegetation Index (NDVI) showing the vegetation values of the region with the help of Sentinel satellite images were used together with the other data sets in the analysis process. The relations between preliminary factors have been revealed using Apriori algorithm. The patterns caused by these factors have been determined using *K*-means algorithm. The results of the *K*-means algorithm has been tested using performance metric methods.

2 Materials and methods

In this section, the study area is introduced and a detailed explanation is given about the data and methods used.

2.1 Study area and data

The landslide occurred in the village of Karahacılı and the surrounding six villages, namely Emirler, İnsu, Değirmençay, Uzunkaş and Turunçlu, in January 2019 after heavy rains that fell in December 2018. The total area of the village of Karahacılı, which is located at 36°.813081 N and 34°.482932E (Figure 1), is 786.03 ha. Karahacılı has agricultural land on which pomegranates, black grapes and figs are grown. As a result of the landslide that had occurred in the area, energy transmission lines collapsed and the main transportation roads and agricultural lands were destroyed. In the investigation of field conducted after the landslide occurred, it was determined that, in general, there were cultivated mixed agricultural products and fruit trees on the slopes in the area.

Various data sets were used in the paper (Table 1).

The landslide area was comprised of Langhian-Serravallian age clayey limestone. The Lower-Middle Miocene age Kaplankaya Formation, which covered the entire region, was digitized from a 1:250,000 scale

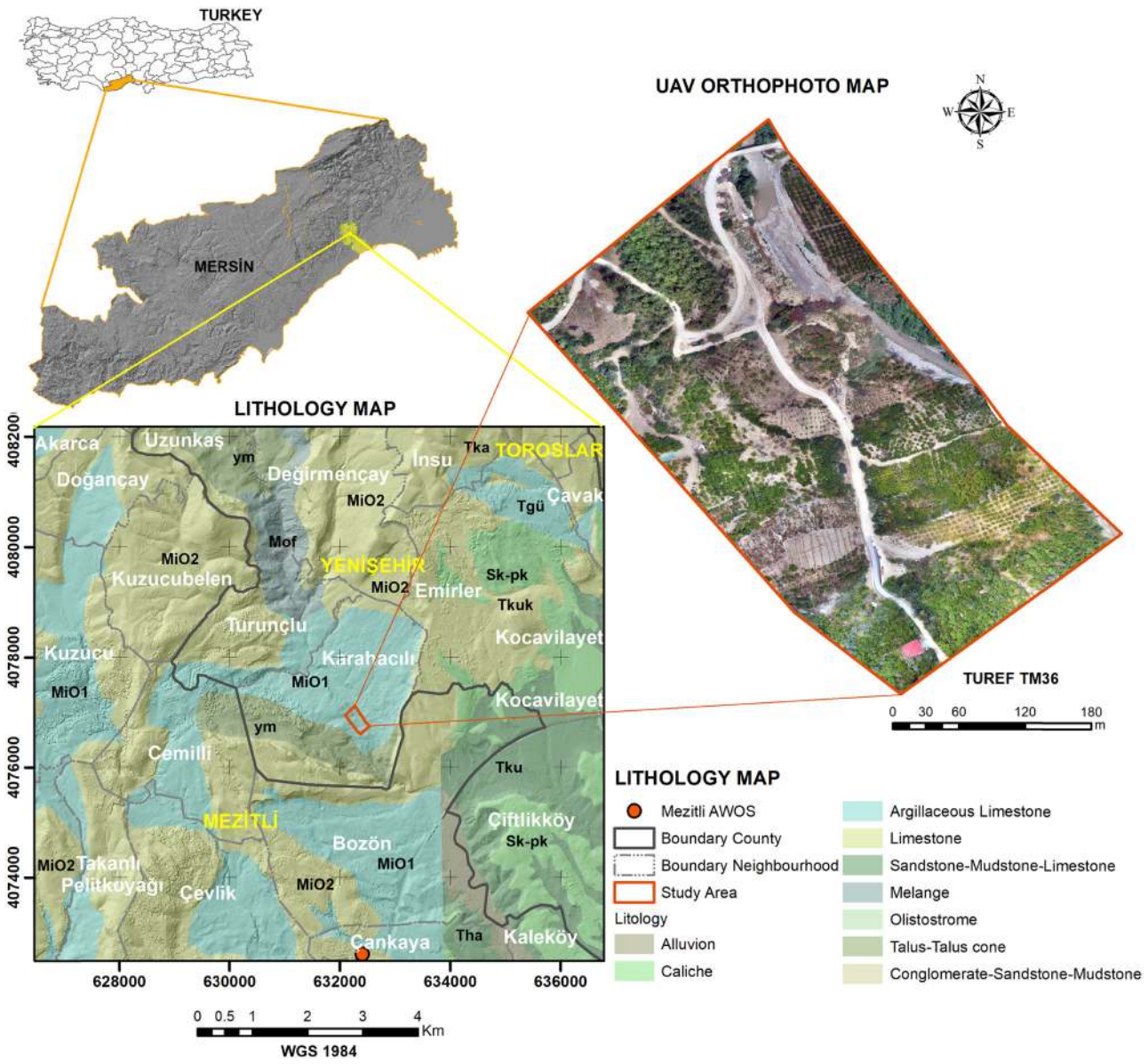


Figure 1: Study area location.

geology map. This formation was comprised of Langhian-Serravallian age clayey limestone. In addition to limestone, non-cohesive material consisting of sand and gravel was observed in the areas that were affected by the landslide. No active fault line was observed in the immediate vicinity of the study area. In addition, debris, which increases the risk of landslides, was found in the landslide area.

When the studies carried out on the region were examined, it was observed that the study area was close to active landslide zones [74,75]. The rubble located behind the crown of the area caused landslide as a result of heavy rainfall.

Table 1: Data and features used in the study

Data name	Data features
2018 Mosaic Orthophoto	30 cm, WGS84
Digital Elevation Model (DEM)	5 m, WGS84
2019 UAV Orthophoto	Approximately 2 cm, TUREF_TM36
2019 Sentinel	15 m, WGS84
October 2018, November 2020 Google Earth images	WGS84
Lithology map	1:100,000, WGS84
Boundary of Mersin	ITRF 96
Meteorology data	2018–2019, Daily rainfall

Event landslide inventories are the records that show the landslides that occur as a result of a certain trigger factor such as excessive rainfall [15], snowmelts, earthquakes and volcanic factors [76]. Excessive rainfall was determined as the triggering factor of the landslide that occurred in the village of Karahacılı in January 2019. Rainfall data were obtained from the Mezitli Automatic Weather Observing System, which is nearly 4 km away from the landslide area (Figure 1). 24-months rainfall data belongs to 2018 and 2019 was shown in a graphic and polynomial trendlines were drawn, R^2 values are 0.72 and 0.84 respectively (Figure 2). The region takes rainfall only in winter as seen in the graphic. It was determined that region received more than 400 kg/m^2 precipitation in December 2018. In January 2019, it is seen that the precipitation increased more than 2 times compared to the previous year.

In this paper, orthomosaic map prepared in 2018, and orthophoto produced in 2019 via UAV were used to create landslide inventory map. The flight, which was performed on 28 June 2019 by using Anafi Parrot, was planned with 65 m height, 1.86 cm/pixel GSD and 80% side 70% frontal overlap [77]. A total of 447 photos were obtained from the flight, which covered an area of 21.1 ha. The TUREF (ITRF96) 36-3° projection system was selected for the area. Out of the 21.1 ha flight area [41] 10.3 ha was used as the study area. In this paper, a total of five Ground Control Points (GCPs) were used. Although the GCPs were taken into consideration to show a homogeneous distribution, full homogeneity could not be achieved when the points were established due to the high-level difference in the area with landslides, the inability to descend to the creek bank and the fact that the crown of the landslide area was dangerous and sliding was still continuing. The coordinates and elevation information of the GCPs were determined by Continuously Operating Reference Stations. Total error amounts of *X*, *Y* and *Z* are 2.39, 1.59 and 1.57 mm respectively.

The determination of the highest slope angle of the area obtained by using DEM is one of the most important topographic factors for the risk assessment of landslide zones [78]. In the study carried out by Capitani et al. [79], it was revealed that slope factor in superficial and clayey deposits and maintenance for shallow surfaces could be effective in landslide formation [79]. Drainage networks are among the hydrographic factors that are effective in the formation of landslides by reaching the saturation of the water on the lower and upper surfaces [5,80,81]. For this reason, slope, aspect and stream maps were created from the 5 m resolution DEM data to evaluate the preliminary factors such as topographic and hydrographic before the landslide occurred and to be used in spatial data mining.

The other preliminary factor used in this paper was NDVI. NDVI widely used to describe vegetation cover, is the well-known normalized ratio between red (R) and near-infrared (NIR) and is one of the most used vegetation indices (equation (1)).

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{R})}{(\mathrm{NIR} + \mathrm{R})}. \tag{1}$$

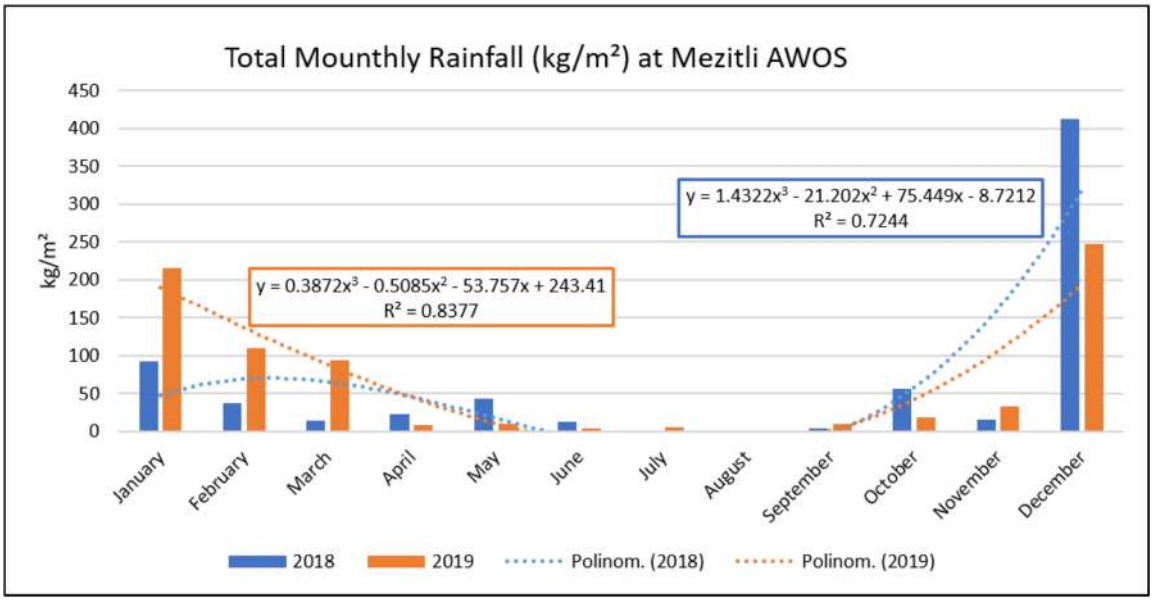


Figure 2: Rainfall graph.

