Automatic Recognition of Long Period Events From Volcano Tectonic Earthquakes at Cotopaxi Volcano

Román A. Lara-Cueva, *Member, IEEE*, Diego S. Benítez, *Senior Member, IEEE*, Enrique V. Carrera, *Member, IEEE*, Mario Ruiz, and José Luis Rojo-Álvarez, *Senior Member, IEEE*

***Abstract*—Geophysics experts are interested in understanding the behavior of volcanoes and forecasting possible eruptions by monitoring and detecting the increment on volcano-seismic activity, with the aim of safeguarding human lives and material losses. This paper presents an automatic volcanic event detection and classification system, which considers feature extraction and feature selection stages, to reduce the processing time toward a reliable real-time volcano early warning system (RT-VEWS). We built the proposed approach in terms of the seismicity presented in 2009 and 2010 at the Cotopaxi Volcano located in Ecuador. In the detection stage, the recordings were time segmented by using a nonoverlapping 15-s window, and in the classification stage, the detected seismic signals were 1-min long. For each detected signal conveying seismic events, a comprehensive set of statistical, temporal, spectral, and scale-domain features were compiled and extracted, aiming to separate long-period (LP) events from volcano-tectonic (VT) earthquakes. We benchmarked two commonly used types of feature selection techniques, namely, wrapper (recursive feature extraction) and embedded (cross-validation and pruning). Each technique was used within a suitable and appropriate classification algorithm, either the support vector machine (SVM) or the decision trees. The best result was obtained by using the SVM classifier, yielding up to 99% accuracy in the detection stage and 97% accuracy and sensitivity in the event classification stage. Selected features and their interpretation were consistent among different input spaces in simple terms of the spectral content of the frequency bands at 3.1 and 6.8 Hz. A comparative analysis showed that the most relevant features for automatic discrimination between LP and VT events were one in the time domain, five in the frequency domain, and nine in the scale domain. Our study provides the framework for an event classification system with high accuracy and reduced computational requirements, according to the orientation toward a future RT-VEWS.**

***Index Terms*—Decision trees (DT), feature extraction, feature selection, seismic event classification, seismic event detection, support vector machines (SVM), volcanic events.**

Manuscript received April 6, 2015; revised September 8, 2015, January 22, 2016, and April 8, 2016; accepted April 17, 2016. This work was supported in part by the Universidad de las Fuerzas Armadas-ESPE under Grant 2013-PIT-014 and Grant 2015-PIC-004, by the Spanish Government through Research Grants PRINCIPIAS under Grant TEC2013-48439-C4-1-R, and by the Comunidad de Madrid through PRICAM under Grant S2013/ICE-2933.

R. A. Lara-Cueva and E. V. Carrera are with the Departamento de Eléctrica y Electrónica, Universidad de las Fuerzas Armadas-ESPE, 171-5-231B Sangolquí, Ecuador (e-mail: ralara@espe.edu.ec).

D. S. Benítez is with the Colegio de Ciencias e Ingenierías El Politécnico, Universidad San Francisco de Quito, 117-1200-841 Quito, Ecuador.

M. Ruiz is with the Instituto Geofísico, Escuela Politécnica Nacional, E11-253 Quito, Ecuador.

J. L. Rojo-Álvarez is with the Department of Signal Theory and Communications, Rey Juan Carlos University, 28943 Fuenlabrada, Spain.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2016.2559440

I. Introduction

In THE case of natural disasters or anthropogenic hazards, early warning systems are necessary to save human lives [1], [2]. One of the main natural hazards are volcanic eruptions [3], and in this regard, several monitoring systems have been previously developed and deployed around the world for trying to better understand this phenomenon [4]. The seismic signature of volcanic events are registered today by powerful seismometers or geophones, and the main types of seismic signals on active volcanoes are volcano-tectonic (VT) earthquakes, long-period (LP) events, tremors (TRE), and hybrid (HYB) events [5], [6]. Fig. 2(c) and (d) shows signal examples of VT and LP events originated at Cotopaxi Volcano. VT events are earthquakes taking place in a volcanic environment with a variable time duration that is usually less than 30 s and a wide spectral content that is typically above 5 Hz, whereas LP events, also known as low-frequency events, show lack of distinguishable phases and a typical time duration that is below 90 s, with spectral content limited at narrow frequency bands between 2 and 5 Hz. HYB earthquakes share the features of VT and LP events, and they are characterized by high-frequency signals, usually with a wide spectral bandwidth above 10 Hz. Finally, TRE are characterized by their constant amplitude and long duration, and thus they are the most distinctive signals generated by volcanoes, and they are widely considered the most complex types of events, with duration ranging from a few minutes to several days [7].

Our main interest is to help geophysicists to understand the behavior of volcanoes and to try to predict eruptions when possible. For this purpose, we designed a volcano monitoring system that includes a high-accuracy event detection followed by a classification component. A volcano early warning system (VEWS) based on our proposal may allow authorities to alert the population, as soon as possible, about the risk of eventual disasters. In this setting, and as a first step to the future development of a real-time VEWS (RT-VEWS), the aim of this paper is to build two connected systems, i.e., an automatic detection system for volcanic seismicity, and subsequently followed by a high-accuracy event classification system. Our hypothesis is that the use of carefully designed detector and classifier systems, both of them based on appropriate feature extraction and machine learning techniques, will reduce the processing time currently required by human experts for the examination of the seismic signals. Specifically, we addressed the design of an automatic system for first separating LP and VT events from the background noise (BN), and then identifying LP from VT

with the highest possible accuracy. We are mainly interested in LP and VT events since these types of events are crucial for helping to forecast eruptions [8]–[10]. The classification of other types of events is beyond the scope of this paper, but it could be readily addressed with our proposed approach.

After the literature review, a general procedure was defined and designed to process, analyze, and identify volcano-seismic signals. We used a preprocessing stage, which removes noise and nonvolcanic originated signals such as thunder, by using bandpass filters [11]. Then, a stage capable of detecting volcanic events is proposed [12], and it is followed by a feature extraction [13] and feature selection stage [14]. Following the state-of-the-art of preceding studies in geophysics, the use of filter and embedded methods allowed us to determine the main features in time, frequency, and scale domains. Finally, a classification stage is used to identify VT and LP events [15]. In this setting, we benchmarked two commonly used feature selection techniques, i.e., wrapper and embedded methods, each of them in a suitable and appropriate classification algorithm, i.e., support vector machines (SVMs) and decision trees (DTs), as discussed later in detail.

There are two main contributions from this paper. First, the whole system (i.e., event detection and classification) is based on machine learning techniques, and our proposed approach applies a two-stage solution, consisting of event detection followed by event classification. Previous studies reported in the literature [16]–[19] did not apply any detection strategy based on classification, but instead they just used sliding windows (with fixed or variable length) and a ratio triggering algorithm for event detection. Therefore, our proposed specific event detector based on classification techniques is novel for this application. In this setting, we also employ feature extraction and feature selection techniques to reduce the processing time toward a reliable real-time system. As a second contribution, we focus our analysis to the data obtained from the Cotopaxi volcano. Although there are several studies about seismological events at Cotopaxi volcano, including lahar studies or very-LP and LP events activity analysis [20], [21], none of them focused on algorithms for automatic real-time detection and classification of seismic events. To our best knowledge, this is the first try to create an automatic process of signals from such volcano.

The rest of this paper is organized as follows. Section II summarizes previous studies and results about the automatic classification of seismic events by using machine learning techniques. Section III describes the proposed approach and the experimental study including event detection, feature extraction, feature selection, and event classification. Section IV presents the results obtained in the automatic classification process when considering different classifiers. Finally, Section V presents the discussion and the main conclusions.

II. Related Work

Our study is referred to Cotopaxi, an active volcano which is part of the so-called Ring of Fire, located at latitude 0°41'05" S and longitude 78°25'54.8" W in the Andean mountain region of Ecuador. On this volcano, a monitoring system has been previously deployed by the Instituto Geofísico de la Escuela

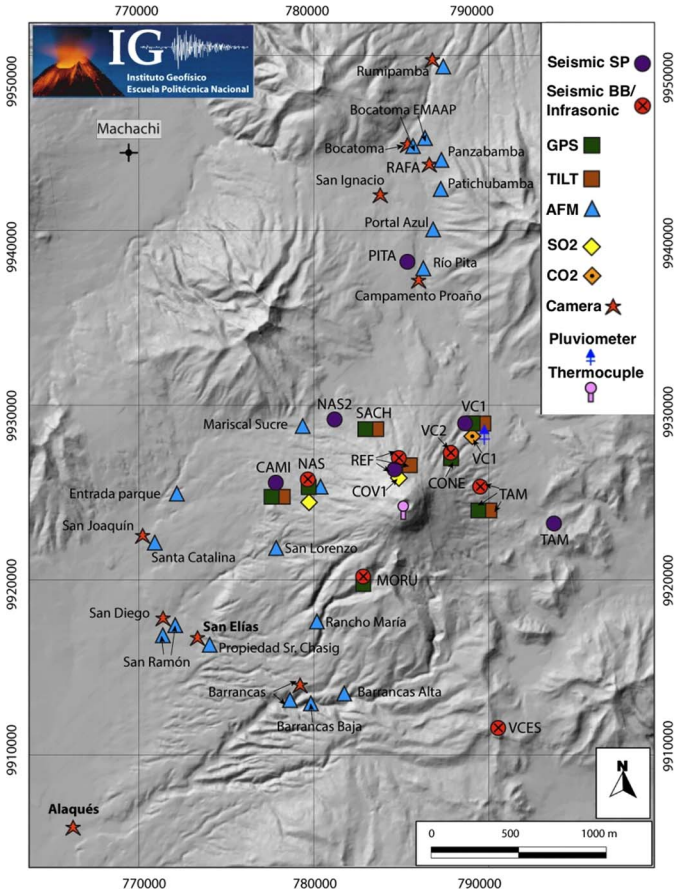

Fig. 1. Seismological stations deployed at Cotopaxi Volcano.

