We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Machine Learning in Volcanology: A Review

Roberto Carniel and Silvina Raquel Guzmán

Abstract

A volcano is a complex system, and the characterization of its state at any given time is not an easy task. Monitoring data can be used to estimate the probability of an unrest and/or an eruption episode. These can include seismic, magnetic, electromagnetic, deformation, infrasonic, thermal, geochemical data or, in an ideal situation, a combination of them. Merging data of different origins is a non-trivial task, and often even extracting few relevant and information-rich parameters from a homogeneous time series is already challenging. The key to the characterization of volcanic regimes is in fact a process of data reduction that should produce a relatively small vector of features. The next step is the interpretation of the resulting features, through the recognition of similar vectors and for example, their association to a given state of the volcano. This can lead in turn to highlight possible precursors of unrests and eruptions. This final step can benefit from the application of machine learning techniques, that are able to process big data in an efficient way. Other applications of machine learning in volcanology include the analysis and classification of geological, geochemical and petrological "static" data to infer for example, the possible source and mechanism of observed deposits, the analysis of satellite imagery to quickly classify vast regions difficult to investigate on the ground or, again, to detect changes that could indicate an unrest. Moreover, the use of machine learning is gaining importance in other areas of volcanology, not only for monitoring purposes but for differentiating particular geochemical patterns, stratigraphic issues, differentiating morphological patterns of volcanic edifices, or to assess spatial distribution of volcanoes. Machine learning is helpful in the discrimination of magmatic complexes, in distinguishing tectonic settings of volcanic rocks, in the evaluation of correlations of volcanic units, being particularly helpful in tephrochronology, etc. In this chapter we will review the relevant methods and results published in the last decades using machine learning in volcanology, both with respect to the choice of the optimal feature vectors and to their subsequent classification, taking into account both the unsupervised and the supervised approaches.

Keywords: machine learning, volcano seismology, volcano geophysics, volcano geochemistry, volcano geology, data reduction, feature vectors

1. Introduction

Pyroclastic density currents, debris flow avalanches, lahars, ash falls can affect dramatically the life of people living close to volcanoes, and other volcanic products such as lava flows can severely affect properties and infrastructures. Several volcanoes

lie close to highly populated areas and the impact of their eruptions could be eco-nomically very strong. Stochastic forecasts of volcanic eruptions are difficult [1, 2], but deterministic forecasts (i.e., specifying when, where, how an eruption will occur) are even harder. Many volcanoes are monitored by observatories that try to estimate at least the probability of the different hazardous volcanic events [3]. Different time series can be monitored and hopefully used for forecasting, including seismic data [4], geomagnetic and electromagnetic data [5], geochemical data [6], deformation data [7], infrasonic data [8], gas data [9], thermal data from satellite [10] and from the ground [11]. Whenever possible, a multiparametric approach is always advis-able. For instance, at Merapi volcano, seismic, satellite radar, ground geodetic and geochemical data were efficiently integrated to study the major 2010 eruption [12]; a multiparametric approach is essential to understand shallow processes such as the ones seen at geothermal systems like e.g., Dallol in Ethiopia [13]. Although many time series may be available, seismic data remain always at the heart of any monitoring system, and should always include the analysis of continuous volcanic tremor [14]; tremor has in fact a great potential [15] due to its persistence and memory [1, 2] and its sensitivity to external triggering such as regional tectonic events [16] or Earth tides [17]. Moreover, its time evolution can be indicative of variations in other parameters, such as gas flux [18]. Other information-rich time series can be built looking at the time evolution of the number of the different discrete volcano-seismic events that can be recorded on a volcano. These include volcano-tectonic (VT) earthquakes, rockfall events, long-period (LP) and very-long-period (VLP) events, explosions, etc. Counting the overall number of events is not enough: one has to detect them and classify them, because they are linked to different processes, as detailed below. For this reason it is important to generate automatically different time series for each type of volcano-seismic event.

VT can be described as "normal" earthquakes which take place in a volcanic environment and can indicate magma movement [19, 20]. LP events have a great potential for forecasting [21]. Their debated interpretation involves the repeated expansion and compression of sub-horizontal cracks filled with steam or other ash-laden gas [22], stick–slip magma motion [23], fluid-driven flow [24], eddy shedding, turbulent slug flow, soda bottle analogues [25], deformation acceleration of solidified domes [26] and slow ruptures [27]. Explosion quakes are generated by sudden magma, ash, and gas extrusion in an explosive event, often associated to VLP events [28]. In many papers also "Tremor episodes" (TRE events) are described and counted, usually associated to magma degassing [20]. However, a volcano with any activity produces a continuous "tremor" which detectability only depends on the seismic instrumentation sensitivity [29, 30]. So, the class "TRE" should be better defined as "tremor episode that exceeds the detection limits". Of course, at volcanoes we can also record natural but non-volcanic seismic signals such as far tectonic earthquakes, far explosions, etc., and also anthropogenic signals e.g., due to industries, ground vehicles, helicopters used for monitoring, etc.

Most volcano observatories rely on manual classification and counting of such seismic events, which suffers from human subjectivity and can become unfeasible during an unrest or a seismic crisis [31, 32]. For this reason, manual classification should be substituted by an automated processing, and here is where machine learning (ML) comes into place. The same reasoning applies of course also to the automated processing of other monitoring time series, such as deformation, gas and water geochemistry, etc. Moreover, ML in volcanology is not restricted to monitoring active volcanoes but has demonstrated to be useful also when dealing with other large datasets. Examples include correlating volcanic units in general e.g., [33], of tephra e.g., [34, 35] and ignimbrites e.g., [36], a task which may become very difficult especially when many deposits of similar ages and geochemical and

petrographic characteristics crop out in a given area. ML is also effective for discriminating tectonic settings of volcanic rocks [34, 37]. Recently it has been used also for the prediction of trace elements in volcanic rocks [38].

2. Machine learning

ML is a field of computer science dedicated to the development of algorithms which are based on a collection of examples of some phenomenon. These examples can be natural, human-generated or computer-generated. From another point of view, ML can be seen as the process of solving a problem by building and using a statistical model based on an existing dataset [39]. ML can also be defined as the study of algorithms that allow computer programs to automatically improve through experience [40]. ML is only one of the ways we expect to achieve Artificial Intelligence (AI). AI has in fact a wider, dynamic and fuzzier definition, e.g., Andrew Moore, former Dean of the School of Computer Science at Carnegie Mellon University, defined it as "the science and engineering of making computers behave in ways that, until recently, we thought required human intelligence". ML is usually characterized by a series of steps: data reduction, model training, model evaluation, model final deployment for classification of new, unknown data (see **Figure 1**). The training (which is the proper learning phase) can be supervised, semi-supervised, unsupervised or based on reinforcement.

More data does not necessarily imply better results. Low quality and irrelevant data can instead lead to worse classification performances. If for each datum we have a very high number of columns, we may wonder how many of those are really informative. A number of techniques can help us with this process of **data reduction**. The simplest include column variance estimations and evaluating correlations between columns. Each of the components of the vector that "survive" this phase is called a feature and is supposed to describe somehow the data item, hopefully in a way that makes it easier to associate the item to a given class. There are dimensionality reduction algorithms [41] where the output is a simplified feature vector that is (almost) equally good at describing the data. There are many techniques to find a smaller number of independent features, such as Independent Component Analysis (ICA) [42], Non-negative Matrix
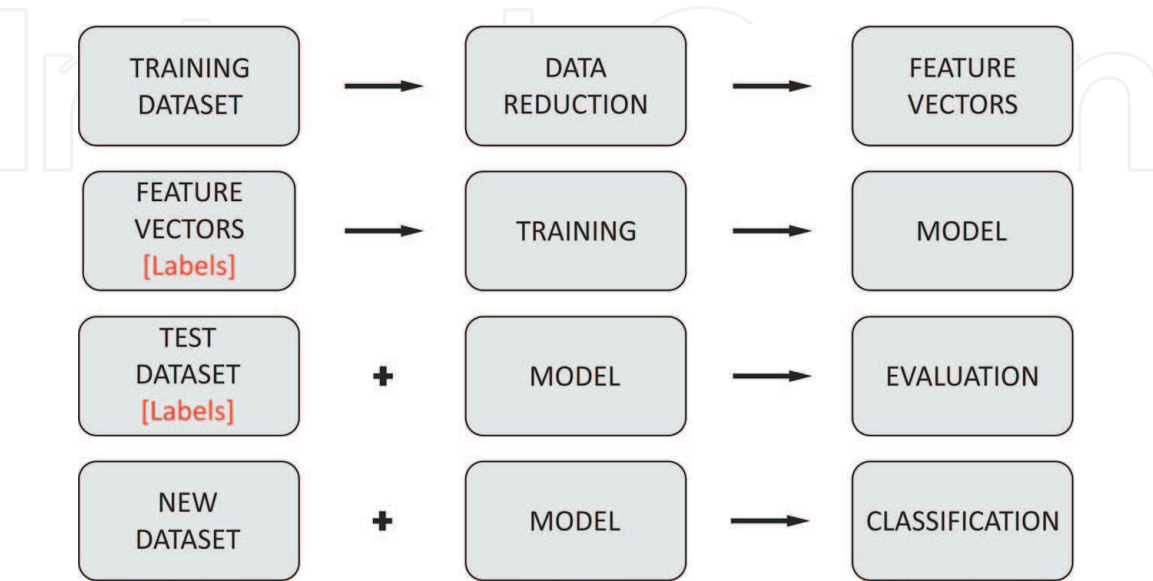
Figure 1.
ML can be divided in several steps, from top to bottom. Raw data have first to be reduced by extracting short and information-rich feature vectors. These can then be used to build models that are trained, analyzed and finally used for classification of new data. The [labels] are present only in a (semi-)supervised approach.

Factorization (NMF) [43], Singular Value Decomposition [44], Principal Component Analysis (PCA) [45] and Auto-encoders [46]. Linear Discriminant Analysis (LDA) [47] uses the training samples to estimate the between-class and within-class scatter matrices, and then employs the Fisher criterion to obtain the projection matrix for feature extraction (or feature reduction).

