ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

ACCEPTED ON ANNALS OF GEOPHYSICS, 61, 2018; Doi: 10.4401/ag-8049

Classification of Mount Etna (Italy) Volcanic Activity by Machine Learning Approaches

Alireza Hajian[1], Flavio Cannavo'[2], Filippo Greco[2], Giuseppe Nunnari[3]

(1) Department of Physics, Najafabad Branch, Islamic Azad University, Najafabad, Iran, PO Box 8514143131

(2) Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, Osservatorio Etneo, Piazza Roma, 2, 95123, Catania, Italy

(3) Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, Universita degli Studi di Catania, Viale A. Doria, 6, 95125, Catania, Italy

Classification of Mount Etna (Italy) Volcanic Activity by Machine Learning Approaches

Alireza Hajian[(1)], Flavio Cannavò[(2)], Filippo Greco[(2)], Giuseppe Nunnari[(3)]

(1) Department of Physics, Najafabad Branch, Islamic Azad University, Najafabad, Iran, PO Box 8514143131

(2) Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, Osservatorio Etneo, Piazza Roma, 2, 95123, Catania, Italy

(3) Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, Università degli Studi di Catania, Viale A. Doria, 6, 95125, Catania, Italy

*Corresponding author: Giuseppe Nunnari

E-mail address: giuseppe.nunnari@dieei.unict.it

Tel. +39 095 7382306

Subject Classification:

04.08.06. Volcano Monitoring, 04.08.07. Instruments and Techniques, 05.01.01. Data Processing, 05.01.04 Statistical Analysis, 05.02.99 General or Miscellaneous.

Abstract

Assessment of the ongoing activity of volcanoes is one of the key factors to reduce volcanic risks. In this paper, two Machine Learning (ML) approaches are presented to classify volcanic activity using multivariate geophysical data, namely the Decision Tree (DT) and K-Nearest Neighbours (KNN). The models were implemented using a data set recorded at Mount Etna (Italy), in the period 01 January 2011 – 31 December 2015, encompassing lava fountain events and intense Strombolian activity. Here a data set consisting of five geophysical features, namely the root-mean-square of seismic tremor (RMS) and its source depth, counts of clustered infrasonic events, radar RMS backscattering power and tilt derivative, was considered. Model performances were assessed by using a set of statistical indices commonly considered for classification approaches. Results show that between the investigated approaches the DT model is the most appropriate for classification of volcano activity and is suitable for early warning systems applications. Furthermore, the comparison with a different classifier approach, reported in literature, based on Bayesian Network (BN), is performed.

1. Introduction

One of the key challenge to modern volcanology is to identify and characterize volcano activity, based on parameters recorded by the monitoring network that might be useful for hazard assessment and risk mitigation. This topic has maximum priority especially for volcanoes located close to densely urbanized areas, such as Mount Etna volcano, Italy. Mount Etna is nowadays one of the best-monitored volcanoes worldwide. A wide data set ranging from seismic, geodetic, gravimetry, geochemical, video etc., most of them collected in real time, were able to set strict constraints on the timing of the paroxysmal events which occurred in the last decade at Etna volcano [i.e. Aloisi et al. 2018; Bonaccorso et al. 2011; Greco et al. 2016].

Changes in the state of activity as well as in the eruptive style of the volcano can be recognized in the geophysical and geochemical data collected by this monitoring network. The problem is that any change in these parameters does not especially imply a significant change in volcanic activity. Hence, the characterization of volcanic activity based only on one geophysical and/or geochemical parameter, may be lead to ambiguous forecasts. One solution to this problem is through using a combination of different parameters to reduce the level of ambiguity and to enhance the quality of interpretations of the volcano activity.

Volcano monitoring aims at the recognition of changes in observable parameters before hazardous activity develops in order to alert the Civil Defence Authorities. However, up to now, the lack of models for automatically recognizing the significance of changes in multiparametric real-time data, makes their integration difficult (or in many cases impossible). Using expert tools for real-time volcano monitoring can be helpful in enhancing the quality of decision-making processes in order to reduce the volcano risk [Cannavò et al. 2017].

In order to integrate geophysical parameters there are some statistical methods in which a pre-assumption about the model is applied and attempts are made to fit the data to that model; this means that they are model-based. The other new solution is through using supervised intelligent methods, which can learn from experimental data without any pre-assumption about the model. In the last decades, there has been an increasing tendency to use intelligent methods in geophysical volcanology studies, especially for monitoring the volcano activity. For instance, Langer et al. [2009] used Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) as pattern classifiers for volcanic tremor data, which is extremely useful for monitoring volcanic activity at any moment and in whatever condition. They investigated the tremor features and their relationship to regimes of volcanic activity. Messina and Langer [2011] combined various unsupervised pattern recognition techniques, self-organizing maps, classic k-means cluster analysis and fuzzy c-means clustering for pattern recognition of volcanic tremor data on Mount Etna. Brancato et al. [2016] used different kinds of Artificial Neural Networks (ANNs) for pattern recognition of Etna flank eruption forecasting. Malfante et al. [2017] used Machine Learning for Automatic Classification of Volcano-Seismic signatures and presented an automatic classification of volcano-seismic events, based on a comprehensive signal representation with a large feature set. Cannavò et al. [2017] proposed a multivariate probabilistic graphical model for real-time volcano monitoring on Mount Etna, based on an extension of a

Bayesian Network (BN). The resulting probabilistic model was able to encoding the conditional dependencies between measurements and volcanic states.

In this paper, we compare two different ML approaches, namely the Decision Tree (DT) and the K-Nearest-Neighbour (KNN) algorithms, in order to find a classification rule capable of mapping a set of five real-valued geophysical features against observed volcanic activity, represented in categorical form. The data set considered is the one used by Cannavò et al. [2017], so that the performances will be easily compared. We focus on the relationships between geophysical parameters patterns and the volcano status and discuss the methodological differences among these classifiers and their results.

The paper is organized as follows: in section 2, the problem is stated and the volcanological framework, in which the data set was recorded, is described. In section 3, the methods considered for the problem solution are presented. In section 4, we describe how the data set is organized for training the classifiers and how the model performances will be assessed. In section 5, numerical results are presented and the DT and KNN approaches compared. The selected classifier will be finally compared with the BN model proposed by Cannavò et al. [2017]. In section 6, the capabilities of the proposed classifier will be illustrated analysing selected episodes occurred on Mount Etna during 2011-2015. Finally, in section 7, the conclusions are drawn.

2. Problem statement

The main purpose of this paper is to implement a classifier capable of recognizing different phases of the volcanic activity which occurred at Mount Etna between 01 January 2011 and 31 December 2015, through a set of geophysical parameters (features) recorded in this area by the multidisciplinary monitoring network (Fig. 1).