The index varies between –1 and 1, with higher values relating to healthy photosynthetic vegetation and lower values relating to stressed vegetation or no vegetation (bare soil).

NDVI is used to determine the changes in the vegetation cover of landslide areas [70]. In this paper, Sentinel-2 image on 25 December 2018 was used.

2.2 Methods

The methods used were introduced in this section. First of all, information about the inventory maps was given. In addition, relative frequency and spatial data mining methods were discussed and their formulas were explained.

2.2.1 Landslide inventory map properties

Landslide inventories include information regarding the location, class, morphometry, triggering factors, volume, run-out of distance, activity, recurrence intervals and the date of the landslide, various statistical information, and the damage caused by the landslide in varying degrees of detail [12,15,24,25,73,82]. Landslide inventory maps can be examined under four main headings: landslide archive maps, event landslide inventory maps, historical landslide inventory maps [83] and multi-time landslide inventory maps [12]. In addition to field investigation, remote sensing data obtained using platforms such as spaceborne, airborne and ground-based are preferred in the creation of landslide inventories [84]. In the evaluation of the obtained data, different approaches are used considering the data set.

Field investigation is performed shortly after the event occurs. Detailed information on the location, volume, factors contributing to the landslide, type and destruction are recorded [15]. Field studies are generally applied in the mapping of small landslides that have occurred naturally or occurred on a specific date due to a specific event.

The landslide areas formed before and after the landslides in the region are drawn by stereoscopic or manual drawing and visual assessments [15,17,31,73,85] and then processed on topographic maps [12,24,86]. In addition to visual interpretation, approaches such as automatic image evaluation approach [12,19,33,37,87–89] are used in the evaluation of obtained images. These methods can be used alone or in combination for the production of landslide inventory maps.

2.2.2 Relative frequency distribution

A relative frequency distribution shows the proportion of the total number of observations associated with each value or class of values and is related to a probability distribution, which is extensively used in statistics [90]. The general relative frequency distribution formula was adapted to slope (equation (2)) and aspect (equation (3)) data for this paper.

$$F_{N_i} = \frac{f_{N_i}}{\sum f_{N_i}}, \tag{2}$$

where:

i = {L, NL, VGS, GS, MS, SS, VSS, ES, StS, and VStS}

F_{N_i}: Percentage of relative frequency of slope type

f_{N_i}: Frequency of slope type

$\sum f_{N_i}$: Total frequency of slope type

$$F_{N_j} = \frac{f_{N_j}}{\sum f_{N_j}}, \tag{3}$$

j = {N, NE, E, SE, S, SW, W, and NW}

F_{N_j}: Percentage of relative frequency of aspect type

f_{N_j}: Frequency of aspect type

$\sum f_{N_j}$: Total frequency of aspect type

2.2.3 Apriori algorithm for association rules

Association rule is a model that defines a method that examines the co-occurrence of events and data relationships of certain types [91]. Association rules reveal the occurrence of events in terms of probability and define their correlations. The main purpose of this model is to reveal interesting relationships between data. Today, this method is used in Web usage mining, intrusion detection, continuous production, bioinformatics and many fields. There are Apriori [92], Eclat and FP-growth, SETM, Partition, RARM – Rapid Association Rule Mining and CHARM algorithms for creating rules.

The processes of the Apriori algorithm are following;

1. Association rule mining is defined as: $I = \{i_1, i_2, \ldots, i_n\}$ and $D = \{t_1, t_2, \ldots, t_m\}$, I where I is a set of n binary attributes called items and D is a set of transactions called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I.
2. A rule is defined as an implication of the following form: $X \Rightarrow Y$ $X, Y \subseteq I$. A rule is defined only between a set and a single item, $X \Rightarrow i_j$ for $i_j \in I$. Every rule is composed by two different sets of items, also known as

itemsets, X and Y, where X is called antecedent and Y consequent.

3. In order to select interesting rules from the set of all possible rules, constraints on various measures of significance and interest are used. The best-known constraints are minimum thresholds on support and confidence.
4. X, Y be itemsets, $X \Rightarrow Y$ an association rule and T a set of transactions of a given database. Support, confidence and lift are three measures of the relationship or rule [91]. From these, three commonly used measures of association can be estimated.
5. Support (equation (4)) is an indication of how frequently the itemset appears in the data set. The support of X with respect to T is defined as the proportion of transactions, t, in the data set which contains the itemset X.

$$\text{Supp}(X) = \frac{|\{t \in T;\ X \subseteq t\}|}{|T|}. \tag{4}$$

6. Confidence (equation (5)) is an indication of how often the rule has been found to be true. The confidence value of a rule, $X \Rightarrow Y$, with respect to a set of transactions, T, is the proportion of the transactions that contain X and also contain Y. Confidence is defined as:

$$\text{conf}(X \Rightarrow Y) = \frac{\text{Supp}(X \cup Y)}{\text{Supp}(X)}. \tag{5}$$

7. The lift (equation (6)) of a rule is defined as:

$$\text{Lift}(X \Rightarrow Y) = \frac{\text{Supp}(X \cap Y)}{\text{Supp}(X) * \text{Supp}(Y)}, \tag{6}$$

or the ratio of the observed support to that expected if X and Y were independent.

The Apriori algorithm [48] which is based on the principle of the frequency of repetition between events, was used for the landslide area in this paper by adapting it to the spatial data mining field. When the literature was examined, it was observed that there were landslide studies that had used the Apriori algorithm [56]. For the study area, the Apriori algorithm was applied to elevation, slope, aspect, hydrologic and NDVI categorize data using WEKA software.

2.2.4 *K*-means algorithm for clustering

Cluster analysis is the process of dividing data into groups. Unlike the classification method, the classes are not predetermined. The groups and clusters that the data will be divided into, and even how many different groups they will be divided into are determined according to the similarity of the available data. Cluster analysis is used in many fields such as biology, medicine, anthropology and marketing.

The K-means partitional-clustering algorithm is the most preferred algorithm employing a square-error criterion. K-means was first used by MacQueen [93]. The general aim is to get the partition that, for a fixed number of clusters, minimizes the total square error. The processes of K-means is following:

1. The data set has N samples in an n-dimensional space.
2. The number of cluster (K) has been determined by one of the methods such as Elbow, Silhouette. N samples has been partitioned into K clusters $\{C_1, C_2, \ldots, C_k\}$.
3. The distribution of the members to the previously determined K number of centers starts randomly [50,59,94]. Each C_k has n_k samples and each sample is in exactly one cluster so that $\sum n_k = N$, where $k = 1, 2, \ldots, K$.
4. The mean vector M_k (equation (7)) of cluster C_k is defined as the centroid of the cluster or where x_{ik} is the ith sample belonging to cluster C_k.

$$M_k = \left(\frac{1}{n_k}\right)\sum_{i=1}^{n_k} x_{ik}. \tag{7}$$

5. The square error (equation (8)), which is also called the within-cluster variation (e_k^2), for cluster C_k is the sum of the squared differences between each sample in C_k and its centroid [95].

$$e_k^2 = \sum_{i=1}^{n_k} (x_{ik} - M_k)^2. \tag{8}$$

6. The square error for the entire clustering space containing K clusters is the sum of the within-cluster variations (equation (9)), which is named within clusters sum of square (E_k^2).

$$E_k^2 = \sum_{k=1}^{K} e_k^2. \tag{9}$$

7. The objective of a square-error clustering method is to find a partition containing K clusters that minimize E_k^2 for a given K [95]. The elements of each C_k set and their distances from the M_k center are calculated with the Euclidean distances formula (equation (10)).

$$D(M_k, X_i) = D_{M_k} = \sqrt{\sum_{i=1}^{K} (M_k - x_{ik})^2}. \tag{10}$$

8. For example, whichever of the two cluster centers (M_k or M_{k+1}) the sample is close to the sample (x_{ik})

is considered as an element of that cluster. So $D(M_k, X_i) < D(M_{k+1}, X_i)$ is $X_i \in C_k$.

9. Euclidean distances are calculated for all cluster elements and it is determined which cluster they belong to.
10. The transactions are updated iteratively, depending on the distances between cluster members and centers, until they reach a position where the centers will not change [50,94].

Although *K*-means is not used in the analysis of categorical data, it has been made available with various analyses [96]. With the system provided by WEKA, *K*-means algorithm can also be used for categorical data [97]. Clustering is used in spatial data mining [98] while the *K*-means algorithm is used in landslide studies [25,63,67]. In the *K*-means algorithm, elevation, slope, aspect, stream and NDVI categorize data were evaluated with WEKA software and used for the landslide area. The results obtained were visualized in ArcGIS 10.5 software.

2.2.5 Performance Metrics

The performance metrics method was used for the verification of the clustering analysis. For each cluster, the number of samples to represent the cluster was determined using the sample size method (equation (11)).

$$n = \frac{Nt^2p(1-p)}{d^2(N-1) + t^2p(1-p)}, \tag{11}$$

where:

n: Sample size

N: Population size

t: Standard value of 1.96 in confidence level at 95%

p: population proportion (0.5)

d: the margin of error

To assess the clustering algorithms performance metrics such as accuracy (equation (12)), recall (equation (13)), specificity (equation (14)), precision (equation (15)) and *F*1 score (equation (16)) were used [99–103].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{12}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{13}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}, \tag{14}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{15}$$

$$F1\text{-score} = 2.\frac{\text{Precision} - \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{16}$$

where:

TP: True Positive (Landslides correctly observed in 2019 and continuing in 2020).

TN: True Negative (Non-landslide correctly observed in 2019 and 2020).

FP: False Positive (Landslide correctly observed in 2019 and Non-landslides correctly observed in 2020).

FN: False Negative (Non-landslide correctly observed in 2019 and new landslides correctly observed in 2020).

The points belonging to each cluster were selected randomly for performance metric analysis. The metric values of these points were calculated using 2018 Google Earth, 2019 UAV, 2020 Google earth images.

