Combining Filter-Based Feature Selection Methods and Gaussian Mixture Model for the Classification of Seismic Events From Cotopaxi Volcano

Pablo Venegas, Noel Pérez, *Member, IEEE*, Diego Benítez, *Senior Member, IEEE*, Román Lara-Cueva, *Senior Member, IEEE*, and Mario Ruiz

***Abstract*—This paper proposes an exhaustive evaluation of five different filter-based feature selection methods in combination with a Gaussian mixture model classifier for the classification of long-period (LP) and volcano-tectonic (VT) seismic events recorded at Cotopaxi volcano in Ecuador. The experimentation included both exploring and ranking search spaces of seismic-signal-based features, and selecting subsets of optimal features for event classification. The evaluation was carried out by using an experimental dataset formed by 587 LP and 81 VT feature vectors, each composed of 84 statistical, temporal, spectral, and scale-domain features extracted from the original seismic signals. The best result in accuracy, precision, recall, and processing time for LP seismic event classification was obtained by using the Chi2 discretization method with five features, achieving 95.62%, 99.08%, 95.94%, and 3.7 ms, respectively, whereas for VT seismic event classification, the uFilter method with five features reached the scores of 96.71%, 85.23%, 96.00%, and 4.1 ms, respectively. For the classification of both seismic events simultaneously, the uFilter method with five features yielded 96.70%, 97.77%, 96.7%, and 4.1 ms, respectively. According to the Wilcoxon statistical test, these classification schemes provide competitive seismic event classification, while reducing the required processing time.**

***Index Terms*—Feature selection methods, Gaussian mixture model (GMM) classifier, redundancy analysis, seismic events classification.**

I. Introduction

NOWADAYS, developing technologies and systems capable of helping scientists for forewarning the population in the case of volcanic eruptions is of great importance for saving human lives and minimizing possible consequences. In such a sense, the analysis and recognition of seismic signals is an essential activity for studying the volcanic dynamic process and its intrinsic structure. Currently, volcano observatories deploy a large number of monitoring sensor networks (broadband and short-period seismometers) capable of detecting low-intensity seismic signals or microseisms, where each seismogram may contain several types of seismic signals, such as long-period (LP), volcano-tectonic (VT), tremors (TRE), very long-period (VLP), explosion (EXP), and hybrid (HYB) seismic events [1]. Given the high sensibility of the recording instruments, in some cases, signals of nonvolcanic origin, such as icequakes, lightnings, or tectonic earthquakes, can also be recorded by the seismograph. In this paper, we would focus our study to classify LP and VT seismic signal events, since they are critical for helping to forecast eruptions [2], and they are the main types of events recorded at Cotopaxi [3]. LP events have spectral content mostly restricted to a narrow band between 2 and 5 Hz and have typical time duration shorter than 90 s, whereas VT events have higher frequency content (>6 Hz), and its time duration is usually less than 30 s [4]. Examples of LP and VT seismic signals and their respective spectrograms are illustrated in Fig. 1.

In this sense, several systems for automated analysis of microseisms have been developed and proposed in the literature to detect [5], classify [6]–[9], or even to find the most appropriate subset of features inside a feature space computed from seismic signals of volcanic origin [4], [6], [10]–[12]. Detection approaches like the one developed in [5] focused on detecting a valid seismic event over background noise in the seismogram.

Classification-based studies, on the other hand, use machine learning techniques, such as artificial neural network (ANN) [6]–[8], hidden Markov models (HMMs) [9], etc., and mostly concentrate their effort in constructing a more accurate and robust method; however, the algorithm complexity and the required processing time are always a matter of concern. Other research works have focused on determining the most relevant features for the seismic event classification problem, for example, in [10], discriminative feature selection techniques were used with a Gaussian mixture model (GMM)-based classifier to select the features with more discriminatory information between different types of seismic events extracted from the Colima volcano in Mexico. In [11], a method that used feature selection strategies in combination with a k-nearest neighbor

Manuscript received March 1, 2019; accepted April 30, 2019. Date of publication May 28, 2019; date of current version July 17, 2019. This work was supported in part by the Universidad San Francisco de Quito USFQ (Poligrants Program) under Grants 10100 and 12494, and in part by the Universidad de las Fuerzas Armadas ESPE under Grants 2013-PIT-014 and 2016-EXT-038. *(Corresponding author: Diego Benítez.)*

P. Venegas, N. Pérez, and D. Benítez are with the Colegio de Ciencias e Ingenierías "El Politécnico," Universidad San Francisco de Quito, Quito 17-1200-841, Ecuador (e-mail: pablo.venegas@estud.usfq.edu.ec; nperez@usfq.edu.ec; dbenitez@usfq.edu.ec).

R. Lara-Cueva is with the Grupo de Investigación en Sistemas Inteligentes (WiCOM-Energy) and the Centro de Investigaciones de Redes Ad-Hoc, Departamento de Eléctrica, Electrónica y Telecomunicaciones, Universidad de las Fuerzas Armadas ESPE, Sangolquí 171-5-231B, Ecuador (e-mail: ralara@espe.edu.ec).

M. Ruiz is with Instituto Geofísico, Escuela Politécnica Nacional, Quito 170517, Ecuador (e-mail: mruiz@igepn.edu.ec).

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/JSTARS.2019.2916045

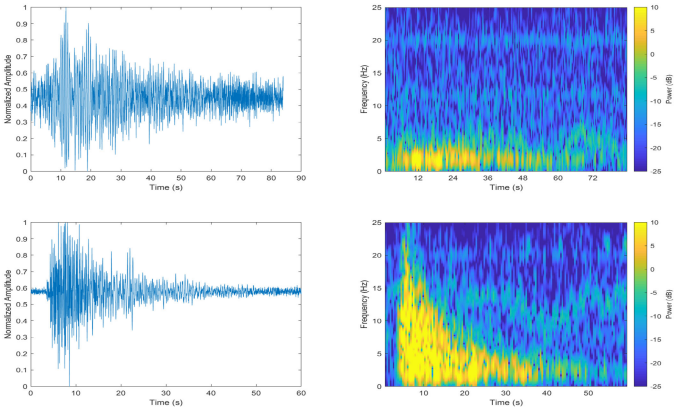

Fig. 1. Example of seismic signals and their spectrogram for LP (top row) and VT (bottom row) events originated at Cotopaxi volcano.

(kNN) classifier was proposed for considering the relevance of time-variant features extracted from the Nevado del Ruiz volcano in Colombia. Similarly, in [4] and [12], the use of feature selection techniques in combination with kNN and decision tree (DT) classifiers was proposed for the classification of seismic events from the Cotopaxi volcano in Ecuador, respectively. The use of evolutionary computation techniques such as genetic algorithms [6] has also been explored for selecting a relevant subset of features, for example, in [7], a genetic algorithm was used in conjunction with an ANN classifier to classify seismic signals from the Villarrica volcano in Chile and, thus, to determine the most adequate set of features for the classification. Nevertheless, evolutionary computation techniques have some limitations such as requiring a high computational cost, since they usually involve a large number of iterations, and being unstable, since the algorithms often select different features from different runs [13]. Despite good results obtained in the past, determining the most appropriate subset of features for rapid and accurate seismic event classification remains a difficult task. Usually, taking a satisfactory subset of features, instead of the optimal one, leads to a good decision, but studies about feature selection strategies for this problem are scarce.

The selection of an optimal subset of features is related to the reduction or exploration of the feature space. Reduction techniques involve the use of methods based on the transformation of the original space, such as the fast Fourier transform (FFT) to the frequency domain, wavelet transform to the scale domain, principal components analysis (PCA) to the orthogonal main components that express the data difference of the data matrix, etc. In all cases, the transformation of the original space requires a large amount of computation time from standard computers [workstations without graphics processing unit (GPU)], which is not suitable for the development of real-time applications. On the other hand, exploring the feature space is linked to feature selection techniques, which are represented by three different paradigms, according to the type of feature searching algorithm: wrapper and embedded methods, which incorporate a machine learning classifier (MLC) for deciding the merit of features, and filter methods, which use the data characteristics as main heuristics rather than MLCs to assess the features importance [14]. The feature space is explored exhaustively in all cases, which increases the processing time consumed by the model.