Formally, the input-output representation of the classifier, that we aim to identify, can be expressed as follows:

$$Y(t) = f(X(t)) \tag{1}$$

Where:

- $Y(t)$ indicates the categorical model output at time *t;*
- $X(t)$ is the real-valued vector of features measured at the same time;
- f is a map representing the unknown classification rule.

In a similar manner to what was done by Cannavò et al. [2017], the volcano activity is expressed by a set of three different digit classes $G = \{0,1,2\}$, where 0 stands for no activity, 1 for Strombolian activity and 2 for paroxysmal explosive activity, while the feature vector X consists of geophysical data coming from the monitoring networks. Here, with the term Strombolian activity we intend burst frequency lower than a fraction of minutes and magma jets usually not exceeding tens to hundreds meters and weak effusive activity. Moreover, with the term paroxysmal events, we intend lava fountains and strong Strombolian eruptions, characterized by high burst frequency, up to a few per second, and magma jets up to several hundreds of meters, usually accompanied by the emission of lava flows [Andronico et al. 2013].

2.1. Volcanological framework and features description

After the flank eruption of 2008-2009, lava fountain events and intense Strombolian activity, have characterized activity at Mount Etna. Between January 2011 and December 2013, a sequence of 44 lava fountain events, characterized by powerful gas emission, together with ejection of lava fragments to heights ranging from tens to hundreds of meters, occurred from a pit on the east flank of the South-East Crater (SEC), one of the four summit craters of Etna, as well as Strombolian activity and intra-crater lava flow emissions at Bocca Nuova Crater (BNC) and Voragine (VOR). The location of such craters and the area where the measuring stations are installed is shown in Figure 1. The main explosive activity built a new summit cone, the New South-East Crater (NSEC) [Behncke et al. 2014]. No major flank eruptions have accompanied this activity. Generally, all the 2011-2013 lava fountain episodes from the NSEC showed similar main characteristics, with the height of the lava fountain reaching 300-1000 m, ash columns reaching 5-9 km and associated lava flows 4-6 km long, descending the eastern flank of the volcano.

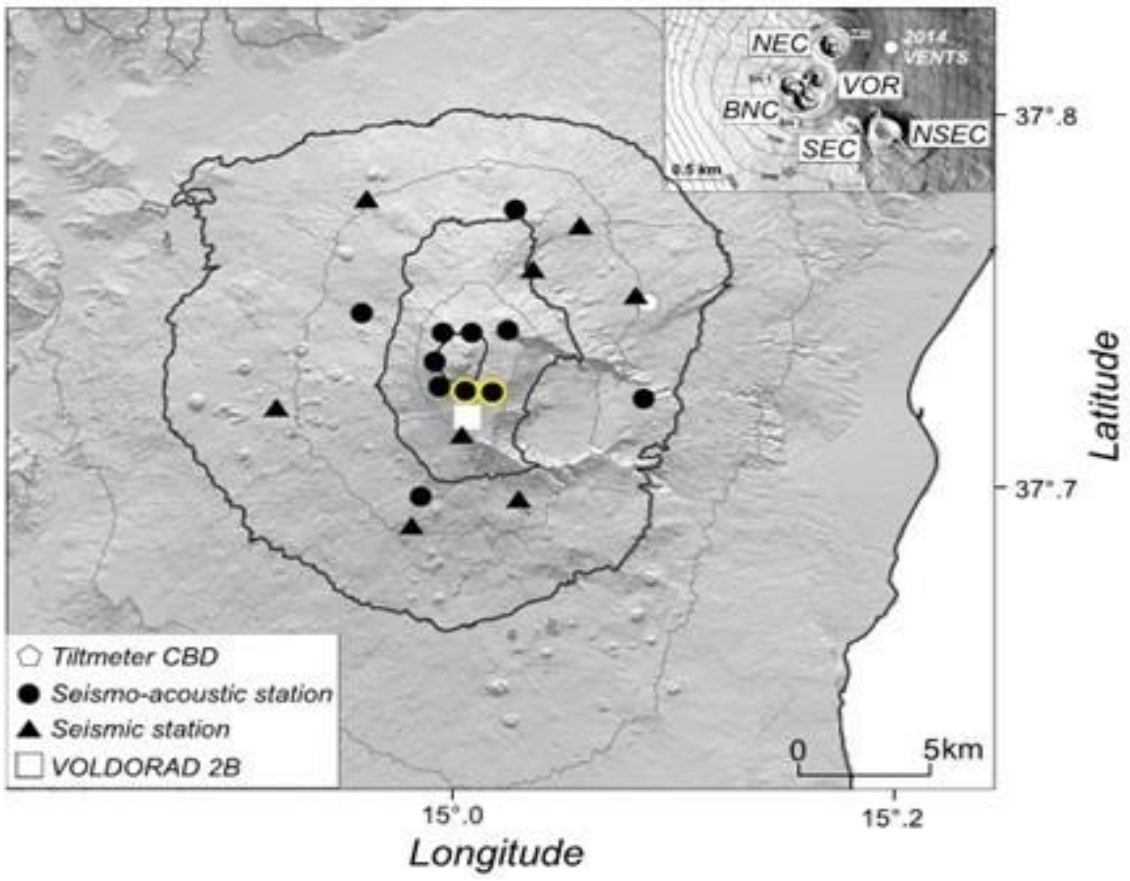

Figure 1. Location map of the seismic, acoustic, and tiltmeter stations used in this research. The position of the Doppler radar (VOLDORAD 2B) is also shown. The yellow circles indicate the stations destroyed during paroxysmal episodes on 28 February and 11 November 2013 [after Cannavò et al. 2017].

The average total density rock equivalent (DRE) volume of magma emitted during each of the NSEC lava fountains, including both pyroclastic products and lava flows, was about $2.5 \times 10^6 m^3$ per event [De Beni et al. 2015]. After the 2011-2013 NSEC lava fountains, the eruptive activity switched to moderate lava effusion from the NSEC toward NE during January-April and July-August 2014. On 28 December 2014, a further lava fountain episode occurred at the NSEC [Gambino et al. 2016].

