



# Hybrid Computational Intelligence Modeling of Coseismic Landslides' Severity

Anastasios Panagiotis Psathas<sup>1</sup> , Antonios Papaleonidas<sup>1</sup> ,  
George Papathanassiou<sup>1</sup>, Lazaros Iliadis<sup>1</sup> , and Sotirios Valkaniotis<sup>2</sup>

<sup>1</sup> Department of Civil Engineering, Democritus University of Thrace, 67100 Xanthi, Greece  
{anpsatha, papaleon, gpapatha, liliadis}@civil.duth.gr

<sup>2</sup> Koronidos str 9, 42131 Trikala, Greece

**Abstract.** Coseismic Landslides (COLA) are one of the most widespread and destructive hazards to result from earthquakes in mountainous environments. They are always associated with almost instantaneous slope collapse and spreading, posing significant hazards to human lives and lifeline facilities worldwide. Current methods to identify COLA immediately after an earthquake using optical imagery are too slow to effectively inform emergency response activities. Their realistic prediction is crucial for the design of key infrastructure and to protect human lives in seismically active regions. Forecasting their severity could be extremely beneficial for the effective treatment of disastrous consequences. The goal of this research is to propose a hybrid model that takes into consideration only three of the most affordable factors to acquire, the *Slope* of the active areas, the *Aspect* and the *Geological Form*. Determination of their correlation could predict the severity of COLA phenomena. The dataset used in this research, comprises of 421 records for year 2003 and 767 for 2015 from the Greek island of Lefkada. The introduced hybrid model employs *Fuzzy c-Means*, *Ensemble Adaptive Boosting* and *Ensemble Subspace k-Nearest Neighbor* algorithms. The model managed to successfully classify the *Coseismic Landslides* according to their severity. The performance was high especially for the classes of major severity.

**Keywords:** Forecasting landslides · Fuzzy c-Means · T-norms · k-Nearest-Neighbors · Ensemble Subspace k-NN · Adaptive Boosting · Ensemble AdaBoost

## 1 Introduction

Landslides occur in a variety of environments, characterized by either steep or gentle slope gradients, from mountain ranges to coastal cliffs or even underwater. Gravity is the primary driving force for a landslide to occur, but there are other factors affecting slope stability that produce specific conditions, which make a slope prone to failure. In many cases, the landslide is triggered by a specific event, such as a heavy rainfall, an earthquake, a slope cut to build a road and many others. Earthquake-induced landslides are one of the most catastrophic effects of earthquakes, as evidenced by many historic events over the past decades, especially in countries with high seismicity [1]. As a few examples,

the 1994 Northridge earthquake triggered more than 11,000 landslides. In the 2008 Wenchuan earthquake in China, the Tangjiashan landslide with over 20.37 million m<sup>3</sup> mass movement, blocked the main river channel and formed a landslide dam, putting millions of people downstream at risk. According to Jibson et al. [18], there are cases that the consequences of landslides, triggered by an earthquake, have a massive impact in human lives and facilities. The correlation of the pattern of COLA with geological and topographical variables i.e. *lithology*, *slope angle* and *slope aspect* with the *volume of landslides* has been investigated by several researchers. The density of the mapped landslide concentration is normally associated with the seismic shaking 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].

The pinpoint of the areas that are most vulnerable in Coseismic Landslides is vital in order to take actions in time and reduce the risk in those areas. Realistic prediction of COLA is crucial for the design of key infrastructure and to protect human lives in seismically active regions. Among many existing methods for landslide assessment, the Newmark sliding mass model has been extensively utilized to estimate earthquake-induced displacements in slopes, earth dams and landfills since the 1960s. As technology develops, new methods and techniques were proposed for assessing the degree of danger within an area. Such, instruments are satellite imagery and Geographic Information System technology (GIS), especially using statistical analyses of geo-environmental and seismologic factors into GIS software [29]. In particular, the characteristics of the land sliding area is statistically related to control factors such as topographic, geologic and seismic parameters e.g. *slope angle*, *slope aspect*, *curvature*, *lithology*, *PGA*, *seismic intensity distribution* and *distance from the seismic fault or epicenter* [27]. 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]. Coupling effect between topography and soil amplification, leads to complex wave propagation patterns due to scattering and diffracting of waves within the low velocity near surface layers. These ground motion effects have significant impact on COLA assessment, but only limited efforts have considered them in empirical models. There is a clear need to develop innovative numerical schemes to address the above challenges.

In existing literature, there is a lack of models that can predict the severity of Coseismic Landslides using only the slope angle, slope aspect and the geological form of a specific area. This work represents an extended version of the previous research of the authors [31]. It is very common, to use more features for the forecasting, which derive from the use of expensive equipment. Thus, it is essential for the deployment of a model, that it is not based on large financial funds and be equally effective. Being able to predict the severity of an upcoming landslide after an earthquake, it could be extremely beneficial for the effective treatment of disastrous consequences. The development of such a model, could assist risk management organizations, public agencies and stakeholders, or even governments, to apply a better distribution of the staff and financial resources to each area, for the confrontation of potential corollaries, or even develop appropriate mitigation plans, increasing the resilience of the community.

