Classification of Coseismic Landslides Using Fuzzy and Machine Learning Techniques

Anastasios Panagiotis Psathas[1(✉)] , Antonios Papaleonidas[1] ,
George Papathanassiou[1], Sotiris Valkaniotis[2], and Lazaros Iliadis[1]

[1] Department of Civil Engineering, School of Engineering, Democritus
University of Thrace, University Campus, PC 67100 Xanthi, Greece
{anpsatha, papaleon, gpapatha, liliadis}@civil.duth.gr
[2] Koronidos Street 9, 42131 Trikala, Greece

Abstract. Seismically generated landslides represent one of the most damaging hazards associated with earthquakes in countries with high seismicity. The delineation of prone to coseismic landsliding areas is crucial in order to predict the occurrence of earthquake-induced landslides and consequently reduce the relevant risk. The goal of this study is to investigate the correlation of the pattern of coseismic landslides with geological and topographical variables i.e. lithology, slope angle and slope aspect with the volume of landslides based on fuzzy logic and machine learning techniques. For this task, a real dataset of 421 and 767 instances for years 2003 and 2015 respectively from the island of Lefkada was used. A new approach based on *Fuzzy C-Means Algorithm* and *Ensemble Subspace k-Nearest-Neighbors (Ensemble Subspace k-NN)* is proposed. Landslides were classified according to their severity with a success rate of 99.5% and 98.7% for 2003 and 2015 respectively. The performance of the proposed approach was evaluated using "One Versus All" Strategy, calculating Accuracy, Sensitivity, Specificity, Precision and F-1 Score for each cluster.

Keywords: Landslides · Lefkada · Fuzzy C-Means · S-Nome · k-Nearest-Neighbors · Ensemble Subspace k-NN · Clustering · Classification

1 Introduction

It is well known that landslide can be triggered by rainfall, earthquakes, volcanic eruption and man-made activities. Experience has shown that seismically induced landslides represent one of the most damaging hazards associated with earthquakes in countries with high seismicity [1]. According to Jibson et al. [18], the effect of seismically-induced landslides on human lives and facilities may exceed in some cases the damage directly connected to the shaking. The correlation of coseismic landslides with the seismic and morphological parameters has been investigated by several researchers, mainly after the devastating 2008 Wenchuan, China earthquake. The outcome arisen by this correlation is that both the volume and number of landsliding phenomena are relevant to earthquake magnitude. In particular, it was shown that

landslide frequencies are higher in areas of highest peak ground acceleration (PGA) and that landslide density decays with the epicentral or fault distance [2, 20, 25, 31].

The delineation of prone to coseismic landsliding areas is crucial in order to predict the occurrence of earthquake-induced landslides and consequently reduce the relevant risk. Nowadays, satellite imagery and GIS technology are considered as basic tools that are used by earth scientist for evaluating the hazard and the risk within an area, initially by the introduction and statistical analyses of geo-environmental and seismologic factors into GIS software [29]. In particular, the characteristics of the landsliding area is statistically related to control factors such as topographic, geologic and seismic parameters e.g. slope angle, slope aspect, curvature, lithology, Peak Ground Acceleration (PGA) and seismic intensity distribution, and distance from the seismic fault or epicenter [27, 39]. These correlations can provide crucial information that can be used for seismic landslide hazard analysis and planning mitigation measures for prone earthquake-induced landslides regions [6, 7, 18, 30].

Frequently, the statistical analysis is based on bivariate and multivariate approaches. The goal of this study is to investigate the correlation of the pattern of coseismic landslides with geological and topographical variables i.e. lithology, slope angle and slope aspect with the volume of landslides based on fuzzy logic and machine learning techniques. In particular, *Fuzzy C-Means Algorithm* [3, 12] was used for data clustering and Ensemble Subspace k-Nearest-Neighbors (Ensemble Subspace k-NN) was used for the classification [10, 15, 37]. Existing bibliography like [23] and [33] does not exploit the combination of the above algorithms.

The motivation for the development of this research was the need for a more flexible model. More specifically, the existing approaches (e.g. the bivariate analysis) are using crisp values for the determination of slope angle or slope aspect classes. Thus, borderline values can be easily misclassified.

2   Area of Research

The geology of the Lefkada island, comprises: (1) a carbonate sequence of the Ionian zone, (2) limestone of Paxos (Apulia) zone restricted in the SW peninsula of the island, (3) few outcrops of ionian flysch (turbidites) and Miocene marls-sandstones mainly in the northern part of the island [9, 34]. The boundary between the two different geological zones – Ionian and Paxos, runs in an approximate NW- SE direction through this region and outcrops onshore south-central Lefkada Island near Hortata village, in the form of a buried thrust fault by scree and late Quaternary deposits [28]. Pleistocene and especially Holocene coastal deposits are extended in the northern edge of Lefkada, where the homonym capital town is founded, in the valley of Vassiliki and in the coast Nydri.