Politécnica Nacional (IGEPN) (see Fig. 1), which has currently installed: (a) six short-period seismological stations (i.e., PITA, NAS2, VC1, REF, CAMI, and TAM), four of them with vertical-axis sensors, and two with three-axis sensors, and all having a frequency response range of 1–50 Hz; and (b) six broadband stations (i.e., VC2, REF, NAS, TAM, MORU, and VCES), with a frequency response range of 0.1–50 Hz [22]. Currently, expert scientists must analyze seismograms of volcanic signals by visual inspection to label and classify events, which represents a subjective and hard-work-demanding process. This task also requires a significant amount of time and noteworthy experience. A reliable automatic classification system can significantly reduce the effort required to make a faster and objective classification [16], [17].

In this context, a strategy to classify VT from other nonvolcanic originated events, such as quarry blasts (QBs) and thunder, was proposed for the Vesuvius Volcano (Italy) [18]. In that work, a feature extraction strategy based on linear predictive coding was developed, yielding a set of 60 parameters of temporal and spectral characteristics. By using artificial neural networks (ANNs) as classifiers, the system reached a 100% accuracy rating in separating VT from QB. Similarly, in [23], a hidden Markov model (HMM) classifier was applied to the data from the San Cristóbal Volcano (Nicaragua), to identify LP from two types of explosions, and BN in raw seismic data, yielding an 80% accuracy rating. Another system was proposed in [24], in which a fuzzy algorithm was used for automatic

classification of local and regional earthquakes, and two non-volcanic originated events, i.e., QB and machinery noise. Six main features were considered in the time and frequency domains, yielding a 96% classification accuracy.

A feature selection strategy was developed in [25] for Nevado del Ruiz Volcano (Colombia), considering VT, LP, and ice quakes. Two classifiers were considered, based on the parametric power spectral density (PSD) estimation of events. A Bayesian classifier built on dissimilarity representations and k-nearest neighbor (k-NN) classifier were used, obtaining accuracy rates of 81% and 84%, respectively. Whereas in [26], the authors considered HB, LP, TRE, and VT events and worked on the stochastic variability of a wide set of time-variant features. With this approach, the classification rate improved from 78% to 88% when k-NN was used instead of HMM. In [27], several feature sets proposed in [28]–[32] were benchmarked to discriminate HYB, LP, TRE, and VT events, yielding the best accuracy of 83% when a k-NN classifier was used.

Meanwhile in [28], a segmentation window of 30 s and a combination of ANN with genetic algorithms were applied to the data from Villarica Volcano (Chile), obtaining a baseline recognition rate of 93%, by using an input space of temporal and spectral features when considering LP, TRE, and energetic TRE events. In [33], the seismic signals were segmented with a rectangular window of 1 min, and a feature extraction strategy was applied by using circular statistics, obtaining the instantaneous phase and the mean energy of the events with Hilbert and wavelet transforms, to identify LP events and VT earthquakes at Llaima Volcano (Chile). A linear discriminant classifier reached a 92.5% accuracy rating. In this setting, an SVM classifier obtained a baseline recognition accuracy of 80% by using a set of features related to the amplitude, frequency, and phase of the seismic signals [19]. Ten percent of the entire record was used as segmentation window with 50% overlapping, which was distinct in different recordings, due to the variable lengths for the entire seismic events at Villarica Volcano.

According to this review, previous studies have demonstrated the possibilities of using machine learning techniques in this setting. Nevertheless, the literature still lacks supportive evidence about which are the main design parameters to be considered in each stage, from preprocessing to classification, to satisfy RT requirements.

III. Proposed Methodology

As already mentioned, our main interest in the middle term is to develop a VEWS, which can accomplish RT capabilities for monitoring and decision-making. In this paper, we focus on developing an automatic classifier to distinguish LP and VT events as a supporting tool that can be used by experts to quantify seismic activity and to alert in case of emergency situations. The proposed system consists of several stages.

1) Acquired signals are initially treated by a preprocessing subsystem.
2) Next, a subsystem previously proposed by our research group [34] is adapted here to detect the presence of an (any type) event from BN.
3) Then, an additional subsystem is given by a machine learning approach, and it is specifically devoted to classify the previously detected events into one of two target classes, i.e., VT and LP. In designing this subsystem, attention has to be placed on the stages of feature extraction and feature selection.

These subsystems, together with their performance evaluation, are described in this section.

A. Preprocessing Subsystem

The data used in this paper were provided by the IGEPN. Data consisted of $N = 914$ independent volcano-seismic recordings (i.e., 759 LP, 116 VT, 30 HYB, and 9 TRE) sampled at 100 Hz. These particular recordings have been extracted, identified, and labeled by experts at the IGEPN from continuous monitoring recordings of seismograms. Each available recording conveyed a single volcanic event, and it was preprocessed with a 128th-order bandpass finite-impulse response filter. The filter had a passband between 1 and 15 Hz, given that the undesired sea microseisms generate spectral power content of about 0.2 Hz [35]. In addition, it is known that the main spectral content for LP events is expected to be in the frequency range of 2–4 Hz and for VT earthquakes, in the 5–10 Hz bandwidth [36]. The resulting filtered record was then normalized to have a zero-mean and unit-variance signal. Fig. 2 shows an example of a recording segment including a VT event, both for the raw recording as registered by the deployed sensors (a), and for its corresponding filtered and normalized signal (b).

B. Event Detection and Segmentation Subsystem

Volcanic events in the database need to be extracted from BN as the first stage. In previous work, we developed an LP event detector with high detection capabilities with respect to BN (see [34] for details). Given that our current system aims to distinguish between LP and VT events, we adapted that method to provide with an accurate event detector in the presence of LP, VT, HYB, and TRE events available in the database, before addressing the event classification.

For this purpose, a nonoverlapping sliding window of length $w = 1500$ samples (15 s) was first applied for the segmentation of each signal. This produced a data matrix that was used to train a classifier for distinguishing between BN and the presence of any event, according to the event labeling provided by the IGEPN experts. We used supervised learning, specifically with the DT as a machine learning technique. The output of the DT classifier predicted a class y for each segment, which was labeled with $y = +1$ in the case of existing events of any kind, and $y = -1$ otherwise. The performance obtained with this event detection stage is analyzed in Section IV.

The sequence of ± 1's generated by the detector and applied to each sliding window in the same recording was post-processed for event delimitation. The average time position of consecutive detected segments was computed and denoted t_d. We decided to use a time window of 60 s since this value is a good tradeoff for conveying LP and VT event duration at the Cotopaxi Volcano [22]. Therefore, the starting t_s and ending t_e time stamps of each event were obtained as $t_s = t_d - 30$ and

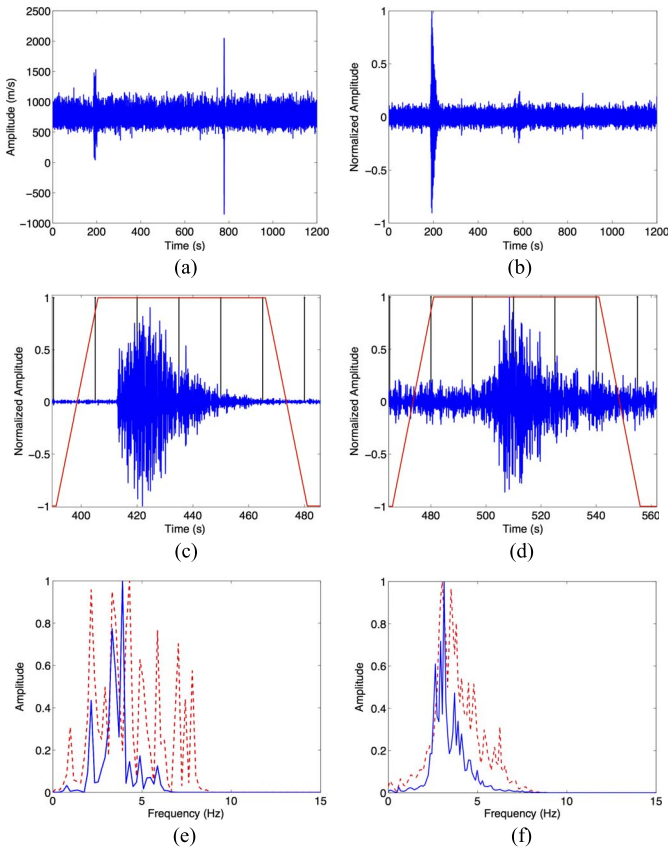

Fig. 2. Examples for preprocessing and feature extraction. (a) Raw recorded waveform conveying a VT event. (b) Same signal after filtering and normalizing. (c) and (d) Examples of VT and LP events after detection and segmentation. (e) Welch periodogram for the example events (LP in blue continuous, VT in red discontinuous). (f) Wavelet decomposition at level 6 for the example events (LP in blue continuous, VT in red discontinuous).