The statistical analysis of geo-environmental and seismologic factors is performed by bivariate and multivariate approaches. The purpose of this study is the recommendation of a hybrid algorithm, which could find the association of three main factors of COLA, (*slope, aspect and geological form*), with their severity. The proposed hybrid model uses *Fuzzy c-Means* clustering [3, 12], *Ensemble Adaptive Boosting* (ENAB) and *Ensemble Subspace k-Nearest Neighbor* (ES\_k-NN) classifiers [10, 15]. The existing literature, like [23, 32] does not exploit the combination of the above algorithms. Current methods identify COLA after an earthquake, consider optical imagery. They are too slow to effectively inform emergency response activities. There is a need for a fast and flexible model to consider more affordable factors. All current approaches are using crisp values for the determination of involved features. This could lead to misclassification of COLA for values close to the borderline.

## 2 Area of Research

Lefkada, is a Greek island in the Ionian Sea on the west coast of Greece, connected to the mainland by a long causeway and floating bridge. Lefkada measures 35 km from north to south, and 15 km from east to west. The area of the island is about 302 km<sup>2</sup>, the area of the municipality (including the islands Kalamos, Kastos and several smaller islets) is 333.58 km<sup>2</sup>. The basic geological forms for of the island are: a) A carbonate sequence of the Ionian zone. b) Limestone of Paxos (Apulia) zone restricted in the SW peninsula of the island. c) Few outcrops of ionian flysch (turbidites) and Miocene marls-sandstones mainly in the northern part of the island [9]. The geological Zones of “Ioanian” and “Paxos” are separated by a bountary which is located in the NW-SE direction of the region and projects onshore Southcentral Lefkada, near “Hortata” Village, in the form of a burried thrust fault by scree and late Quaternary deposits [28]. Pleistocene and especially Holocene coastal deposits are sprawling in the Nothern edge of Lefkada, where its capital is located, in the valley of “Vassiliki” and in the Coast “Nydri”. Due to its location in the Ioanian sea and to its complex crustal deformation resulting from the subduction of the African Plate towards NE and to the Apulian platform continental collision further to the Northwest, Lefkada is one of the most tectonically active areas in the European continent [14, 16]. The principal active tectonic structure is 140 km long dextral strike-slip Cephalonia-Lefkada transform fault CTF [24]. It has a GPS slip-rate bracketed between 10 and 25 mm/yr. Most of the slope failure cases have been reported on the western part of the island, which owes its deep morphology to this offshore CTF and its onshore sub-parallel fault; the “Athani-Dragano” fault [9]. The latter is a NNE-SSW striking fault, forming a narrow elongated continental basin, precicely depicted in the region’s morphology and indicated on satellite images and aerial photos.

There is thorough and detailed record about at least 23 events, with crucial impact on the ground of Lefkada [26]. A first conclusion drawn by the events is that earthquakes occur in pairs (twin ore 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). Therefore, it is of great importance to pinpoint the location of Coseismic Landslides since it will be useful in order to reduce the hazards and increase the resilience at the island.

## 2.1 Coseismic Landslides at the Island of Lefkada

The most recent and well examined earthquakes are those of 2003 and 2015. The penultimate earthquake caused massive slope failures at the western part of the island. The amount of the debris material that arose was remarkably larger than the one of 2015. Numerous landslides occurred 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. Among the most indicative rock falls with diameters up to 4m, 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 frequent occurrence of this failures led to the closure of the road network which lasted for more than two 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 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]. 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. 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. The short period between each satellite imagery pair (2–3 years) is believed to include only the COLA, 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. 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<sup>2</sup> and 1.6 km<sup>2</sup> for the 2015 and 2003 events. These planar-oriented areas are obtained as projected measurements. The minimum and maximum landslide areas were 40.4 m<sup>2</sup> and 42,940 m<sup>2</sup> for the 2015 earthquake, while for the penultimate event the relevant values were 129.8 m<sup>2</sup> and 98,300 m<sup>2</sup> [29]. The minimum and maximum

landslide areas for 2015 in an area of 1.78 km<sup>2</sup>, were 51.30 m<sup>2</sup> and 58.330 m<sup>2</sup>. They were found by considering the DEM for the delineation of the landslide area. The values for the 2003 earthquake were 140.9 m<sup>2</sup> and 148.469 m<sup>2</sup> in an area of 2.28 km<sup>2</sup> [29].

### 3 Dataset Pre-processing

Five features were related to COLA, 4 of which were numeric (*Planar Area, Average Slope, Area, Average Aspect, Id*) and 1 was nominal (*Geologic Form*). Given the fact that there are several landslides with 2 or more geological forms, the Coseismic Landslides were reassigned. This resulted in 421 instances for 2003 and 767 for 2015. For both years, the features are the same. Thus, the same data preprocessing was applied for both datasets. Nonetheless, in the 2003 COLA, two additional geological forms were observed, compared to ones of 2015. The data preprocessing for these two years was done independently. This resulted in an overall evaluation of the proposed algorithm, that proves its consistency and efficiency regardless the year. For data handling, 3 steps were followed. The 1<sup>st</sup> was related to the manual processing of the *Average Slope, Average Aspect* and *Geological Forms*. The 2<sup>nd</sup> and 3<sup>rd</sup> steps were performed by developing novel code in *Matlab R2019a*.