Regarding the seismicity, it is pointed out that the island of Lefkada is considered as one of the most tectonically active areas in Europe being part of the high seismicity Ionian Sea area, and particularly due to the complex crustal deformation resulting from the subduction of the African plate towards NE and the Apulian platform continental collision further to the northwest [14, 16]. The main active tectonic structure, is the 140 km long dextral strike-slip Cephalonia-Lefkada Transform fault (Fig. 1) (CTF;

[24, 36, 38]), which has a GPS slip-rate bracketed between 10 and 25 mm/yr. The steep morphology on the western part of the island, where most of slope failure cases are reported is due to this offshore CTF and its onshore sub-parallel fault; the Athani-Dragano fault [9, 35]. The latter one is a NNE-SSW striking fault forming a narrow elongated continental basin, very well expressed in the region's morphology and marked on satellite images and aerial photos.

There is reliable detailed information for at least 23 events, since 1612 which induced ground failures at the island of Lefkada [26]. A first conclusion arising from the list of historical events is that earthquakes appear in couples (twin or cluster events) with time period of occurrence ranging between 2 months and 5 years e.g. 1612–1613 (16 months); 1625–1630 (5 years); 1722–1723 (10 months); 1767–1769 (2 years); 1783–1783 (2 months, possible aftershock); 1867–1869 (2 years); 1914–1915 (2 months); 1948–1948 (2 months). Thus, it is crucial to determine the location of coseismic landslides since it will be beneficial for reducing the risk and increasing the resilience at the island.

Fig. 1. Map of the island of Lefkada showing the Cephalonia-Lefkada Transform Fault CTF

2.1 Coseismic Landslides at the Island of Lefkada

The most recently occurred and well-studied earthquakes are the ones of 2003 and 2015. The penultimate event triggered extensive slope failures at the western part of the island. The volume of the debris material that moved downwards was larger than the one of the 2015 earthquake. Rock falls were widespread on the whole island and especially in the northwestern and central area, on both natural and cut slopes, as well as, on downstream road embankment slopes. The most characteristic rock falls, with diameters up to 4 m, were observed along the 6 km long road of Tsoukalades-Agios Nikitas, which is within the epicentral area, and are accompanied by gravel, small rock and soil slides [26]. The massive occurrence of these failures is the reason for the closure of the road network at this area of the island for more than 2 years. The reported

rock falls followed the trace of a 300 m high morphological scarp, and especially a 10–40 m high artificial slope [26].

Regarding the 2015 earthquake, the dominant geological effects were related to slope failures i.e. rock falls and slides, and shallow and deep-seated landslides on both natural and cut slopes [28]. These failures were documented on the western part of the island, while the most densely concentration of these phenomena was reported on the coastal zone from Porto Katsiki to Egremnoi-Gialos beach and along the 6 km long coastal road of Tsoukalades - Agios Nikitas [28]. Shallow landslides and rock slides were mainly generated in areas where the clastic material covered the bedrock, and particularly in places where the rock mass was heavily jointed. Deep-seated landslides were mainly documented at the area of Egremnoi [29]. At this area, deep-seated landslides were reported, and large amount of debris material moved downslope inducing severe damages to the road network and to residential houses. The debris consists of coarse-grained size material with significant amount of large-diameter gravels and few boulders.

In order to investigate the earthquake-induced landslide density, event-based inventories were developed by taking into account aerial and satellite imagery in Google Earth in order to enrich and update existing landslide datasets, previously compiled for the two earthquakes [27]. In particular, Google Earth imagery of June 12, 2003 and December 19, 2005 was used for mapping 2003 earthquake landslides, and November 15, 2013 and April 15, 2016 for 2015 earthquake, respectively. Landslide activity along the western part of Lefkada is considered as minimal between major earthquakes, as observed on multi-date satellite imagery and confirmed by local residents. Considering this, the short period between each satellite imagery pair (2–3 years) is believed to include only the coseismic landslides, with very few if any at all landslides triggered by other factors. In total, 301 and 596 coseismic landslides were mapped for the 2003 and 2015 earthquakes, respectively. For the extraction of morphological and terrain parameters of the compiled landslide datasets, a detailed digital elevation model (DEM) with spatial resolution of 5 m was used. The 5 m DEM was obtained from Hellenic Cadastre and it was extracted from aerial imagery stereo-pairs, having a vertical accuracy of 4 m [29].

Having completed the polygon-based inventories, a statistical analysis of landslide distribution took place. In total, 596 and 301 landslides were identified covering (planar area) 1.29 km^2 and 1.6 km^2 for the 2015 and 2003 events, respectively. It is pointed out that these planar-oriented areas are obtained as projected measurements. The minimum and maximum landslide area were evaluated as 40.4 m^2 and 42940 m^2 for the 2015 earthquake, while for the penultimate event the relevant values are 129.8 m^2 and 98300 m^2, respectively [29]. The relevant values of minimum and maximum landslide area for the 2015 event, which were evaluated by taking into account the digital elevation model for the delineation of the landslide area, are 1.78 km^2 total area, 51.30 m^2 minimum and 58330 m^2 maximum area, while for the 2003 earthquake the total landsliding area covered 2.28 km^2 with minimum area of 140.9 m^2 and maximum 148469 m^2 [29].