$t_e = t_d + 30$, respectively. This postprocessing stage provided us with data matrix $\mathbf{T}$, which is given by

$$\mathbf{T} = \left[\mathbf{t}_1^T, \mathbf{t}_2^T, \ldots, \mathbf{t}_i^T, \mathbf{t}_{i+1}^T, \ldots, \mathbf{t}_M^T\right]^T \quad (1)$$

where M is the number of vectors, and $\mathbf{t}_i$ are the signal segments in each recording containing some event within a 60-s window. We had available 350 LP and 116 VT events in this matrix since initial recordings only have a single event per record. To avoid bias in the classifier, we used balanced classes for training, including 116 segments of each class (hence $M = 232$ signal vectors, and each vector $\mathbf{t}_i$ conveyed 6000 elements or time samples).

Fig. 2(c) and (d) shows two examples of waveforms conveying VT and LP events after detection and segmentation. In these panels, black vertical lines denote the segments (or sliding windows) used for the event detector, whereas the red line indicates the detection (i.e., $+1$) or not (i.e., -1) of an event in each window. Since the whole delimiting time window has 60 s, the VT event is fully included in the window (four segments), whereas the LP event, although having larger duration, is still mostly included in the window.

C. Design of the Feature Extraction Stage

The VT—LP classification subsystem was also designed in a machine learning framework. The input vector to the classifier was given either by the raw signal in vector $\mathbf{t}_i$ or by a set of features (measurements on specific properties of the signal) aimed to enhance the relevant information to be fed to the classifier. We scrutinized several feature extraction strategies to extract a set of relevant features from each row of data matrix $\mathbf{T}$, yielding a feature vector which can be denoted by $\mathbf{x}_i = g(\mathbf{t}_i)$, where $g(\cdot)$ is the feature extraction operator to be defined for each feature extraction strategy. Two different feature extraction strategies were benchmarked, and they are summarized in the following.

(a) *Features from PSD*. The PSD was calculated for each signal by using its Welch periodogram [37], yielding feature vectors $x_i^F = g_F(t_i)$, where g_F denotes the PSD computing operator. The Welch periodogram used a *boxcar* window with 50% overlapping unit amplitude and length of 512 samples. Each resulting feature vector had a number of features given by the number of samples in the Fourier representation of the Welch periodogram, yielding, therefore, 257 features. Fig. 2(e) shows the PSD features for the example events in Fig. 2(c) and (d).

(b) *Features from Signal Parameters in Several Domains*. Not only the spectral representation but also many other approaches have been proposed in the literature to analyze volcanic events with a number of signal parameters. In this setting, we can define a parameter as a measurement of some signal quality, either in the time, frequency, or scale domain. In this case, the number of parameters used as features were 80 (see [38] and references therein for details), and they were compiled as follows.

- First, a total of 12 features were obtained in the time domain, i.e., mean time μ_t, standard deviation σ_t, signal entropy En_t, concentration of the signal around the mean value of the distribution kurtosis$_t$, multiscale entropy MSE$_t$, time to reach the maximum peak $t_{t,\mathrm{mp}}$, the difference between the maximum and minimum peaks P2P$_t$, rms, difference between the maximum and minimum peaks in the rms signal $\mathrm{P2P}_t^{\mathrm{rms}}$, signal energy in rms E_t^{rms}, zero-crossing number Z_t, and peaks density in the rms signal.
- Second, a total of 19 features were obtained in the frequency domain by applying the fast Fourier transform (FFT) to each signal in the data matrix. Thus, we have the following: the number of peaks ψ_f over a threshold of 0.9, the maximum detected frequency $\max_f$, the mean frequency μ_f, the frequency standard deviation σ_f, the spectral entropy En_f, the spectral energy E_f, the spectral kurtosis kurtosis$_f$, the spectral multiscale entropy MSE$_f$, the maximum peak value U, the maximum frequency in the frequency bands from 10 to 20 Hz $\max_f^{10-20}$ and from 20 to 30 Hz $\max_f^{20-30}$, the rms rms$_f$, the difference between the maximum and minimum peaks P2P$_f$, the difference between the maximum and minimum peaks in rms $\mathrm{P2P}_f^{\mathrm{rms}}$, the signal energy in rms E_f^{rms}, the peak density in rms, the highest peak value in rms H_f, and the second (H_f') and the third (H_f'') highest peaks in rms. In all these features, the subscript f indicates their frequency-domain dependence.

- Finally, a total of 49 features in the scale domain were extracted, where a wavelet transform was applied to each signal by using a tenth-order *symlet* mother wavelet. Note that the use of wavelet transform aimed to overcome the resolution limitations exhibited by the Fourier transform [39]. Knowing the main frequency bands where events are usually present (from 2 to 4 Hz for LP and from 5 to 10 Hz for VT), we decided to work with a decomposition δ at the level 6 of the wavelet transform, obtaining the corresponding approximation cA and detailed cD wavelet coefficients. Hence, we retrieved seven features, i.e., the mean value μ_w, the frequency of the maximum value $(f_{\max,w})$, the energy percentage $E^{\delta}_{\%w}$, the difference between the maximum and minimum peaks $\mathrm{P2P}^{\delta}_{w}$, the maximum frequency V^{δ}_{w}, the RMS $\mathrm{rms}^{\delta}_{w}$, and the difference between the maximum and minimum peaks in rms $\mathrm{P2P}^{\delta,\mathrm{rms}}_{w}$, for each coefficients $cA6$, $cD6$, $cD5$, $cD4$, $cD3$, $cD2$, and $cD1$. Fig. 2(e) and (f) allows the comparison of the Welch periodogram with the level 6 wavelet coefficients, respectively, for the preceding examples of VT and LP events; the subscript w indicates their scale-domain dependence.

D. Classification Algorithms and Feature Selection Strategies

For the event classification stage, we also considered a supervised machine learning approach. Thus, we scrutinized the suitability of two well-known classification algorithms, i.e., DT [15], [40] and SVM [41].

DT is a nonparametric supervised learning method used for approximating discrete-valued target functions in classification and regression. The aim of this method is to create a model that can predict the c possible values of a target variable by learning simple decision rules inferred from the data features. Each node in the tree specifies a rule for some attribute of the instance, and each branch descending from that node corresponds to one of the possible values for that attribute. An instance is classified by starting at the root node of the tree, testing the attribute specified by this node, and then moving down the branch corresponding to the value of the attribute. This process is then repeated for the subtree rooted at the new node. The algorithm promotes small trees instead of large ones, which yields classifiers with good generalization capabilities. The depth or *leafiness* of the tree is the free parameter for this machine learning technique, which is measured in terms of the information actually contained by the child nodes. Here, we used two indexes for this purpose, i.e., with the average amount of information contained in each event (as given by the entropy $E(t)$), and with the statistical dispersion and inequality (as given by Gini index $G(t)$) [42] as follows:

$$E(t) = -\sum_{i=0}^{c-1} p(i|t)\log_2 p(i|t) \tag{2}$$

$$G(t) = 1 - \sum_{i=0}^{c-1} \left[p(i|t)\right]^2 \tag{3}$$

where $p(i|t)$ denotes the fraction of records belonging to output value i at a given node t, and c is the number of possible values of the target variable.

Conventional machine learning classifiers, such as Gaussian maximum likelihood or ANN, can be strongly affected by the high dimensionality of input observation vectors, and they can tend to overfit to the data in the presence of noise or to perform poorly with a low number of available training samples [43]. In the last few years, the use of SVMs [44], [45] for machine learning practical applications has received wide attention because the method integrates in the same classification procedure: 1) a simple way of dealing with nonlinear classification boundaries, as samples that are nonlinearly separable in the input space are mapped to a higher dimensional space where a simpler (linear) classification is performed; 2) an intrinsic regularization procedure, which controls, with efficiency, the model complexity; and 3) the minimization of an upper bound of the generalization error, thus following the structural risk minimization principle.