The main objective of our long-term work roadmap is to develop a framework for a real-time volcano early warning system. In this context, the work presented in this paper aims to study the behavior of five different feature selection methods: Information Gain (IG) [15], [16], One Rule (1R) [16], [17], RELIEF [18], Chi2 Discretization [16], [19], and uFilter [20], in combination with a GMM [21] for classifying LP and VT seismic events, as quickly as possible, in an experimental dataset of features extracted from seismic signals collected at the Cotopaxi volcano in Ecuador. Although there are several studies on seismological events at Cotopaxi volcano [4], [12], to our best knowledge, this is the first attempt to comprehensively evaluate filter-based feature selection methods for the automatic classification of seismic events from that volcano.

The considered methods belong to the filter paradigm (independently of the MLC) and, since they avoid the MLCs training-test phase, are less complex and perform much faster than wrapper (as in [4], [6]–[12], and [22]) or embedded methods [14], or any dimension reducing procedure like the PCA. The latter not only increases the running time of the model by computing the singular value decomposition of the data matrix, but also the computed principal components do not safely point out which variables are the most important in terms of conserving the information. Instead of that, PCA identifies a subset of variables related to the first few principal components, which could summarize the relevant information expressed by the entire dataset [23]. According to the work presented in [24], this procedure assumes that vectors with larger magnitude coefficients represent the most appropriate linear combination (principal component), which could be incorrect, since this assumption ignores the relative position (variable correlations) and magnitude (standard deviation of variables) of vectors.

The selection of the GMM, as the chosen classifier for this study, was based on two heuristic reasons: 1) the GMM statistical nature, which could lead to a good matching with the selected filter-based feature selection methods; and 2) good performance results previously obtained by the GMM in speaker recognition problems. Some approaches have found that voice and seismic signals share similar behavior at some frequency threshold [25], [26]. Others MLCs approaches like HMM have also reported good voice and seismic signal classification results [9], but the computation of the Viterbi algorithm for finding the most probable states of the HMM in the training phase, using a forward-based strategy for efficiently computing the probability of a set of observations, demands a considerable amount of computation time, which is proportional to the increase in the input size [27].

The remainder of this paper is organized as follows. Section II describes the seismic events dataset used in this study, the five filter-based feature selection methods previously mentioned, the GMM classifier, and the experimental methodology used. Section III covers and discusses the experimental results obtained from the selection of the best classification scheme for each scenario (before and after the redundancy analysis) as well as from the global comparison against other previously

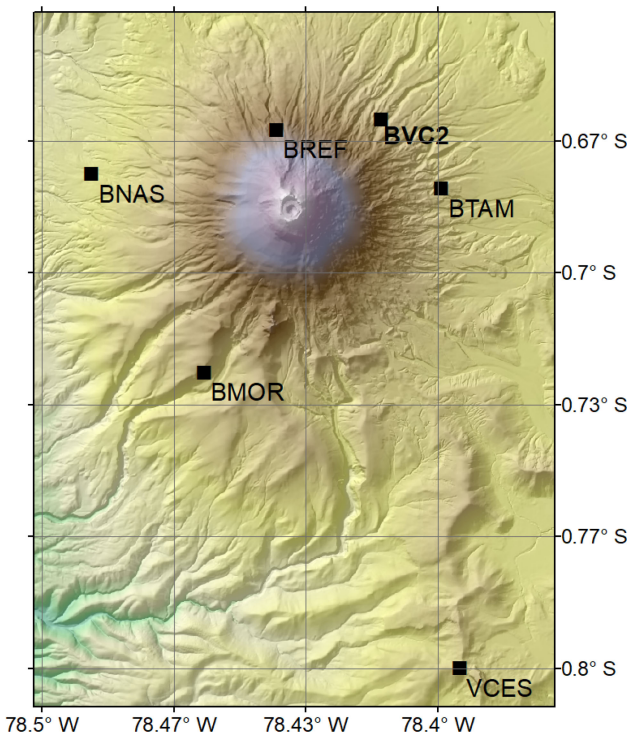

Fig. 2. Network of permanent broad band seismic stations around Cotopaxi volcano; the dataset for this study has been recorded at the BVC2 broadband seismic station. Image provided by IGEPN.

developed methods. Finally, in Section IV, we outline the principal achievements of this paper and our plans for future work.

II. Materials and Methods

A. Database

The database used in this paper corresponds to several seismic events recorded at the Cotopaxi volcano located in the Andean mountain region of Ecuador (latitude 0°41'05" S and longitude 78°25'54.8" W). Cotopaxi is an active snow-capped volcano constantly monitored by the Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN), institution responsible for monitoring and analyzing volcanic activity in Ecuador. Monitoring of seismic activity at Cotopaxi is based mainly on a high-resolution network of six broadband seismic stations (BREF, BVC2, BTAM, BNAS, BMOR, and VCES) with a frequency response range of 0.1–50 Hz [28], located above the volcano flanks, as illustrated in Fig. 2. Seismic events arrive at different seismological stations with different characteristics, since the internal routes within the volcano are not homogeneous, so we decided to work with the data provided by one of the most stable broadband stations (BVC2), which is located at just 3 km from the Cotopaxi crater.

The database used in this study consisted of a total number of 668 seismic records (587 LP and 81 VT). The seismograms were recorded between the months of January and March 2009 at the BVC2 station by using a three-axial seismometer (CMG-40T Guralp) with a sensitivity of 1600 V/m·s^{-1}, and were sampled at 50 Hz using a 24-bit analog-to-digital converter (Geotech Smart 24D digitizer). The acquisition system used the STA/LTA algorithm [29] to detect seismic events; seismograms are recorded in separated files of 12 000 s of duration. The labeling and identification of events was done manually by analysts at the IGEPN using the signals from all the available stations considering the waveform, spectrogram, and location.

Inside the database, each seismic record contains one independent microseism signal, which was used for creating a benchmark dataset containing a set of 84 features computed from the time (13 features), frequency (21 features), and scale domains (50 features). The frequency and scale domains were obtained by the application of the FFT and wavelet transform, respectively. A brief description of the employed features is shown in Table XIII in the Appendix. More detailed information about the features and their calculation can be found in [4], [12], and [30]. For the purpose of this study, one experimental dataset was created, containing a total of 587 LP and 81 VT feature vectors (the maximum number of available instances).

B. Feature Selection Methods

As mentioned before, five different low-complexity feature selection methods were evaluated in this paper. A brief description of each of these methods is given as follows.

1) Information Gain: The IG measurement, normalized with the symmetrical uncertainty coefficient [15], [16], is a symmetrical measure, in which the amount of information gained about Y after observing X is equal to the amount of information gained about X after observing Y (a measure of feature–feature intercorrelation). This model is used to estimate the value of an attribute Y for a novel sample (drawn from the same distribution as the training data) and to compensate for IG bias toward attributes with more values.

2) One Rule: This method estimates the predictive accuracy of features by creating single-feature-based rules (similar to a single-level DT) [16], [17]. Thus, it is possible to calculate a classification accuracy for each rule (feature). After that, a ranked list of features is obtained according to their classification scores. This approach is unusual, due to the fact that no search is conducted over the feature space. However, experiments choosing the highest ranked features to feed a common MLC have shown that, on average, when using only the top ten features, the obtained results are as accurate as with other filter methods that perform the feature space exploration [20].

3) RELIEF: This method uses instance-based learning to assign a relevance weight to each feature [16], [18]. The weight of each feature reflects its ability to distinguish between class values. Feature weights are updated according to how well their values distinguish the sampled instance from its nearest hit (instance of the same class) and nearest miss (instance of a opposite class). The feature will receive a high weight if it differentiates between instances from different classes and has the same value for instances of the same class. For nominal features, it is defined as either 1 (the values are different) or 0 (the values are the same), whereas for numeric features, the difference is the actual difference normalized in the interval [0..1].