3 Results

In this section, Apriori, *K*-Means, performance metric results of the study were obtained and maps were created.

3.1 Inventory mapping preparation of karahacili landslide

Landslide inventory mapping was performed using the orthomosaic maps of 2018 and UAV orthophoto of 2019; and five different areas, namely the creek (water) at the junction of the valley, the road affected by the landslide and three different agricultural lands affected by the landslide, were selected. Unlike other studies in which places that have become scars after being affected by landslides were determined, the present study placed importance on selecting landslide areas that directly damaged the socio-economic structure. The boundaries showing the positions of these areas before and after the landslides were determined (Figure 3).

When the photographs taken after the landslide were examined, the effect of the landslide on the area could be seen clearly. In the first photo, the creek was filled with the debris that dropped from above, in the third photo the road line had cracked, and in the last three photos the agricultural lands had shifted and the crops had been destroyed (Figure 3a).

The before and after of the areas affected by the landslides in the study area are shown in Figure 3b. The areal changes and slip amounts of these areas were calculated.

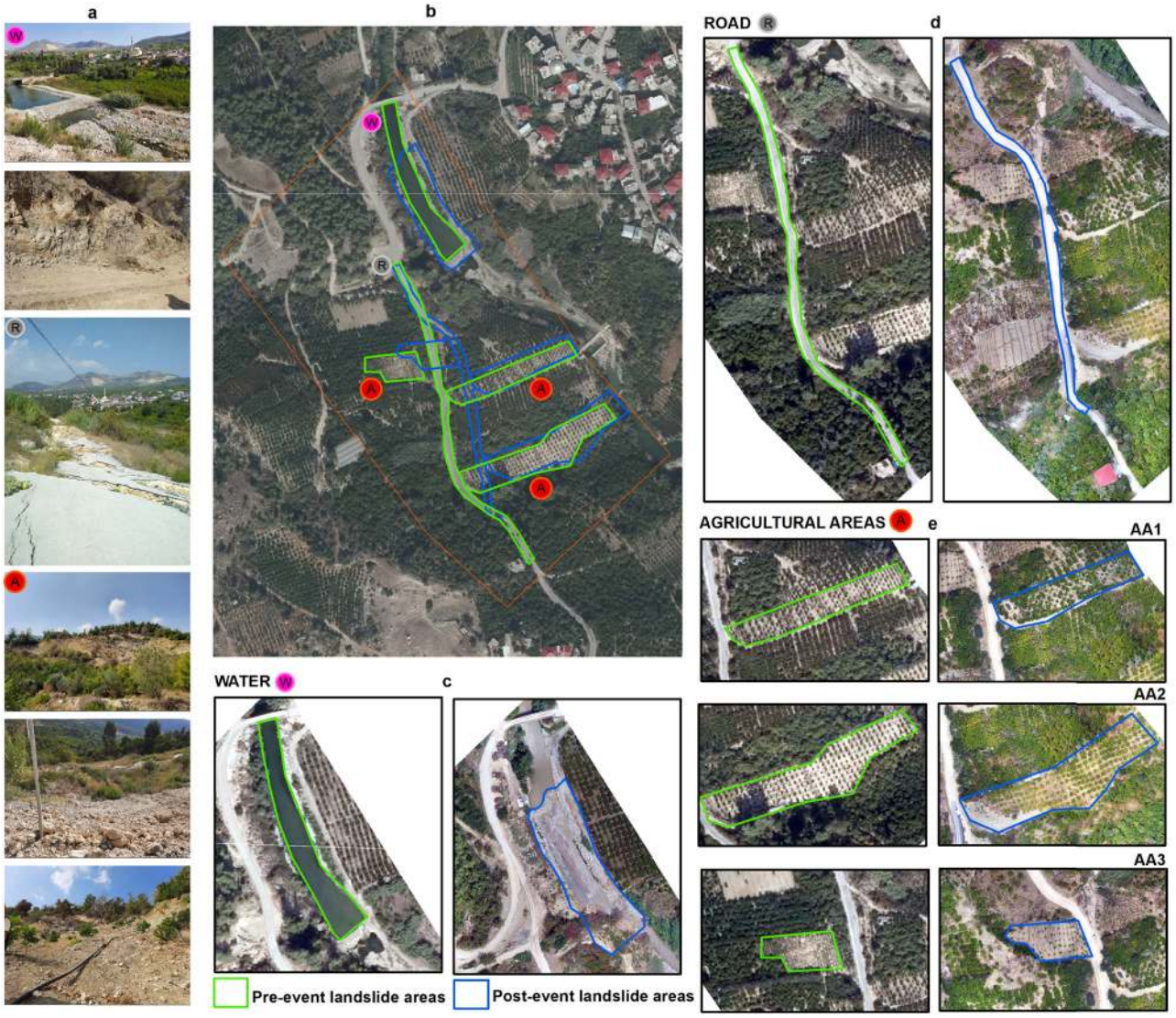


Figure 3: (a) Post event photo of the study area. (b) Landslide areas. (c) Change detection of the water area. (d) Change detection of the road. (e) Change detection of the agricultural areas.

The rubble that slid from the upper part of the valley blocked the creek, causing it to expand and decrease the water area (Figure 3c). The road axis shifted, resulting in the change of route. In the land discovery after the landslide, it was observed that the new road was asphalted and opened to village transportation, however, the sliding still continued (Figure 3d). The shape of the agricultural lands affected by the landslides were generally preserved after the landslide, however, in some places they expanded, shortened and moved in the direction of the landslide (Figure 3e).

The area and change distance of the drawn landslide areas are shown in Table 2. It was found that all of the areas slid towards the creek in the north-east direction after the landslide occurred. Accordingly, it was determined that an increase of 1387.63 m^2 and a shift of 12.24 m had occurred due to the decrease in the water area that resulted from the blockage due to debris. The

Table 2: Area and distance difference in the sliding direction of landslide areas

	Pre-event (m^2)	Post-event (m^2)	Area difference (m^2)	Distance difference (m)
Water (W)	2680.04	4067.67	−1387.63	12.24
Road (R)	2182.02	1907.73	274.29	29.49
AA1	2246.76	2263.47	−16.71	28.54
AA2	4031.54	4189.35	−157.81	28.79
AA3	1220.18	1195.81	24.37	33.75

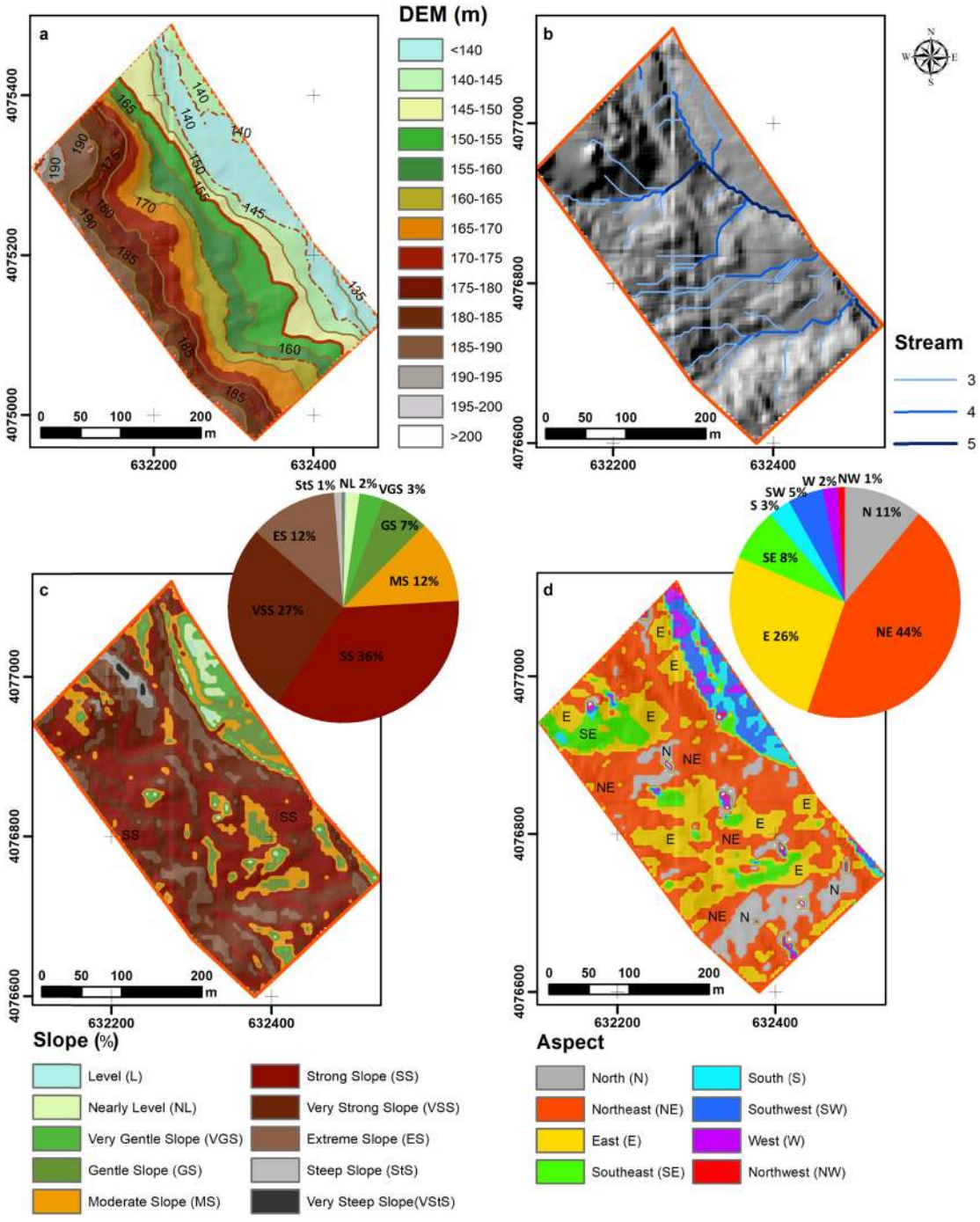


Figure 4: (a) DEM. (b) Stream. (c) Slope% (Level: 0–0.5, Nearly Level: 0.5–2, Very Gentle Slope: 2–5, Gentle Slope: 5–9, Moderate Slope: 9–15, Strong Slope: 15–30, Very Stroge Slope: 30–45, Extre-me Slope: 45–70, Steep Slope: 70–100, Very Steep Slope: >100). (d) Aspect (Flat: –1, North: 0–22.5, Northeast: 22.5–67.5, East: 67.5–112.5, Southeast: 112.5–157.5, South: 157.5–202.5, Southwest: 202.5–247.5, West: 247. 5–292.5, Northwest: 292.5–337.5, North: 337.5–360).

road area shrank by 274.29 m^2 and shifted 29.49 m from the road axis. In terms of the agricultural lands, the AA1 and AA2 areas increased by 16.71 and 157.81 m^2, respectively, and shifted an average of 28.66 m. On the other hand, the AA3 area decreased by 24.37 m^2 and shifted 33.75 m.