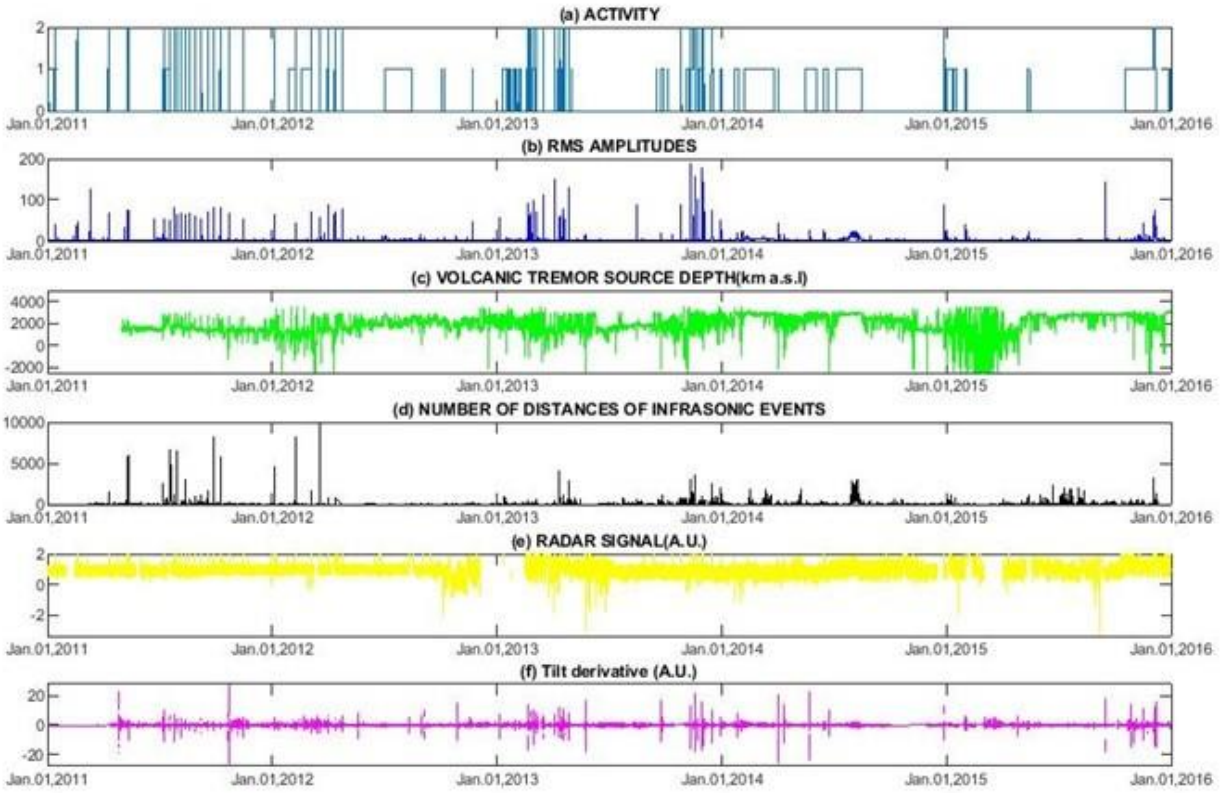

Figure 2. From top to bottom: a) time series showing the state of Mount Etna during 01 January 2011 – 31 December 2015; b) seismic RMS amplitudes; c) depth of volcanic tremor source centroid; d) number of distances of spatially clustered infrasonic events; e) signal obtained by the Doppler radar VOLDORAD 2B; f) tilt derivative of the signal recorded by the CBD station [after Cannavò et al. 2017].

As mentioned in the previous section, for the purpose of the present work, the activity of the Etna volcano has been categorized by using three different classifications:

- "0" to indicate quiet, i.e. no relevant volcanic activity or unknown activity, characterized by low amplitudes of all considered features;
- "1" to indicate Strombolian activity, i.e. mildly explosive activity, characterized by medium amplitude of seismic tremor RMS, shallower source of the seismic tremor, clustered infrasonic events, no eruption column but possible ash emission detectable by the Doppler radar;

- "2" to indicate paroxysm, i.e. an energetic activity with lava fountains, characterized by high amplitude of seismic tremor RMS, shallower source of the seismic tremor, eruption column formation and ash emission detectable by the Doppler radar.

For consistency, the observed volcanic activity and the corresponding features were re-sampled with a 10 min sampling rate.

The main reasons for choosing the five features listed in Table 1 are: (i) they are available in real time at a rate compatible with early warning purposes; (ii) they are usually sensitive to Etna volcanic activity; (iii) they roughly represent a summary of all available information from the monitoring networks. In Figure 2, the time series of activity at Mount Etna during 01 January 2011 – 31 December 2015, together with the five features used are shown. We select this period because Mount Etna exhibited an exceptional variety of explosive activity from Strombolian to lava fountain events. Follows, a description of the five features used:

- The feature x_1 (weighted normalized median of seismic tremor; RMS) is based on seismic RMS amplitudes which has been routinely used at Istituto Nazionale di Geofisica e Vulcanologia (INGV), Osservatorio Etneo, for real-time monitoring [Cannata et al. 2013]. This feature was calculated on the vertical component of the signal within 10 min long time windows, recorded by 19 stations, belonging to the permanent seismic network (sampling rate of 100 Hz). The signal was filtered in the band 0.5 − 5.5 Hz, which generally includes the relevant volcanic tremor information. In order to reduce the dimensionality of the model, we averaged the normalized RMS time series calculated for each station as performed by [Cannavò et al. 2017].
- The feature x_2 (normalized volcanic tremor source depth) was obtained by a grid search method, based on a spatial seismic amplitude distribution and assuming the propagation in a homogeneous medium, within 30 min long time windows. The use of relatively long time windows is necessary to reduce the effects of transient events (such as long period events, volcano-tectonic earthquakes, and regional earthquakes) on the volcanic tremor locations [Cannavò et al. 2017; and references therein].
- The feature x_3 (normalized number of infrasonic events) was obtained by locating about 160.000 infrasonic events, recorded by the permanent infrasonic network during 2011-2015.
- The feature x_4 (normalized radar RMS) was obtained by processing the total radar echo backscattered by particles in the atmosphere crossing the beam of the 1.274 GHz Doppler radar, called VOLDORAD 2B and installed in the upper southern flank of the volcano (2600 m a.s.l.). This signal is dominated by reflections from particles, and hence it is sensitive to the formation of eruptive columns; it can be used to detect abundant pyroclastic particles in atmosphere crossing the radar beam [Donnadieu et al. 2016]. In particular, periods of intense explosive activity producing abundant ash, are characterized by sharp increases in the considered time series.
- The feature x_5 (normalized tilt derivative) was obtained by using the data recorded by the tiltmeter station referred to a CDB (Fig. 1), installed at 10 m depth and equipped with a high resolution (< 0.005 μ*rad*) tiltmeter

[Ferro et al. 2011]. Since tilt data is affected by several noise sources (e.g., temperature, Earth tides, local earthquakes, and teleseism), we considered its derivative as input to the model.

Overall, the parameters related to volcanic tremor, infrasonic events, and ground deformation are sensitive to explosive activity, while the radar signal is sensitive to the development of sustained ash-rich eruption columns. Since the radar signal and infrasound are not affected by pure effusive activity, and because lava effusion during the considered period was always accompanied by explosive activity with different intensities, we avoided modelling a state of pure effusive activity and included it in the eruption states. Table 1 shows the summary of the input/output of the model classifier.