4) Chi2 Discretization: This method consists of a justified heuristic for supervised discretization [16], [19]. Numerical features are initially sorted by placing each observed value into its own interval. Then, the chi-square statistic is used to determine whether the relative frequencies of the classes in adjacent intervals are similar enough to justify merging. The extent of the merging process is controlled by an automatically set chi-square threshold. The threshold is determined by attempting to maintain the fidelity of the original data.

5) uFilter: This method is based on the nonparametric Mann–Whitney U-test [20]; thus, it tests whether two independent observation samples are drawn from the same or identical distributions. The basic idea is that a particular pattern exhibited when *m* number of X random variables, and *n* number of Y random variables, are arranged together in increasing order of magnitude, therefore providing information about the relationship between their parent populations. The baseline method is improved by computing one Z-score for each class and assigning the final weight to the feature based on the computation of the absolute value of the numerical difference between Z scores.

C. GMM Classifier

The GMM is a probabilistic model that represents subpopulations within a multimodal global population. The GMM inherits from the Gaussian distribution its calculation efficiency and fast computing speed; for this reason, it is used for speech recognition [31] and object tracking [32]. Given the similarity between the voice signals and the seismic signals [25], [26], the use of the GMM as a classifier for seismic events is proposed here.

The GMM is made up of two types of values: the mixtures weights and the variables of each component (subpopulation). The variables of each k-component are the mean ($\vec{\mu}_k$) and the covariance matrix (Σ_k). The mixture weight (ϕ_k) represents the percentage of presence of a k-component within the total population [21].

The formulation of the Gaussian density function of a single component is given by

$$\mathcal{N}(\vec{x} \mid \vec{\mu}_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_i|}} \times \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu}_i)^{\mathrm{T}} {\Sigma_i}^{-1} (\vec{x} - \vec{\mu}_i)\right) \quad (1)$$

where $\vec{x}$ is the set of data vectors with D dimension (number of variables).

The sum of all the mixture weights of all the k-components of the GMM is normalized to 1, fulfilling the constraint given by

$$\sum_{i=1}^{k} \phi_i = 1. \quad (2)$$

The GMM is given by the sum of the Gaussian density functions of all the components; this is represented by

$$p(\vec{x}) = \sum_{i=1}^{k} \phi_i \mathcal{N}(\vec{x} \mid \vec{\mu}_i, \Sigma_i) \quad (3)$$

where k is the total number of components within the population.

The GMM fits data using an iterative algorithm known as the expectation–maximization algorithm. Initially, random values are established for the mean ($\vec{\mu}_k$), the covariance matrix (Σ_k), and mixture weight (ϕ_k) of each component; then, the expectation–maximization algorithm performs the following steps.

1) *Expectation step:* The algorithm calculates the posterior probability (expectation) of each data point x_i belonging to one of the k-components that make up the GMM, given the model parameters $\vec{\mu}_k$, Σ_k, and ϕ_k. This step is expressed as

$$\hat{\gamma}_{ik} = \frac{\hat{\phi}_k \mathcal{N}(x_i \mid \hat{\mu}_k, \hat{\Sigma}_k)}{\sum_{j=1}^{K} \hat{\phi}_j \mathcal{N}(x_i \mid \hat{\mu}_j, \hat{\Sigma}_j)} \quad (4)$$

where $\hat{\gamma}_{ik}$ is the probability that x_i belongs to a k-component.

2) *Maximization step:* By using the component-membership posterior probability (expectation) calculated in the expectation step, the algorithm updates the model parameter values $\vec{\mu}_k$, Σ_k, and ϕ_k by applying maximum likelihood. The maximization equations for the model parameters are

$$\hat{\phi}_k = \sum_{i=1}^{N} \frac{\hat{\gamma}_{ik}}{N} \quad (5)$$

$$\hat{\mu}_k = \frac{\sum_{i=1}^{N} \hat{\gamma}_{ik} x_i}{\sum_{i=1}^{N} \hat{\gamma}_{ik}} \quad (6)$$

$$\hat{\Sigma}_k = \frac{\sum_{i=1}^{N} \hat{\gamma}_{ik} (x_i - \hat{\mu}_k)^2}{\sum_{i=1}^{N} \hat{\gamma}_{ik}} \quad (7)$$

where N is the total number of data points.

The algorithm then iterates over these two steps until convergence [33].

For the problem at hand, we used the GMM classifier with parameter $k = 2$, meaning the LP and VT classes. The covariance matrix parameter (Σ_k for $k = 1..2$) was set to diagonal constraint [34]. This configuration has shown to be satisfactory (based on previous experimentation) when compared to the full constraint.

D. Experimentation

This section outlines the experimental evaluation of the selected filter-based feature selection methods in combination with a GMM classifier on the experimental dataset, for classifying first LP events, then VT events, and, finally, both seismic events together. The overall procedure involved the following steps.

1) Normalizing the values of the experimental dataset using the min–max normalization method [35], in order to bring all values into the range [0, 1] (see step 1 in Fig. 3).

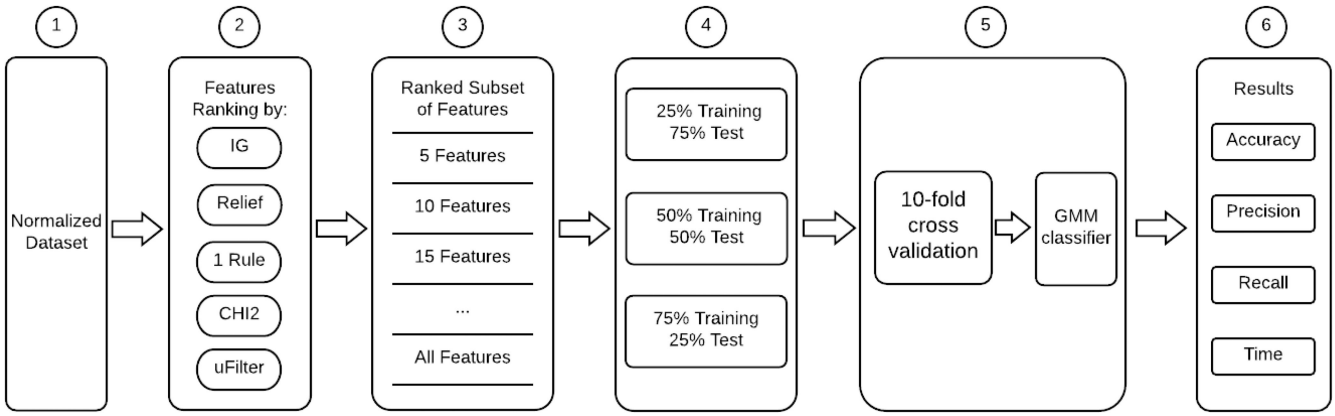

Fig. 3. Applied experimental workflow.

2) Applying the five filter-based feature selection methods, IG [15], [16], Relief [16], [18], 1R [16], [17], CHI2 [16], [19], and uFilter [20], on the experimental dataset to produce five different features rankings (see step 2 in Fig. 3).
3) Creating several ranked subsets of features from the rankings formed in the previous step. The number of characteristics included in each subset varies from five to the total number of characteristics, with increments of five as an empirical threshold (see step 3 in Fig. 3).
4) Splitting each ranked subset of features using three different training-test partitions: 25–75%, 50–50%, and 75–25%. Since each ranked subset of features contains both seismic events, the partition's percentage was applied to each single event, i.e., 25% of LP vectors plus 25% of VT vectors formed the test set of the 25–75% split scenario. This process was performed until forming all the training-test sets (see step 4 in Fig. 3).
5) Applying a tenfold cross-validation (CV) method [36] on formed training-test sets before training and classification steps of the GMM classifier. Therefore, samples will not appear simultaneously in the training and test dataset, thus guaranteeing disjoint test partitions and preventing the classifier from overfitting [36]. The number of times of the tenfold CV method application was empirically selected for gathering the mean of performance results and standard deviation over 100 runs (see step 5 in Fig. 3).
6) Establishing a comparative analysis based on metrics such as accuracy, precision, recall, and processing time scores, in order to select the best classification schemes for LP, VT, and both seismic events together (see step 6 in Fig. 3). All comparisons were made using the Wilcoxon statistical test [37], [38] in order to evaluate the statistical difference between classification schemes. In the case of a tie (according to the statistical test) among the classification schemes, the criterion used to select the most appropriate scheme was to select the one with the least number of features and, therefore, the shortest processing time required. It should be noted that the accuracy, precision, and recall metrics were calculated individually according to the type of events: for LP or VT events (as disjoint units), the output of the confusion matrix was used, whereas for the case of both events together (as a binary classification problem), the use of the weighted average technique [39] based on the confusion matrix was mandatory.