It was observed during the experiments that there are some variations in landslides that have the same severity but appear to have quite different slopes. For this reason, the natural logarithm function  $\ln(x)$  was applied on the values of the slopes in order to smooth out any spikes. Regarding the *average aspect*, the initial elaboration was to transform it from nominal to numeric in a scale from 1 to 8, according to Fig. 1.

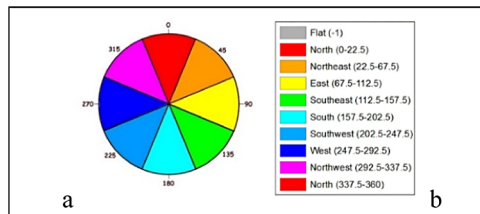


Fig. 1. Aspect and corresponding degrees

Nonetheless, it was considered more efficient to separate the landslides that have aspect 10° from the ones that have 350° (Fig. 1b). For this purpose, for each landslide the actual aspect in degrees was used.

### 3.1 Labeling Geological Forms

The Geological form feature was transformed to numeric, in a scale from 1 to 20 for 2003, from 1 to 18 for 2015. Table 1 presents the numeric labels for each Form.

**Table 1.** Geological form types for 2003 and 2015

Geological Form	al	C	Ci	Cs	Csd	E	J1	J1d	Jar	Jc
Form Label 2003	1	2	3	4	5	6	7	8	9	10
Geological Form	Jm	Js	M	Mb	Pc	Qc	Qp	Qt	Tc	Tg
Form Label 2003	11	12	13	14	15	16	17	18	19	20
Geological Form 2015	al	C	Ci	Cs	E	J1	J1d	Jar	Jc	
Form Label 2015	1	2	3	4	5	6	7	8	9	
Geological Form 2015	Jm	Js	M	Mb	Pc	Qc	Qp	Qt	Tg	
Form Label 2015	10	11	12	13	14	15	16	17	18	

*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 Fuzzy C-Means Clustering of Landslides

After applying a statistical analysis of the datasets, it was pinpointed that some values of the *area* and the *planar area*, had high standard deviation and a non-representative mean value. FCM is the fuzzy equivalent of the “hard” clustering algorithm. It has been employed, due to the fact that it allows an individual to be partially classified into more than one cluster, with different degrees of membership. It was performed on the features *Area* and *Planar Area*, to provide the labels required for the development of the hybrid machine learning model [3, 4, 12]. The process of fuzzy partitioning is performed through a repetitious optimization of the objective function 1 with the update of membership  $u_{ij}$  and the cluster centers  $c_j$  (function 2):

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

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

where  $m$  is the fuzzifier (determining the level of cluster fuzziness),  $u_{ij}$  is the degree of membership of  $x_i$  in the cluster  $j$ ,  $x_i$  is the  $i^{\text{th}}$  of  $d$ -dimensional measured data,  $c_j$  is the  $d$ -dimension center of the cluster,  $\|*\|$  is any norm expressing the similarity between any measured data and the center. The final condition for stopping the iterations is when  $\max_{ij} \left\{ \left| u_{ij}^{(k+1)} - u_{ij}^k \right| \right\} < \varepsilon$ , where  $\varepsilon$  is a termination criterion between 0 and 1, and  $k$  indicates the iteration steps [21, 22].

This paragraph presents the hyperparameters and their values. The corresponding membership degrees  $u_{ij}$  for large  $m$  values are small, which implies clusters with smaller bounties. The *maxIterations* (assigned the default value 100) is the maximum number of optimization iterations and *minImprovement* (assigned the default value 0.00001) is the minimum improvement in the Objective function (OBJ) per iteration. The value of  $m$  was chosen (after trial and error) to be equal to 2.

The developed *FCM.m* script in Matlab, transfers the content of the data files into Matlab Tables for further processing and applies the FCM. The *FCM.m* script creates the clusters of the COLA according to their severity and assigns the labels. The number of clusters was chosen to be 6. The *Linguistic* of the respective clusters is a combination of the 3 potential states (*Low Medium, High*) of the *Planar area* and *Area* (Tables 2 and 3).

**Table 2.** Clusters with their corresponding labels (*planar area, area*)

Cluster#	1	2	3	4	5	6
Linguistic	Low, Low	Low, Medium	Medium, Medium	Medium, High	High, High	Extreme, Extreme

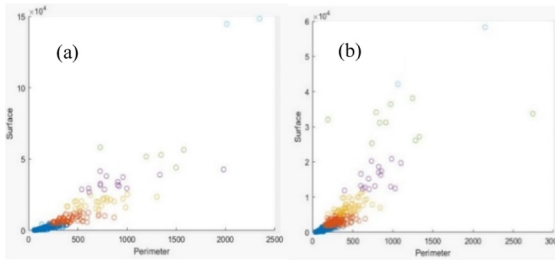
**Table 3.** Respective landslides for each cluster for 2003 and 2015

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

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

**Algorithm 1.** The *FCM.m* Matlab script

<b>Inputs:</b> 421 incidents of 2003 and 767 incidents of 2015 exported from *.xlsx files.
<b>Step 1:</b> Read and convert each *.xlsx file to a Matlab table.
<b>Step 2:</b> For each Table the column <i>Id</i> was used to retrieve distinct Landslides. 2 new tables constructed
<b>Step 3:</b> For each new table <i>Planar Area</i> and <i>Area</i> column is chosen
<b>Step 4:</b> FCM algorithm is applied. Centers of clusters and MEV's of each observation are calculated.
<b>Step 5:</b> Each instance is classified in the cluster on which the MEV is the highest.
<b>Step 6:</b> Clusters pass to the original data. Clusters are plotted in Figs 2a and 2b