The ν–SVM algorithm for classification was used in this paper. It is defined in summary as follows, and the interested reader is recommended to see, e.g., [45] for details. Given a labeled training data set $\{\boldsymbol{x}_i, y_i\}_{i=1}^{n}$, where $\boldsymbol{x}_i \in \mathbb{R}^N$ and $y_i \in \{-1,\ +1\}$, and given a nonlinear mapping $\boldsymbol{\phi}(\cdot)$, the ν–SVM method solves the following:

$$\min_{\boldsymbol{w},\xi_i,b,\rho} \left\{ \frac{1}{2}\|\boldsymbol{w}\|^2 + \nu\rho + \frac{1}{n}\sum_{i=1}^{n}\xi_i \right\} \tag{4}$$

subject to the following:

$$y_i\left(\langle \boldsymbol{\phi}(\boldsymbol{x}_i), \boldsymbol{w}\rangle + b\right) \geq \rho - \xi_i \qquad \forall i = 1,\ldots,n \tag{5}$$

$$\rho \geq 0, \xi_i \geq 0 \qquad \forall i = 1,\ldots,n \tag{6}$$

where $\boldsymbol{w}$ and b define a linear classifier in the feature space, and ξ_i are positive slack variables enabling to deal with errors. The appropriate choice of nonlinear mapping $\boldsymbol{\phi}$ guarantees that the transformed samples are more likely to be linearly separable in the (higher dimensional) feature space. In this formulation, variable ρ is controlled with coefficient ν, which adds another degree of freedom to the margin, with the size of the margin increasing linearly with ρ. Therefore, the tradeoff between the training error and the generalization error is controlled in the ν–SVM formulation by adjusting ν in the range [0, 1], which acts as an upper bound on the fraction of margin errors, and it is also a lower bound on the fraction of support vectors (SVs).

Primal problem (4) is solved by using its dual problem counterpart, yielding $\boldsymbol{w} = \sum_{i=1}^{n} y_i\alpha_i\boldsymbol{\phi}(\boldsymbol{x}_i)$ (see [45] for further theoretical and algorithmic details), and the decision function for any test vector $\boldsymbol{x}_*$ is finally given by the following:

$$f(\boldsymbol{x}_*) = \mathrm{sgn}\left(\sum_{i=1}^{n} y_i\alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}_*) + b\right) \tag{7}$$

where α_i are Lagrange multipliers corresponding to constraints in (4), with the SVs being those training samples $\boldsymbol{x}_i$ with nonzero Lagrange multipliers $\alpha_i \neq 0$. The bias term b is calculated by using the *unbounded* Lagrange multipliers as $b = 1/k\sum_{i=1}^{k}(y_i - \langle\boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{w}\rangle)$, where k is the number of

unbounded Lagrange multipliers $(0 < \alpha_i < C)$. Note that a particularity of SVM is that decision function $f(\mathbf{x})$ is a function of a small subset of the training examples, which are the SVs. Those are the examples that are the closest to the decision boundary and lie on the margin, as well as those wrong-class examples. The existence of such SVs is at the origin of the computational properties of SVM and their competitive classification performance. The interested reader can see [46] for more details about the algorithm with linear and nonlinear SVMs.

Another key point is the use of Mercer kernels $K(\boldsymbol{x}_i, \boldsymbol{x}_*) = \langle \boldsymbol{x}_i, \boldsymbol{x}_* \rangle$, to handle the nonlinear algorithm implementations. In this paper, we use the two well-known Mercer kernels given by the linear kernel $K(\boldsymbol{x}, \mathbf{z}) = \langle \boldsymbol{x}, \mathbf{z} \rangle$, and the Gaussian kernel $K(\boldsymbol{x}, \mathbf{z}) = \exp(-\|\boldsymbol{x} - \mathbf{z}\|^2 / 2\sigma^2)$, where σ is the width-free parameter, to be tuned together with ν free parameter during the training and validation stage.

We used a feature selection strategy to identify the p most relevant features that improved t_p and its overall performance. We worked with recursive feature extraction (RFE), which is a well-known wrapper method that adds or removes features by generating the ranking of them using backward feature elimination [47], [48], to find the optimal combination that maximizes model performance. RFE-SVM is a weight-based method, in which at each step, the coefficients of the weight vector of SVM are used as the feature ranking criterion.

Free parameters were tuned for both classification techniques by following the usual cross-validation process. For DT, the tree depth was determined; meanwhile for SVM, ν free parameter was tuned to control the number of SVs.

E. Performance

Both the detection and the classification performance were measured in terms of accuracy A, precision P, sensitivity or recall R, specificity S criterion, and balanced error rate BER, which are defined as follows:

$$A(\%) = \frac{N_C}{N_T} \times 100 \tag{8}$$

$$P(\%) = \frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FP}}} \times 100 \tag{9}$$

$$R(\%) = \frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FN}}} \times 100 \tag{10}$$

$$S(\%) = \frac{N_{\mathrm{TN}}}{N_{\mathrm{TN}} + N_{\mathrm{FP}}} \times 100 \tag{11}$$

$$\mathrm{BER} = 1 - \frac{R + S}{2 \times 100} \tag{12}$$

where N_C is the number of correctly classified events, N_T is the total number of events used to feed the classifier, N_{TP} is the number of true positives, N_{FN} is the number of false negatives, N_{TN} is the number of true negatives, and N_{FP} is the number of false positives. We calculated these performance measures using training and testing folds. The time consumption of the process in the system, denoted by t_p, was also considered.

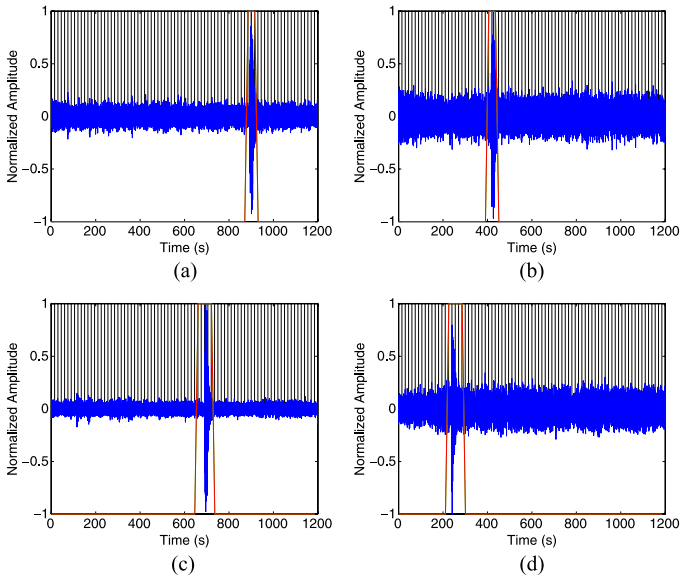

Fig. 3. Event detection representation with time duration of 1 min for each excerpt (red), and with segmentation window $w = 15$ s (black), for (a) and (b) LP events, and (c) and (d) VT earthquakes.

IV. Results

Here, we present the results obtained by applying the methodology proposed to our data set from the Cotopaxi Volcano. The entire data set was divided into training and testing sets, and the independence among all the records in each of these sets was ensured. The experiments were carried out in Matlab, on a Core I5 personal computer with 3.1 GHz and 4-GB RAM.

The detector stage showed a correct event detection rate of 99%, for discriminating LP and VT events from BN. Fig. 3 shows examples of four detected events. As described, a time duration of 1 min was set for each analysis window w_d since this kind of events at Cotopaxi Volcano have this mean time duration. Fig. 3(c) and (d) shows that each excerpt comprised entirely the VT earthquakes since VT earthquakes have time duration lower than 1 min.

A. Results Using DT

In the frequency domain, the DT algorithm obtained the model by using a training set containing $N = 116$ (58 instances for LP and VT), which made possible to induce a tree shape taking the top-down form, as depicted in Fig. 4. The DT algorithm selected three key features X_i, where i is the number of the selected feature corresponding to the amplitude of the PSD in a determined frequency, beginning from the top node with the rule $X_{16} \geq -0.42$ (corresponding to the amplitude value at 3 Hz), $X_{21} \geq -0.58$ (corresponding to the amplitude value at 4 Hz), and $X_{36} \geq 0.47$ (corresponding to the amplitude value at 6.8 Hz), which made possible to classify into any of the six possible leaves. Fig. 4(d)–(f) shows a mesh representation of ten segments, where from 1 to 5 (from 6 to 10) correspond to examples of LP (VT) events. We can observe that DT algorithm retrieved about three effective features in each segment to predict the outcome.