In **supervised learning**, the dataset is a collection of example couples of the type (data, label) $\{(x_i, y_i)\}_{i=1..N}$. Each element x_i is called a feature vector and has a companion label y_i. In the supervised learning approach the dataset is used to derive a model that takes a feature vector as input and outputs a label that should describe it. For example, the feature vector of volcano-seismic data could contain several amplitude-based, spectral-based, shape-based or dynamical parameters and the label to be assigned could be one of those described above, i.e., VT, LP, VLP. In a volcanic geochemical example, feature vectors could contain major elements weight percentages, and labels the corresponding rock type. The reliability of the labels is often the most critical issue of the setup of a supervised ML classification scheme. Labels should therefore be assigned carefully by experts. In general, it is much better to have relatively few training events with reliable labels than to have many more, but not so reliable, labeled examples.

In **unsupervised learning**, the dataset is a collection of examples without any labeling, i.e., containing only the data $\{x_i\}_{i=1..N}$. As in the previous case, each x_i is a feature vector, and the goal is to create a model that maps a feature vector x into a value (or another vector) that can help solving a problem. Typical examples are all the clustering procedures, where the output is the cluster number to which each datum belongs. The choice of the best features to use is a difficult one, and several techniques of Unsupervised Feature Selection were proposed, with the capability of identifying and selecting relevant features in unlabeled data [48]. Unsupervised outlier detection methods [49] can also be used, where the output indicates if a given feature vector is likely to describe a "normal" or "anomalous" member of the dataset.

The **semi-supervised learning** approach stands somehow in the middle, and the dataset contains both labeled (usually a few) and unlabeled (usually many more) feature vectors. The basic idea is similar to supervised learning, but with the possibility to exploit also the presence of (many more) unlabeled examples in the training phase.

In **reinforcement learning**, the machine is "embedded" in an environment, which state is again described by a feature vector. In each state the machine can execute actions, which produce different rewards and can cause an environmental state transition. The goal in this case is to learn a policy, i.e., a function or model that takes the feature vector as input and outputs an optimal action to execute in that state. The action is optimal if it maximizes the expected average reward. We can also say that reinforcement learning is a behavioral learning model. The algorithm receives feedback from the data analysis, guiding the user to the best outcome. Here the main point is that the system is not trained with a sample dataset but learns through trial and error. Therefore, a sequence of successful decisions will result in that process being reinforced, because it best solves the problem at hand. Problems that can be tackled with this approach are the ones where decision making is sequential, and the goal is long-term, such as game playing, robotics, resource management, or logistics. Time is therefore explicitly used here, contrary to other approaches, in which in most of the cases data items are analyzed one by one without taking into account the time order in which they arrive.

In some domains (and volcanology is a good example) training data are scarce. In this case we can profit from knowledge acquired in another domain using

techniques known as **Transfer Learning** (TL) [50]. The basic idea here is to train a model in one domain with abundant data (original domain) and then use it as a pretrained model in a different domain (with less data). There is a successive fine-tuning phase using domain-specific available data (in the target domain). This approach was applied for instance at Volcán de Fuego de Colima (Mexico) [51], Mount St. Helens (USA) and Bezymianny (Russia) [52].

Among the computer languages that are most used for implementing ML techniques we can cite Python [53], R [54], Java [55], Javascript [56], Julia [57] and Scala [58]. Many dedicated, open source libraries are available for each of them, and many computer codes, also specialized for volcanic and geophysical data, can be found in open access repositories such as GitHub [59].

3. Machine learning techniques

Extracted feature vectors can become inputs to several different techniques of machine learning. We can cite among others Cluster Analysis (CA) [60], Self-Organizing Maps (SOM) [61–63], Artificial Neural Networks (ANN) and Multi Layer Perceptrons (MLP) [64–66], Support Vector Machines (SVM) [67], Convolutional Neural Networks (CNN) [51], Recurrent Neural Networks (RNN) [68], Hidden Markov Models (HMM) [3, 31, 69–71] and their Parallel System Architecture (PSA) based on Gaussian Mixture Models (GMM) [72].

CA (**Figure 2a**) is an unsupervised learning approach aimed at grouping similar data while separating different ones, where similarity is measured quantitatively using a distance function in the space of feature vectors. The clustering algorithms can be divided into hierarchical and non-hierarchical. In the former a tree-like structure is built to represent the relations between clusters, while in the latter new clusters are formed by merging or splitting existing ones without following a tree-like structure but just grouping the data in order to maximize or minimize some evaluation criteria. CA includes a vast class of algorithms, including e.g., K-means, K-medians, Mean-shift, DBSCAN, Expectation–Maximization (EM), Clustering using Gaussian Mixture Models (GMM), Agglomerative Hierarchical, Affinity Propagation, Spectral Clustering, Ward, Birch, etc. Most of these methods are described and implemented in the open-source Python package scikit-learn [73]. The use of six different unsupervised, clustering-based methods to classify volcano seismic events was explored at Cotopaxi Volcano [32]. One of the most difficult issues is the choice of the number of clusters into which the data should be divided; this number in most of the cases has in fact to be fixed a priori before running the code. Several techniques exist in order to help with this choice, such as elbow, silhouette, gap statistics, heuristics, etc. Many of them are described and included in the R package NbClust [74]. Problems arise when the estimates that each of them provides are contradictory.

Another approach to unsupervised classification is SOM (**Figure 2b**) or Kohonen maps [75, 76], a type of ANN trained to produce a low dimensional, usually 2D, discretized representation of the feature vector space. The training is based on competitive and collaboration learning, using a neighborhood function to preserve the input topological properties.

A very common type of ANN, often used for supervised classification, is MLP, which consists of at least three layers of nodes (**Figure 2c**): an input layer, (at least) one hidden layer and an output layer. Nodes use nonlinear activation functions and are trained through the backpropagation mechanism. If the number of hidden layers of an ANN becomes very high, we talk of Deep Neural Networks (DNN), which are also used mainly in a supervised fashion. Among DNN, the CNN (**Figure 2d**) contain
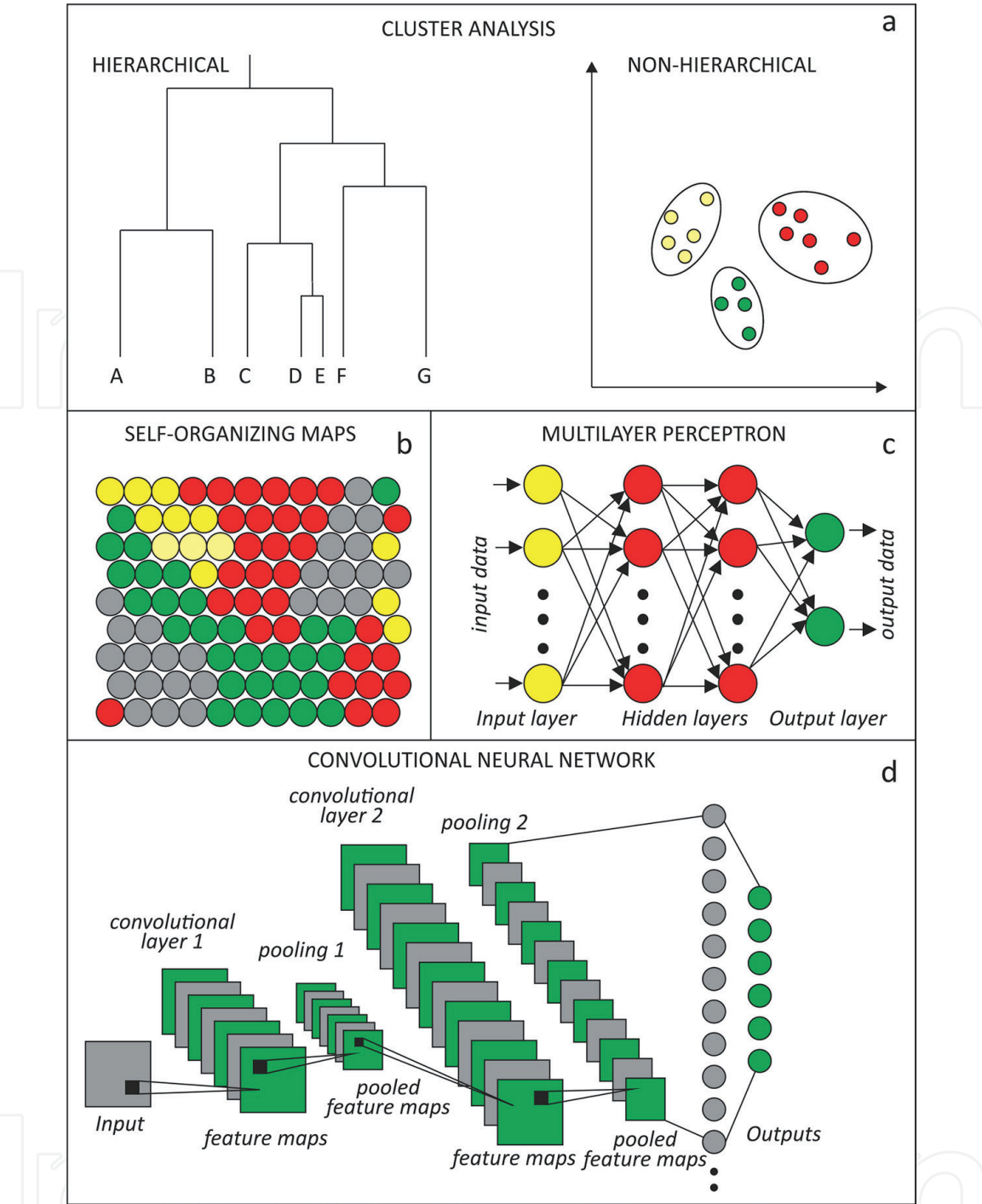
Figure 2.
Schematic illustration of some of the ML techniques described in the text. (a) Cluster analysis in its hierarchical and non-hierarchical versions. (b) Self-organizing maps (c) multilayer perceptron (d) convolutional neural network.

at least some convolutional layers, that convolve their inputs with a multiplication or other dot product. The activation function in the case of CNN is commonly a rectified linear unit (ReLU), and there are also pooling layers, fully connected layers and normalization layers.