TABLE I
LP EVENT CLASSIFICATION PERFORMANCE RESULTS IN THE 25–75% TRAINING-TEST SPLIT SCENARIO

Best Scheme				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (ms)
uFilter+5f	**96,71**	**99,48**	**97,00**	**4,1**
Other Schemes				
IG+5f	96,04	99,55	95,98	4,1

Acc.: Accuracy; Pre.: Precision; Rec.: Recall; f: Features.

In the last step of the experiment, a redundancy analysis was also carried out on the most appropriate classification schemes using the Pearson correlation [40], in order to separate the redundant features from the relevant ones and, thus, to produce the final optimal subset of features.

III. RESULTS AND DISCUSSION

A total of 240 ranked subsets of features containing time, frequency, and scale-based features computed from segmented seismic signals were analyzed using the proposed method. The statistical comparison based on the mean of accuracy, precision, recall, and processing time required over 100 runs highlighted interesting results for classifying LP, VT, and both seismic events together on the experimental datasets using three different training-test split scenarios. For better interpretation of the obtained results, we only included the set of classification schemes that did not represent statistical difference in terms of performance (considering the assessment metrics) among them.

A. LP Event Classification

The highest metric scores in the 25–75% training-test split scenario were obtained by using the GMM classifier in combination with the uFilter feature selection method, with five features. This result was statistically superior to the majority of evaluated schemes. However, there was another combination with similar statistically performance at $p = 0.05$. According to the results shown in Table I, it is possible to consider both

TABLE II
LP EVENT CLASSIFICATION PERFORMANCE RESULTS IN THE 50–50% TRAINING-TEST SPLIT SCENARIO

Best Scheme				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (ms)
Chi2+5f	**95,63**	**99,08**	**95,94**	**3,7**
Other Schemes				
1R+5f	95,21	99,21	95,34	3,8
Chi2+15f	94,88	99,44	94,72	4,6

Acc.: Accuracy; Pre.: Precision; Rec.: Recall; f: Features.

TABLE III
LP EVENT CLASSIFICATION PERFORMANCE RESULTS IN THE 75–25% TRAINING-TEST SPLIT SCENARIO

Best Scheme				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (ms)
1R+5f	**94,47**	**98,46**	**95,23**	**4,8**
Other Schemes				
uFilter+5f	93,65	98,42	94,32	5,2

Acc.: Accuracy; Pre.: Precision; Rec.: Recall; f: Features.

combinations as the most appropriate classification schemes in this scenario. Thus, we selected the GMM classifier and the uFilter method with five features.

In the 50–50% training-test split scenario, the highest metric scores were obtained by the GMM classifier using the Chi2 discretization method with five features (see Table II). Although there were other combinations with similar performance at $p = 0.05$, the one with the shortest processing time was selected as the most appropriate classification scheme in this scenario.

For the 75–25% training-test split scenario, the highest metric scores were achieved by the combination of the GMM classifier with the 1R method using five features, as shown in Table III. This result showed a statistical superiority compared to other combinations. However, the scheme using the uFilter method with five features revealed to be statistically similar ($p = 0.05$) in terms of performance, although the processing time required was the worst obtained. Therefore, the most appropriate classification scheme in this scenario was the one with the highest metric performance.

Regarding the LP events classification, it seems that the GMM classifier matched very well with three different feature selection methods: the uFilter (see Table I), Chi2 discretization (see Table II), and 1R methods (see Table III). Therefore, any of this methods could be selected as the most appropriate classification scheme for classifying LP events. However, the Chi2 discretization in the 50–50% training-test split scenario performed slightly better in terms of processing time, ensuring it as the best selection.

B. VT Event Classification

The highest metric scores in the 25–75% training-test split scenario were obtained by the GMM classifier in combination with the uFilter method with five features (see Table IV). This result was statistically superior to the majority of remaining classification schemes. However, there was another scheme formed by the IG method with five features that performed statistically

TABLE IV
VT EVENT CLASSIFICATION PERFORMANCE RESULTS IN THE 25–75% TRAINING-TEST SPLIT SCENARIO

Best Scheme				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (ms)
uFiler+5f	**96,71**	**85,23**	**96,00**	**4,1**
Other Schemes				
IG+5f	96,04	82,23	96,50	4,1

Acc.: Accuracy; Pre.: Precision; Rec.: Recall; f: Features.

TABLE V
VT EVENT CLASSIFICATION PERFORMANCE RESULTS IN THE 50–50% TRAINING-TEST SPLIT SCENARIO

Best Scheme				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (ms)
Chi2+5f	**95,63**	**78,78**	**93,25**	**3,7**
Other Schemes				
1R+5f	95,21	76,27	94,30	3,8
Chi2+15f	94,88	74,57	96,00	4,6

Acc.: Accuracy; Pre.: Precision; Rec.: Recall; f: Features.

TABLE VI
VT EVENT CLASSIFICATION PERFORMANCE RESULTS IN THE 75–25% TRAINING-TEST SPLIT SCENARIO

Best Scheme				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (ms)
1R+10f	**93,35**	**69,22**	**91,29**	**4,8**
Other Schemes				
1R+15f	93,54	69,18	91,47	4,3

Acc.: Accuracy; Pre.: Precision; Rec.: Recall; f: Features.

similar at $p = 0.05$. Despite the good performance reached by both classification schemes, the selected scheme using the uFilter method remains as the most appropriate for this scenario.

For the 50–50% training-test split scenario, the highest metric scores were obtained by the GMM classifier using the Chi2 discretization method with five features (see Table V). The results obtained for this combination outperformed the majority of evaluated classification schemes. However, there were two other configurations (one with the 1R method and another with the Chi2 discretization method) that also reached satisfactory results without losing statistical significance at $p = 0.05$, but with an increased processing time required. Therefore, the Chi2 discretization method with five features stands as the most appropriate classification scheme for this scenario.

In the case of the 75–25% training-test split scenario, the GMM classifier using the 1R method produced the most effective classification schemes. The combinations using a total number of ten and 15 features exceeded the remaining combinations (see Table VI). However, the first classification scheme was selected as the most appropriate in this scenario because it reached a higher performance, while decreasing the number of features used in the final subset.

Regarding VT event classification, good performance obtained by the GMM classifier using the uFilter, Chi2 discretization, and 1R methods demonstrates a satisfactory example of the generalization of the classifier when combined with filter-based feature selection methods. However, the combination in the

TABLE VII
LP AND VT EVENT CLASSIFICATION PERFORMANCE RESULTS IN THE 25–75% TRAINING-TEST SPLIT SCENARIO

Best Scheme				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (ms)
uFilter+5f	**96,71**	**97,77**	**96,70**	**4,1**
Other Schemes				
IG+5f	96,04	97,47	96,04	4,1

Acc.: Accuracy; Pre.: Precision; Rec.: Recall; f: Features.

TABLE VIII
LP AND VT EVENT CLASSIFICATION PERFORMANCE RESULTS IN THE 50–50% TRAINING-TEST SPLIT SCENARIO

Best Scheme				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (ms)
uFilter+5f	**95,88**	**96,76**	**95,87**	**4,7**
Other Schemes				
1R+5f	95,21	96,64	95,62	4,7
Chi2+15f	95,63	96,76	95,87	4,9

Acc.: Accuracy; Pre.: Precision; Rec.: Recall; f: Features.

25–75% training-test split scenario (see Table IV) outperformed the remaining classification schemes in terms of precision. Therefore, it was selected as the best method for classifying VT seismic events.