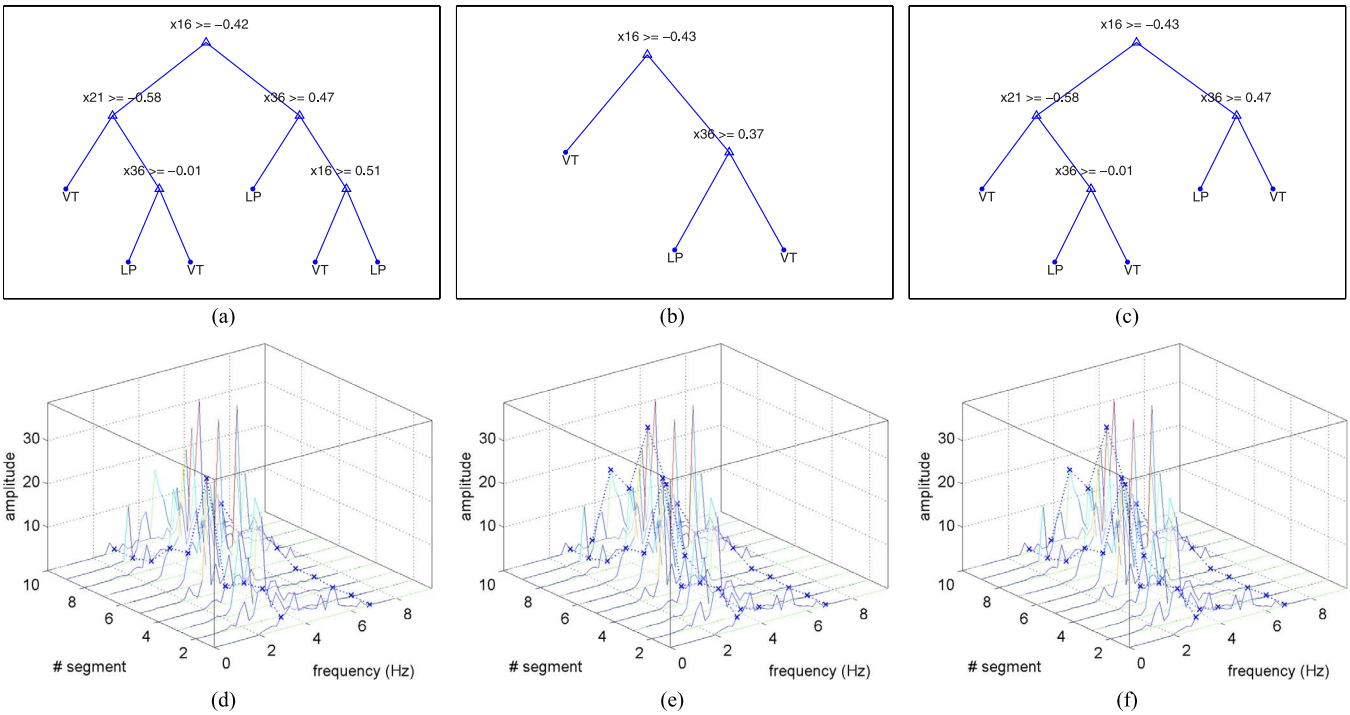

Fig. 4. Tree representation considering frequency features for cross-validation and pruning feature selection methods in DT. Segments from 1 to 5 (from 6 to 10) correspond to examples of LP (VT) events: (a) main features in the tree and (d) features verified in each signal; (b) cross-validation and (c) pruning feature selection methods, with main features in the tree; and (d)–(f) features retrieved in each signal.

Cross-validation and pruning methods were performed to control *leafiness*. Cross-validation selected two key features, beginning from the top node, with rule $X_{16} \geq -0.43$, and $X_{36} \geq 0.37$, for classifying into one of the three possible leaves, as depicted in Fig. 4(b). By pruning the selected three key features, the top node was kept with the same rule, and two more rules were defined, given by $X_{21} \geq -0.58$ and $X_{36} \geq 0.47$, which corresponded to 4 and 6.8 Hz, respectively, and they were classified into one of the five possible leaves, as depicted in Fig. 4(c). We observed in Fig. 4(d)–(f) that VT earthquakes have the main spectral content above 4 Hz, which was retrieved by DT algorithm in 4 Hz (X_{21}) and 6.8 Hz (X_{36}), whereas LP events present their main spectral component in 3 Hz (X_{16}) and above 6 Hz has negligible values.

The DT algorithm retrieved 15 key features by considering the parameter features, beginning from the top node, with rule $X_{27} \geq 0.3$, and making possible to classify into any of the 20 possible leaves, as depicted in Fig. 5(a). By using the cross-validation method, the DT algorithm selected one key feature, i.e., the top node with rule $X_{27} \geq 0.3$, classified into one of the two possible leaves, as depicted in Fig. 5(b), while pruning the selected five key features corresponding to E_f^{rms} (X_{27}), $\text{P2P}_w^{\delta_5}$ (X_{67}), $E_{\%w}^{\delta_3}$ (X_{57}), H_p'' (X_{31}), and μ_t (X_1). The top node was kept with the same rule and five more rules were defined, which allowed to classify into one of the six possible leaves, as depicted in Fig. 5(c).

Table I shows the results obtained with the testing set containing $N = 116$ independent cases. The pruning method identified five key features, which made possible to reach the best system performance. The main differences with the original features were in the terms of the BER and t_p.

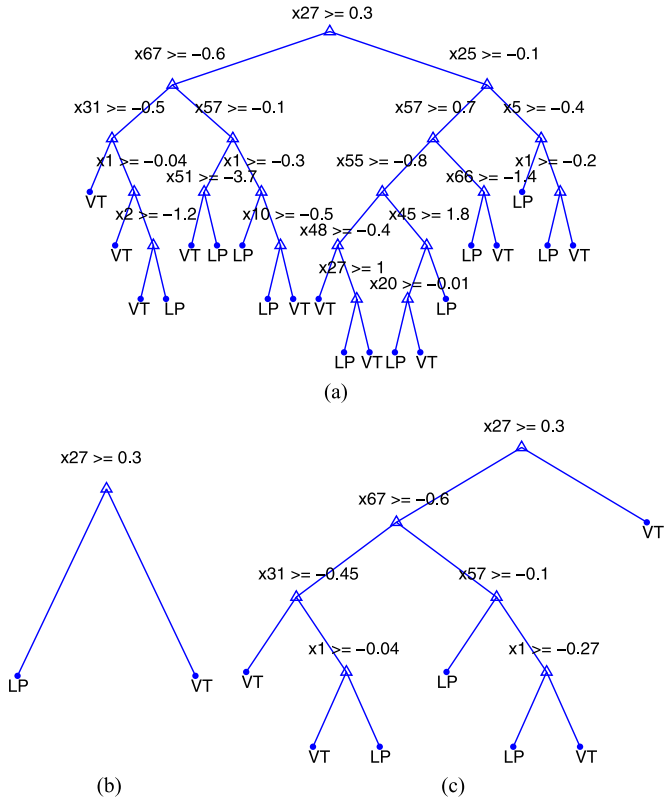

Fig. 5. Tree representation considering parameter features, cross-validation, and pruning feature selection methods. (a) Main features in the tree, (b) cross validation, and (c) pruning feature selection methods.

B. Results Using Linear and Nonlinear SVM Classifiers

SVM is often sensitive to class unbalance, and for this reason, we first determined the training percentage for our database

TABLE I
EXPERIMENTAL PERFORMANCE RESULTS FOR CLASSIFICATION WHEN APPLYING FEATURE SELECTION STAGE WITH DT CLASSIFIER

Matrix	N. Features–Method	A (%)	P (%)	R (%)	S (%)	BER	t_p (s)
Frequency / parameters	4/15–Default	90/94	96/90	83/98	96/89	0.1/0.06	1.5/1
Frequency / parameters	2/1–Cross-Val.	83/84	78/78	92/91	74/78	0.17/0.16	0.7/0.6
Frequency / parameters	3/**5–Pruning**	90/**94**	96/**90**	83/**98**	96/**89**	0.1/**0.06**	1/**0.7**

TABLE II
EXPERIMENTAL RESULTS FOR CLASSIFICATION BY USING LINEAR AND NONLINEAR[1] SVM CLASSIFIERS

Matrix	N. Features–Method	A (%)	P (%)	R (%)	S (%)	BER	t_p (s)
Frequency / parameters	257/80–Default	92/97	93/99	91/96	93/99	0.08/0.03	0.23/0.08
Frequency / parameters	4/**15–RFE**	93/**97**	94/**97**	92/**96**	94/**98**	0.07/**0.03**	0.06/**0.06**
Frequency / parameters	257/80–Default[1]	84/97	86/96	81/98	87/96	0.16/0.03	7/3.9
Frequency / parameters	15/15–RFE[1]	93/93	96/93	91/94	96/93	0.06/0.06	1.9/1.9

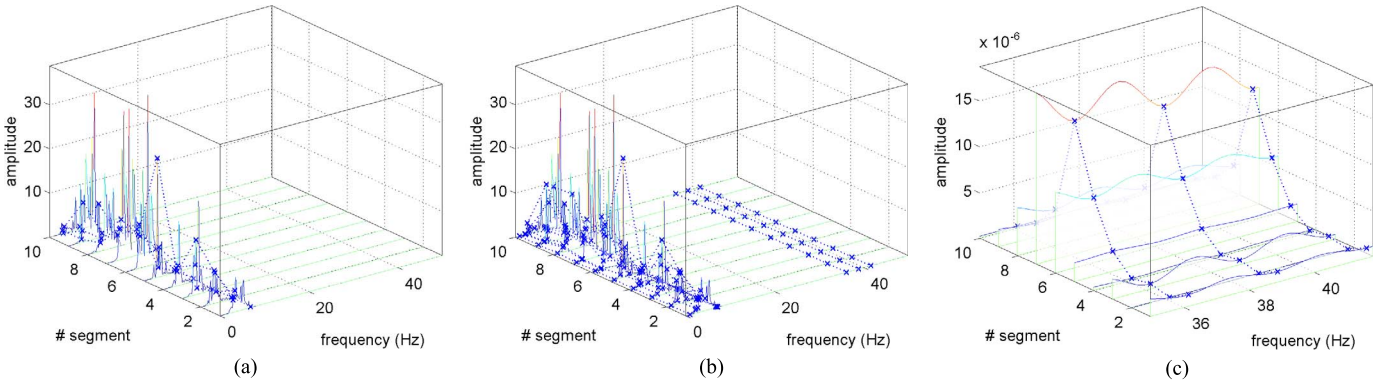

Fig. 6. Tree representation considering frequency features and RFE feature selection method. Features retrieved in each $\mathbf{g_i}$ for (a) linear, (b) nonlinear, and (c) zoomed nonlinear SVMs above 35 Hz.