A RNN is a type of ANN with a feedback loop (**Figure 3a**), in which neuron outputs can also be used as neuron inputs in the same layer, allowing to maintain some information during the training process. Long Short Term Memory networks (LSTM) are a subset of RNN, capable of learning long-term dependencies [77] and better remember information for long periods of time. RNN can be used for both supervised and unsupervised learning.
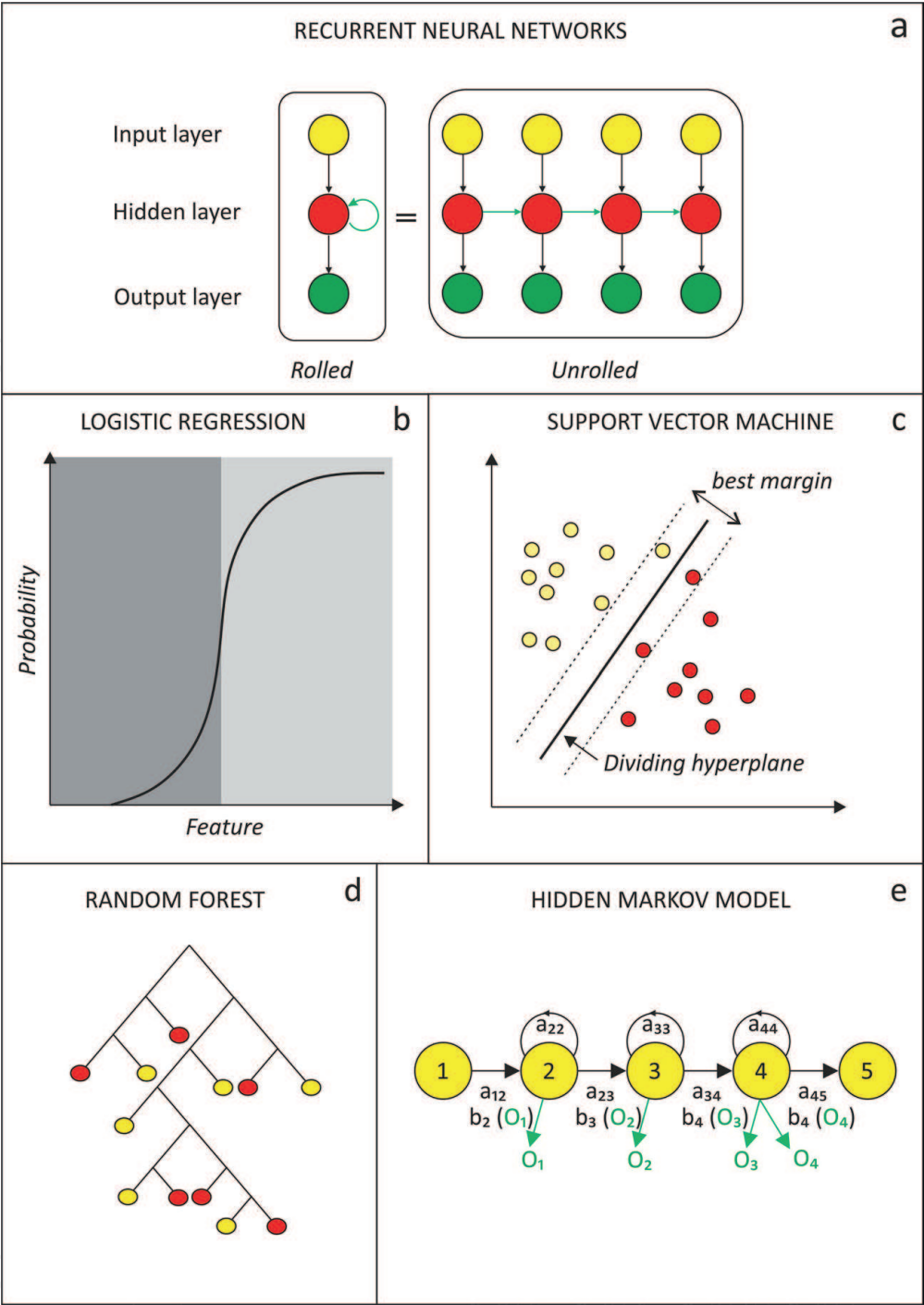
Figure 3.
Schematic illustration of some of the ML techniques described in the text. (a) Recurrent neural network (b) logistic regression (c) support vector machine (d) random forest (e) hidden Markov model.

Logistic regression (LR) (**Figure 3b**) is a supervised generalized linear model, i.e., the classification (probability) dependence on the features is linear [78]. In order to avoid the problems linked to high dimensionality of the data, techniques such as the Least Absolute Shrinkage and Selection Operator (LASSO) can be applied to reduce the number of dimensions of the feature vectors which are input to LR [79].

SVM (**Figure 3c**) constitute a supervised statistical learning framework [80]. It is most commonly used as a non-probabilistic binary classifier. Examples are seen as points in space, and the aim is to separate categories by a gap that is as wide as possible. Unknown samples are then assigned to a category based on the side of the gap on which they fall. In order to perform a non-linear classification, data are mapped into high-dimensional feature spaces using suitable kernel functions.

Sparse Multinomial Logistic Regression (SMLR) is a class of supervised methods for learning sparse classifiers that incorporate weighted sums of basis functions with sparsity-promoting priors encouraging the weight estimates to be either significantly large or exactly zero [81]. The sparsity concept is similar to the one at the base of Non-negative Matrix Factorization (NMF) [82]. The sparsity-promoting priors result in an automatic feature selection, enabling to somehow avoid the so-called "curse of dimensionality". So, sparsity in the kernel basis functions and automatic feature selection can be achieved at the same time [83]. SMLR methods control the capacity of the learned classifier by minimizing the number of basis functions used, resulting in better generalization. There are fast algorithms for SMLR that scale favorably in both the number of training samples and the feature dimensionality, making them applicable even to large data sets in high-dimensional feature spaces.

A Decision Tree (DT) is an acyclic graph. At each branching node, a specific feature x_i is examined. The left or right branch is followed depending on the value of x_i in relation to a given threshold. A class is assigned to each datum when a leaf node is reached. As usual, a DT can be learned from labeled data, using different strategies. In the DT class we can mention Best First Decision Tree (BFT), Functional Tree (FT), J48 Decision Tree (J48DT), Naïve Bayes Tree (NBT) and Reduced Error Pruning Trees (REPT). Ensemble learning techniques such as Random SubSpace (RSS) can be used to combine the results of the different trees [84].

The Boosting concept, a kind of ensemble meta-algorithm mostly (but not only) associated to supervised learning, uses original training data to create iteratively multiple models by using a weak learner. Each model would be different from the previous one as the weak learners try to "fix" the errors made by previous models. An ensemble model will then combine the results of the different weak models. On the other side, Bootstrap aggregating, also called by the contracted name Bagging, consists of creating many "almost-copies" of the training data (each copy is slightly different from the others) and then apply a weak learner to each copy and finally combine the results. A popular and effective algorithm based on bagging is Random Forest (RF). Random Forest (**Figure 3d**) is different from the standard bagging in just one way. At each learning step, a random subset of the features is chosen; this helps to minimize correlation of the trees, as correlated predictors are not efficient in improving classification accuracy. Particular attention has to be taken in order to best choose the number of trees and the size of the random feature subsets.

A Hidden Markov Model (HMM) (**Figure 2e**) is a statistical model in which the system being modeled is assumed to be a Markov process. It describes a sequence of possible events for which the probability of each event depends only on the state occupied in the previous event. The states are unobservable ("hidden") but at each state the Model emits a "message" which depends probabilistically on the current state. Applications are wide in scope, from reinforcement learning to temporal pattern recognition, and the approach works well when time is important; speech [85], handwriting and gesture recognition are then typical fields of applications, but also volcano seismology [69, 86].

4. Applications to seismo-volcanic data

Eruptions are usually preceded by some kind of change in seismicity, making seismic data one of the key dataset in any attempt to forecast volcanic activity [4]. As we mentioned before, manual detection and classification of discrete events can be very time consuming, up to becoming unfeasible during a volcanic crisis.

An automatic classification procedure becomes therefore highly valuable, also as a first step towards forecasting techniques such as material Failure Forecast Method (FFM) [87, 88]. Feature vectors should be built in order to provide most information about the source, minimizing e.g., path and site effects. In many cases features can be independent from a specific physical model describing a phenomenon. This allows ML to work well even when there is no scientific agreement on the generation of a given seismic signal. A good example in volcano seismology is given by the LP events. Standardizing data, making them independent from unwanted variables is also in general a convenient approach [31]. Time-domain and spectral-based amplitudes, spectral phases, auto- and cross-correlations, statistical and dynamical parameters have been considered as the output of data reduction procedures that can be included into feature vectors [14]. In the literature, these have included linear predictor coding for spectrograms [66], wavelet transforms [89], spectral autocorrelation functions [90], statistical and cepstral coefficients [91]. Extracted feature vectors become then the input to one or another ML method.

CA is probably the most used class of unsupervised techniques and the applications to volcano seismology follow this general rule. Spectral clustering was applied e.g., to seismic data of Piton de la Fournaise [60]. The fact that e.g., LP seismic signals can be clustered into families indicates that the family members are very similar to each other. The existence of similar events implies similar location and similar source process, i.e., it means the presence of a source that repeats over time in an almost identical way. Clustering data after some kind of normalization forces CA algorithms to look for similar shapes, independently of size. If significant variations in amplitude are then seen within families, this can indicate that the source processes of these events are not only repeatable but also scalable in size, as observed e.g., at Soufrière Hills Volcano, Montserrat [92] or at Irazú, Costa Rica [93]. The similarity of events in the different classes can then be used to detect other events, e.g., for the purpose of stacking them and obtain more accurate phase arrivals; this was done e.g., at Kanlaon, Philippines [94]. For this purpose, an efficient open-source package is available, called Repeating Earthquake Detector in Python (REDPy) [95].