Table 1. The Machine Learner inputs and output variables with related descriptions. AU standing for arbitrary unit.

Representation	Description
x_1	Weighted Normalized Median of seismic tremor RMS (A.U.)
x_2	Normalized seismic tremor source depth (A.U.)
x_3	Normalized number of distances of Infrasonic events (A.U.)
x_4	Normalized Radar backscatter power at Montagnola station (A.U.)
x_5	Normalized Tilt derivative (A.U.)
Y	State of the volcano activity (A.U.)

3. Methods

In this section, a short description of considered classification approaches is provided.

3.1. Decision Learning Tree

Decision learning trees are popular ML algorithms belonging to the class of supervised learning algorithms. The main motivations to use Decision Tree (DT) is that they mimic the human level thinking and, differing from SVM and NN, are not black box models. Another reason for choosing these algorithms is because they are less time consuming with respect to other tested clustering approaches.

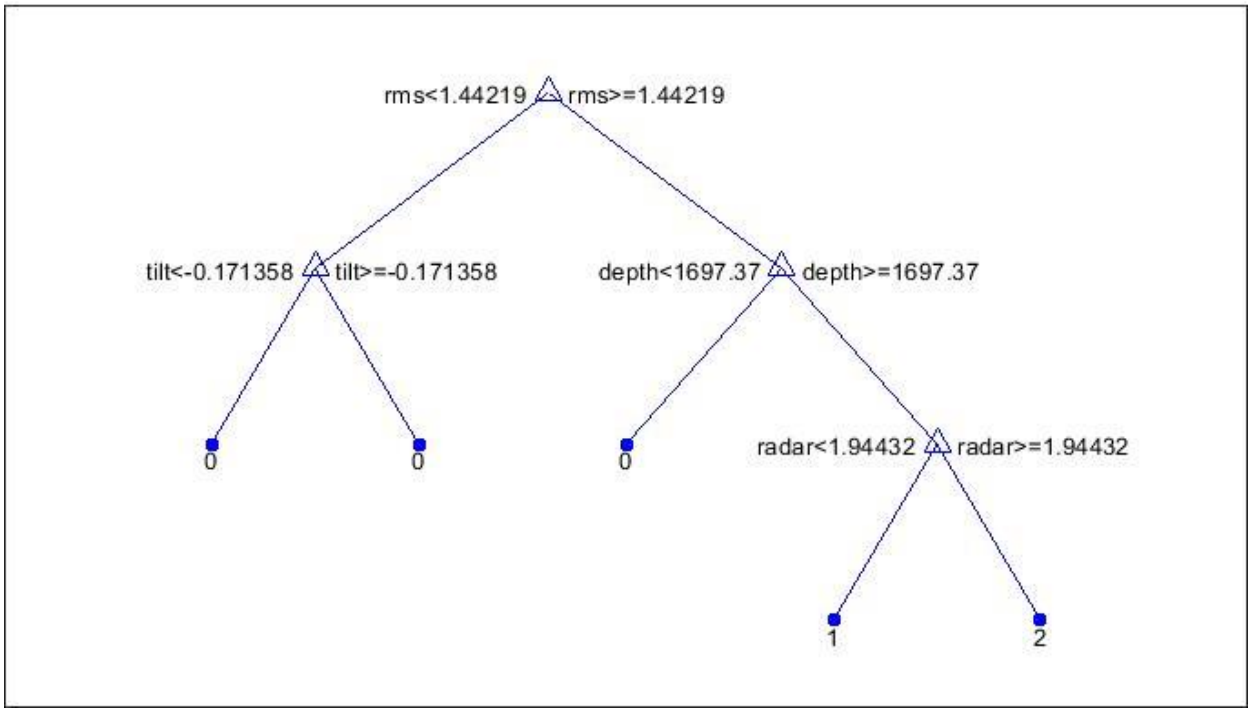

Figure 3. Example to classify the volcanic activity in terms of three features: average seismic RMS amplitude (rms); radar signal (radar); tilt derivative (tilt).

In order to illustrate how a DT algorithm works, we report in Figure 3 an attempt to classify the volcanic activity based on three features, namely the average seismic RMS amplitude, the radar signal and the tilt derivative. The classification tree was obtained training the model with a small number of splits (i.e. the number of levels of the tree), in order to obtain a rough, but easily, interpretable classification rule. In this example, the classification algorithm gives a prominent role to the rms feature, which is thus chosen as the root of the tree. Visiting the tree start from the root; in each node of the tree a binary condition is indicated by which it is possible to make the tree descent, until a leaf is reached. Thus, in this example, the class 2 will be assigned to a set of features which satisfies the condition *rms* ≥ 1.44219 and depth ≥ 1697.37 and *radar* ≥ 1.94432. However, it is trivial to understand that the interpretability of the classification rule can be done only for trees with a limited number of levels and features. In this paper, classification trees will be referred to as fine, when the number of splits is of the order of hundreds, medium, when the number of splits is of the order of tens and coarse, when this number is of the order of units.

3.2. K-Nearest-Neighbours

The K-Nearest-Neighbours (KNN) method uses observations in the training set closest to *X* to form *Y*. Specifically, the K-Nearest-Neighbour fits for *Y* defined as

$$\hat{Y} = \frac{1}{K} \sum_{x_i \in N_K(x)} y_i \qquad (2)$$

where $N_K(x)$ is the neighbourhood of x, defined as the K closest point x_i in the training sample. Closeness implies the use of a metric, which in the simplest case can be the traditional Euclidean distance. However, several alternatives are possible, such as the Chebshev distance or the Minkowsky distance just to mention a few. Of course, the performance of a K-Nearest-Neighbours classifier depends on the number K of elements in the $N_K(x)$ neighbour. Similar to the tree classifiers, KNN can be referred to as fine, medium or coarse depending on wheter K is of the order of units, tens or hundreds, respectively.

4. Training of the classification models and performance assessment

In order to identify the unknown learning rule (see eq. 1), described in section (2), the available data set, which spans over five years, was organized into two subsets: four years of data was considered for training the classifier while the remaining one year was considered for the test. However, in order to avoid bias of the model error, the testing year was rotated using the leave-one-out approach, described in Figure 4. In the Figure, it is schematically indicated that the model error E was averaged over five iterations. During each iteration, one year of the overall data set was left out from the training set and considered only for the testing. The performance indices to evaluate the error model E, are described in the next section (4.1).