TABLE III
EXPERIMENTAL RESULTS FOR CLASSIFICATION BY USING LINEAR SVM CLASSIFIER, BY CONSIDERING INDEPENDENT FEATURES IN THE TIME $\mathfrak{T}$, FREQUENCY $\mathfrak{F}$, AND SCALE $\mathfrak{W}$ DOMAINS

Domain	N. Features	A (%)	P (%)	R (%)	S (%)	BER	t_p (s)
$\mathfrak{T}$	1	90	94	85	94	0.10	0.04
$\mathfrak{F}$	**5**	**97**	**97**	**96**	**97**	**0.03**	**0.05**
$\mathfrak{W}$	9	93	92	95	92	0.07	0.05

which maximized the system performance. These values were set to 35% and 50% for linear and nonlinear SVM, respectively. Then, we set the ν free parameter to 0.05 for linear SVM classifier, and 0.1 and 0.15 for nonlinear SVM classifier, with frequency and parameter features. Parameter γ was set to 8.86e-5 and 4.00e-3, respectively, by making an iterative process to maximize the system performance. Finally, we selected the main features by using RFE. The results are summarized in Table II. For both feature sets, i.e., linear and nonlinear SVMs, the use of RFE improved the performance in terms of the t_p. RFE selected 4 key features for frequency features and 15 for parameter features in the case of linear SVM, whereas for nonlinear SVM, 15 key features were selected for both feature sets. We obtained the best results with linear SVM and by using RFE method, which identified 15 key features considering parameter features, and the main difference with the other methods was in terms of t_p and BER.

Considering frequency features with the linear SVM, the algorithm recognized the relevance of frequency bands corresponding to 3.11, 3.3, 3.5, and 7 Hz, as depicted in Fig. 6(a). Meanwhile, the nonlinear SVM identified 15 bands, which corresponds to 1.17, 2.5, 3, 3.11, 3.3, 3.5, 4.5, 6.8, 7, 7.2, 7.4, 22.3, 36.2, 39, and 41.6 Hz. We observed that not only low frequencies were considered but also high frequency bands were identified, as shown in Fig. 6(b). Therefore, we observed in Fig. 6(c) that nonlinear SVM considers frequencies above 35 Hz. In general, linear and nonlinear SVMs were able to identify LP from VT in simple terms of amplitude and spectral content.

By using parameter features and RFE feature selection method, linear SVM retrieved 15 key features, i.e., Z_t (X_{11}), ψ_f (X_{13}), MSE_f (X_{20}), U_{10-20} (X_{21}), rms_f (X_{24}), E_f^{rms} (X_{27}), $\text{V}_w^{\delta_5}$ (X_{39}), $\mu_w^{\delta_5}$ (X_{41}), $\mu_w^{\delta_1}$ (X_{53}), $E_{\%w}^{\delta_5}$ (X_{56}), $E_{\%w}^{\delta_2}$ (X_{59}), $\text{P2P}_w^{\delta_5}$ (X_{68}), $\text{P2P}_w^{\delta_2}$ (X_{76}), $\text{P2P}_w^{\delta_2,\text{RMS}}$ (X_{77}), $\text{P2P}_w^{\delta_1}$ (X_{80}), and nonlinear SVM retrieved also 15 key features, i.e., Z_t (X_{11}), ψ_f (X_{13}), E_f^{rms} (X_{19}), MSE_f (X_{20}), $\text{rms}_f(X_{24})$, $E_f^{\text{rms}}(X_{27})$, $\text{V}_w^{\delta_5}$ (X_{39}), $\text{P2P}_w^{\delta_6}$ (X_{42}), $\text{rms}_w^{\delta_5}$ (X_{51}), $\mu_w^{\delta_1}$ (X_{53}), $\text{P2P}_w^{\delta_5}$ (X_{68}), $\text{P2P}_w^{\delta_2}$ (X_{76}), $\text{P2P}_w^{\delta_2,\text{RMS}}$ (X_{77}), $\text{V}_w^{\delta_5}$ (X_{78}), $\text{rms}_w^{\delta_5}$ (X_{79}), $\text{P2P}_w^{\delta_1}$ (X_{80}). We identified 11 matching features, which corresponded to Z_t, ψ_f, MSE_f, rms_f, E_f^{rms}, $\text{V}_w^{\delta_5}$, $\mu_w^{\delta_1}$, $\text{P2P}_w^{\delta_5}$, $\text{P2P}_w^{\delta_2}$, $\text{P2P}_w^{\delta_2,\text{rms}}$, and $\text{P2P}_w^{\delta_1}$.

In the feature selection block, the RFE method presented an improvement in all metrics in at least 1% by considering parameter features, and using 4 of the 257 features. Moreover, the pruning method got similar results with the DT classifier considering both matrices, whereas cross-validation got worse results in at least 10% in each metric. This strategy permitted to reduce the t_p from 1500 to 700 ms. Linear SVM-RFE identified 15 main features: one in the time domain, i.e., $\mathfrak{T}$ (Z_t), five in the frequency domain, i.e., $\mathfrak{F}$ $(\psi_f$, MSE_f, U_{10-20}, rms_f, $E_f^{\text{rms}})$, and nine in the scale domain, i.e., $\mathfrak{W}$ $(\text{V}_w^{\delta_5}$, $\mu_w^{\delta_5}$, $\mu_w^{\delta_1}$, $E_{\%w}^{\delta_5}$,

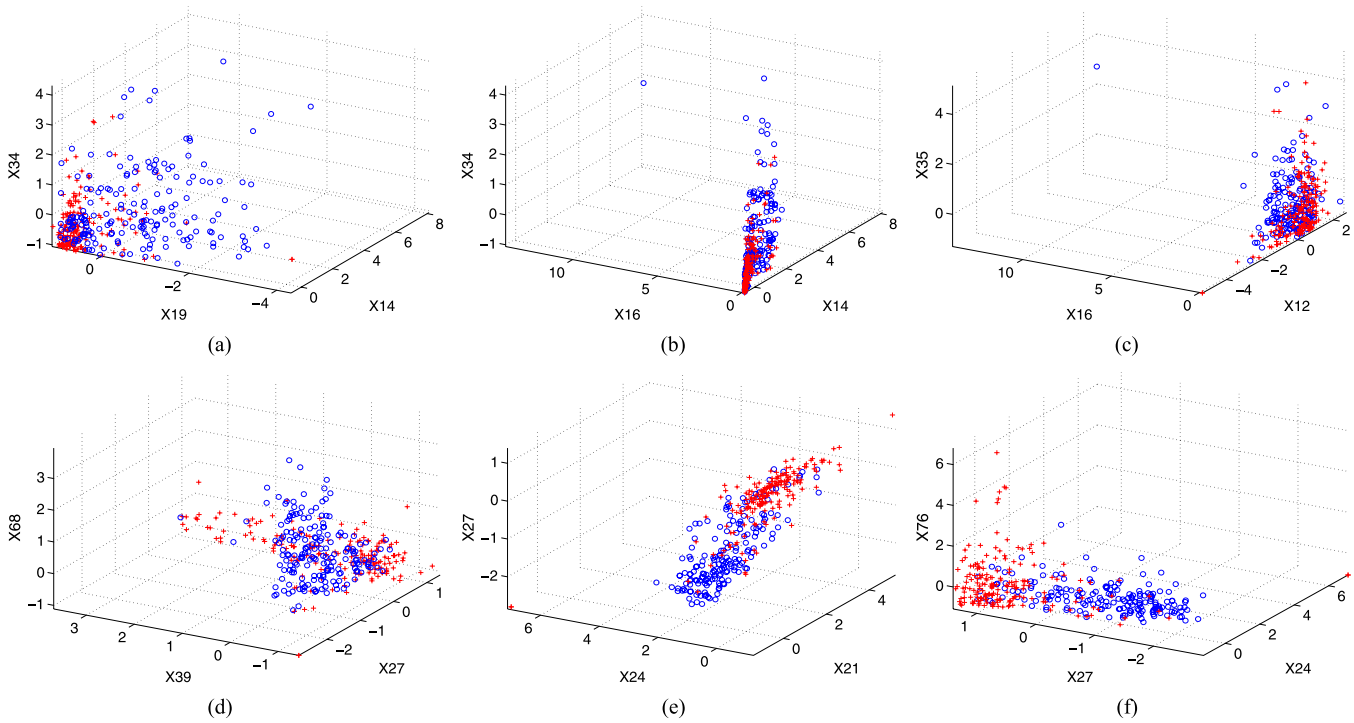

Fig. 7. Scatter plots of selected main features by considering (a)–(c) G matrix and (d)–(f) H matrix using (a)–(d) DT classifier, (b)–(e) linear SVM, and (c)–(f) nonlinear SVM, respectively. Red dots represent VT, whereas blue dots represent LP events.