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

### 3.3 Fuzzy Clustering with FCM Algorithm and T-Norm

Some instances had similar membership values (MEV) for two clusters, thus the need to create new clusters has proved imperative. Landslides with MEV (to their dominant class) below a certain threshold, were re-sorted. Thus, each factor was assigned a weight [5, 33] based on Eq. (3) that implements a fuzzy conjunction  $f(\mu_i, w_i)$  between many fuzzy sets. Each MEV  $\mu_i$  is assigned a weight  $w_i$ .

$$\mu_S(x_i) = \text{Agg}(f(\mu_A(x_i), w_1), f(\mu_A(x_i), w_2), \dots, f(\mu_A(x_i), w_n)) \quad (3)$$

$i = 1, 2, \dots, k$ . Function  $f$  is defined as follows:  $f(a, w) = a^{\frac{1}{w}}$ ,  $a$  is the MEV [13, 33]. The *Hamacher T-Norm* was used as an Aggregation function.

$$A \cap B = \frac{\mu_A(x) + \mu_B(x) - 2\mu_A(x)\mu_B(x)}{[1 - \mu_A(x) + \mu_B(x)]} \quad (4)$$

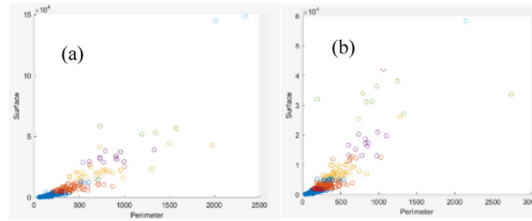
The script *T-Norm\_Clustering.m* was deployed in Matlab, to apply the FCM with the *Hamacher* fuzzy conjunction. It is presented in natural language form in Algorithm 2.

**Algorithm 2.** The *T-Norm\_Clustering.m* Matlab Script

<b>Inputs:</b> All membership degrees of each observation for 2003 and 2015.
<b>Part 1: Step 1:</b> Threshold is defined at 0.7
<b>Step 2:</b> Weights definition. The weight of the highest MEV is 4, for the 2 <sup>nd</sup> highest is 2, and 0 for the rest
<b>Part 2: Step 1:</b> If a MEV to a cluster is higher than 0.7 the incident belongs to this cluster.
<b>Step 2:</b> If a MEV is under 0.7 the incident belongs to a new cluster for which the MEV is calculated by Eq. (3), (4) and (5).

The *T-Norm\_Clustering.m* script developed new clusters between the already existing ones, for both 2003 and 2015. Four new clusters were created for 2003 and 5 for 2015. If a cluster was created between clusters 4 and 5, then its name would be 4.5. The result of clustering with FCM and *Hamacher* is presented in Fig. 3a and b. The T-Norm optimized the classification of each instance according to its severity. Table 4, presents the exact number of instances that correspond to each cluster.





**Fig. 3.** Clusters for original data for 2003 (3a) and 2015 (3b) after applying T-Norm

**Table 4.** Total landslides of each cluster for 2003 and 2015 after T-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

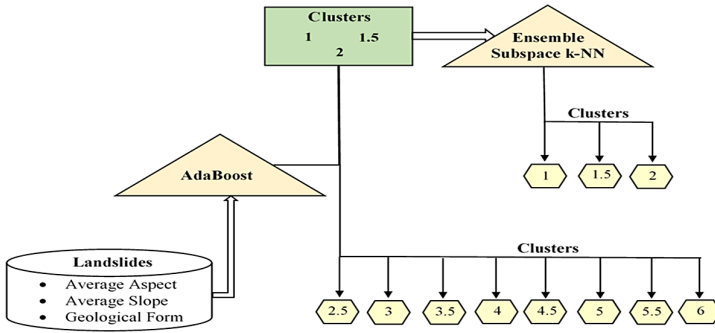
## 4 Classification Methodology

Following the clustering of *Coseismic Landslides* using the features “*planar area*” and “*area*”, a classification based on 3 factors was performed. The three independent variables used for the classification are *Average Slope*, *Average Aspect* and *Geological Form* of landslides, which was labeled as indicated in Table 1. The cluster obtained by the FCM with *Hammacher Aggregation*, is the target variable.

A total of 25 classification algorithms were employed: *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*, *Coarse KNN*, *Cubic KNN*, *Weighted KNN*, *Fine KNN*, *Medium KNN*, *Gaussian Naive Bayes*, *Kernel Naive Bayes*, *Boosted Trees*, *Bagged Trees*, *Subspace Discriminant*, *Subspace KNN*, *RUSBoost Trees*, *Ensemble Adaptive Boosting*. The one with the highest performance was the *Ensemble Adaptive Boosting Algorithm (AdaBoost)*. AdaBoost, has proved to be very efficient for all classes except the first three (clusters 1, 1.5, and 2). Thus, if an observation was classified in the first 3 clusters, another algorithm was applied in order to classify it more accurately. The best algorithm for this approach, was the *Ensemble Subspace k-NN* (Fig. 4).