In volcano-seismology SOM were applied e.g., to Raoul Island, New Zealand [61]. A hierarchical clustering was applied to results of SOM tremor analysis at Ruapehu [62] and Tongariro [96] in New Zealand, using the Scilab environment. A similar combined approach was applied in Matlab to Etna volcanic tremor [97]. Several geometries of SOM were used, with rectangular or hexagonal nearest neighbors cells, planar, toroidal or spherical maps, etc. [61]. The classic ANN/MLP approach was applied e.g., to seismic data recorded at Vesuvius [66], Stromboli [98], Etna [99], while DNN architectures were applied e.g., to Volcán de Fuego, Colima [100]. The use of genetic algorithms for the optimization of the MLP configuration was proposed for the analysis of seismic data of Villarrica, Chile [101]. CNN were applied e.g., to Llaima Volcano (Chile) seismic data, comparing the results to other methods of classification [102]. RNNs were applied, together with other methods, to classify signals of Deception Island Volcano, Antarctica [68]. The architectures were trained with data recorded in 1995–2002 and models were tested on data recorded in 2016–2017, showing good generalization accuracy.

Supervised LR models have been applied in the estimation of landslide susceptibility [103] and to volcano seismic data to estimate the ending date of an eruption at Telica (Nicaragua) and Nevado del Ruiz (Colombia) [104]. SVM were applied many times to volcano seismology e.g., to classify volcanic signals recorded at Llaima, Chile [105] and Ubinas, Peru [106]. Multinomial Logistic Regression was used, together with other methods, to evaluate the feasibility of earthquake prediction using 30 years of historical data in Indonesia, also at volcanoes [107].

RF was applied to the discrimination of rockfalls and VT recorded at Piton de la Fournaise in 2009–2011 and 2014–2015. 60 features were used, and excellent results were obtained. However, a RF trained with 2009–2011 data did not perform well on data recorded in 2014–2015, demonstrating how difficult it is to generalize models even at the same volcano [108]. RF, together with other methods, was recently used on volcano seismic data with the specific purpose to determine when an eruption has ended [104], a problem which is far from being trivial. RF was also used to derive ensemble mean decision tree predictions of sudden steam-driven eruptions at Whakaari (New Zealand) [109].

Most of the methods described so far try to classify discrete seismic events that were already extracted from the continuous stream, i.e., already characterized by a given start and end. There are therefore in general two separated phases: detection and classification [106]. Continuous HMM on the other side are able to process continuous data and can therefore extract and classify in a single, potential real-time, step. HMM are finite-state machines and model sequential patterns where time direction is an essential information. This is typical of (volcano) seismic data. For instance, P waves always arrive before S waves. HMM-based volcanic seismic data classifiers have therefore been used by many authors [87, 110–113]. HMM are also used routinely in some volcano observatories e.g., at Colima and Popocatepetl in Mexico [71]. Etna seismic data was processed by HMM applied to characters generated by the Symbolic Aggregate approXimation (SAX) which maps seismic data into symbols of a given alphabet [114]. HMM can be also combined with standardization procedures such as Empirical Mode Decomposition (EMD) when classifying volcano seismic data [31].

Another characteristic common to many of the applications published in the literature is the fact that feature vectors are extracted from data recorded at a single station. There are relatively few attempts to build multi-station classification schemes. At Piton de la Fournaise a system based on RF was implemented [115]. At the same volcano, a multi-station approach was used to classify tremor measurements and identify fundamental frequencies of the tremor associated to different eruptive behavior [60]. A scalable multi-station, multi-channel classifier, using also the empirical mode decomposition (EMD) first proposed by [31] was applied to Ubinas volcano (Peru). The principal component analysis is used to reduce the dimensionality of the feature vector and a supervised classification is carried out using various methods, with SVM obtaining the best performance [116]. Of course, with a multi-station approach particular care has to be taken in order to build a system which is robust with respect to the loss of one or more seismic stations due to volcanic activity or technical failures.

Open source software and open access papers are luckily becoming more and more common. If we consider the processing and classification of volcano seismic data, several tools are now available for free download and use, especially within the Python environment. Among the most popular, we can cite ObsPy [117] and Msnoise [118], with which researchers and observatories can easily process big quantities of continuous seismic data. Once these tools have produced suitable feature vectors, we can look for open source software to implement the different ML approaches described in this contribution. Many generic ML libraries are available e.g., on GitHub [59] but very few are dedicated specifically to the classification of volcano seismic data. Among these, we can cite the recent package Python Interface for the Classification of Seismic Signals (PICOSS) [119]. It is a graphical, modular open source software for detection, segmentation and classification of seismic data. Modules are independent and adaptable. The classification is currently based on two modules that use Frequency Index analysis [120] or a multi-volcano pre-trained neural network, in a transfer learning fashion [52]. The concept of a multi-volcano

recognizer is also at the core of the EU-funded VULCAN.ears project [31, 121]. The aim is to build an automatic Volcano Seismic Recognition (VSR) system, conceptually supervised (as it is based on HMM) but practically unsupervised, because once it is trained on a number of volcanoes with labeled sample data, it can be used on volcanoes without any previous data in an unsupervised fashion. The idea is in fact to build robust models trained on many datasets recorded by different teams on different volcanoes, and to integrate these models on the routinely used monitoring system of any volcano observatory. Also in this case, the open source software is made freely available; this includes a command interface called PyVERSO [122] based on HTK, a speech recognition HMM toolkit [123], a graphical interface called geoStudio and a script called liveVSR, able to process real-time data downloaded from any online seismic data server [124], together with some pre-trained ML models [125].

As we mentioned before, in order to train supervised models for classifying seismic events, few events with reliable labels are better than many unreliably labeled examples. Just to give a rough idea, 20 labeled events per class is a good starting point, but a minimum of 50 labeled events per class is recommended. Labelling discrete events is enough for many methods, but for approaches like HMM, where the concept is to run the classification on continuous data, it is essential to have a sufficient number of continuously labeled time periods, in order to "show" the classifier enough examples of transition from tremor to a discrete event, and then back to tremor. It is important to have many examples also of "garbage" events, i.e., events we are not interested in, so that the classifier can recognize and discard them. Finally, it is advisable to have a wide variability of events within each given class rather than having many very similar events. There is not yet an agreement on a single file format to store these labels. As speech recognition is much older and more developed than seismic recognition, it is suggested to adopt standard labelling formats of that domain, i.e., the transcription MLF files, which are normal text files that include for each event the start time, the end time and of course the label. These files can be created manually with a simple text editor, or by using a program with a GUI, such as geoStudio [124] or Seismo_volcanalysis [126]. Other graphical software packages like SWARM [127] use other formats to store the labels, such as CSV, but it is always possible to build scripts that convert the resulting label files into MLF format, which remains the recommended one.

5. Applications of machine learning to geochemical data

ML applications to geochemical data of volcanoes are increasing in the last years, although most of them are limited to the use of cluster analysis. CA has been used for example to identify and quantify mixing processes using the chemistry of minerals [128], also for the study of volcanic aquifers [129, 130] or to differentiate magmatic systems e.g., [131]. Platforms used to carry out these analyses include the Statistical Toolbox in Matlab [132], or the R platform [54]; some geochemical software made in this last platform include the CA as the GCDkit [33]. In most ML analyses on geochemical samples it is common to use whole rock major elements and selected trace elements; some applications also include isotopic ratios. Many ML applications to geochemical data use more than one technique, frequently combining both unsupervised and supervised approaches.

A combination of SVM, RF and SMLR approaches were used by [37] to account for variations of geochemical composition of rocks from eight different tectonic settings. The authors note that SVM used to discriminate tectonic settings as used by [34] is a powerful tool. The RF approach is shown to have the advantage, with

respect to SVM, of providing the importance of each feature during discrimination. The weakness of applying the RF for tectonic setting discrimination is that the evaluation based only on a majority vote of multiple decision trees often makes the obtained quantitative geochemical interpretation of these elements and isotopic ratios difficult. The authors suggest that the best quantitative discriminant is that of SMLR, as it allows to assign to each sample a probability of belonging to a given group (tectonic setting in this case), with still the possibility of identifying the importance of each feature. This tool is a notable step forward in the discrimination of the geochemical signature of the different tectonic settings, which is commonly assessed based on binary or ternary diagrams e.g., [133, 134] which are useful with many samples but are not able to differentiate a tectonic setting where a complex evolution of magmas has occurred. In the last decade multielement variation diagrams were proposed e.g., [135] and also the use of Decision Trees e.g., [136] or LDA e.g., [137] to accurately assign a tectonic setting based on rock geochemistry. Based on rock sample geochemistry, [37] show that a set of 17 elements and isotopic ratios is needed to clearly identify the tectonic setting. Two new discriminant functions were recently proposed to discriminate the tectonic settings of mid-ocean ridge (MOR) and oceanic plateau (OP). 10 datasets (original concentrations as well as isometric log-ratio transformed variables; all 10 major elements as well as all 10 major and 6 trace elements) were used to evaluate the quality of discrimination from LDA and canonical analysis [138].

The software package Compositional Data Package (CoDaPack) [139] and a combination of unsupervised (CA) and supervised (LDA) learning approaches was used by [36] to identify compositional variation of ignimbrite magmas in the Central Andes, trying to use these methods as a tool for ignimbrite correlation. They have used the Statistica software [140] for both CA and LDA.

Correlating tephra and identifying their volcanic sources is a very difficult task, especially in areas where several volcanoes had explosive eruptions in a relatively short period of time. This is particularly challenging when volcanoes have similar geochemical and petrographic compositions. Electron microprobe analysis of glass compositions and whole-rock geochemical analyses are used frequently to make these correlations. However, correlations may not be so accurate when using only geochemical tools that may mask diagnostic variability; sometimes one of the most important advantages of ML in this regard is the speed at which correlations can be made, rather than the accuracy [35]. Other contributions however demonstrate how ML techniques can make these correlations also accurate. Some highly accurate results of ML techniques applied to tephra correlation include those of LDA [141, 142] and SVM e.g., [143]; however, SVM may fail in specific cases and for the case study of tephra from Alaska volcanoes, the combination of ANN and RF are the best ML techniques to apply [35]. The authors use the R software [54] to apply these methods, and they underline the advantage of producing probabilistic outputs.