3   Description of Dataset Pre-processing

The initial datasets (*.xlsx files) consist of 6 columns, 4 of which are numeric (Perimeter, Average Slope, Surface, Id) and 2 are nominal (Average Aspect and Geological Form). The 5th column is used only for data processing to determine the distinct incidents. Each vector represents a landslide with a specific Geological Form on the island of Lefkada. As it is already mentioned, 301 and 596 landslides were identified for years 2003 and 2015 respectively. However, numerous landslides are related to more than one types of geological forms. Taking into account this fact, the landsliding areas have been reassigned based on the geological form upon which they were delineated, resulting in 421 and 767 observations for years 2003 and 2015, respectively. The same data pre-processing approach was used for both datasets. However, data handling for each year was applied independently from the other, given that 2003 observations have been mapped on two additional geological forms com-pared to the case of 2015. Therefore, evaluation of the proposed approach per year has proven its efficiency and consistency. Data processing was performed in three steps. In particular, the first step was applied manually using both *.xlsx files, while second and third steps was achieved by developing code in *Matlab R2019a*.

3.1   Labeling of Nominal Values

Initially, Average Aspect was transformed from nominal to numeric in a scale from one to eight. Geological Form was similarly transformed to a scale from one to twenty for year 2003 and from one to eighteen for year 2015. The transformations can be seen in Tables 1 and 2.

Table 1. Average aspect with the corresponding label for 2003 and 2015.

Average aspect	North	North-East	East	South-East	South	South-West	West	North-West
Average aspect label	1	2	3	4	5	6	7	8

Table 2. Geological form type with the corresponding label for 2003.

Geological form	al	C	Ci	Cs	Csd	E	J1	J1d	Jar	Jc
Geological form label	1	2	3	4	5	6	7	8	9	10
Geological form	Jm	Js	M	Mb	Pc	Qc	Qp	Qt	Tc	Tg
Geological form label	11	12	13	14	15	16	17	18	19	20

Table 3. Geological form type with the corresponding label for 2015.

Geological form	al	C	Ci	Cs	E	J1	J1d	Jar	Jc
Geological form label	1	2	3	4	5	6	7	8	9
Geological form	Jm	Js	M	Mb	Pc	Qc	Qp	Qt	Tg
Geological form label	10	11	12	13	14	15	16	17	18

where al: alluvial; C, Jm, Jc, Jar, J1: limestones of Ionian; Ci, Cs: limestones of Paxos; Csd limestones; E: limestones Eocene; Js: limestone of Paxos; J1d: dolomites; M, Mb: Miocene sandstones; Pc: Pliocene conglomerate; Qc, Qp, Qt: Quaternary sediments; Tc: limestones and dolomites of Triassic; Tg: evaporites.

3.2 Landslides Fuzzy C-Means Clustering

After a statistical pre-processing of data, authors noticed that there were few high values causing a high standard deviation. Therefore, clustering of landslides is deterrent with conventional methods. To overcome this hardship, Fuzzy C-Means Clustering was employed to classify landslides according to their severity. Fuzzy C-Means was used because it offers a very flexible methodology, as each data point can be assigned to more than one clusters with different degrees of membership. This task was performed in order to develop the labeled dataset required for the deployment of the machine-learning model. A subset of available data, perimeter and surface, was used in order to apply the Clustering.

Fuzzy clustering (a well-known soft computing method [4]) is an approach in which each data point can belong to more than one cluster. One of the most widely used fuzzy clustering algorithms is the Fuzzy C-means clustering (FCM) Algorithm. This method, developed by Dunn in 1973 [12] and improved by Bezdek [3], is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2, \ 1 \leq m < \infty \tag{1}$$

Where m is the fuzzifier (the fuzzifier m determines the level of cluster fuzziness), u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}, \quad c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{2}$$

This iteration will stop when $\max_{ij} \left\{ \left| u_{ij}^{(k+1)} - u_{ij}^{k} \right| \right\} < \varepsilon$, where ε is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

Parameters used for the FCM algorithm are presented in Table 4 while columns Perimeter and Surface, from source data set, were chosen as input parameters. Exponent (*m fuzzier)* controls the degree of fuzzy overlap between clusters. A large m results in smaller membership values, u_{ij}, and hence, fuzzier clusters. In the limit $m = 1$, the memberships, u_{ij}, converge to 0 or 1, which implies a crisp partitioning. maxIterations is the maximum number of optimization iterations and minImprovement is the minimum improvement in the objective function between successive iterations When the objective function improves by a value below this threshold, the optimization stops. A smaller value produces more accurate clustering results, but the clustering can take longer to converge. For parameters' values, the ones most used in the relevant literature were selected [21, 22].

Table 4. Options for FCM algorithm

Exponent (*m fuzzier*)	2
maxIterations	100
minImprovement	0.00001

A script *FCM.m* was developed in Matlab, aiming the transformation of *.xlsx file in Matlab tables. It has already been mentioned that the goal of this step is to create clusters for the severity of lansdlides. The chosen number of clusters is 6 (calculated as the existing combinations of 2 parameters with 3 states each). Clusters with their labels and their names are presented in Table 5. The first part of the name corresponds to Perimeter and the Second to Surface

Table 5. Cluster with the corresponding labels and names

Cluster	Name of cluster
1	Low Low
2	Low Medium
3	Medium Medium
4	Medium High
5	High High
6	Extreme Extreme

The *FCM.m* Script is presented in the form of natural language, in Algorithm 1.

Algorithm 1. The FCM.m Matlab Script

Script 1: *FCM.m*

Inputs: 421 incidents of 2003 and 767 incidents of 2015 exported from *.xlsx files.

 Part 1:

  Step 1: Read and convert each *.xlsx file to a Matlab table.