$E^{\delta_2}_{\%w}$, $\text{P2P}^{\delta_5}_{w}$, $\text{P2P}^{\delta_2}_{w}$, $\text{P2P}^{\delta_2,\text{rms}}_{w}$, $\text{P2P}^{\delta_1}_{w}$). Table III shows the results with each set tested independently, where the best result were obtained with five features in the frequency domain, reaching similar performance values, by reducing the t_p to 500 ms. Additionally, we identified two bands corresponding to detail level 2 $f_{s2} \in (0, 25)$ Hz and level 5 $f_{s5} \in (0, 3.13)$ Hz in the scale domain, where by just considering some parameters allowed us to discriminate VT from LP events with acceptable accuracy. Fig. 7 shows possible feature combination, by considering the three most relevant features in each classifier. We observed in Fig. 7(e), for example, two well-defined clusters, made possible by linear separability, to recognize LP from VT. This was possible by considering linear SVM classifier and U_{10-20}, rms_f, and E^{rms}_f, which are in the frequency domain.

V. Discussion and Conclusions

LP and VT events have proven to be keys for monitoring any volcano activity, including Cotopaxi since they provide important information about the volcano internal status [8]–[10]. With appropriate RT constraints for monitoring, detecting, and classifying volcanic signals, VEWS may be used as a tool for detecting *potencial anormal* increments in seismic activity. Such information, in combination with other parameters such as the external activity of the volcano, may be used by the local authorities to launch an effective early warning to the population.

Our proposed approach applies a two-stage solution, consisting of event detection followed by event classification. Previous studies did not apply any detection strategy based on classification, but instead, a rectangular sliding window of fixed or variable size was used [19], [28], [33]. In our setting, the use of the detection stage has simplified the learning problem of distinguishing LP, VT, and nonvolcanic origin events. In fact, the first stage detects seismic events from BN with high accuracy, therefore reducing the computational requirements of the system. The detection accuracy was near 99%. In our study, the proposed event detector showed superior performance in relation to previous studies, however, further studies with standard data sets should be conducted to be able to give a quantitative comparison analysis.

In the classification stage, the DT algorithm showed an accuracy near 90% by considering the frequency features, whereas a linear SVM has reached 97% considering the parameter features. These values are comparable to the best results obtained in previous studies by applying different methodologies to the classification problem, such as Hilbert and wavelet transforms [33], or Bayesian and k-NN classifiers [25].

Additionally, our experiments have shown that the best results were obtained with the parameter features, where a set of 80 features in time, frequency, and scale domains was considered. We took into account well-known features, with some of them given by experts of IGEPN, which had been obtained by heuristic and observation processes. An error rate of 3% was obtained using 15 main features retrieved by RFE and linear SVM classifier. The main difference between linear and nonlinear SVMs was t_p equal to 60 and 1900 ms, respectively. In the frequency domain, the linear SVM classifier was able to reach a performance similar to the one obtained when using the optimized features with RFE, but by considering only five main features ($\psi_f, \text{MSE}_f, U_{10-20}, \text{rms}_f, E^{\text{rms}}_f$). This proposed strategy has been developed to use it in RT analysis since it reduces the processing time to values approximately lower than 1 min.

Although the experimental results show the feasibility and reliability of the proposed system, specific aspects at the detection

and classification stages still have to be considered to generalize the system to identify all volcano-seismic signals. As future work, we are interested in classifying LP, VT, HYB, TRE, and nonvolcanic origin signals like thunders, by using linear or nonlinear multiclass SVM. It should be possible to improve the detection stage by using well-known automatic detection algorithms applied in speech recognition. These algorithms could determine the starting and ending time stamps of an event, hence, a better analysis can be performed in the entire duration of the volcanic event, while maintaining or improving the accuracy rate.

Regarding feature selection, we have addressed the problem in this paper from the machine learning theory point of view, and most of the main features found by the algorithms were consistent with common features used in other works for event classification in other volcanoes [24], [28], [30]. Although the volcanologist significance of the most important features related to Cotopaxi volcano is still unclear at this moment, and it requires deeper and further investigation, feature selection techniques were used in this paper to simplify the problem of classification. We are planning to address the problem of volcano-specific characteristic significance in future studies. Furthermore, analyzing and combining data from multiple stations may help us to improve the performance of the system, this will also be addressed in future works. Finally, benchmarking our algorithm with seismological data acquired with similar equipment for other volcanoes is to be addressed among the next steps.

ACKNOWLEDGMENT

The authors would like to thank the Instituto Geofísico de la Escuela Politécnica Nacional for providing the data set used in this paper.

REFERENCES

[1] G. Berz *et al.*, "World map of natural hazards-a global view of the distribution and intensity of significant exposures," *Natural Hazards*, vol. 23, no. 2/3, pp. 443–465, 2001.

[2] K. Hewitt, *Regions of Risk: A Geographical Introduction to Disasters*. Abingdon, U.K.: Routledge, 2014.

[3] S. Makowski Giannoni, R. Rollenbeck, K. Trachte, and J. Bendix, "Natural or anthropogenic? On the origin of atmospheric sulfate deposition in the Andes of southeastern Ecuador," *Atmos. Chem. Phys.*, vol. 14, no. 20, pp. 11 297–11 312, 2014.

[4] F. A. Pfaffl and W.-C. Dullo, "The first ascent to the Volcano Cotopaxi in Ecuador by Wilhelm Reiss (1838–1908)," *Int. J. Earth Sci.*, vol. 103, no. 4, pp. 1175–1179, 2014.

[5] S. R. McNutt, "Volcanic seismology," *Annu. Rev. Earth Planet. Sci.*, vol. 32, pp. 461–491, 2005.

[6] H. Langer, S. Falsaperla, T. Powell, and G. Thompson, "Automatic classification and a-posteriori analysis of seismic event identification at Soufriere Hills Volcano, Montserrat," *J. Volcanol. Geothermal Res.*, vol. 153, no. 1, pp. 1–10, 2006.

[7] J. Ibáñez and E. Carmona, "Sismicidad volcánica," *Curso Internacional de Volcanología y Geofísica Volcánica, Serie Casa de los Volcanes*, vol. 7, pp. 269–282, 2000.

[8] B. A. Chouet, "Long-period volcano seismicity: Its source and use in eruption forecasting," *Nature*, vol. 380, no. 6572, pp. 309–316, 1996.

[9] C. J. Bean *et al.*, "Long-period seismicity in the shallow volcanic edifice formed from slow-rupture earthquakes," *Nature Geosci.*, vol. 7, no. 1, pp. 71–75, 2014.

[10] P. Cusano, M. Palo, and M. West, "Long-period seismicity at Shishaldin Volcano (Alaska) in 2003–2004: Indications of an upward migration of the source before a minor eruption," *J. Volcanol. Geothermal Res.*, vol. 291, pp. 14–24, 2015.

[11] A. M. Baig, M. Campillo, and F. Brenguier, "Denoising seismic noise cross correlations," *J. Geophys. Res., Solid Earth (1978–2012)*, vol. 114, no. B8, 2009, Art. no. B08310.

[12] A. Ruano *et al.*, "Seismic detection using support vector machines," *J. Neurocomput.*, vol. 135, pp. 273–283, 2014.

[13] I. Guyon, *Feature Extraction: Foundations and Applications*. New York, NY, USA: Springer-Verlag, 2006.

[14] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. Boca Raton, FL, USA: CRC Press, 2007.

[15] T. M. Mitchell, *Machine Learning*, vol. 45. Burr Ridge, IL, USA: McGraw-Hill, 1997.

[16] T. S. Newman and A. K. Jain, "A survey of automated visual inspection," *J. Comput. Vis. Image Understand.*, vol. 61, no. 2, pp. 231–262, 1995.

[17] D. Mery and O. Medina, "Automated visual inspection of glass bottles using adapted median filtering," *J. Image Anal. Recog.*, vol. 3212, pp. 818–825, 2004.

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

[19] M. Curilem *et al.*, "Pattern recognition applied to seismic signals of the Llaima Volcano (Chile): An analysis of the events' features," *J. Volcanol. Geothermal Res.*, vol. 282, pp. 134–147, 2014.

[20] E. Aguilera, M. Pareschi, M. Rosi, and G. Zanchetta, "Risk from lahars in the northern valleys of Cotopaxi Volcano (Ecuador)," *Natural Hazards*, vol. 33, no. 2, pp. 161–189, 2004.