SOM was used as an unsupervised neural network approach to analyze geochemical data of Ischia, Vesuvius and Campi Flegrei [144]. The advantage of this method is that there is no need of previous knowledge of geochemical or petrological characteristics and that it allows the use of large databases with large number of variables. The SOM toolbox for Matlab [132] was used by [144] to perform two tests, the first based on major elements and selected trace elements to find similar evolution processes, the second to investigate the magmatic source, so a vector containing a selection of ratios between major and trace elements was adopted. One of the enhancements of this method is that the resulting clusters permitted to differentiate rock samples that were only comparably distinguished by 2D diagrams of isotopic ratios; in other words, similar results were obtained with the limited availability of less expensive geochemical data.

One of the applications of ML techniques that maybe extremely useful in geochemistry is the apparent possibility of predicting the concentration of unknown elements if a large number of data of other elements is known. A combination of ML techniques was used by [38] to predict Rare Earth Elements (REE) concentrations on Ocean Island Basalts (OIB) using RF. They used 1283 analyses of which 80% were used for training and the remaining 20% to validate the results. They found good estimations only in the Light Rare Earth Elements (LREE), suggesting that the results may be improved by using a larger set of input data for training. One possible solution may be the use of not only major elements for training but also of other trace elements obtained through the same analytical method of major elements.

The origin of the volcanoes in Northeast China, analyzed by RF and DNN using the full chemical compositional data, was associated to the Pacific slab, subducting at Japan, reaching ~600-km depth under eastern China, and extending horizontally up to Mongolia. The boundary between volcanoes triggered by fluids and melts from the slab and those not related to it was located at the westernmost edge of the deeply buried Pacific slab [145].

As highlighted by [143] ML methods require the integration with other techniques such as fieldwork, petrographic observations and classic geochemical studies to obtain a clearer picture of the investigated problem. While in other fields, it is relatively easy (and cheap) to acquire big amounts of data (hundreds or more), this is not the case for geochemistry. However, we underline that the application of ML techniques to the geochemistry of volcanic rocks does need a minimum dataset size. In the literature a set of 250 analyses is described as sufficiently large amount of data but, as usual, one can try using the available data (often even less than 50) but thousands of examples would definitely improve the results.

6. Applications of machine learning to other volcanological data

ML appears more and more often in volcanology literature, and specific fields of application span now also other sub-disciplines.

Mount Erebus in Antarctica has a persistent lava lake showing Strombolian activity, but its location is definitely remote. Therefore, automatic methods to detect these explosions are highly needed. A CNN was trained using infrared images captured from the crater rim and "labeled" with the help of accompanying seismic data, which was not used anymore during the subsequent automatic detection [146].

Clast morphology is a fundamental tool also for studies concerning volcanic textures. Texture analysis of clasts provides in particular information about genesis, transport and depositional processes. Here, ML has still to be developed fully but e.g., the application of preprocessing techniques such as the Radon transform can be a first step towards an efficient definition of feature vectors to be used for classification, as shown e.g., at Colima volcano [147].

The Museum of Mineralogy, Petrography and Volcanology of the University of Catania implemented a communication system based on the visitor's personal experience to learn by playing. There is a web application called I-PETER: Interactive Platform to Experience Tours and Education on the Rocks. This platform includes a labeled dataset of images of rocks and minerals to be used also for petrological investigations based on ML [148].

Satellite remote sensing technology is increasingly used for monitoring the surface of the Earth in general, and volcanoes in particular, especially in areas where ground monitoring is scarce or completely missing. For instance, in Latin America

202 out of 319 Holocene volcanoes did not have seismic, deformation or gas monitoring in 2013 [7]. A complex-valued CNN was proposed to extract areas with land shapes similar to given samples in interferometric synthetic aperture radar (InSAR), a technique widely applied in volcano monitoring. An application was presented grouping similar small volcanoes in Japan [149]. InSAR measurements have great potential for volcano monitoring, especially where images are freely available. ML methods can be used for the initial processing of single satellite data. Processing of potential unrest areas can then fully exploit integrated multi-disciplinary, multi-satellite datasets [7]. The Copernicus Programme of the European Space Agency (ESA) and the European Union (EU) has recently contributed by producing the Sentinel-2 multispectral satellites, able to provide high resolution satellite data for disaster monitoring, as well as complementing previous satellite images like Landsat. The free access policy also promotes an increasing use of Sentinel-2 data, which is often processed by ML techniques such as SVM and RF [150]. A transfer learning strategy was applied to ground deformation in Sentinel-1 data [151] and a range of pretrained networks was tested, finding that AlexNet [152] is best suited to this task. The positive results were checked by a researcher and fed back for model updating.

The global volcano monitoring platform MOUNTS (Monitoring Unrest from Space) uses multisensor satellite-based imagery (Sentinel-1 Synthetic Aperture Radar SAR, Sentinel-2 Short-Wave InfraRed SWIR, Sentinel-5P TROPOMI), ground-based seismic data (GEOFON and USGS global earthquake catalogs), and CNN to provide support for volcanic risk assessment. Results are visualized on an open-access website. The efficiency of the system was tested on several eruptions (Erta Ale 2017, Fuego 2018, Kilauea 2018, Anak Krakatau 2018, Ambrym 2018, and Piton de la Fournaise 2018–2019) [153].

Debris flow events are one of the most widespread and dangerous natural processes not only on volcanoes but more in general in mountainous environments. A methodology was recently proposed [154] that combines the results of deterministic and heuristic/probabilistic models for susceptibility assessment. RF models are extensively used to represent the heuristic/probabilistic component of the modeling. The case study presented is given by the Changbai Shan volcano, China [154].

Mapping lava flows from satellite is another important remote sensing application. RF was applied to 20 individual flows and 8 groups of flows of similar age using a Landsat 8 image and a DEM of Nyamuragira (Congo) with 30 m resolution. Despite spectral similarity, lava flows of contrasting age can be well discriminated and mapped by means of image classification [155].

The hazard related to landslides at volcanoes is also significant. DNN models were proposed for landslide susceptibility assessment in Viet Nam, showing considerable better performance with respect to other ML methods such as MLP, SVM, DT and RF [156]. The use of DNN approach could be therefore an interesting approach for the landslide susceptibility mapping of active volcanoes.

Muon imaging has been successfully used by geophysicists to investigate the internal structure of volcanoes, for example at Etna (Italy) [157]. Muon imaging is essentially an inverse problem and it can profit from the application of ML techniques, such as ANN and CA [158].

Combinations of supervised and unsupervised ML techniques have been used to map volcanoes also on other planets. A ML paradigm was designed for the identification of volcanoes on Venus [159]. Other studies have used topographic data, such as DEM and associated derivatives obtained from orbital images, to detect and classify manually labeled Martian landforms including volcanoes [160].

7. Conclusions

ML techniques will have an increasing impact on how we study and model volcanoes in all their aspects, how we monitor them and how we evaluate their hazards, both in the short and in the long term. The increasing number of monitoring equipment installed on volcanoes on one side provides more and more data, on the other often causes their real time processing unfeasible especially when most needed i.e., during unrest and eruptions. Here ML will show its best usefulness, as it can provide the perfect tools to sift through big data to identify subtle patterns that could indicate unrest, hopefully well before eruptions. One important issue is the one of generalization. We must go towards the construction of ML models that can be applied on different volcanoes, for instance when previous data is not available for training specific models. The concepts of transfer learning can be important here.

The routine use of ML tools at the different volcano observatories should be promoted by providing easy installation procedures and easy integration into existing monitoring systems. Open source software should be always chosen whenever possible. On the other hand, observatories should provide good open training data to ML developers, researchers and data scientists in order to improve the models in a virtuous circle. An easy availability of open access data, both from the ground and from satellites should be exploited for building reliable training sets in the different fields of volcanology. This will allow "scientific competition" between research groups using different ML approaches and make a direct comparison of results easier, like it is common in other disciplines where "standard" training datasets are available for download to everybody.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Author details

Roberto Carniel[1*] and Silvina Raquel Guzmán[2]

1 DPIA, University of Udine, Udine, Friuli, Italy

2 IBIGEO, UNSa-CONICET, Rosario de Lerma, Salta, Argentina

*Address all correspondence to: roberto.carniel@uniud.it

IntechOpen

References

[1] O. Jaquet, R. Carniel, S. Sparks, G. Thompson, R. Namar, and M. Di Cecca, "DEVIN: A forecasting approach using stochastic methods applied to the Soufrière Hills Volcano," *J. Volcanol. Geotherm. Res.*, vol. 153, no. 1-2 SPEC. ISS., pp. 97-111, May 2006.

[2] O. Jaquet and R. Carniel, "Multivariate stochastic modelling: Towards forecasts of paroxysmal phases at Stromboli," *J. Volcanol. Geotherm. Res.*, vol. 128, no. 1-3, pp. 261-271, Nov. 2003.

[3] W. P. Aspinall, R. Carniel, O. Jaquet, G. Woo, and T. Hincks, "Using hidden multi-state Markov models with multi-parameter volcanic data to provide empirical evidence for alert level decision-support," *J. Volcanol. Geotherm. Res.*, vol. 153, no. 1-2 SPEC. ISS., pp. 112-124, 2006.

[4] R. Ortiz, A. García, J. M. Marrero, S. la Cruz-Reyna, R. Carniel, and J. Vila, "Volcanic and volcano-tectonic activity forecasting: A review on seismic approaches," *Ann. Geophys.*, vol. 62, no. 1, 2019.

[5] G. Currenti, C. del Negro, V. Lapenna, and L. Telesca, "Multifractality in local geomagnetic field at Etna volcano, Sicily (southern Italy)," *Nat. Hazards Earth Syst. Sci.*, vol. 5, no. 4, pp. 555-559, 2005.

[6] R. M. Green, M. S. Bebbington, S. J. Cronin, and G. Jones, "Geochemical precursors for eruption repose length," *Geophys. J. Int.*, vol. 193, no. 2, pp. 855-873, 2013.

[7] S. Ebmeier *et al.*, "Satellite geodesy for volcano monitoring in the Sentinel-1 and SAR constellation era," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 5465-5467.

[8] E. Marchetti *et al.*, "Long range infrasound monitoring of Etna volcano," *Sci. Rep.*, vol. 9, no. 1, 2019.

[9] A. Aiuppa *et al.*, "Unusually large magmatic CO2 gas emissions prior to a basaltic paroxysm," *Geophys. Res. Lett.*, 2010.