### 4.1 Ensemble AdaBoost

It makes predictions based on a number of different models. By combining individual models, it tends to be less biased and less data sensitive. *Ensemble AdaBoost* works especially well with decision trees. It is the most popular boosting technique, developed for classification. It learns from previous mistakes, e.g. misclassification data points, by increasing their weights. The learner with higher weight has more influence on the final decision.



**Fig. 4.** The hybrid classification model

**4.2 Ensemble Subspace k- Nearest-Neighbors (Ensemble Subspace k-NN)**

The k-nearest neighbors (k-NN) is a lazy and non-parametric Learning algorithm [8].

It is a traditional classification rule, which assigns the label of a test sample with the majority label of its k nearest neighbors, from the training set. Given a set X of n points and a distance function, k-NN search finds the k closest points to a query point or a set of them [17]. Dunami [11] first introduced a weighted voting method, called the *distance-weighted (DW) k-nearest neighbor rule (Wk-NN)*. In this approach, the closer one neighbor is, the greater the weight that corresponds to it, using the DW function.

The neighbor farthest from all the others corresponds to a weight of 0, while the one closest to the observation corresponds to a weight of 1. All neighbors in the middle area get corresponding values between 0 and 1. The most common and most consistent ensemble method for k-NN is the *Ensemble Subspace k-NN* and related works using this algorithm can be found in [10, 15, 17]. Tuning of the hyperparameters was done based on a combination of *10-fold Cross Validation* and *Grid Search*. According to the literature, this combination is one of the most widely strategies used in machine learning.

This was a multiclass classification case, so the “*One Versus All*” Strategy [19, 30] was used for the evaluation. Table 5 shows the 5 performance indices that were used.

**Table 5.** Indices used for the evaluation of the multi-class classification

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	$F1 = 2 * TP / (2 * TP + FP + FN)$

## 5 Experimental Results

The experiments were performed in Matlab R2019a. The initial range of the AdaBoost hyperparameters is the following: *Maximum number of splits* (MNS) takes values in the interval [10, 500], whereas the range interval of the parameter *Number of learners* (NLE) is [1,800] and the respective one for the *Learning Rate* (LR) is [0.001, 1]. The optimal hyperparameters values found were 175, 88, 1, 10, for the MNS, NLE and LR and the *number of Grid Divisions* respectively. The *Learner Type* was a “*Decision Tree*” and the *Optimizer* employed was *Grid Search*. The *Ensemble AdaBoost* achieved an accuracy of 64% and 68% for 2003 and 2015. The following Tables 6 and 7 are the confusion matrices of the above optimal algorithm for 2015 and 2003, whereas Fig. 5a and b are the ROC curves for the respective years. Tables 8 and 9 present the values of all the performance indices for AdaBoost.

It is obvious that AdaBoost successfully classifies all the landslides that belong to Cl<sub>2.5</sub> and above. This is significant, as the algorithm can indicate with high accuracy, COLA that would be the most disastrous. Prediction of the first 3 clusters, requires more attention.

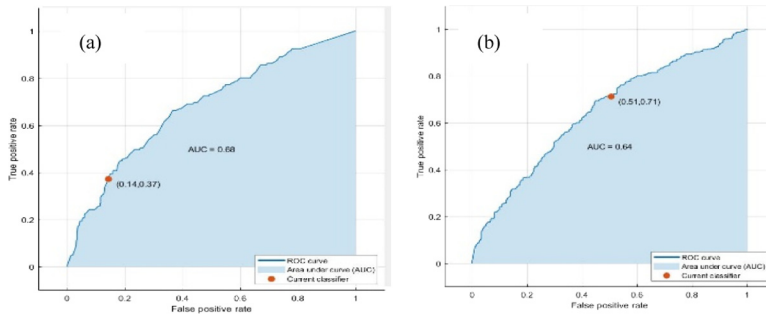
**Table 6.** Confusion Matrix for the Ensemble AdaBoost for 2015

		Predicted Class											
		Class	Cl <sub>1</sub>	Cl <sub>1.5</sub>	Cl <sub>2</sub>	Cl <sub>2.5</sub>	Cl <sub>3</sub>	Cl <sub>3.5</sub>	Cl <sub>4</sub>	Cl <sub>4.5</sub>	Cl <sub>5</sub>	Cl <sub>5.5</sub>	Cl <sub>6</sub>
True Class	Cl <sub>1</sub>	300	42	75	0	0	0	0	0	0	0	0	0
	Cl <sub>1.5</sub>	21	20	10	0	0	1	0	0	0	0	0	0
	Cl <sub>2</sub>	51	9	72	2	9	0	0	0	2	0	0	0
	Cl <sub>2.5</sub>	1	1	6	18	1	0	0	0	0	0	0	0
	Cl <sub>3</sub>	1	0	8	0	45	1	1	0	0	0	0	0
	Cl <sub>3.5</sub>	1	0	1	0	0	22	0	0	0	0	0	0
	Cl <sub>4</sub>	0	0	0	0	0	0	18	0	0	0	0	0
	Cl <sub>4.5</sub>	0	0	0	0	0	0	0	7	0	0	0	0
	Cl <sub>5</sub>	0	0	0	0	0	0	0	0	15	0	0	0
	Cl <sub>5.5</sub>	0	0	0	0	0	0	0	0	0	2	0	0
	Cl <sub>6</sub>	0	0	0	0	0	0	0	0	0	0	0	3