C. Joint LP and VT Events Classification

The results shown in Table VII indicate that the GMM classifier using the uFilter method with five features reached the highest metric scores in the 25–75% training-test split scenario. There was another classification scheme that demonstrated to be statistically similar in terms of performance ($p = 0.05$). However, the GMM classifier using the uFilter method with five features reached the higher scores; therefore, it appeared as the most appropriate classification scheme in this scenario.

Similarly, in the 50–50% training-test split scenario, the highest metric scores were obtained by the GMM classifier using the uFilter method with five features (see Table VIII). This result was statistically superior when compared to other classification schemes ($p = 0.05$). Although the 1R and Chi2 discretization methods with five and 15 features, respectively, achieved similar classification results ($p = 0.05$), the scheme using the Chi2 discretization method was not considered as the most appropriate classification scheme in this scenario due to the total number of features needed for achieving the satisfactory performance. On the other hand, the 1R method performed very well, but slightly lower than the uFilter method. Therefore, the uFilter method was selected as the best one in this scenario.

For the 75–25% training-test split scenario, the GMM classifier and the 1R method with five features achieved the best performance metric. This combination obtained the highest metric scores and statistically overcame the majority of the classification schemes (see Table VII). The combination using the Chi2 discretization method also provided good results without losing statistical importance at $p = 0.05$, but it increased the processing time required. Therefore, the scheme that uses the 1R method was the most appropriate classification scheme for this scenario.

TABLE IX
LP AND VT EVENT CLASSIFICATION PERFORMANCE RESULTS IN THE 75–25% TRAINING-TEST SPLIT SCENARIO

Best Scheme				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (ms)
1R+5f	**94,47**	**95,58**	**94,48**	**4,8**
Other Schemes				
Chi2+5f	93,65	95,05	93,65	5,2

Acc.: Accuracy; Pre.: Precision; Rec.: Recall; f: Features.

Regarding the classification of both seismic events together, the GMM classifier in combination with the uFilter method with five features in the 25–75% training-test split scenario reached the best performance results over most of the scenarios, while improving the required processing time (see Table VII). Therefore, it was the best selection for this scenario.

From the results presented above, the uFilter method achieved the best overall performance in terms of producing good classification schemes (four out of nine possible), followed by the 1R method (three out of nine possible) and the Chi2 discretization method (two out of nine possible). Furthermore, a pattern related to the test-training split scenarios emerged as follows: the 25–75%, 50–50%, and 75–25% training-test split scenarios were dominated by the uFilter (see Tables I, IV, VII, and VIII), Chi2 discretization (see Tables II and V), and 1R (see Tables III, VI, and IX) methods, respectively. These results are extremely related to the intrinsic semantic operation of the feature selection methods and the used dataset. The uFilter method is an improvement of the Mann–Whitney U-Test, where random variables of different classes (two in its initial version) are arranged together in increasing order of magnitude, providing evidence against random mixing, which would tend to discredit the null hypothesis of identical distribution [41]. An advantage of using this method is that the two samples under consideration may not necessarily have the same number of observations, as it was in our case. This could explain the predominance of this method in the 25–75% split scenario. On the other hand, the Chi2 discretization method is based on the X^2 statistical function; therefore, it must comply with the minimum number of observations required to perform the test satisfactorily; this explains its superiority against other methods in the 50–50% split scenario.

Finally, the predominance of the 1R method was expected in the last scenario, since this method is based on the production of rules (one per feature); in this manner, the method takes advantage by using more data for training; therefore, the produced rules were more precise, as well as the selection of features. However, this behavior was not enough for producing the most appropriate classification schemes as the uFilter (in the 25–75% split scenario) or Chi2 discretization (in the 50–50% split scenario) methods did.

On the other hand, since the GMM classifier uses mixed Gaussian distributions to approximate new incoming samples, it is constrained by the statistical assumption of having the minimum samples in the training phase, and training it with the 75% of the data produced satisfactory results, but very often introduced

instability in the leaning model because of overtraining. Furthermore, it needed more iterations for converging to a solution, while increasing the processing time, and sometimes, it failed to reach the solution within the maximum number of iterations allowed.

D. Feature Subset Validation

Filter methods for feature selection are efficient when using high-dimensional data due to their linear time complexity in terms of the dimensionality N (total number of features), but they are unable to remove redundant features because they are individual evaluators, i.e., the weight is assigned according to the degree of relevance [42], and whenever the features are considered relevant for the class. All features will be selected without taking into account the possibility that they may be highly correlated with each other, that is, they could be redundant. Therefore, feature subset validation through a redundancy analysis is very appropriate; the main idea is to reduce the feature subset size, while keeping the prediction accuracy.

In this sense, a two-step procedure was also considered in the analysis: 1) selecting the best subset of features for each scenario (LP, VT, and both seismic events together); and 2) performing a redundancy analysis using the Pearson correlation [40], which will allow us to separate the redundant functions from the relevant ones and, thus, produce the final optimum subset of the features for the experimental dataset.

For a better interpretation of the results, two degrees of feature relevance were used: strong and weak (redundant and nonredundant) [43]. Strong relevance indicates the importance of the feature in such a way that it cannot be removed without losing prediction accuracy. On the other hand, weak relevance refers to a feature that contributes less to the prediction accuracy. Besides, irrelevant features can never contribute to the accuracy of the prediction. Therefore, a subset of features is relevant if it only contains strong and weak characteristics.

As shown in the previous section, the subset of best features selected in each scenario provided discriminant features by eliminating irrelevant features. These subsets were considered as the starting point for the redundancy analysis step; a correlation heuristic value was used as the acceptance threshold (c-Pearson values greater than 0.75 in both positive and negative directions). Table X summarizes the redundancy analysis carried out over the most appropriate subset of features selected in each experimental scenario.

From Table X, it is possible to notice the presence of correlated features in the best feature subsets, which means that there are pairs of features that provide the same information in the group, i.e., they are redundant. Therefore, only one of the features in the correlated pair is selected to form the subset of nonrelevant or weakly relevant features, whereas the uncorrelated features were included in the highly or strongly relevant subset.

As can be seen in Table X, the most important features for the classification of seismic events are f49 (for LP events) and f28 (for VT and both events together). Feature f49 corresponds to the mean frequency value of level 3 detail decomposition of the discrete symlet wavelet transform of the signal, whereas feature f28 corresponds to the ratio between the peak amplitude and the root mean square (rms) of the FFT of the signal.

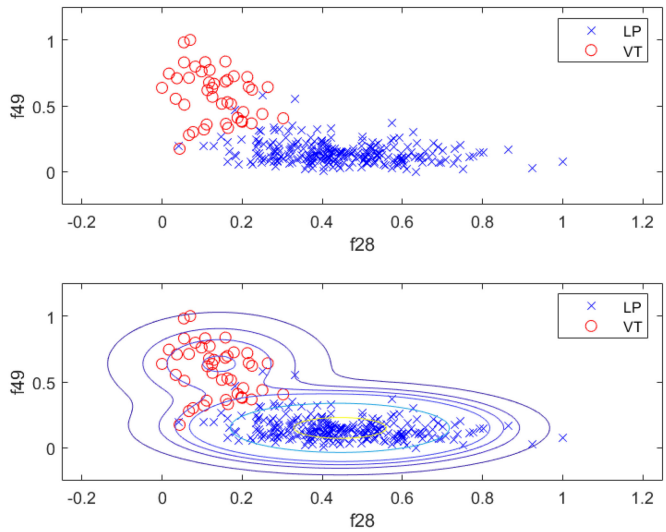

Fig. 4. Strongly relevant feature distribution plot: in the natural feature space (top row) and using the GMM classifier with the expectation–maximization algorithm convergence (bottom row).