  Step 2: For each Table the column Id Numbers was used to retrieve distinct Landslides. 2 new Matlab tables are constructed with the unique values

  Step 3: For each new table Perimeter and Surface column is chosen

 Part 2:

  Step 1: FCM algorithm was applied with options described in Table 4 and clusters described in Table 5. Centers of clusters and membership degrees of each observation were calculated.

  Step 2: Each instance was classified in the cluster on which the membership degree was the highest.

 Part 3:

  Step 1: Clusters pass to the original data

  Step 2: Original data with theirs clusters are plotted in Figures 1a and 1b for 2003 and 2015 respectively.
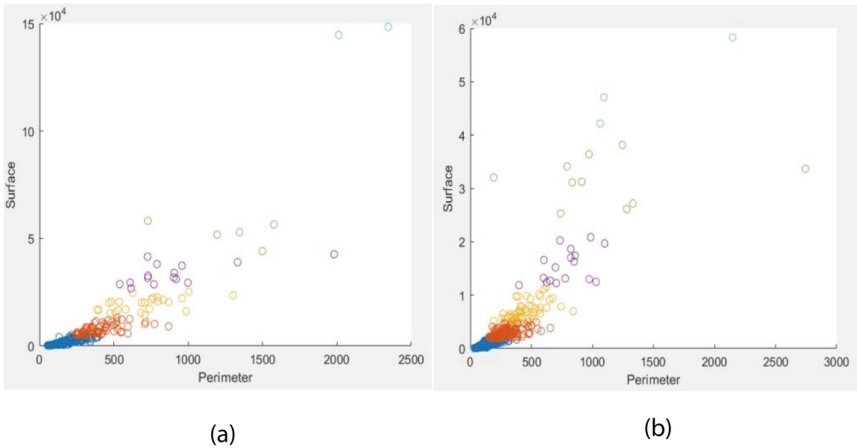

Fig. 2. Clusters for original data for 2003 (1a) and 2015 (1b) respectively.

Total landslides for each cluster for the years 2003 and 2015 are presented in Table 6.

Table 6. Total landslides of each cluster for 2003 and 2015

Clusters	1	2	3	4	5	6	Total instances
2015	448	183	74	35	22	5	767
2003	238	100	44	27	7	5	421

3.3 Fuzzy Clustering with FCM Algorithm and S-Norm

After creating the clusters it was observed that some instances do not belong exclusively to a cluster (e.g. point (1981, $4.2 \cdot 10^4$) presented in Fig. 2a or point (2744, $3.3678 \cdot 10^4$) presented in Fig. 2b), as well as some instances in between 2 clusters with similar membership values for both. Considering the above observation, the re-creation of clusters for all instances is essential. Consequently, we decided that landslides, which have degree of membership, for their dominant class, below a certain threshold will be re-sorted. In order to perform this, it is necessary to use weights on each factor [5, 41].

Equation (3) implements a fuzzy coupling between many fuzzy sets, using a function $f(\mu_i, w_i)$ which assigns the weight w_i to the membership degree μ_i.

$$\mu_{\widetilde{S}}(x_i) = Agg\left(f\left(\mu_{\widetilde{A}}(x_i), w_1\right), f\left(\mu_{\widetilde{A}}(x_i), w_2\right), \ldots, f\left(\mu_{\widetilde{A}}(x_i), w_n\right)\right) \tag{3}$$

Where $i = 1, 2, \ldots, k$ and k the number of instances examined and n factors' number [13]. The function f used in the coupling function (Eq. (3)) can be defined as:

$$f(a, w) = a^{\frac{1}{w}} \tag{4}$$

where a is the membership degree and w is the corresponding weight.

For the Aggregation function was used Hamacher aggregation as S-Norm operator [40].

$$\widetilde{A} \cap \widetilde{B} = \frac{\mu_{\widetilde{A}}(x) + \mu_{\widetilde{B}}(x) - 2\mu_{\widetilde{A}}(x)\mu_{\widetilde{B}}(x)}{[1 - \mu_{\widetilde{A}}(x) + \mu_{\widetilde{B}}(x)]} \tag{5}$$

Another script *S-Norm_Clustering.m* was developed in Matlab. The second script is presented in the form of natural language, in Algorithm 2.

<div align="center">Algorithm 2. The S-Norm_Clustering.m Matlab Script</div>

Script 2: *S-Norm_Clustering.m*

Inputs: All membership degrees of each observation for 2003 and 2015.

> **Part 1:**
>> **Step 1:** Threshold is defined at 0.7
>>
>> **Step 2:** Weights are defined. The highest membership's degree weight is 4, the second highest is 2, and for the others 0.
>
> **Part 2:**
>> **Step 1:** If a membership degree for a cluster is higher than 0.7 the incident belong to this cluster.
>>
>> **Step 2:** Else if the membership degree is under 0.7 the incident belongs to a new cluster, for with the membership degree is calculated through Eq. (3),(4) and (5).