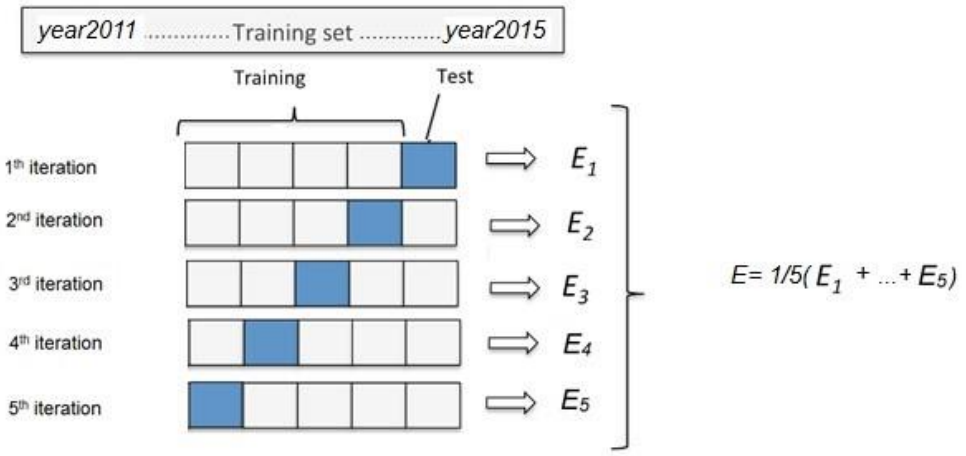

Figure 4. Schematic of the Leave-One-Out cross validation method used in this study.

4.1. Classification performance metrics

In order to assess the performances of a classifier, different metrics can be considered. A popular metric consists in computing a set of indices referred to as the True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), False Negative Rate (FNR), accuracy and so on. Another example of metrics refers to the evaluation of the Precision, the Specificity, and the Fall-out etc. It is obvious that the choice of metrics influences how the

performance of a ML algorithm is measured and compared. In this paper we will refer to the former metrics mentioned above.

Given a classification experiment, let us indicate as *P* and *N* the number of actual positive and actual negative cases for a given class, respectively, and as *TP*, *TN*, *FP* and *FN* the number of true positive, true negative, false positive and false negative cases, respectively, recognized by the classifier, for the considered class. Thus, the following rates are defined:

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} = 1 - FNR \quad (3)$$

$$TNR = \frac{TN}{N} = \frac{TN}{TN + FP} = 1 - FPR \quad (4)$$

$$FNR = \frac{FN}{P} = \frac{FN}{FN + TP} = 1 - TPR \quad (5)$$

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - TNR \quad (6)$$

The meaning of the above indices can be expressed as follows:

- The *TPR* expresses the proportion of actual positives that are correctly classified by the model as belonging to a given class. Best values of TPR approaches to 1, while in the worst case TPR approaches to 0. The *TPR* is referred in literature also as specificity or recall.
- The *TNR* expresses the proportion of actual negatives that are correctly classified as not belonging to a given class. As for the *TPR*, best values of TNR approaches 1, while worst values approaches 0. The *TNR* is referred to also as sensitivity or selectivity.
- The FNR expresses the proportion of false negatives in a given class with respect to all actual positives in the same class. Of course, in the best case FNR approaches 0, while in the worst case approaches 1.
- The FPR expresses the proportion of false positives in a given class with respect to the total number of actual negatives in the same class. Similar to the FNR in the best case FNR approaches 0, while in the worst case approaches 1.

Table 2. Sensitivity (TPR) and specificity (TNR) for different types of tested Tree Classifiers.

Activity	Quiet			Strombolian			Paroxysm		
Type of DT	fine	medium	coarse	Fine	medium	coarse	fine	medium	Coarse
TPR	0.8703	0.8792	0.8445	0.6471	0.5083	0.3054	0.5878	0.3104	0.1456
TNR	0.6782	0.4456	0.4102	0.8701	0.6587	0.7645	0.9989	0.7915	0.8432

Table 3. Sensitivity (TPR) and specificity (TNR) for different types of tested KNN Classifiers.

Activity	Quiet			Strombolian			Paroxysm		
Type of KNN	fine	medium	coarse	Fine	medium	coarse	fine	medium	coarse
TPR	0.8370	0.8616	0.8551	0.3361	0.2031	0.3309	0.5240	0.2310	0.3985
TNR	0.3569	0.1931	0.2210	0.8372	0.5764	0.66424	0.9984	0.6654	0.7361

Together with the four indices above, we also consider the index referred as F_1, defined as:

$$F_1 = \frac{2TP}{2TP + FP + FN} = \frac{2rp}{r + p} \qquad (7)$$

which represents a harmonic mean of precision (p) and recall (r). It is trivial to understand that F_1 approaches 1 in the best case and 0 in the worst case.

5. Numerical results

As mentioned in the section 2, in this study two kinds of classifiers were considered, namely the DTs and the KNNs. In more detail, as a first stage, for each kind of classifier different configurations were considered, meaning that coarse, medium and fine classifiers were trained, in order to select the most appropriate for the application. Results obtained considering the TPR and the TNR for each class and for each configuration are reported in Tables 2 and 3. Looking at these tables, it is possible to infer that the fine Tree classifier is the optimum among the DT classifiers, and similarly, the fine KNN is the optimum among the tested KNNs, particularly for the Strombolian and paroxysmal events. A direct comparison between the fine DT and fine KNN is shown in Figure 5 which shows that the fine DT classifier outperforms the fine KNN for all the classes and indices (from here, in this paper, the fine DT and the fine KNN will be referred to as FDT and FKNN, respectively). In particular, the sensitivity of the FDT (Fig. 5a) alarm is higher than that of FKNN especially for paroxysmal activity (class 2). The FPR (fall out) value for Strombolian class (Fig. 5c) in FKNN is higher than that of the FDT. When the fall out is high for Strombolian class, it means that most of the time when the alarm is 'on', the volcano state is quiet or in paroxysmal state. The TNR (specificity) of FDT and FKNN (Fig. 5b) are very close to each other for the Strombolian and paroxysmal classes, but for the quiet class, the TNR (specificity) of FDT is higher than the FKNN. It is necessary to mention that the FNR index was zero for the class 2, for both the FDT and FKNN (Fig. 5d). The evaluation metric we investigated for the competition between FDT and FKNN was the F_1 score, as defined in (eq. 6). The comparison between the FDT and FKNN in terms of the F_1 index is shown in Figure 6. As it can be seen the F_1 score, i.e. the harmonic mean of precision and recall, for the FDT is higher than that of FKNN for all activity classes. Thus, we selected the FDT as the

optimum between the two compared approaches. The performance of the selected classifier is summarized in Table 4, in terms of a confusion matrix.