3.2 Relative frequency distribution of topographic data

The topographical and hydrographic features of the study area before the landslide were presented with maps

Table 3: Standard slope descriptors [110]

No	Slope (%)	Approximate degrees	Terminology	No	Slope (%)	Approximate degrees	Terminology
1	0–0.5	0	Level (L)	6	15–30	8.5–16.5	Strong slope (SS)
2	0.5–2	0.3–1.1	Nearly level (NL)	7	30–45	16.5–24	Very strong slope (VSS)
3	2–5	1.1–3	Very gentle slope (VGS)	8	45–70	24–35	Extreme slope (ES)
4	5–9	3–5	Gentle slope (GS)	9	70–100	35–45	Steep slope (StS)
5	9–15	5–8.5	Moderate slope (MS)	10	>100	>45	Very steep slope (VStS)

(Figure 4). ARCGIS 10.5 software was used for the generation of the stream, slope and aspect maps in order to interpret the topographic and hydrological properties. When the DEM of the village of Karahacılı was examined, it was observed that the average altitude above sea level varied between 140 and 200 m. The elevation data were classified every 5 m to be used in the analysis. The land rises towards the northwest and southwest directions and the crown zone of the landslide is located in these parts (Figure 4a).

Stream lines were created in five different degrees and were all used in the spatial data mining analysis. However, only the 3rd degree and higher were shown on the maps. When the stream map was examined, it was observed that 3rd degree streams dominated the area (Figure 4b).

The slope map obtained from the DEM data was divided into 10 different classes (Figure 4c) using standard slope descriptors [104] (Table 3). ARCGIS aspect classification was used for the aspect data (Figure 4d).

Relative frequency distribution for the slope and aspect values of the whole study area and the sum of the landslide areas is presented in a table to determine whether the study area reflects the characteristics of the landslide areas. It was evaluated with relative frequency distribution (Table 4).

Table 4: Relative frequency distribution of topographic data

Slope Type	F_{N_i} of the study area	F_{N_i} of the total landslide areas	Aspect Type	F_{N_j} of the study area	F_{N_j} of the total landslide areas
L	0.31	0.10	N	11.05	0.45
NL	1.96	0.31	NE	44.19	4.32
VGS	3.37	0.62	E	26.16	4.92
GS	6.73	1.07	SE	7.35	1.41
MS	11.70	2.48	S	3.20	0.05
SS	35.64	5.13	SW	4.94	0.43
VSS	26.67	1.72	W	2.10	0.29
ES	12.34	0.38	NW	1.00	0.12
StS	1.07	0.00			
VStS	0.21	0.00			

When the slope percentage map of the pre-landslide study area was evaluated, it was determined that approximately 36% of the whole area remained in the strong slope, while approximately 27% remained in the very strong slope (Figure 4c). Landslide formation is caused by extreme rainfall, human and natural activities, land use alterations [105,106] and a large part of the area being in the strong slope and very strong slope. There was a 0.8760 correlation between the study area and the total of the landslide areas. Accordingly, the Strong slope feature was observed as approximately 5% in the areas where landslides were observed, as in the study area (Table 4).

When the aspect map of the pre-landslide study area was evaluated, it was observed that approximately 44% of the whole area was in the northeast (NE), while approximately 26% was in the East (E) (Figure 4d). Capitani et al. [79] reported that aspect was not directly effective for landslides, but was effective in cases of surface landslides, the lithological feature of which is clayey limestone [79]. The study area also triggered the landslide as a result of demonstrating surface landslides and clayey limestone features, and being predominantly in the northeast (NE) aspect and, thus, being less exposed to sun. There was a correlation of 0.8981 between the aspect of the study area and the total landslide areas. Dominant NE and E aspect are seen in landslide areas with ratios of 4 and 5% respectively, like the study area (Table 4).

3.3 Spatial data mining evaluation

In this section, the pre-processing results of the data were obtained. Using the results obtained from the preprocessing, Apriori and *K*-means algorithms were applied. In order to determine the *K*-means accuracy of the study, the performance metric process was performed and the results were presented.

3.3.1 Pre-processing data

Data must be pre-processed to be used in spatial data mining. Topographic, hydrographic and NDVI data used for this purpose were reclassified. The aspect was divided into 9, Slope 10, NDVI 4, stream 5, and DEM 14 classified. From all these reclassified data sets obtained, 4,174 instances were obtained for the study area. Two different data mining methods were used in this paper; Apriori, one of the association rules algorithms, and *K*-means, one of the clustering algorithms.

3.3.2 Apriori algorithm for association rules evaluation

The reason for using the Apriori algorithm is to model the relationship between the preliminary factors of the region and the landslide. The Apriori algorithm was analyzed with the help of WEKA by selecting 4,174 instances and 5 preliminary factors. All data were used for the Associator model (full training set). In the analysis, the minimum support was 10%, minimum confidence was 50% and the number of cycles performed was 18. The best 10 rules obtained are presented in Table 5.

According to the Apriori algorithm results obtained, Rule 1 has the highest confidence (55%). High lift values in Apriori algorithms indicate that the probability of the rules being seen together is high. In this paper, the lift value of Rule 2 and 8 was 1.15. This value was the highest among all of the rules. According to Rule 2, there was a possibility of a 1st Degree Stream Line in the areas with a very strong slope. Similarly, according to Rule 8, there was a possibility that the areas that were of a very strong slope had an East aspect (Table 5).

Apriori rule maps were created and relations between attributes were revealed (Figure 5).

Maps showing rule 1, rule 5 and rule 10, which were created using Apriori algorithm, represent landslide areas in the best manner.

It was determined that 1,470 instances show strong slope property, among which 810 have low vegetation with 55% confidence interval in Rule 1 map. It is seen that this rule has been distributed densely in road and agricultural areas of landslide areas.

It is obtained that 561 instances have low vegetation, among which 1,095 have SE aspect with 51% confidence interval in Rule 5 map. This rule is distrubuted densely especiall in agricultural lands.

It is obtained that 957 instances have low vegetation among which 1,919 have 1th Degree Stream Line with 50% confidence interval in rule 10. This rule also densely distributed in agricultural lands.

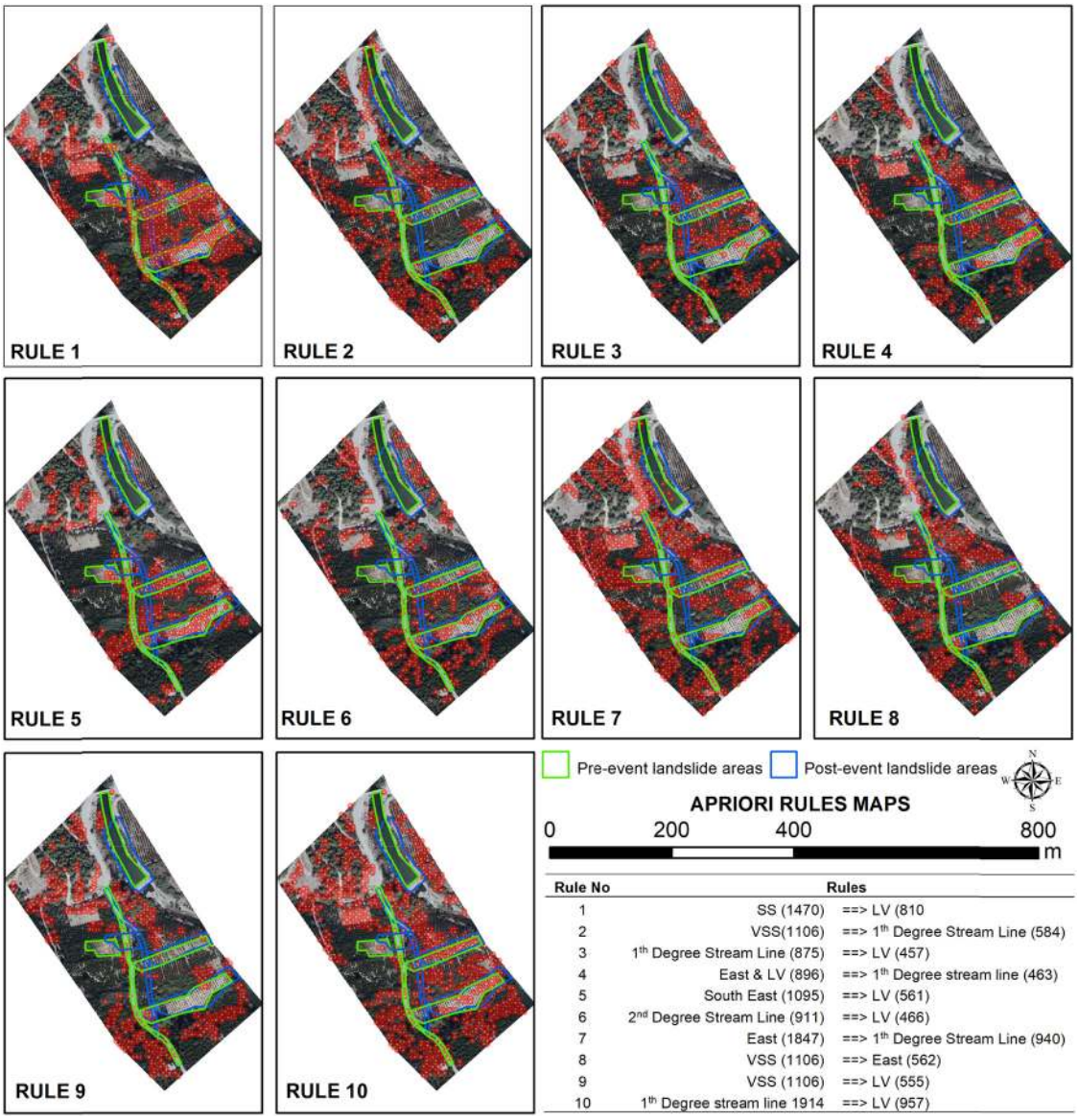
When the rules were evaluated as a whole, it was determined that the topographic and vegetation effects of the 10 different rules that emerged were more prominent in general, followed by the hydrographic effects.