After applied *S-Norm_Clustering.m,* 4 more clusters were created for landslides of 2003 and 5 more for 2015. New clusters were created between the already existing clusters. Therefore, they took their label depending the clusters that are between. For example, cluster between cluster 1 and 2 is labeled as cluster 1.5. Clusters are presented in Fig. 3a, b, and incidents for each cluster in Table 7. It is obvious that new clustering classifies observations more effectively according to severity.
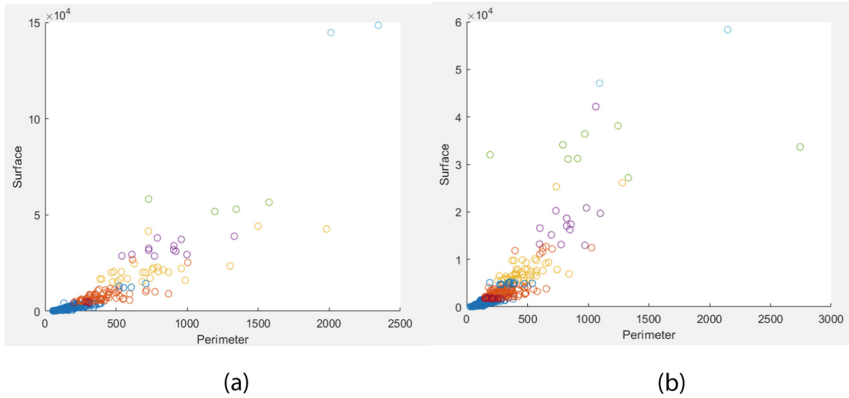

(a)   (b)

Fig. 3. Clusters for original data for 2003 (2a) and 2015 (2b) respectively after S-Norm

Table 7. Total landslides of each cluster for 2003 and 2015 after S-Norm

Clusters	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	Total instances
2015	417	52	145	27	56	25	18	7	15	2	3	767
2003	222	27	84	9	37	5	22	4	6	0	5	421

4   Classification Methodology

After having clustering done, we used classification algorithms to ascertain coseismic landslides' proper classification. The independent variables used for the classification are Perimeter, Average Slope, Surface, Average Aspect and Geological Form of landslides. Average Slope and Average Aspect were labeled as indicated in Tables 1, 2 and 3. The dependable value is the cluster derived from clustering with S-Norm.

A total of 23 classification algorithms have been employed namely: *Fine Tree, Medium Tree, Coarse Tree, Linear Discriminant, Quadratic Discriminant, Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM, Coarse Gaussian SVM, Cosine KNN, Cubic KNN, Weighted KNN, Fine KNN, Medium KNN, Gaussian Naive Bayes, Kernel Naïve Bayes, Boosted Trees, Bagged Trees, Subspace Discriminant, Subspace KNN, RUSBoost Trees.*

However, only the one with the highest values of performance indices will be described herein.

4.1   Ensemble Subspace k-Nearest-Neighbors (Ensemble Subspace k-NN)

Classifying query points based on their distance to specific points (or neighbors) can be a simple but yet effective process. The k-nearest neighbors (k-NN) is a lazy and non-parametric Learning algorithm [8]. It is widely used as a predictive performance benchmark, when we are trying to develop more sophisticated models. Given a set X of n points and a distance function, k-NN search finds the k closest points to a query point or set of them [17]. Dunami [11] first introduced a weighted voting method, called the *distance-weighted (DW) k-nearest neighbor rule* (Wk-NN). According to this approach, the closer neighbors are weighted more heavily than the farther ones, using the DW function. The weight w_i for the *i*-th nearest neighbor of the query x' is defined following function 1:

$$w_i' = \begin{cases} \frac{d(x'x_k^{NN})-d(x'x_i^{NN})}{d(x'x_k^{NN})-d(x'x_1^{NN})} & , \quad if \quad d(x'x_k^{NN}) \neq d(x'x_1^{NN}) \\ 1 & , \quad if \quad d(x'x_k^{NN}) = d(x'x_1^{NN}) \end{cases} \qquad (6)$$

Finally, the classification result of the query is determined by the majority weighted voting as in function 2:

$$y' = \arg\max_y \sum_{(x_i^{NN}, y_i^{NN}) \in T'} w_i' \times \delta(y = y_i^{NN}). \qquad (7)$$

Based on Eq. (7), a neighbor with smaller distance is weighted more heavily than one with greater distance: the nearest neighbor is assigned a weight equal to 1, whereas the furthest one a weight of 0 and the weights of the others are scaled linearly to the interval in between.

Despite its simplicity, k-NN gives competitive results and in some cases even outperforms other complex learning algorithms. However, k-NN is affected by

non-informative features in the data, which is something rather common with high dimensional data. Several attempts have been made to improve the performance of nearest neighbors' classifiers by ensemble techniques. Some related work on ensemble of k-NN classifiers can be found in [10, 15, 17].

Subspace ensembles have the advantage of using less memory than ensembles with all predictors, and can handle missing values (NaNs).

The basic random subspace algorithm uses these parameters.

- m is the number of dimensions (variables) to sample in each learner.
- d is the number of dimensions in the data, which is the number of columns (predictors) in the data matrix X.
- n is the number of learners in the ensemble. Set n using the NLearn input.

The basic random subspace algorithm performs the following steps:

1. Choose without replacement a random set of m predictors from the d possible values.
2. Train a weak learner using just the m chosen predictors.
3. Repeat steps 1 and 2 until there are n weak learners.
4. Predict by taking an average of the score prediction of the weak learners, and classify the category with the highest average score.