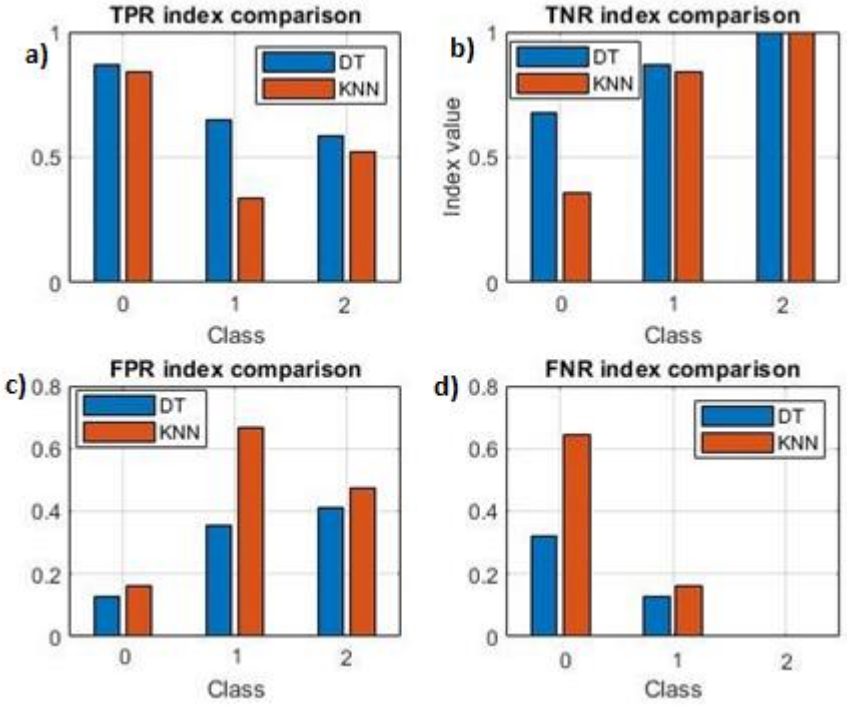

Figure 5. A direct comparison between FDT and FKNN.

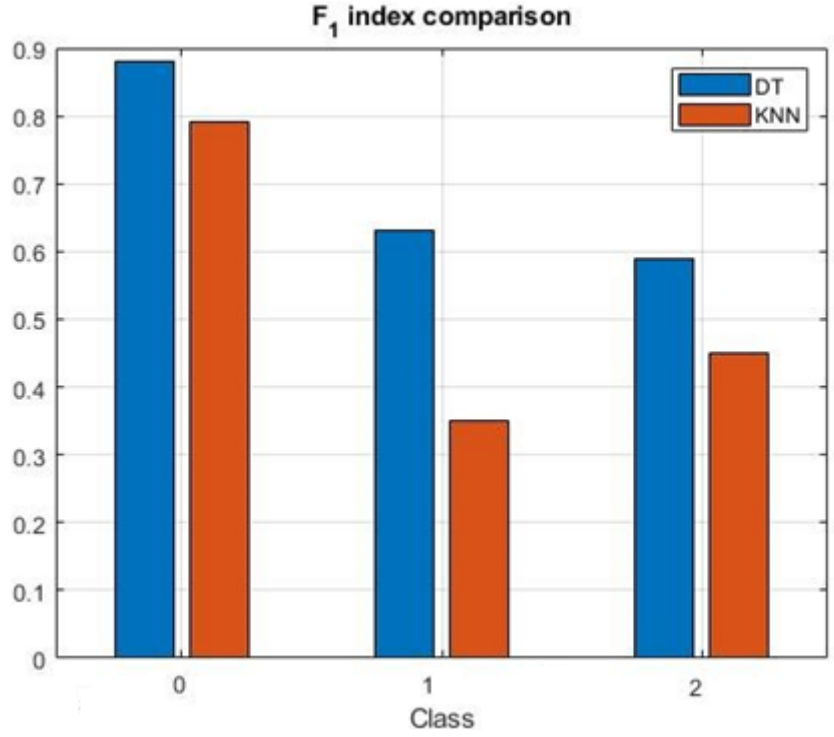

Figure 6. A direct comparison between FDT and FKNN in terms of the F_1 index.

Table 4. Confusion matrix of the FDT classifier.

FDT	Predicted Quiet	Predicted Strombolian	Predicted Paroxysm
Actual Quiet	87.06%	12.94%	0.0%
Actual Strombolian	25.69%	73.25%	1.06%
Actual Paroxysm	0.0%	74.07%	25.93%

The confusion matrix for continuous estimation of the ongoing volcano state shows in its main diagonal the percentage of events, correctly recognized in each class. The table also shows that a percentage of 12.94% of actual Quiet events (class 0), was recognized by the classifier as Strombolian activity, while none event of this class was attributed to class 2. A percentage of 25.69% of events belonging to the Strombolian class, were attributed to the class 0. Finally, a percentage of 1.06% events actually belonging to the Strombolian class, were incorrectly attributed to the paroxysm (class 2). The most relevant drawback of the FDT classifier is that a high percentage (74.07%) of events actually belonging to the paroxysm (class 2), was attributed to class 1 (Strombolian).

Referring to the Strombolian class, the FDT estimated no activity (Quiet state) instead of the correct state for 25.69% of events which is, nevertheless, better than the BN model (about 50%) reported in Cannavò et al. [2017]. The main reason that 25.69% of the Strombolian class events are confused as quiet state, is due to a few but long periods of weak Strombolian activity that is not distinctly reflected in any input variables.

6. Evaluation of the FDT classifier for selected episodes

To test the performance of the FDT classifier, to track the actual episodes, we compared the output model time series with the true one. As a sample, the FDT result for the episodes between 30 July 2011 and 31 July 2011 is shown in Figure 7. This example shows a *delay* between the predicted state of the model and the actual state of the volcano. This means that the alarm is on with a delay after starting of the Strombolian event. Conversely, as it can be seen in Figure 7, the alarm for the paroxysmal event is *on time*. In order to calculate the performance of the model for paroxysmal events detection, we changed the model into a binary classifier, which means '0' for both Quiet and Strombolian and '1' for paroxysmal events. The performance indices of this kind of classifier (binary) were calculated for the considered period 2011-2015, based on the leave-one-out cross validation method described in section 4.1.