3.3.3 *K*-means algorithm for clustering

In this paper, *K*-means algorithm was used for cluster analysis. A total of 4,174 instances and 5 preliminary factors data set were processed. In the creation of the model, three iteration processes were performed. According to the result that it can best reflect the structure of the study area, $k = 5$ was chosen. In the *K*-means algorithms, the number of clusters is given beforehand. For this, the elbow and silhouette methods are used. In the present study, according to the elbow method $k = 5$. Five different clusters were randomly assigned by the central system. The creation of the model took 0.11 seconds and the result of the iteration is presented in Table 6. It was determined that

Table 5: Apriori algorithm analyses best rule results

Rule No	Rules		Conf (%)	Lift
1	Strong slope (1,470)	⇒ Low vegetation (810)	55	1.08
2	Very strong slope (1,106)	⇒ 1th degree stream line (584)	53	1.15
3	1th Degree stream line (875)	⇒ Low vegetation (457)	52	1.03
4	East & low vegetation (896)	⇒ 1th degree stream line (463)	52	1.13
5	South East (1,095)	⇒ Low vegetation (561)	51	1.01
6	2nd Degree stream line (911)	⇒ Low vegetation (466)	51	1.00
7	East (1,847)	⇒ 1th degree stream line (940)	51	1.11
8	Very strong slope (1,106)	⇒ East (562)	51	1.15
9	Very strong slope (1,106)	⇒ Low vegetation (555)	50	0.99
10	1th Degree stream line (1,914)	⇒ Low vegetation (957)	50	0.98



Rule No	Rules	
1	SS (1470)	==> LV (810
2	VSS(1106)	==> 1th Degree Stream Line (584)
3	1th Degree Stream Line (875)	==> LV (457)
4	East & LV (896)	==> 1th Degree stream line (463)
5	South East (1095)	==> LV (561)
6	2nd Degree Stream Line (911)	==> LV (466)
7	East (1847)	==> 1th Degree Stream Line (940)
8	VSS (1106)	==> East (562)
9	VSS (1106)	==> LV (555)
10	1th Degree stream line 1914	==> LV (957)

Figure 5: Apriori algorithm maps.

1,580, 1,051, 311, 406 and 826 instances were belong to C0, C1, C2, C3 and C4 clusters respectively, among 4,174 instances.

The cluster percentage values that were obtained as a result of the analysis and reflect the characteristics of the area were listed from high to low. Accordingly, 38% of the study area was in the C0 cluster, its aspect was South East, slope was Strong Slope, vegetation was Low Vegetation, land height was less than 140 m and there was no stream line. The C1 cluster represented 25% of the study area, the aspect of which was East, the slope of which was Strong Slope, the vegetation of which was Dense Vegetation, the height of which was between 170 and 175 m and the stream line of which was 1st Degree. Twenty percent of the study area was in the C4 cluster, its aspect was East, slope was very Strong Slope, vegetation was Low Vegetation, Stream Line was 1st degree and height was between 140 and 145 m. Cluster C3 represented 10% of the study area, of which the aspect was East, slope was Strong Slope, vegetation was mixed types, height was between 155 and 160 m and no stream line was present. The C2 cluster represented 7% of the study area and therefore was the least distributed cluster in the area. The aspect of this area was South East, slope was Strong Slope, vegetation was Low vegetation, Stream Line was 3rd Degree and height was between 140 and 145 m. Field scenery of before landslide related to five different clusters were shown in Figure 6.

Table 6: *K*-means algorithm results

Cluster Name	C0	C1	C2	C3	C4
Number of instances in clusters and %	1,580 (38%)	1,051 (25%)	311 (7%)	406 (10%)	826 (20%)
Preliminary Factors	Southeast	East	Southeast	East	East
	SS	SS	SS	SS	VSS
	LV	DV	LV	MTV	LV
	Non	1th Degree	3rd Degree	Non	1th Degree
	<140	170–175	140–145	155–160	140–145

SS: strong slope; VSS: very strong slope; LV: low vegetation; DV: dense vegetation; MTV: mixed types vegetation.

When the map was investigated (Figure 6), it was seen that C0 cluster has been densified in the study area and road, water and agricultural land in South east of landslide area. C1 cluster, which is another dominant cluster in the study area, has been densified in crown site of landslide. C4 cluster has been densify in the North Eastern part of the study area.

3.3.3.1 Performance metrics evaluation

Performance metrics of *K*-means algorithm were obtained using 2018 Google Earth, 2019 UAV and 2020 Google Earth images. Average accuracy, recall, specificity, precision and *F*1 score were found as 0.77, 0.69, 0.84, and 0.73 respectively. It was determined that landslide is seen in

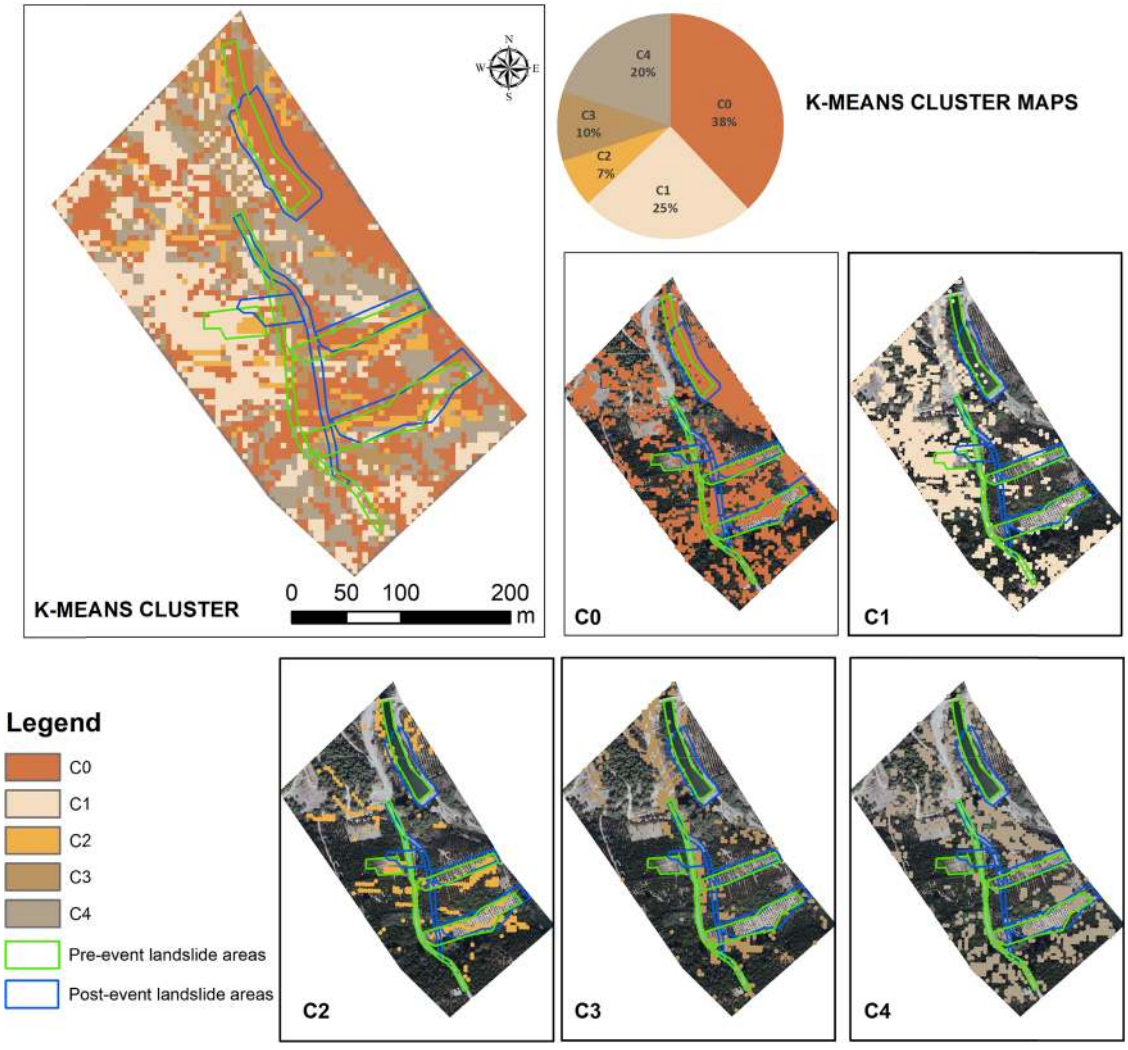


Figure 6: *K*-means algorithm maps.

Table 7: Samplings and performance metrics results

Cluster	*N*	*n*	TP	FP	FN	TN	TP + FN	%	Accuracy	Recall	Specificity	Precision	*F1*-score
C0	1,580	313	70	32	53	158	123	0.39	0.73	0.57	0.83	0.69	0.62
C1	1,051	278	87	24	26	141	113	0.41	0.82	0.77	0.85	0.78	0.78
C2	311	169	47	11	51	60	98	0.58	0.63	0.48	0.85	0.81	0.60
C3	406	196	89	13	11	83	100	0.51	0.88	0.89	0.86	0.87	0.88
C4	826	261	81	26	30	124	111	0.43	0.79	0.73	0.83	0.76	0.74
Avg.									0.77	0.69	0.84	0.78	0.73

70 (TP) from 313 samples, and 53 (FN) samples have no landslide in 2019 but landslide in 2020. Landslides and non-landslides were determined with accuracy of 0.73, 0.82, 0.63, 0.88, 0.79 for C0, C1, C2, C3 and C4, respectively.

Recall metric values were found using equation (13). Recall values C3, C1 and C4 clusters are 0.89, 0.77 and 0.73 respectively. it can be said that landslides continue in these clusters and attention should be paid in terms of landslides, and the East aspect stands out as a common feature.

Except for the accuracy and recall metrics, the points where landslides were detected in 2019 and 2020 (TP) and new landslides in 2020 (FN) were collected for each cluster, and each cluster was divided by the number of samples.

The most striking situation here is the 58% landslide occurrence in the C2 cluster. When the characteristics of the C2 cluster were examined, it was determined that it was located in the southeast, showed LV, and also had a 3rd Degree Stream Line waterline. The regions where this cluster is found should be observed carefully (Table 7).

4 Discussion

Processing landslide areas on maps and preparing inventories is one of the most preferred methods [12]. In this paper, modern mapping methods were used in the preparation of landslide inventories and spatial data mining approaches were used in the interpretation of landslide areas.