[10] F. Marchese, N. Pergola, and L. Telesca, "Investigating the temporal fluctuations in satellite advanced very high resolution radiometer thermal signals measured in the volcanic area of Etna (Italy)," *Fluct. Noise Lett.*, vol. 6, no. 3, pp. L305-L316, 2006.

[11] A. J. L. L. Harris, R. Carniel, and J. Jones, "Identification of variable convective regimes at Erta Ale Lava Lake," *J. Volcanol. Geotherm. Res.*, vol. 142, no. 3-4, pp. 207-223, Apr. 2005.

[12] Surono *et al.*, "The 2010 explosive eruption of Java's Merapi volcano-A '100-year' event," *J. Volcanol. Geotherm. Res.*, vol. 241-242, pp. 121-135, 2012.

[13] R. Carniel, E. M. Jolis, and J. Jones, "A geophysical multi-parametric analysis of hydrothermal activity at Dallol, Ethiopia," *J. African Earth Sci.*, vol. 58, no. 5, pp. 812-819, Dec. 2010.

[14] R. Carniel, "Characterization of volcanic regimes and identification of significant transitions using geophysical data: A review," *Bull. Volcanol.*, vol. 76, no. 8, pp. 1-22, Jul. 2014.

[15] M. Tárraga, J. Martí, R. Abella, R. Carniel, and C. López, "Volcanic tremors: Good indicators of change in plumbing systems during volcanic eruptions," *J. Volcanol. Geotherm. Res.*, vol. 273, pp. 33-40, 2014.

[16] R. Carniel and M. Tárraga, "Can tectonic events change volcanic tremor at Stromboli?," *Geophys. Res. Lett.*, vol. 33, no. 20, 2006.

[17] S. Dumont *et al.*, "The dynamics of a long-lasting effusive eruption modulated by Earth tides," *Earth Planet. Sci. Lett.*, 2020.

[18] G. Tamburello *et al.*, "Periodic volcanic degassing behavior: The Mount Etna example," *Geophys. Res. Lett.*, 2013.

[19] V. M. Zobin, *Introduction to Volcanic Seismology: Third Edition*. 2016.

[20] S. R. McNutt, "Volcano seismology and monitoring for eruptions," *Int. Handb. Earthq. Eng. Seismol.*, pp. 383-406, 2002.

[21] B. Chouet, "Volcano Seismology," *Pure Appl. Geophys.*, vol. 160, no. 3, pp. 739-788, 2003.

[22] I. Molina, H. Kumagai, and H. Yepes, "Resonances of a volcanic conduit triggered by repetitive injections of an ash-laden gas," *Geophys. Res. Lett.*, 2004.

[23] R. M. Iverson, "Dynamics of seismogenic volcanic extrusion resisted by a solid surface plug, Mount St. Helens, 2004-2005," *US Geol. Surv. Prof. Pap.*, no. 1750, pp. 425-460, 2008.

[24] B. R. Julian, "Volcanic tremor: nonlinear excitation by fluid flow," *J. Geophys. Res.*, vol. 99, no. B6, pp. 11, 811-859, 877, 1994.

[25] M. Hellweg, "Physical models for the source of Lascar's harmonic tremor," *J. Volcanol. Geotherm. Res.*, vol. 101, no. 1-2, pp. 183-198, 2000.

[26] J. B. Johnson, J. M. Lees, A. Gerst, D. Sahagian, and N. Varley, "Long-period earthquakes and co-eruptive dome inflation seen with particle image velocimetry," *Nature*, vol. 456, no. 7220, pp. 377-381, 2008.

[27] C. J. Bean, L. De Barros, I. Lokmer, J.-P. Métaxian, G. O'Brien, and S. Murphy, "Long-period seismicity in the shallow volcanic edifice formed from slow-rupture earthquakes," *Nat. Geosci.*, vol. 7, no. 1, pp. 71-75, 2014.

[28] E. Marchetti and M. Ripepe, "Stability of the seismic source during effusive and explosive activity at Stromboli Volcano," *Geophys. Res. Lett.*, vol. 32, no. 3, pp. 1-5, 2005.

[29] R. Carniel, "Comments on the paper 'Automatic detection and discrimination of volcanic tremors and tectonic earthquakes: An application to Ambrym volcano, Vanuatu' by Daniel Rouland, Denis Legrand, Mikhail Zhizhin and Sylvie Vergniolle [J. Volcanol. Geotherm. Res. 181," *J. Volcanol. Geotherm. Res.*, vol. 194, no. 1-3, pp. 61-62, Jul. 2010.

[30] A. Jolly, C. Caudron, T. Girona, B. Christenson, and R. Carniel, "'Silent' Dome Emplacement into a Wet Volcano: Observations from an Effusive Eruption at White Island (Whakaari), New Zealand in Late 2012," *Geosci.*, vol. 10, no. 4, pp. 1-13, 2020.

[31] G. Cortés, R. Carniel, M. A. Mendoza, and P. Lesage, "Standardization of Noisy Volcanoseismic Waveforms as a Key Step toward Station-Independent, Robust Automatic Recognition," *Seismol. Res. Lett.*, vol. 90, no. 2 A, pp. 581-590, 2019.

[32] A. Duque *et al.*, "Exploring the unsupervised classification of seismic events of Cotopaxi volcano," *J. Volcanol. Geotherm. Res.*, p. 107009, 2020.

[33] V. Janoušek, C. M. Farrow, and V. Erban, "Interpretation of whole-rock geochemical data in igneous geochemistry: Introducing Geochemical Data Toolkit (GCDkit)," *J. Petrol.*, vol. 47, no. 6, pp. 1255-1259, 2006.

[34] M. Petrelli and D. Perugini, "Solving petrological problems through machine learning: the study case of tectonic discrimination using geochemical and isotopic data," *Contrib. to Mineral. Petrol.*, 2016.

[35] M. S. M. Bolton *et al.*, "Machine learning classifiers for attributing tephra to source volcanoes: an evaluation of methods for Alaska tephras," *J. Quat. Sci.*, 2020.

[36] M. Brandmeier and G. Wörner, "Compositional variations of ignimbrite magmas in the Central Andes over the past 26 Ma — A multivariate statistical perspective," *Lithos*, 2016.

[37] K. Ueki, H. Hino, and T. Kuwatani, "Geochemical discrimination and characteristics of magmatic tectonic settings: A machine-learning-based approach," *Geochemistry, Geophys. Geosystems*, 2018.

[38] J. Hong, C. Gan, and J. Liu, "Prediction of REEs in OIB by major elements based on machine learning," *Earth Sci. Front.*, 2019.

[39] A. Burkov, *The Hundred-Page Machine Learning Book*. Andriy Burkov (January 13, 2019), 2019.

[40] T. M. Mitchell, "The Discipline of Machine Learning," *Mach. Learn.*, 2006.

[41] G. Cortés, M. Carmen Benitez, L. Garcia, I. Alvarez, and J. M. Ibanez, "A Comparative Study of Dimensionality Reduction Algorithms Applied to Volcano-Seismic Signals," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2016.

[42] G. Cabras, R. Carniel, and J. Wasserman, "Signal enhancement with generalized ICA applied to Mt. Etna Volcano, Italy," *Boll. di Geofis. Teor. ed Appl.*, 2010.

[43] R. Carniel, G. Cabras, M. Ichihara, and M. Takeo, "Filtering wind in infrasound data by non-negative matrix factorization," *Seismol. Res. Lett.*, vol. 85, no. 5, pp. 1056-1062, 2014.

[44] R. Carniel, F. Barazza, M. Tárraga, and R. Ortiz, "On the singular values decoupling in the Singular Spectrum

Analysis of volcanic tremor at Stromboli," *Nat. Hazards Earth Syst. Sci.*, vol. 6, no. 6, pp. 903-909 ST-On the singular values decoupling in, 2006.

[45] A. Tharwat, "Principal component analysis - a tutorial," *Int. J. Appl. Pattern Recognit.*, 2016.

[46] J. Guo, H. Li, J. Ning, W. Han, W. Zhang, and Z. S. Zhou, "Feature dimension reduction using stacked sparse auto-encoders for crop classification with multi-temporal, quad-pol SAR Data," *Remote Sens.*, 2020.

[47] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AI Commun.*, 2017.

[48] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "Ranking based unsupervised feature selection methods: An empirical comparative study in high dimensional datasets," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.

[49] P. Caroline Cynthia and S. Thomas George, "An Outlier Detection Approach on Credit Card Fraud Detection Using Machine Learning: A Comparative Analysis on Supervised and Unsupervised Learning," in *Intelligence in Big Data Technologies---Beyond the Hype*, 2021, pp. 125-135.

[50] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345-1359, 2010.

[51] M. Titos, A. Bueno, L. García, C. Benítez, and J. C. Segura, "Classification of Isolated Volcano-Seismic Events Based on Inductive Transfer Learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 869-873, 2020.

[52] A. Bueno, C. Benitez, S. De Angelis, A. Diaz Moreno, and J. M. Ibanez,

"Volcano-Seismic Transfer Learning and Uncertainty Quantification with Bayesian Neural Networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 892-902, 2020.

[53] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[54] R Core Team, "R: A Language and Environment for Statistical Computing." Vienna, Austria, 2013.

[55] K. Arnold, J. Gosling, and D. Holmes, *The Java programming language*. Addison Wesley Professional, 2005.

[56] D. Flanagan, *JavaScript: the definitive guide*. " O'Reilly Media, Inc.," 2006.

[57] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM Rev.*, 2017.

[58] M. Odersky, L. Spoon, and B. Venners, *Programming in scala*. Artima Inc, 2008.

[59] github, "GitHub." 2020.

[60] C. X. Ren, A. Peltier, V. Ferrazzini, B. Rouet-Leduc, P. A. Johnson, and F. Brenguier, "Machine Learning Reveals the Seismic Signature of Eruptive Behavior at Piton de la Fournaise Volcano," *Geophys. Res. Lett.*, vol. 47, no. 3, p. e2019GL085523, 2020.

[61] R. Carniel, L. Barbui, and A. D. Jolly, "Detecting dynamical regimes by Self-Organizing Map (SOM) analysis: An example from the March 2006 phreatic eruption at Raoul Island, New Zealand Kermadec Arc," *Boll. di Geofis. Teor. ed Appl.*, 2013.