4.2   Evaluation of the Activity Model Classifiers

Accuracy is the overall index that has been used in evaluation of the developed Machine Learning models. However, additional indices have been used to estimate the efficiency of the algorithms. Given the fact that we are dealing with a multi-class classification problem, the "One Versus All" Strategy [19, 32] was used. The calculated validation indices that have been considered are presented in the following Table 8.

Table 8. Calculated indices for the evaluation of the multi-class classification approach

Index	Abbreviation	Calculation
Sensitivity (also known as True Positive Rate or Recall)	SNS	$SNS = TP/(TP + FN)$
Specificity, (also known as True Negative Rate)	SPC	$SPC = TN/(TN + FP)$
Accuracy	ACC	$ACC = (TP + TN)/ (TP + FP + FN + TN)$
F1 Score	F1	$F1 = 2*TP/(2*TP + FP + FN)$
Precision (also known as Positive predictive value	PREC	$PREC = TP/(TP + FP)$

Precision (PREC) is the measure of the correctly identified positive cases from all the predicted positive cases. Thus, it is useful when the cost of False Positives is high.

On the other hand, **Sensitivity** (also known as Recall) is the measure of the correctly identified positive cases from all the actual positive cases. It is important when the cost of False Negatives is high. **Specificity** (SPC) is the true negative rate or the proportion of negatives that are correctly identified. **Accuracy** (ACC) is the measure of all correctly identified from the predicted cases. It represents the closeness of the measurements to a specific value. The **F1** score can be interpreted as the harmonic mean (weighted average) of the Precision and Recall. As it is known from the literature, **Accuracy** can be seriously considered when the class distribution is balanced while F1 score is a better metric when there are imbalanced classes as in the above case. Using it as a metric, we are sure that if its value is high, both precision and recall of the classifier indicate good results. In our case the F1 score is the final overall criterion of good performance evaluation.

5   Experimental Results

The experiments were performed with the use of Matlab R2019a software. The options and hyperparameters set for Ensemble Subspace k-NN are presented in Table 9 below:

Table 9.   Tuning algorithm's hyperparameters

Algorithm	Hyperparameters	Optimal values-functions
Ensemble subspace k-NN	Maximum number of splits	20
	Number of learners	30
	Learning rate	0.1
	Subspace dimension	3

The Ensemble Subspace k-NN achieved an accuracy equal to 99.5% and 98.7% for 2003 and 2015 respectively. The Confusion Matrix for each year is presented in the following Figs. 4 and 5.
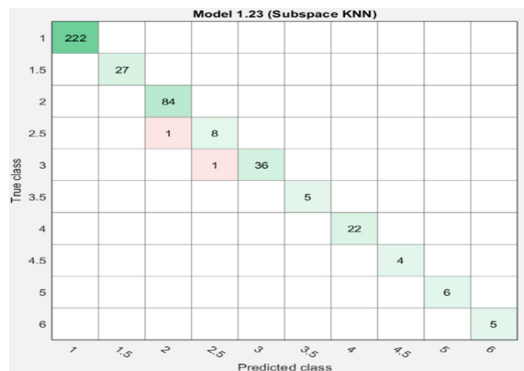

Fig. 4.   Confusion matrix of the ensemble subspace k-NN algorithm for 2003
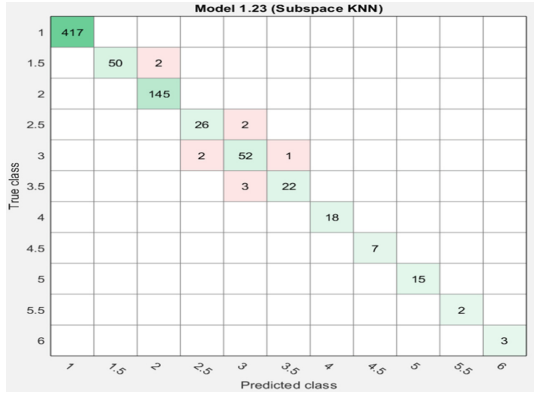
Fig. 5. Confusion matrix of the ensemble subspace k-NN algorithm for 2015

The following two Tables (10, 11), present the values of the performance indices for the above optimal algorithm.

Table 10. Cluster (Cl) classification performance indices for the Ensemble Subspace k-NN Algorithm (2003)

Index	Cl_1	$Cl_{1.5}$	Cl_2	$Cl_{2.5}$	Cl_3	$Cl_{3.5}$	Cl_4	$Cl_{4.5}$	Cl_5	Cl_6
SNS	1	1	1	0.88	0.97	1	1	1	1	0.98
SPC	1	1	0.99	0.99	1	1	1	1	1	0.99
ACC	1	1	0.99	0.99	0.99	1	1	1	1	0.99
PREC	1	1	0.98	0.88	1	1	1	1	1	0.98
F1	1	1	0.99	0.88	0.98	1	1	1	1	0.98

Table 11. Cluster (Cl) classification performance indices for the Ensemble Subspace k-NN Algorithm (2015)

Index	Cl_1	$Cl_{1.5}$	Cl_2	$Cl_{2.5}$	Cl_3	$Cl_{3.5}$	Cl_4	$Cl_{4.5}$	Cl_5	$Cl_{5.5}$	Cl_6
SNS	1	0.96	1	0.93	0.95	0.88	1	1	1	0.97	1
SPC	1	1	0.99	0.99	0.99	0.99	1	1	1	0.99	1
ACC	1	0.99	0.99	0.99	0.98	0.99	1	1	1	0.99	1
PREC	1	1	0.98	0.92	0.91	0.95	1	1	1	0.97	1
F1	1	0.98	0.99	0.92	0.92	0.91	1	1	1	0.97	1

From Tables (10, 11) it is obvious that the values of all indices clearly show a very good performance in both 2003 and 2015's instances.