The *Ensemble k-NN* was employed. Tuning of hyperparameters was applied with the combination of *10-fold cross validation* and *grid search*. The initial range for the hyperparameters of the Ensemble k-NN are: *Maximum number of splits* takes values in [10, 500], the parameter *Number of learners* in [1,800] the *Learning Rate* in [0.001, 1] and *Subspace Dimension* takes values in [2, 10]. The optimal hyperparameters values found were 20, 30, 0.1, 3, 10, for the MNS, NLE, LR, *Subspace Dimension* and the *number of Grid Divisions* respectively. The *Distance Metric* was a “*Euclidean*” and the *Optimizer* employed was *Grid Search*. The *Ensemble Subspace k-NN* achieved an accuracy equal to 70.07% and 72.88% for 2003 and 2015. The Confusion Matrix for each year is presented in Tables 10 and 11 and the corresponding ROC Curves in Fig. 6a and b.

**Table 7.** Confusion matrix for the Ensemble AdaBoost for 2003

		Predicted Class										
		Class	Cl <sub>1</sub>	Cl <sub>1,5</sub>	Cl <sub>2</sub>	Cl <sub>2,5</sub>	Cl <sub>3</sub>	Cl <sub>3,5</sub>	Cl <sub>4</sub>	Cl <sub>4,5</sub>	Cl <sub>5</sub>	Cl <sub>6</sub>
True Class	Cl <sub>1</sub>	151	40	27	4	1	0	0	0	0	0	0
	Cl <sub>1,5</sub>	10	7	7	1	1	1	0	0	0	0	0
	Cl <sub>2</sub>	32	11	28	7	2	3	0	0	0	0	0
	Cl <sub>2,5</sub>	0	1	1	7	0	0	0	0	0	0	0
	Cl <sub>3</sub>	0	1	0	0	0	32	0	2	1	1	0
	Cl <sub>3,5</sub>	0	0	0	0	0	0	5	0	0	0	0
	Cl <sub>4</sub>	0	0	0	0	0	0	0	22	0	0	0
	Cl <sub>4,5</sub>	0	0	0	0	0	0	0	0	4	0	0
	Cl <sub>5</sub>	0	0	0	0	0	0	0	0	0	6	0
	Cl <sub>6</sub>	0	0	0	0	0	0	0	0	0	0	5



**Fig. 5.** ROC Curves for 2015 (5a) and 2003 (5b)

**Table 8.** Classification performance indices for the Ensemble AdaBoost (2015)

Index	Cl <sub>1</sub>	Cl <sub>1,5</sub>	Cl <sub>2</sub>	Cl <sub>2,5</sub>	Cl <sub>3</sub>	Cl <sub>3,5</sub>	Cl <sub>4</sub>	Cl <sub>4,5</sub>	Cl <sub>5</sub>	Cl <sub>6</sub>	Cl <sub>6</sub>
SNS	0.71	0.38	0.49	0.66	0.80	0.91	1	1	1	1	1
SPC	0.74	0.90	0.81	0.99	0.97	0.99	0.99	1	0.99	1	1
ACC	0.73	0.86	0.75	0.97	0.96	0.99	0.99	1	0.99	1	1
PREC	0.80	0.27	0.41	0.90	0.81	0.91	0.94	1	0.88	1	1
F1	0.75	0.32	0.45	0.76	0.81	0.91	0.97	1	0.93	1	1

**Table 9.** Classification performance indices for the Ensemble AdaBoost (2003)

Index	Cl <sub>1</sub>	Cl <sub>1,5</sub>	Cl <sub>2</sub>	Cl <sub>2,5</sub>	Cl <sub>3</sub>	Cl <sub>3,5</sub>	Cl <sub>4</sub>	Cl <sub>4,5</sub>	Cl <sub>5</sub>	Cl <sub>6</sub>
SNS	0.67	0.25	0.33	0.77	0.86	1	1	1	1	1
SPC	0.73	0.83	0.87	0.95	0.98	0.98	0.99	0.99	0.99	1
ACC	0.70	0.78	0.74	0.95	0.96	0.98	0.99	0.99	0.99	1
PREC	0.789	0.11	0.44	0.36	0.88	0.55	0.91	0.80	0.85	1
F1	0.72	0.16	0.38	0.50	0.87	0.71	0.95	0.88	0.92	1

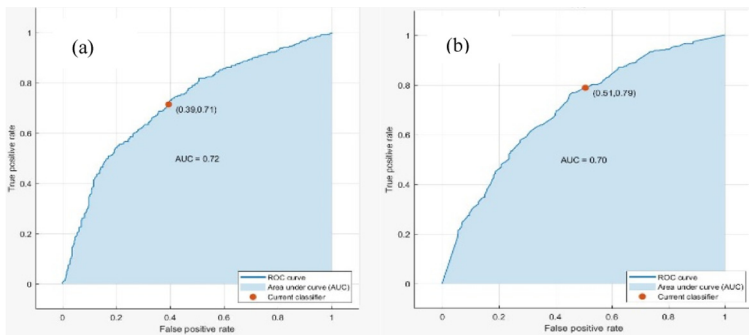
**Table 10.** Confusion matrix for ensemble subspace k-NN algorithm for 2015