These features were considered as strongly relevant features because they are uncorrelated and their absence in the final subset significantly decreased the accuracy, precision, and recall percentage scores as follows: for the classification of LP events, from 95.6257 to 90.9729 ($p < 0.01$), from 99.0805 to 92.3388 ($p < 0.01$), and from 95.9448 to 91.6136 ($p < 0.01$), respectively. Similarly, for VT event classification, these scores were reduced from 96.7059 to 91.2788 ($p < 0.01$), from 85.2333 to 68.7471 ($p < 0.01$), and from 96.0000 to 90.2143 ($p < 0.01$), respectively. Whereas for the classification of both seismic events together, the scores were reduced from 96.7059 to 91.2788 ($p < 0.01$), from 97.7680 to 92.9869 ($p < 0.01$), and from 96.6998 to 90.3057 ($p < 0.01$), respectively. These results were expected, since features f28 and f49 are features taken from the frequency domain, and their values are very different between the seismic event types, which contribute to their differentiation, i.e., the mean value of the f49 feature for LP events is 0.5891, whereas for VT events, it is 0.1488, which represents a value difference of 0.4403. Something similar occurred for the values of the f28 feature. These significant differences can be seen as a clear boundary separation between classes when contrasting each other in the feature space (see the top row in Fig. 4) and when the GMM classifier is used, as illustrated in the bottom row in Fig. 4.

Following the definition of "relevant and irrelevant features" stated in [43], the weakly and strongly relevant features were selected to form the optimal subset of features. These subsets were then used to feed the GMM classifier in order to establish a statistical performance comparison between the best selected subset of features and its corresponding subset of optimal features after the redundancy analysis. Table XI summarizes the results obtained.

TABLE X
SUMMARY OF THE REDUNDANCY FEATURE ANALYSIS

Event Type	Method	Best Features Subset	Redundant Features	c-Pearson	Weakly Relevant	Strongly Relevant
LP	CHI2+5f	f78, f52, f79, f50, f49	f78 = f52, f50, f79	0.98, 0.95, 0.92	f78	f49
VT	UF+5f	f78, f59, f52, f50, f28	f52 = f78, f50, f79	0.98, 0.93, 0.91	f52	f28
LP+VT	UF+5f	f78, f59, f52, f50, f28	f78 = f52, f50, f79	0.98, 0.95, 0.92	f78	f28

c-Pearson: The Pearson correlation value; f: Features.

TABLE XI
STATISTICAL COMPARISON OF FEATURE SUBSETS BASED ON THE USED METRICS BETWEEN THE BEST AND OPTIMAL FEATURE SUBSETS

Event Type	Best Subset	Acc. (%)	Pre. (%)	Rec. (%)	Time (ms)	Optimal Subset	Acc. (%)	Wilcoxon ($\alpha = 0.05$)	Pre. (%)	Wilcoxon ($\alpha = 0.05$)	Rec. (%)	Wilcoxon ($\alpha = 0.05$)	Time (ms)
LP	f78, f52, f79, f50, f49	95.62	99.08	95.94	3.7	f78, f49	94.44	$p < 0.05$	99.00	$p = 0.43$	94.66	$p < 0.05$	3.6
VT	f78, f79, f52, f50, f28	96.71	85.23	96.00	4.1	f52, f28	97.12	$p = 0.84$	87.00	$p = 0.58$	94.50	$p = 0.47$	3.9
LP+VT	f78, f79, f52,f50, f28	96.70	97.77	96.70	4.1	f78, f28	96.86	$p = 0.81$	97.77	$p = 0.66$	96.70	$p = 0.81$	3.2

Acc. = Accuracy; Pre. = Precision; Rec. = Recall; f = Features; The values were rounded to two decimals.

TABLE XII
COMPARISON BASED ON THE ACCURACY, PRECISION, RECALL, AND PROCESSING TIME BETWEEN PREVIOUS WORKS AVAILABLE IN THE LITERATURE AND THE PROPOSED METHOD

Method	Event	No. of Instances	No. of features	Acc. (%)	Pre. (%)	Reca. (%)	Time (ms)
Artificial Neural Network [44]	LP	914	6	97	100	93	230
DT [44]	LP	914	3	96	98	93	31
Multi-Class SVM [45]	LP/VT	300	4	95/91	86/88	95/75	-
Linear SVM [4]	LP	914	5	97	97	96	50
Artificial Neural Networks [6]	LP/VT	2200	6	94/90	81/64	97/96	-
Artificial Neural Networks [7]	LP	1033	8	94	91	96	-
Hidden Markov model [9]	LP/VT	-	5	90/98	94/80	85/98	-
Proposed Method	LP/VT	667	2/2	94*/97*	99*/87*	95*/95*	3.6/3.9

Acc.: Accuracy; Pre.: Precision; Rec.: Recall; f: Features; *Values rounded to the closest integer.

As shown in Table XI, although the number of features was reduced in an optimal feature subset, the performance of the GMM classifier did not increase in the majority of evaluated scenarios according to the assessment metrics. However, the required processing time decreased for the classification of all types of seismic events, which is a very significant result since we are targeting fast seismic event recognition applications. This is due to the minimization of the input vector size before feeding the GMM classifier, which contributes to the GMM model optimization in terms of processing time.

E. Performance Comparison of the Proposed Method Against Other Approaches

The method proposed in this paper focused on the classification of two types of seismic events, LP and VT, whereas previous studies conducted in [4] and [44] were focused only on the classification of a single type of event (LP). By classifying only LP seismic events, those studies reached slightly better performance results than the proposed method, i.e., the accuracy and recall obtained in [4]; using a linear SVM classifier (linear kernel) with five features (obtained using a wrapper-based feature selection method) as input seems to be the best for detecting this type of seismic event; it required 50 ms to fulfill the task. On the other hand, the DT classifier with five features (feature selection obtained based on a combination of wrapper and embedded methods) employed in [44] needed 31 ms to accomplish the task. Despite the good processing time obtained by those approaches, the proposed method achieved its best classification results in 3.6 ms, guarantying the shortest processing time among them. This difference is mainly justified by the fact that previous approaches employed classifier-dependent feature selection methods (as in wrapper or embedded paradigms) instead of filter-based methods (as in the proposed method) for the task of feature selection.

Furthermore, performance comparison of the proposed method (with exception of the processing time variable) against similar seismic classification schemes previously developed in [6], [9], and [45], which use multiclass SVM, ANN, and HMM classifiers, respectively, with four to six features for classifying LP and VT seismic events as disjoint units, highlighted that the proposed method provides competitive results in almost all metrics for both types of seismic events (see Table XII).

TABLE XIII
SUMMARY OF FEATURES

ID	Feature Name	ID	Feature Name	ID	Feature Name
f1	Mean time	f29	Spectral energy	f57	Percentage of energy D5
f2	Standard deviation	f30	Peak Density over rms	f58	Percentage of energy D4
f3	Variance	f31	2nd highest peak value	f59	Percentage of energy D3
f4	Signal entropy	f32	Freq. of 2nd highest peak	f60	Percentage of energy D2
f5	Kurtosis concentration	f33	3rd highest peak value	f61	Percentage of energy D1
f6	Multiscale entropy MSE	f34	Freq. of 3rd highest peak	f62	Percentage of energy A6
f7	Time to reach the max. peak	f35	A6 Vmax	f63	A6 RMS
f8	RMS	f36	A6 Freq. Max.	f64	A6 difference max-min values
f9	Peak to peak value	f37	A6 mean freq.	f65	A6 ratio of the largest absolute value to RMS
f10	Ratio Peak to rms	f38	D6 Vmax	f66	D6 RMS
f11	Signal Energy	f39	D6 Max Freq.	f67	D6 difference max-min values
f12	Zero-crossing number	f40	D6 mean freq.	f68	D6 ratio of the largest absolute value to RMS
f13	Peaks Density over rms	f41	D5 Vmax	f69	D5 RMS
f14	Number of peaks in FFT	f42	D5 Max. Freq.	f70	D5 difference max-min values
f15	Max. detected frequency	f43	D5 mean freq.	f71	D5 ratio of the largest absolute value to RMS
f16	Spectral Entropy	f44	D4 Vmax	f72	D4 RMS
f17	Mean frequency	f45	D4 Max Freq.	f73	D4 difference max-min values
f18	Freq. Standard deviation	f46	D4 mean freq.	f74	D4 ratio of the largest absolute value to RMS
f19	Spectral Variance	f47	D3 Vmax	f75	D3 RMS
f20	Spectral Energy	f48	D3 Max Freq.	f76	D3 difference max-min values
f21	Spectral Kurtosis	f49	D3 mean freq.	f77	D3 ratio of the largest absolute value to RMS
f22	Spectral multiscale entropy MSE	f50	D2 Vmax	f78	D2 RMS
f23	Max. peak in the band 10Hz-20Hz	f51	D2 Max Freq.	f79	D2 difference max-min values
f24	Freq. Max. Peak in the band 10Hz-20Hz	f52	D2 mean freq.	f80	D2 ratio of the largest absolute value to RMS
f25	Max. peak in the band 20Hz-30Hz	f53	D1 Vmax	f81	D1 RMS
f26	Freq. Max. Peak in the band 20Hz-30Hz	f54	D1 mean freq.	f82	D1 difference max-min values
f27	Spectral rms	f55	mean energy	f83	D1 ratio of the largest absolute value to RMS
f28	Ratio peak to rms in freq.	f56	Percentage of energy D6	f84	Energy of signal components