[62] R. Carniel, A. D. Jolly, and L. Barbui, "Analysis of phreatic events at Ruapehu volcano, New Zealand using

a new SOM approach," *J. Volcanol. Geotherm. Res.*, 2013.

[63] A. Köhler, M. Ohrnberger, and F. Scherbaum, "Unsupervised pattern recognition in continuous seismic wavefield records using Self-Organizing Maps," *Geophys. J. Int.*, 2010.

[64] R. Carniel, "Neural networks and dynamical system techniques for volcanic tremor analysis," *Ann. di Geofis.*, vol. 39, no. 2, pp. 241-252, 1996.

[65] A. M. Esposito, L. D'Auria, F. Giudicepietro, T. Caputo, and *M. martini*, "Neural analysis of seismic data: Applications to the monitoring of Mt. Vesuvius," *Ann. Geophys.*, vol. 56, no. 4, 2013.

[66] S. Scarpetta *et al.*, "Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks," *Bull. Seismol. Soc. Am.*, vol. 95, no. 1, pp. 185-196, 2005.

[67] M. Masotti, S. Falsaperla, H. Langer, S. Spampinato, and R. Campanini, "Application of Support Vector Machine to the classification of volcanic tremor at Etna, Italy," *Geophys. Res. Lett.*, vol. 33, no. 20, 2006.

[68] M. Titos, A. Bueno, L. García, M. C. Benítez, and J. Ibañez, "Detection and Classification of Continuous Volcano-Seismic Signals with Recurrent Neural Networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 1936-1948, 2019.

[69] M. Beyreuther, R. Carniel, and J. Wassermann, "Continuous Hidden Markov Models: Application to automatic earthquake detection and classification at Las Canãdas caldera, Tenerife," *J. Volcanol. Geotherm. Res.*, 2008.

[70] M. Beyreuther and J. Wassermann, "Hidden semi-Markov Model based earthquake classification system using

Weighted Finite-State Transducers," *Nonlinear Process. Geophys.*, vol. 18, no. 1, pp. 81-89, 2011.

[71] G. Cortés *et al.*, "Evaluating robustness of a HMM-based classification system of volcano-seismic events at COLIMA and Popocatepetl volcanoes," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2009, vol. 2, pp. II1012-II1015.

[72] G. Cortés, L. García, I. Álvarez, C. Benítez, Á. de la Torre, and J. Ibáñez, "Parallel System Architecture (PSA): An efficient approach for automatic recognition of volcano-seismic events," *J. Volcanol. Geotherm. Res.*, 2014.

[73] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in {P}ython," *J. Mach. Learn. Res.*, vol. 12, pp. 2825-2830, 2011.

[74] M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, "Nbclust: An R package for determining the relevant number of clusters in a data set," *J. Stat. Softw.*, 2014.

[75] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, 1982.

[76] T. Kohonen, *Self-organizing maps*, 3rd ed. Berlin: Springer, 2001.

[77] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, 1997.

[78] P. McCullagh and J. A. Nelder, *Generalized Linear Models, Second Edition (Monographs on Statistics and Applied Probability)*. 1989.

[79] T. Hastie, R. Tibshirani, and J. Friedman, *Elements of Statistical Learning 2nd ed.* 2009.

[80] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, 1995.

[81] B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005.

[82] G. Cabras, R. Carniel, and J. Jones, "Non-negative Matrix Factorization: An application to Erta 'Ale volcano, Ethiopia," *Boll. di Geofis. Teor. ed Appl.*, vol. 53, no. 2, pp. 231-242, 2012.

[83] B. Krishnapuram, L. Carin, and A. J. Hartemink, "Joint classifier and feature optimization for cancer diagnosis using gene expression data," in *Proceedings of the Annual International Conference on Computational Molecular Biology, RECOMB*, 2003.

[84] B. T. Pham *et al.*, "Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers," *Geocarto Int.*, 2020.

[85] L. R. Rabiner and R. W. Schafer, "Introduction to digital speech processing," *Found. Trends Signal Process.*, 2007.

[86] P. Alasonati, J. Wassermann, and M. Ohrnberger, "Signal classification by wavelet-based hidden Markov models: application to seismic signals of volcanic origin," in *Statistics in Volcanology*, 2018.

[87] A. Boué, P. Lesage, G. Cortés, B. Valette, and G. Reyes-Dávila, "Real-time eruption forecasting using the material Failure Forecast Method with a Bayesian approach," *J. Geophys. Res. Solid Earth*, vol. 120, no. 4, pp. 2143-2161, 2015.

[88] M. Tárraga, R. Carniel, R. Ortiz, and A. García, "Chapter 13 The Failure Forecast Method: Review and

Application for the Real-Time Detection of Precursory Patterns at Reawakening Volcanoes," *Dev. Volcanol.*, vol. 10, no. C, pp. 447-469, 2008.

[89] J. P. Jones, R. Carniel, and S. D. Malone, "Subband decomposition and reconstruction of continuous volcanic tremor," *J. Volcanol. Geotherm. Res.*, vol. 213-214, pp. 98-115, 2012.

[90] H. Langer, S. Falsaperla, T. Powell, and G. Thompson, "Automatic classification and a-posteriori analysis of seismic event identification at Soufrière Hills volcano, Montserrat," *J. Volcanol. Geotherm. Res.*, vol. 153, no. 1-2 SPEC. ISS., pp. 1-10, 2006.

[91] J. M. Ibáñez, C. Benítez, L. A. Gutiérrez, G. Cortés, A. García-Yeguas, and G. Alguacil, "The classification of seismo-volcanic signals using Hidden Markov Models as applied to the Stromboli and Etna volcanoes," *J. Volcanol. Geotherm. Res.*, 2009.

[92] D. N. Green and J. Neuberg, "Waveform classification of volcanic low-frequency earthquake swarms and its implication at Soufrière Hills Volcano, Montserrat," *J. Volcanol. Geotherm. Res.*, vol. 153, no. 1-2 SPEC. ISS., pp. 51-63, 2006.

[93] R. Villegas, R. Carniel, I. Petrinovic, and C. Balbis, "Clusters of long-period (LP) seismic events at the Irazú Volcano: what are they telling us?," *J. South Am. Earth Sci.*, no. under final revision, 2020.

[94] W. I. Sevilla, L. A. Jumawan, C. J. Clarito, M. A. Quintia, A. A. Dominguiano, and R. U. Solidum, "Improved 1D velocity model and deep long-period earthquakes in Kanlaon Volcano, Philippines: Implications for its magmatic system," *J. Volcanol. Geotherm. Res.*, 2020.

[95] A. J. Hotovec-Ellis and C. Jeffries, "Near real-time detection, clustering, and analysis of repeating earthquakes Application to Mount St. Helens and Redoubt volcanoes," in *Seismological Society of America Annual Meeting*, 2016.

[96] A. D. Jolly *et al.*, "Seismo-acoustic evidence for an avalanche driven phreatic eruption through a beheaded hydrothermal system: An example from the 2012 Tongariro eruption," *J. Volcanol. Geotherm. Res.*, vol. 286, pp. 331-347, 2014.

[97] A. Messina and H. Langer, "Pattern recognition of volcanic tremor data on Mt. Etna (Italy) with KKAnalysis-A software program for unsupervised classification," *Comput. Geosci.*, vol. 37, no. 7, pp. 953-961, 2011.

[98] S. Falsaperla, S. Graziani, G. Nunnari, and S. Spampinato, "Automatic classification of volcanic earthquakes by using multi-layered neural networks," *Nat. Hazards*, 1996.

[99] H. Langer, S. Falsaperla, M. Masotti, R. Campanini, S. Spampinato, and A. Messina, "Synopsis of supervised and unsupervised pattern classification techniques applied to volcanic tremor data at Mt Etna, Italy," *Geophys. J. Int.*, 2009.

[100] M. Titos, A. Bueno, L. Garcia, and C. Benitez, "A Deep Neural Networks Approach to Automatic Recognition Systems for Volcano-Seismic Events," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2018.

[101] G. Curilem, J. Vergara, G. Fuentealba, G. Acuña, and M. Chacón, "Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms," *J. Volcanol. Geotherm. Res.*, vol. 180, no. 1, pp. 1-8, 2009.

[102] J. P. Canário *et al.*, "In-depth comparison of deep artificial neural network architectures on seismic events

classification," *J. Volcanol. Geotherm. Res.*, vol. 401, 2020.

[103] B. Pradhan and S. Lee, "Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling," *Environ. Model. Softw.*, 2010.

[104] G. F. Manley *et al.*, "Understanding the timing of eruption end using a machine learning approach to classification of seismic time series," *J. Volcanol. Geotherm. Res.*, 2020.

[105] M. Curilem *et al.*, "Pattern recognition applied to seismic signals of Llaima volcano (Chile): An evaluation of station-dependent classifiers," *J. Volcanol. Geotherm. Res.*, vol. 315, pp. 15-27, 2016.

[106] M. Malfante, M. Dalla Mura, J. P. Metaxian, J. I. Mars, O. Macedo, and A. Inza, "Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives," *IEEE Signal Process. Mag.*, 2018.

[107] I. M. Murwantara, P. Yugopuspito, and R. Hermawan, "Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data," *Telkomnika (Telecommunication Comput. Electron. Control.*, 2020.

[108] C. Hibert, F. Provost, J. P. Malet, A. Maggi, A. Stumpf, and V. Ferrazzini, "Automatic identification of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise volcano using a Random Forest algorithm," *J. Volcanol. Geotherm. Res.*, 2017.

[109] D. E. Dempsey, S. J. Cronin, S. Mei, and A. W. Kempa-Liehr, "Automatic precursor recognition and real-time forecasting of sudden explosive volcanic eruptions at Whakaari, New Zealand," *Nat. Commun.*, 2020.

[110] M. C. Benítez *et al.*, "Continuous HMM-based seismic-event classification at deception Island, Antarctica," in *IEEE Transactions on Geoscience and Remote Sensing*, 2007.

[111] M. Bicego, C. Acosta-Munoz, and M. Orozco-Alzate, "Classification of seismic volcanic signals using hidden-markov-model-based generative embeddings," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3400-3409, 2013.