		Predicted Class											
		Class	Cl <sub>1</sub>	Cl <sub>1.5</sub>	Cl <sub>2</sub>	Cl <sub>2.5</sub>	Cl <sub>3</sub>	Cl <sub>3.5</sub>	Cl <sub>4</sub>	Cl <sub>4.5</sub>	Cl <sub>5</sub>	Cl <sub>5.5</sub>	Cl <sub>6</sub>
True Class	Cl <sub>1</sub>	326	26	65	0	0	0	0	0	0	0	0	0
	Cl <sub>1.5</sub>	21	21	9	0	0	1	0	0	0	0	0	0
	Cl <sub>2</sub>	46	4	82	2	9	0	0	0	2	0	0	0
	Cl <sub>2.5</sub>	1	1	6	18	1	0	0	0	0	0	0	0
	Cl <sub>3</sub>	1	0	8	0	45	1	1	0	0	0	0	0
	Cl <sub>3.5</sub>	1	0	1	0	0	22	0	0	0	0	0	0
	Cl <sub>4</sub>	0	0	0	0	0	0	18	0	0	0	0	0
	Cl <sub>4.5</sub>	0	0	0	0	0	0	0	7	0	0	0	0
	Cl <sub>5</sub>	0	0	0	0	0	0	0	0	15	0	0	0
	Cl <sub>5.5</sub>	0	0	0	0	0	0	0	0	0	2	0	0
	Cl <sub>6</sub>	0	0	0	0	0	0	0	0	0	0	0	3

**Table 11.** Confusion matrix for ensemble subspace k-NN algorithm for 2003

		Predicted Class										
		Class	Cl <sub>1</sub>	Cl <sub>1.5</sub>	Cl <sub>2</sub>	Cl <sub>2.5</sub>	Cl <sub>3</sub>	Cl <sub>3.5</sub>	Cl <sub>4</sub>	Cl <sub>4.5</sub>	Cl <sub>5</sub>	Cl <sub>6</sub>
True Class	Cl <sub>1</sub>	161	36	21	4	1	0	0	0	0	0	0
	Cl <sub>1.5</sub>	7	12	5	1	1	1	0	0	0	0	0
	Cl <sub>2</sub>	28	6	38	7	2	3	0	0	0	0	0
	Cl <sub>2.5</sub>	0	1	1	7	0	0	0	0	0	0	0
	Cl <sub>3</sub>	0	1	0	0	32	0	2	1	1	0	0
	Cl <sub>3.5</sub>	0	0	0	0	0	5	0	0	0	0	0
	Cl <sub>4</sub>	0	0	0	0	0	0	22	0	0	0	0
	Cl <sub>4.5</sub>	0	0	0	0	0	0	0	4	0	0	0
	Cl <sub>5</sub>	0	0	0	0	0	0	0	0	6	0	0
	Cl <sub>6</sub>	0	0	0	0	0	0	0	0	0	0	5

Tables 12 and 13, present the values of the performance indices for the above optimal algorithm. After applying and the Ensemble k-NN algorithm, a significant increase of overall accuracy and indexes was observed. The accuracy was 70% and 72% for 2003 and 2015 respectively.



**Fig. 6.** ROC curve for ensemble subspace k-NN for 2015 (6a) and for 2003 (6b)

**Table 12.** Classification performance indices of ensemble subspace k-NN (2015)

Index	Cl <sub>1</sub>	Cl <sub>1.5</sub>	Cl <sub>2</sub>	Cl <sub>2.5</sub>	Cl <sub>3</sub>	Cl <sub>3.5</sub>	Cl <sub>4</sub>	Cl <sub>4.5</sub>	Cl <sub>5</sub>	Cl <sub>6</sub>	Cl <sub>6</sub>
SNS	0.78	0.40	0.56	0.66	0.80	0.91	1	1	1	1	1
SPC	0.76	0.94	0.84	0.99	0.98	0.99	0.99	1	0.99	1	1
ACC	0.77	0.90	0.78	0.98	0.96	0.99	0.99	1	0.99	1	1
PREC	0.82	0.40	0.47	0.90	0.81	0.91	0.94	1	0.88	1	1
F1	0.80	0.40	0.51	0.76	0.81	0.91	0.97	1	0.93	1	1

**Table 13.** Classification performance indices of Ensemble Subspace k-NN (2003)

Index	Cl <sub>1</sub>	Cl <sub>1.5</sub>	Cl <sub>2</sub>	Cl <sub>2.5</sub>	Cl <sub>3</sub>	Cl <sub>3.5</sub>	Cl <sub>4</sub>	Cl <sub>4.5</sub>	Cl <sub>5</sub>	Cl <sub>6</sub>
SNS	0.724	0.44	0.45	0.77	0.86	1	1	1	1	1
SPC	0.78	0.86	0.90	0.95	0.98	0.98	0.99	0.99	0.99	1
ACC	0.75	0.83	0.80	0.95	0.97	0.98	0.99	0.99	0.99	1
PREC	0.821	0.21	0.58	0.36	0.88	0.55	0.91	0.80	0.85	1
F1	0.763	0.28	0.51	0.50	0.87	0.71	0.95	0.88	0.92	1