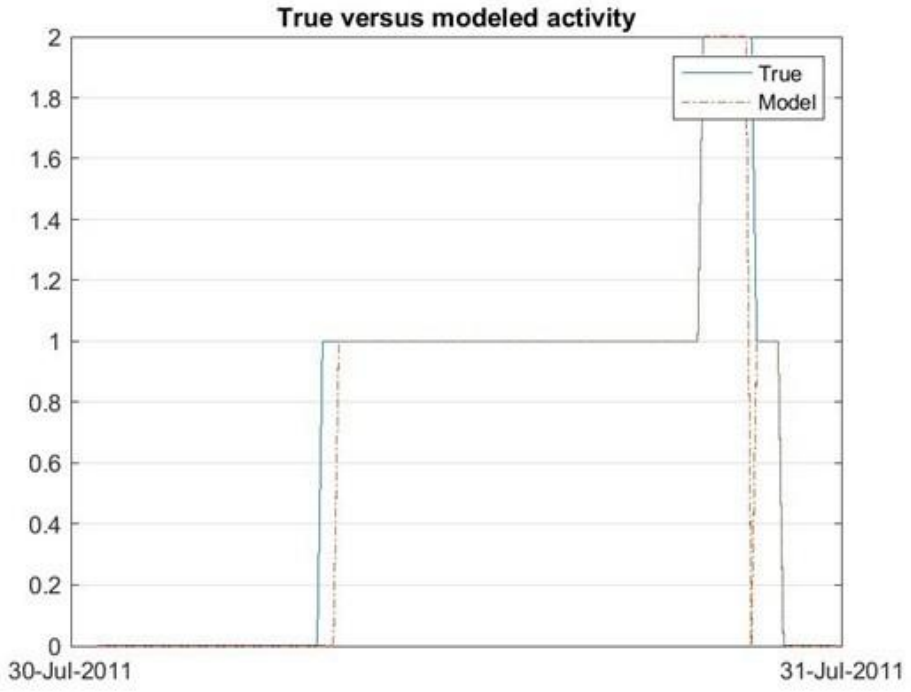

Figure 7. FDT classification estimation for the activity of Etna between 30 July 2011 and 31 July 2011, note: class 0 as Quiet, class 1 as Strombolian and class 2 as Paroxysm.

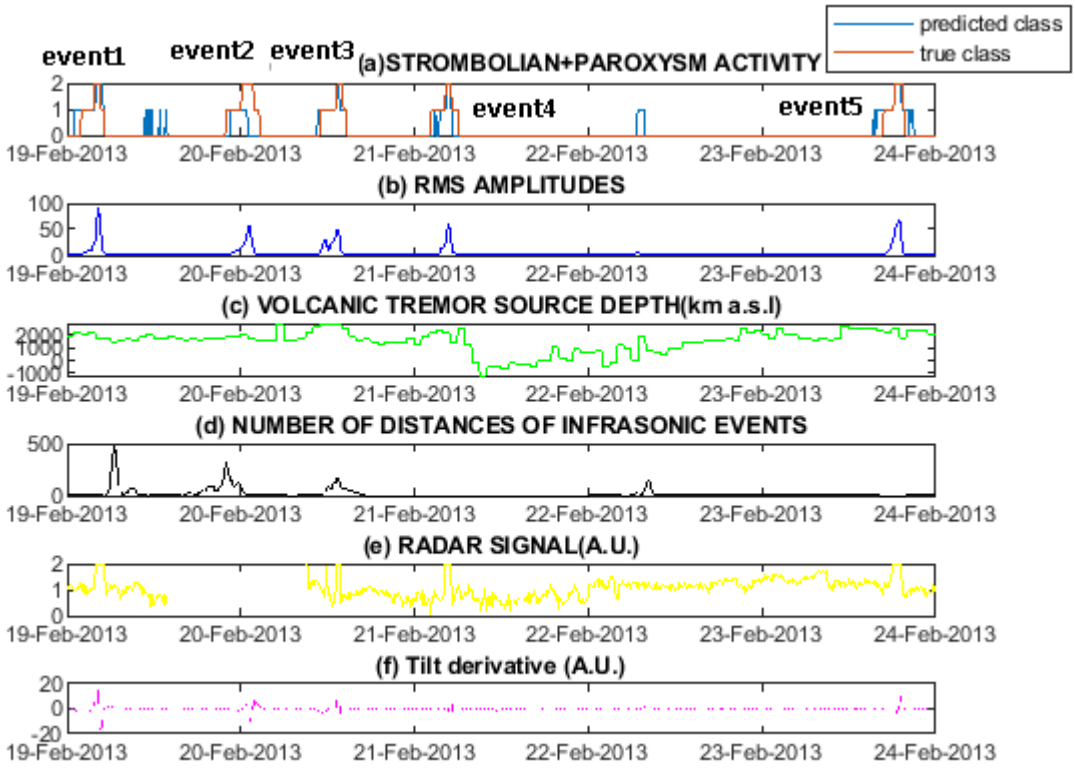

Figure 8. FDT predicted class activity in a time window between 19 February 2013 and 24 February 2013.

Between 19 and 24 February 2013, a series of five compound volcano events (both Strombolian and paroxysm; labelled each with a tag number) occurred, as shown in Figure 8. For each event, the volcano activity starts with a Strombolian class and then leads to a paroxysmal event. In the events 1, 3 and 5 there is an *advanced time* (a time measure of the capability of model to predict an event before it happens). The advance time for events 1 and 5 are greater than that of event 3 (Fig. 8). Conversely, the event 4 is predicted by the model with a *delay*. Only two false alarms are detected. This example highlights the good ability of the FDT model to track all the events starting with Strombolian activity and leading to paroxysmal event.

In order to estimate the FDT classifier capability to detect the beginning of paroxysmal events, we averaged the *advance time* estimated over the whole data set. The comparison between the performances of FDT and BN proposed by Cannavò et al. [2017], both in terms of advance time, TPR and FPR are reported in Table 5. It is possible to see that these models performs similarly in terms of TPR and FPR but the FDT outperforms the BN in terms of advance time, which is especially relevant for early-warning alarm in order to alert the Civil Defence Authorities.

Table 5. Performance for the FDT working style as a binary classifier to detect paroxysm events calculated by leave-one-out cross validation, and the Bayesian Network (BN).

Index	FDT	BN model
TPR(%)	99.66	99.16
FPR(%)	21.34	23.21
Advance Time (min)	55.00	32.16

7. Discussion and conclusions

The search for a reliable and quantifiable method of determining the onset of volcanic eruptions is a key challenge for volcanology and has major implications sociologically and political as well as geologically.

Machine Learning methods make no assumptions about mechanisms behind geophysical and volcanological phenomena, but simply attempt to look for relationships between parameters collected in the volcanoes monitoring network to map these onto a set of outcomes, which are classes of the volcano state.