The result of the field investigation of the study area is the production of orthophoto map with a UAV, used for landslide inventory map. Moreover, NDVI and existing DEM data was used for preparation of topographic and hydrographic maps were carried out by application of spatial data mining methods, respectively. The affected areas before and after the landslide mapping were determined using the UAV and 2018 orthophoto. The chosen method was suitable for monitoring the changes in the areas both visually and spatially and determining the amount of sliding. In contrast to other studies, in which scarred areas affected by landslides were determined, particular attention was paid to the selection of landslide areas that had been directly damaged in terms of the socio-economic structure. For example, the breakage, displacement and spatial change of the road that was most affected by the landslides could be observed very easily with this method. These changes were evaluated in Figure 3. Area and distance changes are presented in Table 2. In addition, visual evaluation was very important in order to monitor the changes in the agricultural areas in terms of economic and ownership aspects and to protect the rights.

Another method used in the analysis of the study area was topographic and hydrographic evaluation. Topographic evaluation was among the most important topographic factors used to analyze the slope, aspect and hydrological condition of the area and determine its effect on the landslide. There are various studies in the literature that have reported the effect of topographic factors on landslides [36,54,56,81,107–109]. In the present paper, the slope, aspect distributions and percentages of the area were determined in order to reveal this effect (Figure 4). In addition, the relative frequency distribution percentage values of the aspect and slope data of the landslide areas of the study area are given (Table 4).

Data mining enables the relationships between data to be analyzed better than classical statistical methods. As was observed in the study, the ability to analyze all data sets at the same time instead of individually brings great advantages in terms of revealing the real feature of the area. The human factor is minimized in the evaluations and increasing the number of attributes does not affect these systems, however, it can provide more effective results in revealing the relationships between the data with these methods. In addition, the model is optimized with the iterative approaches used in the methods.

As association rules, which are widely preferred in the trade and marketing sector, provide a probabilistic correlation between the data, it makes it easier for the probabilities between the attributes of the landslide areas to be examined [56,65–68]. The results of the Apriori algorithm rules of the study area are given in Table 5. Rule 1 had the highest confidence in the area (Figure 5). According to this rule, LV is seen in 55% of the areas with SS in the study area. According to Rule 2, 53% of the fields with VSS have 1th Degree Stream Line. These rules should be taken into account when evaluating the work area.

K-means clustering was also used in landslide assessments [23,56,67,110,111]. Accordingly, the patterns of the landslide area were determined and are shown in Figure 6. Moreover, the results obtained are presented in Table 6.

Performance metrics can be used to evaluate results from clustering. The cluster patterns of the preliminary factors of the study area obtained using the *K*-means algorithm were evaluated for the whole area. The landslides of the clusters at the time of the field investigation and in 2020 were calculated with the performance metrics methods (Table 7).

According to the performance metrics results, the highest values are in the C3 cluster. It can be said that landslides continue where there are C3 clusters, especially in terms of recall values.

5 Conclusion and recommendation

This paper was aimed to evaluate the pre-landslide conditions in Karahacılı District using traditional statistics and data mining methods by creating inventory map.

Relative frequency distribution for aspect and slope in the study area has been evaluated. It was seen that 36% is SS and 44% is NE aspect. This type of information is among the important information recorded in the forms prepared for landslide inventory reports.

The interrelationship of preliminary factors such as topographic, hydrological and vegetation in the region has been determined. All emerging rules are shown with maps. Accordingly, LV was observed in the areas where the region was SS. This resulting Rule 1, which has got 55% confidence, is shown on the map and it is seen that landslide areas are included in this rule. Moreover, Rule 1 is dominant in most of the study area. Apriori algorithm can be used to prepare detailed reports because the relations are very important. Moreover, preliminary factors which have similar properties have been revealed using *K*-means algorithm. Thus, the possibility of multidimensional analysis has emerged for the study area. The distribution of the obtained clusters is visualized using maps. Whether the clusters are in the landslide areas or not was evaluated with performance metrics. While doing this process, 2018 Google Earth, 2019 UAV and 2020 Google Earth images were used. Regions with landslides in 2019 and continuing in 2020 (TP) and new landslides in 2020 (FP) were determined and clusters in these areas were evaluated. The ratio of the sum of the TP and FP values of the C2 cluster to the sample number was 58%, which was higher than the others. This shows us that special attention should be paid to the areas where the C2 cluster is located.

When the pre-landslide orthomosaic map (2018) and Google Earth images (2018) of the area were examined, it was observed that the land use had been changed by the villagers. Previously there had been fruit trees in the crown of the landslide, however, these trees had been cut over time and the areas were converted into agricultural land. The road connecting the village and the city center became unusable after the landslide and, therefore transportation was cut off. People in the area had suffered economic losses due to the shift in agricultural lands. During the interviews conducted with the administrative manager responsible for the area, it was determined that the agricultural lands of nine families were affected by the landslide.

In the later stages of the study, Preliminary and triggering factors can be evaluated together, besides *K*-Means, other ML algorithms can be used and necessary comparisons can be made.

As a result, this paper presents a suggestion on how to evaluate landslides in small areas that cannot be detected due to spatial resolution limitations of satellite images, but seriously affect local life, unlike landslides seen in large areas. Local organizations can create event-based inventory landslide mapping for such landslides by using more detailed, fast and low-cost approaches with the help of ground-based platforms such as UAV.

Decision-makers, on the other hand, can make multidimensional analyzes about the basic structure of the region with the help of algorithms such as Apriori and *K*-means, showing possible risky areas on inventory maps and limiting settlement permits, preventing possible loss of life, property and economic losses.

For landslide zones, databases should be created, monitored and evaluated in systems by making large-scale maps of the zones using modern measurement tools such as UAV and LiDAR. For the analysis of the databases to be created, as the data sets grow, it becomes necessary to use modern methods such as

machine learning and artificial intelligence, which are spatial data mining methods.

Acknowledgments: The authors would like to thank the United States Geological Survey (USGS) for offering the remote sensing data.

Author contributions: L.K. managed the research, analyzed and visualized, L.K., F.B.U., A.A., and M.O.C. considered the research materials. L.K., F.B.U., A.A., and M.O.C. conducted a field survey, and L.K., F.B.U., A.A., M.O.C., and M.Y. reviewed and edited together.

Conflict of interest: Authors state no conflict of interest.

References

[1] Mizutori M, Guha-Sapir D. Economic Losses, Poverty & Disasters (1998–2017) [Internet]. Vol. 4. Centre for Research on the Epidemiology of Disasters (CRED), UNISDR; 2017. Available from: https://www.preventionweb.net/files/61119_credeconomiclosses.pdf

[2] Leoni B, Radford T, Schulman M. Disaster through a different lens behind every effect, there is a cause, a guide for journalists covering disaster risk reduction. Geneva, Switzerland; 2011.

[3] AFAD. Afet Yönetimi Kapsamında 2019 Yılına Bakış ve Doğa Kaynaklı Olay İstatistikleri [Internet]. Ankara, Turkey; 2020. Available from: https://www.afad.gov.tr/kurumlar/afad.gov.tr/e_Kutuphane/Kurumsal-Raporlar/Afet_Istatistikleri_2020_web.pdf

[4] Fidan S, Gorum T. Türkiye'de Ölümcül Heyelanların Dağılım Karakteristikleri ve Ulusal Ölçekte Öncelikli Alanların Belirlenmesi. Türk Coğrafya Derg. 2020;74:123–34.

[5] Anis Z, Wissem G, Vali V, Smida H, Mohamed Essghaier G. GIS-based landslide susceptibility mapping using bivariate statistical methods in North-western Tunisia. Open Geosci. 2019;11(1):708–26.

[6] Milevski I, Dragićević S, Zorn M. Statistical and expert-based landslide susceptibility modeling on a national scale applied to North Macedonia. Open Geosci. 2019;11(1):750–64.

[7] Crawford MM. Kentucky geological survey landslide inventory: from design to application. Lexington: Kentucky Geological Survey, University of Kentucky; 2014. p. 18.

[8] Cruden DM, Varnes DJ. Chapter 3. Landslide types and processes. Landslides: investigation and mitigation, transportation research board special report 247. Washington DC; 1996. p. 36–75 (Bell 1992).

[9] Hungr O, Leroueil S, Picarelli L. The Varnes classification of landslide types, an update. Landslides. 2014;11(2):167–94.

[10] Agca M, Gultekin N, Kaya E. İnsansız Hava Aracından Elde Edilen Veriler ile Kaya Düşme Potansiyelinin Değerlendirilmesi: Adam Kayalar Örneği, Mersin. Geomatik. 2020;5(2):134–45.

[11] Alptekin A, Celik MO, Dogan Y, Yakar M. Mapping of a rockfall site with an unmanned aerial vehicle. Mersin Photogramm J. 2019;1(1):12–6.

[12] Comert R, Avdan U, Gorum T, Nefeslioglu HA. Mapping of shallow landslides with object-based image analysis from unmanned aerial vehicle data. Eng Geol. 2019;260(August):105264. doi: 10.1016/j.enggeo.2019.105264.

[13] Ciampalini A, Raspini F, Bianchini S, Frodella W, Bardi F, Lagomarsino D, et al. Remote sensing as tool for development of landslide databases: the case of the Messina Province (Italy) geodatabase. Geomorphology. 2015;249:103–18. doi: 10.1016/j.geomorph.2015.01.029.

[14] Achour Y, Garçia S, Cavaleiro V. GIS-based spatial prediction of debris flows using logistic regression and frequency ratio models for Zêzere River basin and its surrounding area, Northwest Covilhã, Portugal. Arab J Geosci. 2018;11:181–7.

[15] Hao L, Rajaneesh A, Westen C Van, Sajinkumar KS, Martha TR, Mcadoo JB. Monsoon disaster Kerala, India, land use change Anal. 2020;2(June):1–32.

[16] Shirani K, Pasandi M. Landslide monitoring and the inventory map validation by ensemble DInSAR processing of ASAR and PALSAR images (Case Study: Doab-Samsami Basin in Chaharmahal and Bakhtiari Province, Iran). Geotech Geol Eng. 2021;39(2):1201–22. doi: 10.1007/s10706-020-01554-5.

[17] Çan T, Duman T, Hopa. (Doğu Karadeniz) Bölgesi Heyelan Olay Envanter Haritası ve Yağış Şiddet İlişkisi. In: MÜHJEO'2017: Ulusal Mühendislik Jeolojisi – Jeoteknik Sempozyumu. Adana, Turkey; 2017. p. 12–4.

[18] KGM. Heyelan tanımlama ve veri oluşturma kılavuzu. Ankara, Turkey: Republic of Turkey, The Ministry of Transport and Infrastructure, Karayolları Genel Müdürlüğü (KGM)-General Directorate of Highways; 2015. p. 42.