## 6 Discussion and Conclusion

At first glance, the efficiency of the model is high but not optimal. We must consider that this modeling effort has achieved to effectively classify the severity of the complicated *Coseismic Landslides* phenomenon, using only 3 independent variables. From this point of view, the performance is reliable, and it has a certain level of novelty, as to the best of our knowledge, there does not exist another approach with similar accuracy in the literature. It is a pioneer research, employing state of the art Hybrid Machine Learning algorithms in Geomechanics. Indices, especially the accuracy and F1 score, indicate a very flexible model that can predict the most severe landslides and can handle the landslides that are not so dangerous. The results of 2003 and 2015 are similar, which means that the algorithm is consistent and that it can generalize (not case dependent). The performance of the algorithm for 2015 is better, because of the more contemporary equipment used for the extraction of features. This research addresses one of the most crucial natural hazards. It is of essential importance to urban planning and to the functioning of societies. Future research will focus on predicting the timeframe and the area of a landslide after an earthquake.

## References

1. 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)
2. 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)

3. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer Science & Business Media (2013)
4. Bora, D.J., Gupta, D., Kumar, A.: A comparative study between fuzzy clustering algorithm and hard clustering algorithm. arXiv preprint [arXiv:1404.6059](https://arxiv.org/abs/1404.6059) (2014)
5. Calvo, T., Mayor, G., Mesiar, R. (eds.): *Aggregation Operators: New Trends and Applications*, vol. 97. Physica (2012)
6. 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)
7. 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)
8. Cover, T., Hart, P.: Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13**(1), 21–27 (1967)
9. Cushing, M.: *Evolution structurale de la marge nord-ouest hellenique dans l'île de Lefkada et ses environs (Greece nord-occidentale)*. Ph.D. Thesis. Univ. de Paris-Sud (XI). Centre d'Orsay, France (1985)
10. 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)
11. Dudani, S.A.: The distance-weighted k-nearest neighbor rule. *IEEE Trans. Syst. Man Cybern.* **8**(4), 311–313 (1978)
12. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybern.* **3**(3), 32–57 (1973). <https://doi.org/10.1080/01969727308546046>
13. 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)
14. Ath, G., et al.: 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)
15. 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)
16. Hatzfeld, D., et al.: Microseismicity and strain pattern in northwestern Greece. *Tectonics* **14**(4), 773–785 (1995)
17. Hechenbichler, K., Schliep, K.: *Weighted k-nearest-neighbor techniques and ordinal classification* (2004)
18. 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)
19. Joutsijoki, H., Juhola, M.: Comparing the one-vs-one and one-vs-all methods in benthic macroinvertebrate image classification. In: Perner, P. (ed.) *Machine Learning and Data Mining in Pattern Recognition*, pp. 399–413. Springer Berlin Heidelberg, Berlin, Heidelberg (2011). [https://doi.org/10.1007/978-3-642-23199-5\\_30](https://doi.org/10.1007/978-3-642-23199-5_30)
20. Keefer, D.K.: Landslides caused by earthquakes. *Geol. Soc. Am. Bull.* **95**(4), 406–421 (1984)
21. Khalilia, M.A., Bezdek, J., Popescu, M., Keller, J.M.: Improvements to the relational fuzzy c-means clustering algorithm. *Pattern Recogn.* **47**(12), 3920–3930 (2014)
22. Khalilia, M., Popescu, M.: Fuzzy relational self-organizing maps. In: *2012 IEEE International Conference on Fuzzy Systems*, pp. 1–6. IEEE (June 2012)
23. 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)
24. Louvari, E., Kiratzi, A.A., Papazachos, B.C.: The CTF and its extension to western Lefkada Island. *Tectonophysics* **308**, 223–236 (1999)

25. 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)
26. 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)
27. 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 (2012). <https://doi.org/10.1007/s10346-012-0357-1>
28. Papathanassiou, G., Valkaniotis, S., Ath, G., Grendas, N., El, K.: 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)
29. 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)
30. Psathas, A., Papaleonidas, A., Iliadis, L.: Machine Learning Modeling of Human Activity Using PPG Signals. In: Nguyen, N.T., Hoang, B.H., Huynh, C.P., Hwang, D., Trawiński, B., Vossen, G. (eds.) *Computational Collective Intelligence: 12th International Conference, ICCCI 2020, Da Nang, Vietnam, November 30 – December 3, 2020, Proceedings*, pp. 543–557. Springer International Publishing, Cham (2020). [https://doi.org/10.1007/978-3-030-63007-2\\_42](https://doi.org/10.1007/978-3-030-63007-2_42)
31. Psathas, A., Papaleonidas, A., Papathanassiou, G., Valkaniotis, S., Iliadis, L.: Classification of Coseismic Landslides Using Fuzzy and Machine Learning Techniques. In: Iliadis, L., Angelov, P.P., Jayne, C., Pimenidis, E. (eds.) *Proceedings of the 21st EANN (Engineering Applications of Neural Networks) 2020 Conference: Proceedings of the EANN 2020*, pp. 15–31. Springer International Publishing, Cham (2020). [https://doi.org/10.1007/978-3-030-48791-1\\_2](https://doi.org/10.1007/978-3-030-48791-1_2)
32. Robinson, T.R., et al.: 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)
33. 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)