In this paper, two categories of Machine Learning models were investigated for classification of Mount Etna activities, namely the DTs and KNNs. The global statistical indices show that the FDT and the FKNN are the models with higher performances, compared within their own categories (Coarse and Medium). Furthermore, taking into account the F_1 score as a measurement of the total performance of precision and recall, the FDT outperforms the FKNN. Experimental results show that the FDT model is very sensitive to the paroxysmal activity, being even able to detect in advance the beginning of this kind of activity. We have estimated that the mean advance time is about 55 minutes for the considered set of paroxysmal events. The advance time for events of class 1 (Strombolian activity) has not be estimated, because, from early warning perspective, the tested models are not enough sensitive to the beginning of this weakly explosive activity. A comparison between the results obtained in this work is possible only with the ones in Cannavò et al. [2017], since (to the best of our knowledge) other studies have not used an extensive set of features as in this work. The comparison with Cannavò et al. [2017] showed that the performances among the DT and BN approaches are substantially similar: to a great capacity to detect paroxysms (with a positive rate above 99%) is opposed a not negligible percentage of false alarms (with FPR > 21%, about 1 out of 5 detected paroxsyms was false). This suggests that at this stage of the work the set of features used is more relevant rather than the applied techniques. Therefore, in order to improve the classifier performance, we plan to add some new relevant features to the data set. In particular, we expect to consider the gravity signal, SO_2 emissions and magnetic data.

In conclusion, we can say that FDT methods are promising classification algorithms to associate the categorical volcanic activity to a set of features measured in a given area. The use of such tools in real-time volcano surveillance could be helpful for decision-making processes aimed at reducing volcano risk. Specifically, the tool

could be useful both for civil defence authorities, who can evacuate on time the summit area of the volcano, when volcanic activity changes, and the aviation authorities in order to avoid problems with air traffic.

Acknowledgments

We are indebted to the Authors of the paper Cannavò et al., [2017] for enabling data used in this manuscript. Special thanks are due to the Najafabad Branch, Islamic Azad University President and Research Deputy for supporting the activity of Dott. Alireza Hajian during the sabbatical leave at INGV – Osservatorio Etneo. Prof Giuseppe Nunnari wish to thank the Università degli Studi di Catania for the financial support under the grant Piano della Ricerca 2016-2018, Linea di Intervento 2. We thank the Associated Editor Dr C. Del Negro and the reviewers Prof. C. Bonadonna and Prof. P. Styles for their careful reading of our manuscript and their constructive comments and suggestions.

References

Aloisi M., A. Bonaccorso, F. Cannavò F. and G.M. Currenti (2018). Coupled Short- and Medium-Term Geophysical Signals at Etna Volcano: Using Deformation and Strain to Infer Magmatic Processes From 2009 to 2017, Front. Earth Sci. 6:109, doi: 10.3389/feart.2018.00109.

Andronico, D., M.D. Lo Castro, M. Sciotto, and L. Spina (2013). The 2010 ash emissions at the summit craters of Mt Etna: Relationship with seismo-acoustic signals. J. Geophys. Res. Solid Earth, 118, 51–70, doi: 10.1029/2012JB009895

Behncke, B., S. Branca, R. A. Corsaro, E. De Beni, L. Miraglia, and C. Proietti (2014). The 2011–2012 summit activity of Mount Etna: Birth, growth and products of the new SE crater, J. Volcanol. Geotherm. Res., 270, 10–21, doi:10.1016/j.jvolgeores.2013.11.012.

Bonaccorso, A., A. Cannata, R. A., Corsaro, G., Di Grazia, S., Gambino, F., Greco, L. Miraglia, A., Pistorio, A. (2011). Multidisciplinary investigation on a lava fountain preceding a flank eruption: the 10 491 May 2008 Etna case. Geochem. Geophys. Geosyst. 12:Q07009, doi: 10.1029/2010GC003480.

Brancato, A., P.M. Buscema, G. Massini and S. Gresta (2016). Pattern Recognition for Flank Eruption Forecasting: An Application at Mount Etna Volcano (Sicily, Italy), Open Journal of Geology, 6, 583-597.

Cannata, A., G. Di Grazia, M. Aliotta, C. Cassisi, P. Montalto and D. Patane (2013). Monitoring seismo-volcanic and infrasonic signals at volcanoes: Mt. Etna case study, Pure Appl. Geophys., 170 (11), 1751-1771.

Cannavò, F., A. Cannata, C. Cassisi, G. Di Grazia, P. Montalto, M. Prestifilippo, E. Privitera, M. Coltelli, and S. Gambino (2017). A multivariate probabilistic graphical model for real-time volcano monitoring on Mount Etna, J. Geophys. Res.Solid Earth, 122, 3480–3496, doi: 10.1002/2016JB013512.

De Beni, E., B. Behncke, S. Branca, I. Nicolosi, R. Carluccio, F. D' Ajello Caracciolo and M. Chiappini (2015).The continuing story of Etna new Southeast Crater (2012-2014): Evolution and volume calculations based on field surveys and aerophotogrammetry, J. Volcanol. Geotherm. Res., 2015, 303, 175–186.

Donnadieu, F., P. Freville, C. Hervier, M. Coltelli, S. Scollo, M. Prestifilippo, S. Valade, S. Rivet, and P. Cacault (2016). Near-source Doppler radar monitoring of tephra plumes at Etna, J. Volcanol. Geotherm. Res., 312, 26-39.

Ferro, A., S. Gambino, S. Panepinto, G. Falzone, G. Laudani and B. Ducarme (2011). High precision tilt observation at Mt. Etna volcano, Italy, Acta Geophys., 59(3), 618-632.

Gambino, S., A. Cannata, F. Cannavò, A. La Spina, M. Palano, M. Sciotto, L. Spampinato, and G. Barberi (2016), The unusual 28 December 2014 dike-fed paroxysm at Mount Etna: Timing and mechanism from a multidisciplinary perspective, J. Geophys. Res. Solid Earth, 121(3), 2037–2053, doi:10.1002/2015JB012379.

Greco, F., G. Currenti, M. Palano, A. Pepe, and S. Pepe (2016). Evidence of a shallow persistent magmatic reservoir from joint inversion of gravity and ground deformation data: The 25 – 26 October 2013 Etna lava fountaining event, Geophys. Res. Lett., 43, doi: 10.1002/2016GL068426.

Langer, H., S. Falsaperla, M. Masotti, R. Campanini, S. Spampinato and A. Messina (2009). Synopsis of supervised and unsupervised pattern classification techniques applied to volcanic tremor data at Mt Etna. Italy, Geophys. J. Int., 178, 1132-1144.

Malfante, M., M. Dalla Mura, J. Mars and J. P. Metaxian (2017). Machine Learning for Automatic Classification of Volcano-Seismic Signatures, 25th European Signal Processing Conference, 2457-2461.

Messina, A. and H. Langer (2011). Pattern recognition of volcanic tremor data on Mt. Etna (Italy) with KK Analysis—A software program for unsupervised classification, Computers & Geosciences 37(7):953-961, DOI: 10.1016/j.cageo.2011.03.015.