[112] P. B. Dawson, M. C. Benítez, B. A. Chouet, D. Wilson, and P. G. Okubo, "Monitoring very-long-period seismicity at Kilauea Volcano, Hawaii," *Geophys. Res. Lett.*, 2010.

[113] N. Trujillo-Castrillón, C. M. Valdés-González, R. Arámbula-Mendoza, and C. C. Santacoloma-Salguero, "Initial processing of volcanic seismic signals using Hidden Markov Models: Nevado del Huila, Colombia," *J. Volcanol. Geotherm. Res.*, 2018.

[114] *C. cassisi*, M. Prestifilippo, A. Cannata, P. Montalto, D. Patanè, and E. Privitera, "Probabilistic Reasoning Over Seismic Time Series: Volcano Monitoring by Hidden Markov Models at Mt. Etna," *Pure Appl. Geophys.*, 2016.

[115] A. Maggi, V. Ferrazzini, C. Hibert, F. Beauducel, P. Boissier, and A. Amemoutou, "Implementation of a multistation approach for automated event classification at Piton de la Fournaise volcano," *Seismol. Res. Lett.*, 2017.

[116] P. E. E. Lara *et al.*, "Automatic multichannel volcano-seismic classification using machine learning and EMD," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2020.

[117] M. Beyreuther, R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann, "ObsPy: A python toolbox for seismology," *Seismol. Res. Lett.*, 2010.

[118] T. Lecocq, C. Caudron, and F. Brenguier, "Msnoise, a python package for monitoring seismic velocity changes using ambient seismic noise," *Seismol. Res. Lett.*, 2014.

[119] A. Bueno *et al.*, "PICOSS: Python Interface for the Classification of Seismic Signals," *Comput. Geosci.*, 2020.

[120] H. Buurman and M. E. West, "Seismic precursors to volcanic explosions during the 2006 eruption of Augustine Volcano," *US Geol. Surv. Prof. Pap.*, 2010.

[121] G. Cortés, R. Carniel, P. Lesage, and M. A. Mendoza, "VULCAN. ears: Volcano-seismic Unsupervised Labelling and ClAssificatioN Embedded in A Real-time Scenario," 2020. [Online]. Available: https://cordis.europa.eu/project/id/749249/it. [Accessed: 20-Aug-2020].

[122] G. Cortés, R. Carniel, P. Lesage, and M. A. Mendoza, "pyVERSO - software for building and evaluating Volcano-Seismic Recognition (VSR) system," 2020. .

[123] E. D. Cambridge University, "HTK - Hidden Markov Model Toolkit," 2020. [Online]. Available: http://htk. eng.cam.ac.uk/.

[124] G. Cortés, R. Carniel, P. Lesage, and M. A. Mendoza, "geoStudio & liveVSR software," 2020. .

[125] G. Cortés, R. Carniel, M. A. Mendoza, and P. Lesage, "VSR Databases used in article 'Standardization of noisy volcano-seismic waveforms as a key step towards station-independent, robust automatic recognition.'" Zenodo, 2018.

[126] P. Lesage, "Interactive Matlab software for the analysis of seismic volcanic signals," *Comput. Geosci.*, vol. 35, no. 10, pp. 2137-2144, 2009.

[127] D. Cervelli, P. Cervelli, T. Parker, and T. Murray, "SWARM Seismic Wave Analysis and Real-time Monitor: User Manual and Reference Guide," 2020. [Online]. Available: https://volcanoes. usgs.gov/software/swarm/index.shtml.

[128] J. A. Cortés, J. L. Palma, and M. Wilson, "Deciphering magma mixing: The application of cluster analysis to the mineral chemistry of crystal populations," *J. Volcanol. Geotherm. Res.*, 2007.

[129] U. Morgenstern, C. J. Daughney, G. Leonard, D. Gordon, F. M. Donath, and R. Reeves, "Using groundwater age and hydrochemistry to understand sources and dynamics of nutrient contamination through the catchment into Lake Rotorua, New Zealand," *Hydrol. Earth Syst. Sci.*, 2015.

[130] M. O. Awaleh *et al.*, "Geochemical, multi-isotopic studies and geothermal potential evaluation of the complex Djibouti volcanic aquifer (republic of Djibouti)," *Appl. Geochemistry*, 2018.

[131] F. Barette, S. Poppe, B. Smets, M. Benbakkar, and M. Kervyn, "Spatial variation of volcanic rock geochemistry in the Virunga Volcanic Province: Statistical analysis of an integrated database," *J. African Earth Sci.*, 2017.

[132] The Mathworks Inc., "MATLAB - MathWorks," *www.mathworks.com/ products/matlab*, 2020. .

[133] E. S. Schandl and M. P. Gorton, "Application of high field strength elements to discriminate tectonic settings in VMS environments," *Econ. Geol.*, 2002.

[134] J. A. Pearce and J. R. Cann, "Tectonic setting of basic volcanic rocks determined using trace element analyses," *Earth Planet. Sci. Lett.*, 1973.

[135] C. Li, N. T. Arndt, Q. Tang, and E. M. Ripley, "Trace element

indiscrimination diagrams," *Lithos*. 2015.

[136] C. A. Snow, "A reevaluation of tectonic discrimination diagrams and a new probabilistic approach using large geochemical databases: Moving beyond binary and ternary plots," *J. Geophys. Res. Solid Earth*, 2006.

[137] S. P. Verma and J. S. Armstrong-Altrin, "New multi-dimensional diagrams for tectonic discrimination of siliciclastic sediments and their application to Precambrian basins," *Chem. Geol.*, 2013.

[138] S. P. Verma and L. Díaz-González, "New discriminant-function-based multidimensional discrimination of mid-ocean ridge and oceanic plateau," *Geosci. Front.*, 2020.

[139] CoDaPack, "CoDaPack - Compositional Data Package," 2020. [Online]. Available: http://www. compositionaldata.com/codapack.php. [Accessed: 20-Aug-2020].

[140] C. H. Weiß, "StatSoft, Inc., Tulsa, OK.: STATISTICA, Version 8," *AStA Adv. Stat. Anal.*, 2007.

[141] A. B. Beaudoin and R. H. King, "Using discriminant function analysis to identify Holocene tephras based on magnetite composition: a case study from the Sunwapta Pass area, Jasper National Park.," *Can. J. Earth Sci.*, 1986.

[142] A. J. Bourne *et al.*, "Distal tephra record for the last ca 105,000 years from core PRAD 1-2 in the central Adriatic Sea: Implications for marine tephrostratigraphy," *Quat. Sci. Rev.*, 2010.

[143] M. Petrelli, R. Bizzarri, D. Morgavi, A. Baldanza, and D. Perugini, "Combining machine learning techniques, microanalyses and large geochemical datasets for tephrochronological studies in complex volcanic areas: New age constraints for the Pleistocene magmatism of central Italy," *Quat. Geochronol.*, 2017.

[144] A. M. Esposito, G. Alaia, F. Giudicepietro, L. Pappalardo, and M. D'Antonio, "Unsupervised Geochemical Analysis of the Eruptive Products of Ischia, Vesuvius and Campi Flegrei," in *Smart Innovation, Systems and Technologies*, 2021.

[145] Y. Zhao, Y. Zhang, M. Geng, J. Jiang, and X. Zou, "Involvement of Slab-Derived Fluid in the Generation of Cenozoic Basalts in Northeast China Inferred From Machine Learning," *Geophys. Res. Lett.*, 2019.

[146] B. C. Dye and G. Morra, "Machine learning as a detection method of Strombolian eruptions in infrared images from Mount Erebus, Antarctica," *Phys. Earth Planet. Inter.*, 2020.

[147] G. Moreno Chávez, J. Villa, D. Sarocchi, and E. González-Ramírez, "A method and software solution for classifying clast roundness based on the radon transform," *Comput. Geosci.*, 2020.

[148] D. Sinitò *et al.*, "I-PETER (Interactive platform to experience tours and education on the rocks): A virtual system for the understanding and dissemination of mineralogical-petrographic science," *Pattern Recognit. Lett.*, 2020.

[149] Y. Sunaga, R. Natsuaki, and A. Hirose, "Proposal of complex-valued convolutional neural networks for similar land-shape discovery in interferometric synthetic aperture radar," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.

[150] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 Data for Land

Cover/Use Mapping: A Review," *Remote Sens.*, vol. 12, no. 14, p. 2291, 2020.

[151] N. Anantrasirichai, J. Biggs, F. Albino, P. Hill, and D. Bull, "Application of Machine Learning to Classification of Volcanic Deformation in Routinely Generated InSAR Data," *J. Geophys. Res. Solid Earth*, vol. 123, no. 8, pp. 6592-6606, 2018.

[152] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, vol. 2, pp. 1097-1105.

[153] S. Valade *et al.*, "Towards global volcano monitoring using multisensor sentinel missions and artificial intelligence: The MOUNTS monitoring system," *Remote Sens.*, 2019.

[154] A. Si *et al.*, "Debris flow susceptibility assessment using the integrated random forest based steady-state infinite slope method: A case study in Changbai Mountain, China," *Water (Switzerland)*, 2020.

[155] L. Li, C. Solana, F. Canters, and M. Kervyn, "Testing random forest classification for identifying lava flows and mapping age groups on a single Landsat 8 image," *J. Volcanol. Geotherm. Res.*, 2017.

[156] D. T. Bui, P. Tsangaratos, V. T. Nguyen, N. Van Liem, and P. T. Trinh, "Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment," *Catena*, 2020.

[157] D. Carbone, D. Gibert, J. Marteau, M. Diament, L. Zuccarello, and E. Galichet, "An experiment of muon radiography at Mt Etna (Italy)," *Geophys. J. Int.*, 2013.

[158] G. Yang, D. Ireland, R. Kaiser, and D. Mahon, "Machine Learning for Muon Imaging," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.

[159] M. C. Burl *et al.*, "Learning to recognize volcanoes on Venus," *Mach. Learn.*, 1998.

[160] T. F. Stepinski, S. Ghosh, and R. Vilalta, "Machine learning for automatic mapping of planetary surfaces," in *Proceedings of the National Conference on Artificial Intelligence*, 2007.