6 Discussion and Conclusion

Classification of coseismic landslides according to their severity is a really interesting, important and challenging task. In this paper an approach based on *Fuzzy C-Means Algorithm* and *Ensemble Subspace k-Nearest-Neighbors (Ensemble Subspace k-NN) Algorithm* is proposed and tested. The combination of *Fuzzy C-Means* and Hamacher aggregation as S-Norm operator, sorted the data to 10 clusters for year 2003 and 11 clusters for 2015. Thereafter, *Ensemble Subspace k-NN,* using Average Aspect, Average Slope, Geological Form, Perimeter and Surface as independent input variables, managed to achieve high success rates. The overall accuracy is 99.5% and 98.7% for 2003 and 2015 respectively.

The efficiency of the model is also perceivable from Tables 10 and 11, where indices Accuracy, Sensitivity, Specificity, Perception and F1-score range at high levels. However, some ostensibly inaccurate classifications like in $Cl_{2.5}$ of 2003 or $Cl_{3.5}$ of 2015 do not affect the overall performance of the model as all indicators range from 0.88 to 1.

Concluding, the research described herein managed to correctly classify coseismic landslides according to their severity to the island of Lefkada. Future work will focus on the development of hybrid and ensembles' approaches for the forecasting of landslides' severity or even for forecasting the exact area of a landslide.

References

1. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters (1973)
2. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Springer, New York (2013)
3. Calvo, T., Mayor, G., Mesiar, R. (eds.): Aggregation Operators: New Trends and Applications, vol. 97. Physica, New York (2012)
4. Yager, R.R., Kacprzyk, J.: The Ordered Weighted Averaging Operators. Theory and Applications. Springer, New York (1997)
5. Fan, Z.P., Ma, J., Zhang, Q.: An approach to multiple attribute decision making based on fuzzy preference information on alternatives. Fuzzy Sets Syst. **131**(1), 101–106 (2002)
6. Wei, G., Lu, M.A.O.: Dual hesitant pythagorean fuzzy Hamacher aggregation operators in multiple attribute decision making. Arch. Control Sci. **27**(3), 365–395 (2017)
7. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Trans. Inf. Theory **13**(1), 21–27 (1967)
8. Hechenbichler, K., Schliep, K.: Weighted k-nearest-neighbor techniques and ordinal classification (2004)
9. Dudani, S.A.: The distance-weighted k-nearest neighbor rule. IEEE Trans. Syst. Man Cybern. **8**(4), 311–313 (1978)
10. Grabowski, S.: Voting over multiple k-nn classifiers. In: Modern Problems of Radio Engineering, Telecommunications and Computer Science (IEEE Cat. No. 02EX542), pp. 223–225. IEEE, February 2002

11. Domeniconi, C., Yan, B.: Nearest neighbor ensemble. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, vol. 1, pp. 228–231. IEEE, August 2004

12. Samworth, R.J.: Optimal weighted nearest neighbour classifiers. Ann. Stat. **40**(5), 2733–2763 (2012)

13. Joutsijoki, H., Juhola, M.: Comparing the one-vs-one and one-vs-all methods in benthic macroinvertebrate image classification. In: International Workshop on Machine Learning and Data Mining in Pattern Recognition, pp. 399–413. Springer, Berlin/Heidelberg, August 2011

14. Rifkin, R., Klautau, A.: In defense of one-vs-all classification. J. Mach. Learn. Res. **5**(2), 101–141 (2004)

15. Khalilia, M., Popescu, M.: Fuzzy relational self-organizing maps. In: 2012 IEEE International Conference on Fuzzy Systems, pp. 1–6. IEEE, June 2012

16. Khalilia, M.A., Bezdek, J., Popescu, M., Keller, J.M.: Improvements to the relational fuzzy c-means clustering algorithm. Pattern Recognit. **47**(12), 3920–3930 (2014)

17. Bora, D.J., Gupta, D., Kumar, A.: A comparative study between fuzzy clustering algorithm and hard clustering algorithm. arXiv preprint arXiv:1404.6059 (2014)

18. Robinson, T.R., Rosser, N.J., Densmore, A.L., Williams, J.G., Kincey, M.E., Benjamin, J., Bell, H.J.: Rapid post-earthquake modelling of coseismic landslide magnitude and distribution for emergency response decision support. Nat. Hazards Earth Syst. Sci. Discuss. **17**, 1521–1540 (2017)

19. Kritikos, T., Robinson, T.R., Davies, T.R.: Regional coseismic landslide hazard assessment without historical landslide inventories: a new approach. J. Geophys. Res.: Earth Surf. **120** (4), 711–729 (2015)

20. Ayalew, L., Kasahara, M., Yamagishi, H.: The spatial correlation between earthquakes and landslides in Hokkaido (Japan), a GIS-based analysis of the past and the future. Landslides **8**, 443–448 (2011)