[19] Lv Z, Liu T, Kong XB, Shi C, Benediktsson JA. Landslide inventory mapping with bitemporal aerial remote sensing images based on the dual-path fully convolutional network. IEEE J Sel Top Appl Earth Obs Remote Sens. 2020;13(March):4575–84.

[20] Litoseliti A, Koukouvelas IK, Nikolakopoulos KG. An event-based inventory approach in landslide hazard assessment: the case of the Skolis mountain, northwest Peloponnese, Greece. ISPRS Int J Geo-Information. 2020;9(7):457.

[21] Ghosh S, Westen CJVan, John E, Carranza M, Jetten VG, Cardinali M. Generating event-based landslide maps in a data-scarce Himalayan environment for estimating temporal and magnitude probabilities. Eng Geol. 2012;128:49–62. doi: 10.1016/j.enggeo.2011.03.016.

[22] Gokesch K, Glade T, Schweigl J. Event-based rapid landslide mapping including estimation of potential human impacts on landslide occurrence: a case study in Lower Austria. In: Koboltschnig G, editor. Lucerne, Switzerland: The International Research Society INTERPRAEVENT; 2016. p. 513–24. Available from: http://www.interpraevent.at/palm-cms/upload_files/Publikationen/Proceedings/IP_2016.pdf.

[23] Guzzetti F, Mondini AC, Cardinali M, Fiorucci F, Santangelo M, Chang KT. Landslide inventory maps: new tools for an old problem. Earth-Sci Rev. 2012;112(1–2):42–66. doi: 10.1016/j.earscirev.2012.02.001.

[24] Galli M, Ardizzone F, Cardinali M, Guzzetti F, Reichenbach P. Comparing landslide inventory maps. Geomorphology. 2008;94(3–4):268–89.

[25] Ardizzone F, Basile G, Cardinali M, Casagli N, Del Conte S, Del Ventisette C, et al. Landslide inventory map for the Briga and the Giampilieri catchments, NE Sicily, Italy. J Maps. 2012;8(2):176–80.
[26] Del Ventisette C, Righini G, Moretti S, Casagli N. Multitemporal landslides inventory map updating using spaceborne SAR analysis. Int J Appl Earth Obs Geoinf. 2014;30(1):238–46. doi: 10.1016/j.jag.2014.02.008.
[27] Santangelo M, Gioia D, Cardinali M, Guzzetti F, Schiattarella M. Landslide inventory map of the upper Sinni River valley, Southern Italy. J Maps. 2015;11(3):444–53. doi: 10.1080/17445647.2014.949313.
[28] Achour Y, Boumezbeur A, Hadji R, Chouabbi A, Cavaleiro V, Bendaoud EA. Landslide susceptibility mapping using analytic hierarchy process and information value methods along a highway road section in Constantine, Algeria. Arab J Geosci. 2017;10(8):194.
[29] Fiorucci F, Ardizzone F, Mondini AC, Viero A, Guzzetti F. Visual interpretation of stereoscopic NDVI satellite images to map rainfall-induced landslides. Landslides. 2019;16(1):165–74.
[30] Orhan O, Oliver-Cabrera T, Wdowinski S, Yalvac S, Yakar M. Land subsidence and its relations with sinkhole activity in karapınar region, turkey: a multi-sensor insar time series study. Sens (Switz). 2021;21(3):1–17.
[31] Shao X, Ma S, Xu C, Shen L, Lu Y. Inventory, distribution and geometric characteristics of landslides in Baoshan City, Yunnan Province, China. Sustainability (Switz). 2020;12(6):2433.
[32] Rosi A, Tofani V, Tanteri L, Tacconi Stefanelli C, Agostini A, Catani F, et al. The new landslide inventory of Tuscany (Italy) updated with PS-InSAR: geomorphological features and landslide distribution. Landslides. 2018;15(1):5–19.
[33] Sun W, Tian Y, Mu X, Zhai J, Gao P, Zhao G. Loess landslide inventory map based on GF-1 satellite imagery. Remote Sens. 2017;9(4):1–17.
[34] Đurić D, Mladenović A, Pešić-Georgiadis M, Marjanović M, Abolmasov B. Using multiresolution and multitemporal satellite data for post-disaster landslide inventory in the Republic of Serbia. Landslides. 2017;14:1467–82, Available from: https://link.springer.com/article/10.1007/s10346-017-0847-2?shared-article-renderer
[35] Hadji R, Achour Y, Hamed Y. Using GIS and RS for slope movement susceptibility mapping: comparing AHP, LI and LR methods for the Oued Mellah Basin, NE Algeria. Recent advances in environmental science from the Euro-Mediterranean and surrounding regions; 2018. p. 1853–6. Available from: http://link.springer.com/10.1007/978-3-319-70548-4_536
[36] Görüm T. Tectonic, topographic and rock-type influences on large landslides at the northern margin of the Anatolian Plateau. Landslides. 2018 Feb 1;16(2):333–46. Available from: http://link.springer.com/10.1007/s10346-018-1097-7
[37] Brooks GR. Sensitive clay landslide inventory map and database for Ottawa [Internet]; 2019. Available from: http://geoscan.nrcan.gc.ca/
[38] Conforti M, Mercuri M, Borrelli L. Morphological changes detection of a large earthflow using archived images, lidar-derived dtm, and uav-based remote sensing. Remote Sens. 2021;13(1):1–25.
[39] Guerriero L, Ruzza G, Cusano A, Focareta M, Revellino P, Guadagno FM. Landslide change detection and displacement tracking using nanosatellite imagery: La Montagna landslide, southern Italy. Italian J Eng Geol Environ. 2019;2019(Special Issue 1):53–8.
[40] Lin J, Wang M, Yang J, Yang Q. Landslide identification and information extraction based on optical and multispectral UAV remote sensing imagery. IOP Conf Series Earth Environ Sci [Internet]. 2017 Feb;57(1):012017, Available from: https://iopscience.iop.org/article/10.1088/1755-1315/57/1/012017
[41] Kusak L, Unel FB, Alptekin A, Celik MO, Yakar M. Modelling of a landslide site with satellite and UAV. In: Yakar M, editor. CISET – 2nd Cilicia International Symposium on Engineering and Technology 10–12 October, 2019, Mersin/TURKEY. Mersin: SAGE; 2019. p. 771–5
[42] Torun AT, Orhan O. Investigation of the effect of temporal baseline on DEMs derived with COSMO sky-med data. Int J Eng Geosci. 2020;6(3):157–64.
[43] Senkal E, Kaplan G, Avdan U. Accuracy assessment of digital surface models from unmanned aerial vehicles' imagery on archaeological sites. Int J Eng Geosci. 2021;6(2):81–9.
[44] Yemenicioglu C, Kaya S, Seker DZ. Accuracy of 3D (three-dimensional) terrain models in simulations. Int J Eng Geosci. 2016;1(1):30–3.
[45] Akar A. Evaluation of accuracy of dems obtained from Uav-point clouds for different topographical areas. Int J Eng Geosci. 2017;2(3):110–7.
[46] Plank S, Twele A, Martinis S. Landslide mapping in vegetated areas using change detection based on optical and polarimetric SAR data. Remote Sens. 2016;8(4):307.
[47] Zeybek M, Şanlıoğlu İ. Investigation of landslide detection using radial basis functions: a case study of the Taşkent landslide, Turkey. Environ Monit Assess. 2020;192:41–9.
[48] Han J, Kamber M, Pei J. Data mining: Data mining concepts and techniques. Data Min Concepts Tech. 2012;3:740.
[49] Olson DL, Delen D. Advanced data mining techniques [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 180. Available from: http://link.springer.com/10.1007/978-3-540-76917-0
[50] Maimon O, Rokach L. Data mining and knowledge discovery handbook. In: Maimon O, Rokach L. editors. Data mining and knowledge discovery handbook. Boston, MA: Springer US; 2010. p. 1306. http://link.springer.com/10.1007/978-0-387-09823-4
[51] Kusak L. Web Madenciliği ve Mekânsal İçerik Tespiti. Geomatik. 2019;4(1):14–22.
[52] Li D, Wang S, Li D. Spatial data mining theory and application [Internet]. Vol. 1. Berlin, Heidelberg: Springer-Verlag; 2015. p. 329. http://www.springerreference.com/index/doi/10.1007/SpringerReference_62913
[53] Mezaa MR, Pradhan B. Data mining-aided automatic landslide detection using airborne laser scanning data in densely forested tropical areas. Korean J Remote Sens. 2018;34(1):45–74.
[54] Lai JS. Separating landslide source and runout signatures with topographic attributes and data mining to increase the quality of landslide inventory. Appl Sci (Switz). 2020;10(19):1–23.

[55] Erener A, Mutlu A, Sebnem Düzgün H. A comparative study for landslide susceptibility mapping using GIS-based multi-criteria decision analysis (MCDA), logistic regression (LR) and association rule mining (ARM). Eng Geol. 2016;203:45–55. doi: 10.1016/j.enggeo.2015.09.007

[56] Althuwaynee OF, Aydda A, Hwang I-T, Lee Y-K, Kim S-W, Park H-J, et al. Uncertainty reduction of unlabeled features in landslide inventory using machine learning t-SNE clustering and data mining apriori association rule algorithms. Appl Sci. 2021;11(2):556.

[57] Wang X, Niu R. Spatial forecast of landslides in three gorges based on spatial data mining. Sensors. 2009 Mar 18;9(3):2035–61, http://www.mdpi.com/1424-8220/9/3/2035.

[58] Tien D, Ho BT, Pradhan B, Pham B. GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. Environ Earth Sci. 2016;75(14):1–22.

[59] Wang Q, Wang Y, Niu R, Peng L. Integration of information theory, K-Means cluster analysis and the logistic regression model for landslide susceptibility mapping in the three gorges area, China. Remote Sens. 2017;9(9):938.

[60] Keyport RN, Oommen T, Martha TR, Sajinkumar KS, Gierke JS. A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. Int J Appl Earth Obs Geoinf. 2018;64(February):1–11.

[61] Tran CJ, Mora OE, Fayne JV, Gabriela Lenzano M. Unsupervised classification for landslide detection from airborne laser scanning. Geosci (Switz). 2019;9(5):221.

[62] Tehrani FS, Santinelli G, Herrera Herrera M. Multi-regional landslide detection using combined unsupervised and supervised machine learning. Geomatics Nat Hazards Risk. 2021;12(1):1015–38. doi: 10.1080/19475705.2021.1912196.