Regarding the relevance of feature types, the best selected subset used in [4] was formed by features from the time domain (one), the frequency domain (five), and the scale domain (nine); in [44], the best features (five) were all extracted from the frequency domain, whereas the proposed method only considered two features from the scale domain for the classification of LP events, and one feature from the frequency domain and one from the scale domain for the classification of VT events and both seismic events together. It is important to notice that feature f28 (the amplitudes and frequencies of the second and third peaks of the FFT of the signal) was considered as one of the most relevant features by the method developed in [44] and also by the proposed method. This corroborates the importance of using frequency-domain features for seismic event classification. In general, these results provided clear experimental evidence that the proposed method produces good classification results.

IV. CONCLUSION AND FUTURE WORK

In this paper, different filter-based feature selection methods were studied in combination with a GMM classifier over three different training-test split scenarios, using a dataset of seismic signals from the Cotopaxi volcano, in Ecuador. For LP seismic event classification, the Chi2 discretization method showed the best performance, whereas for just VT events, and both events together (VT and LP), the uFilter method outperformed the others. In the majority of cases, the GMM classifier using the Chi2 discretization method produced faster classification performance versus those produced by the GMM in combination with uFilter or 1R methods. However, the GMM using the uFilter method produced better precision scores. From an overall perspective, the uFilter method was able to produce more useful classification schemes (four out of nine possible) against the others employed feature

selection methods. Thus, the classification schemes presented in the 25–75% training-test split scenario ensured the best selection.

Further redundancy analysis allowed us to obtain an optimal set of features around 50% smaller than those obtained by the feature selection methods before the redundancy analysis in all cases. The most significant features for classification of VT events alone and VT and LP events simultaneously are in the frequency domain, whereas those considered to be strongly relevant for LP events are in the scale domain. According to the Wilcoxon statistical test, the proposed method demonstrated that it can provide competitive schemes for seismic event classification, while reducing the required processing time.

Future work will focus on extending the GMM classifier scheme for use in multiclass classification problems to include several types of seismic events such as VLP events, which indicate magma intrusion, HYB events that are the mixture of VT with LP events and indicate a process of fracture and resonance at the same time, EXP events (VT or LP like signals followed by an infrasound pulse), TRE, and signals of nonvolcanic origin such as icequakes, lahars, or snow avalanches, and regional and teleseismic activity. The low algorithm complexity of the proposed schema makes it suitable for embedded systems implementation; therefore, we also want to explore this possibility.

Furthermore, we are also considering implementing the proposed method on a distributed processing environment with GPUs availability to use more complex MLCs and to reduce the required processing time without losing effectiveness. We also want to exhaustively explore the ranked subset of features sequentially, increasing the number of features in each subset by one instead of the empirical threshold of five used in this paper, in such a way that the whole feature space can be comprehensively explored. Combining and analyzing data from multiple stations may help to improve the performance of the method; this will also be addressed in next steps.

Finally, it should be pointed out that, at this time, it is very difficult to claim that any of the best feature selection methods obtained in our study will be as good in other datasets as it was during our experiment. The only way to know this is by doing cross-dataset experimentation, which is difficult to achieve today, since there are not enough benchmark datasets available for seismic scientific research. However, it is also our intention to reassemble the experiments using seismological data acquired with similar equipment for other volcanoes and other datasets to determine to what extent the best feature selection methods obtained in our approach can be generalized.

APPENDIX

Features f1–f13 were obtained from the time domain; features f14–f34 were obtained in the frequency domain by applying the FFT to each signal; features f35–f84 were obtained in the scale domain by applying the wavelet transform to each signal by using a tenth-order *symlet* mother wavelet up to a decomposition level 6, obtaining the corresponding approximation A and detailed D wavelet coefficients.

ACKNOWLEDGMENT

The seismic data used in this study were provided by Instituto Geofísico, Escuela Politécnica Nacional. The BVC2 seismic station is part of a seismic-infrasound network installed through the project JICA-IGEPN for enhancing volcano monitoring in Ecuador.

REFERENCES

[1] J. Wassermann, "Volcano seismology," in *IASPEI New Manual of Seismological Observatory Practice 2 (NMSOP-2)*, P. Bormann, Ed. Potsdam, Germany: Deutsches GeoForschungsZentrum GFZ, 2012, pp. 1–77.

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

[3] M. Ruiz, B. Guillier, J. L. Chatelain, H. Yepes, M. Hall, and P. Ramon, "Possible causes for the seismic activity observed in Cotopaxi volcano, Ecuador," *Geophysical Res. Lett.*, vol. 25, no. 13, pp. 2305–2308, 1998.

[4] R. A. Lara-Cueva, D. S. Benítez, E. V. Carrera, M. Ruiz, and J. L. Rojo-Álvarez, "Automatic recognition of long period events from volcano tectonic earthquakes at Cotopaxi Volcano," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5247–5257, Sep. 2016.

[5] N. Langet, A. Maggi, A. Michelini, and F. Brenguier, "Continuous kurtosis-based migration for seismic event detection and location, with application to Piton de la Fournaise volcano, La Réunion," *Bull. Seismol. Soc. Amer.*, vol. 104, no. 1, pp. 229–246, 2014.

[6] 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. Geothermal Res.*, vol. 153, no. 1, pp. 1–10, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377027305003793

[7] G. Curilem, J. Vergara, G. Fuentealba, G. Acua, and M. Chacn, "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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377027308006355

[8] M. Orozco, M. E. Garcia, R. P. Duin, and C. A. G. Castellanos, "Dissimilarity-based classification of seismic signals at Nevado del Ruiz volcano," *Earth Sci. Res. J.*, vol. 10, pp. 57–66, Dec. 2006. [Online]. Available: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-61902006000200001&nrm=iso

[9] M. C. Benitez *et al.*, "Continuous HMM-based seismic-event classification at deception island, Antarctica," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 138–146, Jan. 2007.

[10] I. Alvarez, L. Garcia, G. Cortes, C. Benítez, and A. D. 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.

[11] D. Cárdenas-Peña, M. Orozco-Alzate, and G. Castellanos-Dominguez, "Selection of time-variant features for earthquake classification at the Nevado-del-Ruiz volcano," *Comput. Geosci.*, vol. 51, pp. 293–304, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0098300412002919

[12] R. Lara-Cueva, D. Benitez, E. Carrera, M. Ruiz, and J. L. Rojo-Álvarez, "Feature selection of seismic waveforms for long period event detection at Cotopaxi volcano," *J. Volcanol. Geothermal Res.*, vol. 316, pp. 34–49, Mar. 2016.

[13] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, Aug. 2016.

[14] I. Guyon and A. Elisseeff, *An Introduction to Feature Extraction*. Berlin, Germany: Springer, 2006, pp. 1–25. [Online]. Available: https://doi.org/10.1007/978-3-540-35488-8_1

[15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. New York, NY, USA: Cambridge Univ. Press, 1988.