[21] I. Molina, H. Kumagai, A. García-Aristizábal, M. Nakano, and P. Mothes, "Source process of very-long-period events accompanying long-period signals at Cotopaxi Volcano, Ecuador," *J. Volcanol. Geothermal Res.*, vol. 176, no. 1, pp. 119–133, 2008.

[22] H. D. Ortiz Erazo, "Estudio de los Efectos de Sitio Para la Construcción de un Índice de Actividad Sísmica en el Volcán Cotopaxi," M.S. thesis, Dept. of Phys., Escuela Politécnica Nacional, Quito, Ecuador, 2013.

[23] L. Gutiérrez *et al.*, "Volcano-seismic signal detection and classification processing using Hidden Markov Models. Application to San Cristóbal Volcano, Nicaragua," in *Proc. IEEE Symp. Geosci. Remote Sens.*, 2009, vol. 4, pp. 522–525.

[24] E. H. A. Laasri, E.-S. Akhouayri, D. Agliz, D. Zonta, and A. Atmani, "A fuzzy expert system for automatic seismic signal classification," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1013–1027, 2015.

[25] M. Orozco, M. E. García, R. P. Duin, and C. G. Castellanos, "Dissimilarity-based classification of seismic signals at Nevado del Ruiz Volcano," *Earth Sci. Res. J.*, vol. 10, no. 2, pp. 57–66, 2006.

[26] D. Cárdenas-Peña, M. Orozco-Alzate, and G. Castellanos-Domínguez, "Selection of time-variant features for earthquake classification at the Nevado del Ruiz Volcano," *Comput. Geosci.*, vol. 51, pp. 293–304, 2013.

[27] P. A. Castro-Cabrera *et al.*, "A comparison between time-frequency and cepstral feature representations for the classification of seismic-volcanic signals," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. New York, NY, USA: Springer-Verlag, 2014, pp. 440–447.

[28] 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. Geothermal Res.*, vol. 180, no. 1, pp. 1–8, 2009.

[29] M. Ibs-von Seht, "Detection and identification of seismic signals recorded at Krakatau Volcano (Indonesia) using artificial neural networks," *J. Volcanol. Geothermal Res.*, vol. 176, no. 4, pp. 448–456, 2008.

[30] I. Álvarez, L. García, G. Cortés, C. Benítez, and A. De la Torre, "Discriminative feature selection for automatic classification of volcano-seismic signals," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 151–155, Mar. 2012.

[31] R. Avesani, A. Azzoni, M. Bicego, and M. Orozco-Alzate, "Automatic classification of volcanic earthquakes in HMM-induced vector spaces," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. New York, NY, USA: Springer-Verlag, 2012, pp. 640–647.

[32] M. Bicego, C. Acosta-Muñoz, 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, Jun. 2013.

[33] C. San-Martín *et al.*, "Feature extraction using circular statistics applied to volcano monitoring," in *Progress Pattern Recog., Image Anal., Comput. Vis., Appl.* New York, NY, USA: Springer-Verlag, 2010, pp. 458–466.

[34] R. Lara-Cueva, D. Benítez, E. Carrera, M. Ruiz, and J. Rojo-Álvarez, "Feature selection of seismic waveforms for long period event detection

at Cotopaxi Volcano," *J. Volcanol. Geothermal Res.*, vol. 316, pp. 34–49, 2016.
[35] B. Kenneth, *The Seismic Wavefield, Volume I: Introduction and Theoretical Development*. Cambridge, MA, USA: Cambridge Univ. Press, 2001.
[36] J. Lahr, B. Chouet, C. Stephens, J. Power, and R. Page, "Earthquake classification, location, and error analysis in a volcanic environment: Implications for the magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska," *J. Volcanol. Geothermal Res.*, vol. 62, no. 1, pp. 137–151, 1994.
[37] S. M. Kay and S. L. Marple, Jr., "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.
[38] R. Lara-Cueva, P. Bernal, G. Saltos, D. Benítez, and J. L. Rojo-Álvarez, "Time and frequency feature selection for seismic events from Cotopaxi Volcano," in *Proc. IEEE APCASE*, 2015, pp. 1–6.
[39] L. Grafakos, *Classical and Modern Fourier Analysis*. Upper-Saddle River, NJ, USA: Prentice-Hall, 2004.
[40] C. M. Bishop *et al.*, *Pattern Recognition and Machine Learning*, vol. 1. New York, NY, USA: Springer-Verlag, 2006.
[41] B. Scholkopf *et al.*, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2758–2765, Nov. 1997.
[42] G. Dougherty, *Pattern Recognition and Classification: An Introduction*. New York, NY, USA: Springer-Verlag, 2012.
[43] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.
[44] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines: and Other Kernel-Based Learning Methods*. New York, NY, USA: Cambridge Univ. Press, 2000.
[45] B. Schölkopf and A. Smola, *Learning With Kernels*, 1st ed. Cambridge, MA, USA: MIT Press, 2002.
[46] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, no. 1–3, pp. 389–422, 2002.
[47] G. H. John *et al.*, "Irrelevant features and the subset selection problem," in *Proc. 11th Int. Conf. Mach. Learn.*, 1994, pp. 121–129.
[48] R. Kohavi and D. Sommerfield, "Feature subset selection using the wrapper method: Overfitting and dynamic search space topology," in *Proc. KDD*, 1995, pp. 192–197.

Román A. Lara-Cueva (M'08) received the B.Eng. degree in electronic and telecommunications engineering from Escuela Politécnica Nacional, Quito, Ecuador, in 2001; the M.Sc. degree in wireless systems and related technologies from Politecnico di Torino, Turin, Italy, in 2005; and the M.Sc. and the Ph.D. degrees in telecommunication networks for developing countries from Rey Juan Carlos University, Fuenlabrada, Spain, in 2010 and 2015, respectively.

In 2002, he joined the Departamento de Eléctrica y Electrónica, Universidad de las Fuerzas Armadas-ESPE, Sangolquí, Ecuador, where he has been an Associate Professor since 2005. He is the author or coauthor in 11 publicly funded research projects, and directed 6 of them. His research interests include digital signal processing, smart cities, wireless systems, and machine learning theory.

Diego S. Benítez (SM'14) received the B.Eng. degree in electrical engineering from Escuela Politécnica Nacional, Quito, Ecuador, in 1994 and the M.S. and the Ph.D. degrees in electrical engineering from the University of Manchester, Manchester, U.K., in 1997 and 2001, respectively.

From 2005 to 2007, he was a Postdoctoral Research Associate with the University of Manchester. From 2007 to 2012, he was a Senior Research Engineer with the Bosch Research and Technology Center, USA, and an Academic Visitor with the Institute for Complex Engineered Systems and the INFERLab, Carnegie Mellon University, Pittsburgh, PA, USA. From 2012 to 2014, he was a Visiting Research Scholar at the Universidad de las Fuerzas Armadas-ESPE under the "Prometeo Program" of SENESCYT, Ecuador. He is currently a Full-Time Faculty with the Colegio de Ciencias e Ingenierías El Politécnico, Universidad San Francisco de Quito, Quito. His research interests include intelligent instrumentation and measurement systems for energy management, security, and smart buildings applications.

Enrique V. Carrera (M'99) received the B.E. degree in electronic engineering from the Univesidad de las Fuerzas Armadas-ESPE, Sangolquí, Ecuador, in 1992, the M.Sc. degree in electrical engineering from the Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil, in 1996, and the Ph.D. degree in computer engineering from the Federal University of Rio de Janeiro, Rio de Janeiro, in 1999.

From 2000 to 2004, he was a Postdoctoral Associate with the Department of Computer Science, Rutgers University, New Brunswick, NJ, USA. He is currently a Full Professor with the Departamento de Eléctrica y Electrónica, Universidad de las Fuerzas Armadas-ESPE, Sangolquí. His research interests include context-aware systems and distributed mobile computing.

Mario Ruiz received the B.Eng. degree in geophysical engineering from the Escuela Politécnica Nacional, Quito, Ecuador, in 1992, the M.Sc. degree from the New Mexico Institute of Mining and Technology, Socorro, NM, USA, in 2004, and the Ph.D. degree in geological sciences from the University of North Carolina, Chapel Hill, NC, USA, in 2007.

He currently works as the Chair with the Instituto Geofísico, Politécnica Nacional, Quito. His research interests include volcano seismology.

José Luis Rojo-Álvarez (SM'12) received the B.Sc. degree in telecommunication engineering from the University of Vigo, Galicia, Spain, in 1996 and the Ph.D. degree in telecommunications engineering from the Polytechnic University of Madrid, Madrid, Spain, in 2000.

In 2015, he was with Persei Vivarium, Spain, as a Chief Scientific Officer for building bridges between academia and industry for scientific results transfer. Since 2016, he has been a Full Professor with the Department of Signal Theory and Communications, Rey Juan Carlos University, Madrid. He is the author or coauthor of more than 85 indexed papers and more than 150 national and international conference communications. His current research interests include statistical learning theory, digital signal processing and machine learning, complex system modeling, and electronic health recordings.