[63] Gorsevski PV, Jankowski P, Gessler PE. Spatial prediction of landslide hazard using fuzzy k-means and Dempster-Shafer theory. Trans GIS. 2005;9(4):455–74.

[64] Barbu M, Radoi A, Suciu G. Landslide monitoring using convolutional autoencoders. Proceedings of 12th International Conference on Electrical Computers Artif Intelligence, ECAI 2020. Vol. 826452; 2020. p. 1–6.

[65] Guo W, Zuo X, Yu J, Zhou B. Method for mid-long-term prediction of landslides movements based on optimized Apriori algorithm. Appl Sci (Switz). 2019;9(18):3819.

[66] Wu X, Benjamin Zhan F, Zhang K, Deng Q. Application of a two-step cluster analysis and the Apriori algorithm to classify the deformation states of two typical colluvial landslides in the Three Gorges, China. Environ Earth Sci. 2016;75(2):1–16.

[67] Ma J, Tang H, Hu X, Bobet A, Zhang M, Zhu T, et al. Identification of causal factors for the Majiagou landslide using modern data mining methods. Landslides. 2017;14(1):311–22. doi: 10.1007/s10346-016-0693-7.

[68] Li L, Liu R, Yang X, Yang M, Yang Y. Selection of landslide affecting factors based on strong association analysis. In: IOP Conference Series: Materials Science and Engineering [Internet]. Vol. 780. Beijing China: IOP Publishing; 2020. p. 1–6. Available from: https://iopscience.iop.org/article/10.1088/1757-899X/780/7/072051.

[69] Conforti M, Muto F, Rago V, Critelli S. Landslide inventory map of north-eastern Calabria (South Italy). J Maps. 2014;10(1):90–102. doi: 10.1080/17445647.2013.852142.

[70] Khan MH, Sarkar SK. Landslides: an inventory analysis of chattogram city inventory analysis. Proceedings on International Conference on Disaster Risk Management; 2019. p. 440–3.

[71] García-Palomo A, Carlos-Valerio V, López-Miguel C, Galván-García A, Concha-Dimas A. Landslide inventory map of Guadalupe range, north of the Mexico basin. Bol Soc Geol Mex. 2006;58(2):195–204.

[72] Achour Y, Pourghasemi HR. How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? Geosci Front. 2020;11(3):871–83. doi: 10.1016/j.gsf.2019.10.001.

[73] Ghosh T, Bhowmik S, Jaiswal P, Ghosh S, Kumar D. Generating substantially complete landslide inventory using multiple data sources: a case study in Northwest Himalayas, India. J Geol Soc India. 2020;95(1):45–58.

[74] MTA. Türkiye Heyelan Envanteri Projesi (1998–2009) [Internet]; 2021. Available from: https://www.mta.gov.tr/v3.0/bilgi-merkezi/muskovit

[75] Can T, Duman TY, Cil E, Mazman T. Mersin Merkez ve Erdemli ilçeleri Kuzeyinin Coğrafi Bilgi Sistemleri Tabanlı Heyelan Envanter, Duyarlılık, Olası Tehlike ve Risk Değerlendirmesi; 2009.

[76] Legorreta Paulín G, Bursik M, Hubp JL, Mejía LMP, Aceves Quesada FA. GIS method for landslide inventory and susceptibility mapping in the Río El Estado watershed, Pico de Orizaba volcano, México. Nat Hazards. 2014;71(1):229–41.

[77] Alptekin A, Celik MO, Kusak L, Bunyan Unel F. Availability of Anafi Parrot in disaster site modelling. CISET – 2nd Cilicia International Symposium on Engineering and Technology 10–12 October, 2019, Mersin/TURKEY. Mersin; 2019. p. 272–5.

[78] Fenton GA, McLean A, Nadim F, Griffiths DV. Landslide hazard assessment using digital elevation models. Can Geotech J. 2013;50(6):620–31.

[79] Capitani M, Ribolini A, Bini M. The slope aspect: a predisposing factor for landsliding. Comptes Rendus - Geosci. 2013;345(11–12):427–38. doi: 10.1016/j.crte.2013.11.002.

[80] Demir G, Aytekin M, Akgün A, İkizler SB, Tatar O. A comparison of landslide susceptibility mapping of the eastern part of the North Anatolian Fault Zone (Turkey) by likelihood-frequency ratio and analytic hierarchy process methods. Nat Hazards. 2013;65(3):1481–506.

[81] Park S, Choi C, Kim B, Kim J. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. Environ Earth Sci. 2013;68(5):1443–64.

[82] Highland LM, Bobrowsky P. The landslide handbook — a guide to understanding landslides. Reston, Virginia: U.S. Geological Survey Circular 1325; 2008. p. 129.

[83] Duman TY, Çan T, Emre Ö, Keçer M, Doğan A, Ateş Ş, et al. Landslide inventory of northwestern Anatolia, Turkey. Eng Geol. 2005;77(1–2):99–114.

[84] Van Den Eeckhaut M, Hervás J. Landslide inventories in Europe and policy recommendations for their interoperability and harmonisation-A JRC contribution to the EU-FP7

SafeLand project [Internet]. Vol. 1; 2012. Available from: http://eusoils.jrc.ec.europa.eu/library/themes/Landslides/
[85] Sukristiyanti S, Wikantika K, Sadisun IA, Yayusman LF, Pamela P. Utilization of google maps for depicting landslide pattern in Indonesia. IOP Conf Series: Earth Environ Sci. 2020;500:1012042.
[86] Hölbling D, Eisank C, Albrecht F, Vecchiotti F, Friedl B, Weinke E, et al. Comparing manual and semi-automated landslide mapping based on optical satellite images from different sensors. Geosci (Switz). 2017;7(2):37.
[87] Ramos-Bernal RN, Vázquez-Jiménez R, Romero-Calcerrada R, Arrogante-Funes P, Novillo CJ. Evaluation of unsupervised change detection methods applied to landslide inventory mapping using ASTER imagery. Remote Sens. 2018;10(12):1987.
[88] Moosavi V, Talebi A, Shirmohammadi B. Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by Taguchi method. Geomorphology [Internet]. 2014;204:646–56. doi: 10.1016/j.geomorph.2013.09.012.
[89] Piralilou ST, Blaschke T, Ghorbanzadeh O. An integrated approach of machine-learning models and Dempster-Shafer theory for landslide detection. ESA Living Planet Symp. 2018;15(2010):2018.
[90] Freund RJ, Wilson WJ, Mohr DL. Data and statistics. In: Rudolf J. Freund, William J. Wilson DLM, editors. Statistical methods. 3rd ed. Canada: Academic Press; 2010. p. 1–65. Available from: https://www.sciencedirect.com/science/article/pii/B9780123749703000019.
[91] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. ACM SIGMOD Rec. 1993;22(2):207–16.
[92] Agrawal R, Srikant R. Fast algorithms for mining association rules. In: Proceedings 20th Internation Conference on Very Large Data Bases, VLDB; 1994. p. 487–99.
[93] MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability; 1967. p. 281–97.
[94] Ville B de. Microsoft® data mining integrated business intelligence for e-commerce and knowledge management. USA: Digital Press; 2001. p. 320.
[95] Kantardzic M. Data mining. IEEE transactions. Vol. 36. Wiley; 2019. p. 495–6. Available from: http://www.tandfonline.com/doi/abs/10.1080/07408170490426107
[96] Huynh V, San O, Nakamori Y. An alternative extension of the k-means algorithm for clustering categorical data. Int J Appl Math Computer Sci. 2004;14(2):241–7.
[97] Sharma R, Rani A. K-means clustering in spatial data mining using weka interface. Int Conf Adv Commun Comput Technol. 2012;2012:26–30.
[98] Ng RT, Han J. Efficient and effective clustering methods for spatial data mining. Proceedings of VLDB; 1994. p. 144–55
[99] Piralilou ST, Shahabi H, Jarihani B, Ghorbanzadeh O, Blaschke T, Gholamnia K, et al. Landslide detection using multi-scale image segmentation and different machine learning models in the higher himalayas. Remote Sens. 2019;11(21):2575.
[100] Steinbach M, Karypis G, Kumar V. A comparison of document clustering techniques. KDD Workshop on Text Mining. Boston: 2000.
[101] Palacio-Niño J-O, Berzal F. Evaluation metrics for unsupervised learning algorithms; 2019. Available from: http://arxiv.org/abs/1905.05667
[102] Liu P, Wei Y, Wang Q, Xie J, Chen Y, Li Z, et al. A research on landslides automatic extraction model based on the improved mask R-CNN. ISPRS Int J Geo-Information. 2021;10(3):168.
[103] Bharathi BS, Swamy KV. Effective image segmentation using modified K-means technique. In: Proceedings of the 4th International Conference on Trends in Electronics and Informatics, ICOEI 2020; 2020. p. 757–62.
[104] Geography. Measuring Slope Steepness [Internet]. Barcelona Field Studies Centre; 2021. Available from: https://geographyfieldwork.com/SlopeSteepnessIndex.htm
[105] Wubalem A, Meten M. Landslide susceptibility mapping using information value and logistic regression models in Goncha Siso Eneses area, northwestern Ethiopia. SN Appl Sci. 2020;2(5):1–19. doi: 10.1007/s42452-020-2563-0.
[106] Wubalem A. Landslide susceptibility mapping using statistical methods in uatzau catchment area, northwestern Ethiopia. Geoenviron Disasters. 2020;8(1):1–21.
[107] Alkhasawneh MS, Ngah UK, Tay LT, Mat Isa NA, Al-Batah MS. Determination of important topographic factors for landslide mapping analysis using MLP network. Sci World J. 2013;2013:415023.
[108] Le TTT, Kawagoe S. Landslide detection analysis in north Vietnam base on satellite images and digital geographical information. Landsat 8 satellite and historical data approaches. J Jpn Soc Civ Eng Ser G (Environ Res). 2017;73(5):I_239–49. Available from https://www.jstage.jst.go.jp/article/jscejer/73/5/73_I_239/_article/-char/ja/
[109] Preti F, Letterio T. Shallow landslide susceptibility assessment in a data-poor region of guatemala (Comitancillo municipality). J Agric Eng. 2015;46(3):85–94.
[110] Gorsevski PV, Gessler PE, Jankowski P. Integrating a fuzzy k-means classification and a Bayesian approach for spatial prediction of landslide hazard. J Geogr Syst. 2003;5(3):223–51.
[111] Souza FT, Ebecken NFF. A data mining approach to landslide prediction. Manag Inf Syst. 2004;10:423–32.