21. Barlow, J., Barisin, I., Rosser, N., Petley, D., Densmore, A., Wrigth, T.: Seismically-induced mass movements and volumetric fluxes resulting from the 2010 Mw = 7.2 earthquake in the Sierra Cucapah, Mexico. Geomorphology **230**, 138–145 (2015)

22. Chang, K.T., Chiang, S.H., Hsu, M.L.: Modeling typhoon- and earthquake induced landslides in a mountainous watershed using logistic regression. Geomorphology **89**(3–4), 335–347 (2006)

23. Collins, B.D., Kayen, R., Tanaka, Y.: Spatial distribution of landslides triggered from the 2007 Niigata Chuetsu-Oki Japan Earthquake. Eng. Geol. **127**, 14–26 (2012)

24. Cushing, M.: Evoluation structurale de la marge nord-ouest hellenique dans l'ile de Lefkada et ses environs (Greece nord-occidentale). Ph.D. Thesis, University de Paris-Sud (XI), Centre d' Orsay, France (1985)

25. Ganas, A., Marinou, A., Anastasiou, D., Paradissis, D., Papazissi, K., Tzavaras, P., Drakatos, G.: GPS-derived estimates of crustal deformation in the central and north Ionian Sea, Greece: 3-yr results from NOANET continuous network data. J. Geod. **6**, 62–71 (2013)

26. Hatzfeld, D., Kassaras, I., Panagiotopoulos, D., Amorese, D., Makropoulos, K., Karakaisis, G., Coutant, O.: Microseismicity and strain pattern in northwestern Greece. Tectonics **14**(4), 773–785 (1995)

27. Jibson, R.W., Harp, E.L., Michael, J.A.: A method for producing digital probabilistic seismic landslide hazard maps. Eng. Geol. **58**, 271–289 (2000)

28. Keefer, D.K.: Landslides caused by earthquakes. Geol. Soc. Am. Bull. **95**(4), 406–421 (1984)

29. Louvari, E., Kiratzi, A.A., Papazachos, B.C.: The CTF and its extension to western Lefkada Island. Tectonophysics **308**, 223–236 (1999)

30. Meunier, P., Hovius, N., Haines, A.J.: Regional patterns of earthquake-triggered landslides and their relation to ground motion. Geophys. Res. Lett. **34**(20), L20408 (2007)

31. Papathanassiou, G., Pavlides, S., Ganas, A.: The 2003 Lefkada earthquake: field observation and preliminary microzonation map based on liquefaction potential index for the town of Lefkada. Eng. Geol. **82**, 12–31 (2005)

32. Papathanassiou, G., Valkaniotis, S., Ganas, A., Pavlides, S.: GIS-based statistical analysis of the spatial distribution of earthquake-induced landslides in the island of Lefkada, Ionian Islands. Greece. Landslides **10**(6), 771–783 (2013)

33. Papathanassiou, G., Valkaniotis, S., Ganas, A., Grendas, N., Kollia, E.: The November 17th, 2015 Lefkada (Greece) strike-slip earthquake: field mapping of generated failures and assessment of macroseismic intensity ESI-07. Eng. Geol. **220**, 13–30 (2017)

34. Papathanassiou G., Valkaniotis S., Ganas, A.: Spatial patterns, controlling factors and characteristics of landslides triggered by strike-slip faulting earthquakes; case study of Lefkada island, Greece. Bull. Eng. Geol. Environ. (Submitted, 2020)

35. Parise, M., Jibson, R.W.: A seismic landslide susceptibility rating of geologic units based on analysis of characteristics of landslides triggered by the 17 January, 1994 Northridge, California earthquake. Eng. Geol. **58**(3–4), 251–270 (2000)

36. Parker, R.N., Densmore, A.L., Rosser, N.J., de Michele, M., Li, Y., Huang, R.Q., Whadcoat, S., Petley, D.N.: Mass wasting triggered by 2008 Wenchuan earthquake is greater than orogenic growth. Nat. Geosci. **4**(7), 449–452 (2011)

37. Rondoyanni-Tsiambaou, Th.: Les seismes et l'environnement géologique de l'île de Lefkade, Grèce: Passe et Futur. In: Marinos, P., et al. (eds.) Engineering Geology and the Environment, Balkema, pp. 1469–1474 (1997)

38. Rondoyianni, Th., Mettos, A., Paschos, P., Georgiou, Ch.: Neotectonic Map of Greece, scale1:100.000, Lefkada Sheet. I.G.M.E., Athens (2007)

39. Sachpazi, M., Hirn, A., Clement, C., Haslinger, F., Laigle Kissling, E., Charvis, P., Hello, Y., Lepine, J.C., Sapin, M., Ansorge, J.: Western Hellenic subduction and Cephalonia transform: local earthquakes and plate transport and strain. Tectonophysics **319**, 301–319 (2000)

40. Scordilis, E.M., Karakaisis, G.F., Karacostas, B.G., Panagiotopoulos, D.G., Comninakis, P. E., Papazachos, B.C.: Evidence from transform faulting in the Ionian Sea: the Cephalonia island earthquake sequence of 1983. PAGEOPH **123**, 387–397 (1985)

41. Tian, Y., Xu, C., Chen, J., Zhou, Q., Shen, L.: Geometrical characteristics of earthquake-induced landslides and correlations with control factors: a case study of the 2013 Minxian, Gansu, China, Mw 5.9 event. Landslides **14**, 1915–1927 (2017)