[16] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *Proc. 38th Int. Conv. Inf. Commun. Technol., Electron. Microelectron.*, May 2015, pp. 1200–1205.

[17] R. C. Holte, "Very simple classification rules perform well on most commonly used datasets," *Mach. Learn.*, vol. 11, no. 1, pp. 63–90, Apr. 1993. [Online]. Available: https://doi.org/10.1023/A:1022631118932

[18] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proc. 9th Int. Workshop Mach. Learn.*, 1992, pp. 249–256. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781558602472500371

[19] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes," in *Proc. 7th IEEE Int. Conf. Tools Artif. Intell.*, Nov. 1995, pp. 388–391.
[20] N. P. Pérez, M. A. G. López, A. Silva, and I. Ramos, "Improving the Mann–Whitney statistical test for feature selection: An approach in breast cancer diagnosis on mammography," *Artif. Intell. Med.*, vol. 63, no. 1, pp. 19–31, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0933365714001419
[21] D. Reynolds, "Gaussian mixture models," in *Encyclopedia of Biometrics*. New York, NY, USA: Springer, 2015, pp. 827–832.
[22] 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. Geothermal Res.*, vol. 340, pp. 130–142, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377027316303948
[23] J. Cadima, J. Cerdeira, and M. Minhoto, "Computational aspects of algorithms for variable selection in the context of principal components," *Comput. Statist. Data Anal.*, vol. 47, no. 2, pp. 225–236, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016794730300272X
[24] J. F. Cadima and I. T. Jolliffe, "Variable selection and the interpretation of principal subspaces," *J. Agricultural, Biol., Environ. Statist.*, vol. 6, no. 1, p. 62, 2001.
[25] M. Ohrnberger, "Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia," doctoral dissertation, Math.-Naturwissenschaftliche Fakultät, Univ. Potsdam, Potsdam, Germany, Jul. 2001.
[26] P. Goupillaud, A. Grossmann, and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis," *Geoexploration*, vol. 23, no. 1, pp. 85–102, 1984.
[27] J. Wang, C. Xu, E. Chng, and Q. Tian, "Sports highlight detection from keyword sequences using HMM," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. 2004, vol. 1, pp. 599–602.
[28] H. D. Ortiz Erazo, "Study of the site effects for the construction of a seismic activity index at the Cotopaxi volcano," master's thesis, Escuela Politécnica Nacional, Quito, Ecuador, 2013.
[29] Y. Vaezi and M. Van der Baan, "Comparison of the STA/LTA and power spectral density methods for microseismic event detection," *Geophys. J. Int.*, vol. 203, no. 3, pp. 1896–1908, 2015. [Online]. Available: http://dx.doi.org/10.1093/gji/ggv419
[30] R. Lara-Cueva, P. Bernal, M. G. Saltos, D. S. Benítez, and J. L. Rojo-Álvarez, "Time and frequency feature selection for seismic events from Cotopaxi volcano," in *Proc. Asia-Pacific Conf. Comput. Aided Syst. Eng.*, Jul. 2015, pp. 129–134.
[31] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 1, pp. 72–83, Jan. 1995.
[32] P. KaewTraKulPong and R. Bowden, *An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection*. Boston, MA, USA: Springer, 2002, pp. 135–144. [Online]. Available: https://doi.org/10.1007/978-1-4615-0913-4_11
[33] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.
[34] G. J. McLachlan and D. Peel, *Finite Mixture Models* (Wiley Series in Probability and Statistics). New York, NY, USA: Wiley, 2000.
[35] Y. K. Jain and S. K. Bhandare, "Min max normalization based data perturbation method for privacy protection," *Int. J. Comput. Commun. Technol.*, vol. 2, no. 8, pp. 45–50, 2011.
[36] F. G. López, M. G. Torres, B. M. Batista, J. A. M. Pérez, and J. M. Moreno-Vega, "Solving feature subset selection problem by a parallel scatter search," *Eur. J. Oper. Res.*, vol. 169, no. 2, pp. 477–489, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221704005491
[37] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Statist.*, vol. 18, no. 1, pp. 50–60, Mar. 1947. [Online]. Available: https://doi.org/10.1214/aoms/1177730491
[38] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.
[39] T. Hesterberg, "Weighted average importance sampling and defensive mixture distributions," *Technometrics*, vol. 37, no. 2, pp. 185–194, 1995. [Online]. Available: https://amstat.tandfonline.com/doi/abs/10.1080/00401706.1995.10484303
[40] J. D. Gibbons and S. Chakraborti, "Nonparametric statistical inference," in *International Encyclopedia of Statistical Science*. New York, NY, USA: Springer, 2011, pp. 977–979.
[41] J. Marques de Sá, "Estimating data parameters," in *Applied Statistics Using SPSS, STATISTICA, MATLAB and R*. New York, NY, USA: Springer, 2007, pp. 81–109.
[42] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944968
[43] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in *Proc. 11th Int. Conf. Mach. Learn.*, 1994, pp. 121–129. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781558603356500234
[44] R. Lara-Cueva, E. V. Carrera, J. F. Morejon, and D. Benítez, "Comparative analysis of automated classifiers applied to volcano event identification," in *Proc. IEEE Colombian Conf. Commun. Comput.*, Apr. 2016, pp. 1–6.
[45] R. Lara-Cueva, D. S. Benítez, V. Paillacho, M. Villalva, and J. L. Rojo-Álvarez, "On the use of multi-class support vector machines for classification of seismic signals at Cotopaxi volcano," in *Proc. IEEE Int. Autumn Meeting Power, Electron. Comput.*, Nov. 2017, pp. 1–6.

Pablo Venegas received the B.Eng. degree in electronics engineering from the Universidad San Francisco de Quito, Quito, Ecuador, in 2018.

His research interests include digital signal processing, real-time systems, and machine learning.

Noel Pérez (M'19) received the B.Eng. degree in computer science engineering and the M.Sc. degree in applied computer science, in the field of digital image processing, pattern recognition and machine learning, from the Universidad de Ciego de Avila, Ciego de Avila, Cuba, in 2005 and 2007, respectively, and the Ph.D. degree in data mining, pattern recognition, and machine learning from the University of Porto, Porto, Portugal, in 2015.

From 2008 to 2015, he was a Fellow Researcher with the Institute of Mechanical Engineering and Industrial Management, Faculty of Engineering, University of Porto. From 2015 to 2017, he was a Postdoctoral Fellow with the Instituto de Telecomunicaes, Faculty of Sciences, University of Porto. Since 2017, he has been a Full-Time Professor with the Universidad San Francisco de Quito, Quito, Ecuador. His research interests include digital image processing, data mining, pattern recognition, and machine learning.

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

From 2005 to 2007, he was a Postdoctoral Research Associate with the Sensing, Imaging and Signal Processing Research Group, School of Electrical and Electronic Engineering, University of Manchester. From 2007 to 2012, he was a Senior Research Engineer with the Bosch Research and Technology Center, Pittsburgh, PA, USA, where he was also an Academic Visitor with the Institute for Complex Engineered Systems and INFERLab, Carnegie Mellon University. From 2012 to 2014, he was a Visiting Research Scholar with the Universidad de las Fuerzas Armadas ESPE under the "Prometeo Program" of SENESCYT, Ecuador. He is currently a Full-Time Faculty with the Universidad San Francisco de Quito, Quito, Ecuador. His research interests include signal and image processing, and intelligent instrumentation and measurement systems for medical, energy management, security, and smart buildings applications.

Román Lara-Cueva (SM'18) 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 the Politecnico di Torino, Turin, Italy, in 2005, and the M.Sc. and 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 and a Full Professor since 2016. He has authored or coauthored 12 publicly funded research projects and directed seven of them. He currently heads the Ad Hoc Network Research Center and the Smart Systems Research Group (WiCOM-Energy). His research interests include digital signal processing, smart cities, wireless systems, machine learning theory, and volcano seismology.

Mario Ruiz received the B.Eng. degree in geophysical engineering from 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 is currently the Chair of the Instituto Geofísico, Escuela Politécnica Nacional, Quito. His research interests include volcano seismology.