UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

Combining Filter Based Feature Selection Methods and Gaussian Mixture Model for Real-Time Seismic Event Classification at Cotopaxi Volcano

Artículo Académico

.

Pablo Esteban Venegas Gallardo

Ingeniería Electrónica

Trabajo de titulación presentado como requisito
para la obtención del título de
Ingeniero Electrónico

Quito, 30 de noviembre de 2018

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

COLEGIO DE CIENCIAS E INGENIERÍAS

HOJA DE CALIFICACIÓN
DE TRABAJO DE TITULACIÓN

Combining Filter Based Feature Selection Methods
and Gaussian Mixture Model for Real-Time Seismic
Event Classification at Cotopaxi Volcano

Pablo Esteban Venegas Gallardo

Calificación:

Nombre del profesor, Título académico Diego Benítez, Ph.D.

Firma del profesor

Quito, 30 de noviembre de 2018

Derechos de Autor

Firma del estudiante: ____

Nombres y apellidos: Pablo Esteban Venegas Gallardo

Código: 00117268

Cédula de Identidad: 1724194509

Lugar y fecha: Quito, 30 de noviembre de 2018

RESUMEN

Este artículo propone una evaluación exhaustiva de cinco diferentes métodos de selección de características basados en filtrado, en combinación con el Modelo Mixto Gaussiano como clasificador para la categorización prácticamente en tiempo real de eventos sísmicos Largo-Período y Vulcano-Tectónico grabados en el volcán Cotopaxi en Ecuador entre 2009 y 2010. La experimentación incluyó la exploración y clasificación de espacios de búsqueda de características basadas en señales sísmicas, y la selección del subconjunto de características óptimas para la clasificación de eventos. La evaluación se llevó a cabo utilizando una base de datos balanceada de eventos sísmicos, formada por 80 vectores de características por clase, cada uno compuesto por 84 características estadísticas, temporales, en el dominio de la frecuencia y de la escala extraídas de las señales sísmicas originales. El mejor resultado en exactitud, precisión, exhaustividad y tiempo de procesamiento, para la clasificación de eventos sísmicos tipo Largo-Período se obtuvo utilizando el método de Information Gain con 10 características, logrando 91.75%, 91.04%, 92.75% y 0.0027s, respectivamente. Mientras que para la clasificación de eventos sísmicos Vulcano-Tectónicos, el método de Discretización CHI2 con 10 características alcanzó los puntajes de 90.81%, 91.12%, 90.63% y 0.0025s, respectivamente. Para la clasificación de ambos eventos sísmicos juntos, el método de Information Gain con 10 características, obtuvo 91.88%, 91.94%, 91.86% y 0.0027s, respectivamente. De acuerdo con la prueba estadística de Wilcoxon, estos esquemas de clasificación demostraron proporcionar una categorización competitiva de los eventos sísmicos, a la vez que reducen el tiempo de procesamiento.

Palabras clave: Métodos de selección de características, clasificador GMM, análisis de redundancia, clasificación de eventos sísmicos.

ABSTRACT

This paper proposes an exhaustive evaluation of five different filter-based feature selection methods in combination with a Gaussian Mixture Model classifier for almost real time classification of Long-Period and Volcano-Tectonic seismic events recorded at Cotopaxi volcano in Ecuador between 2009 and 2010. 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 using a balanced seismic event dataset formed by 80 feature vectors per class, each composed by 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 Long-Period seismic event classification was obtained by using the Information Gain method with 10 features, achieving 91.75%, 91.04%, 92.75% and 0.0027s, respectively. While for Volcano-Tectonic seismic event classification, the CHI2 discretization method with 10 features reached the scores of 90.81%, 91.12%, 90.63% and 0.0025s, respectively. For the classification of both seismic events together, the Information Gain method with 10 features, yields 91.88%, 91.94%, 91.86% and 0.0027s, respectively. According to the Wilcoxon statistical test, these classification schemes demonstrated to provide competitive seismic event classification while reducing the processing time.

Key words: Feature Selection Methods, GMM Classifier, Redundancy Analysis, Seismic Events Classification.

TABLA DE CONTENIDO

Combining Filter Based Feature Selection Methods and Gaussian Mixture Model for Real-Time Seismic Event Classification at Cotopaxi Volcano

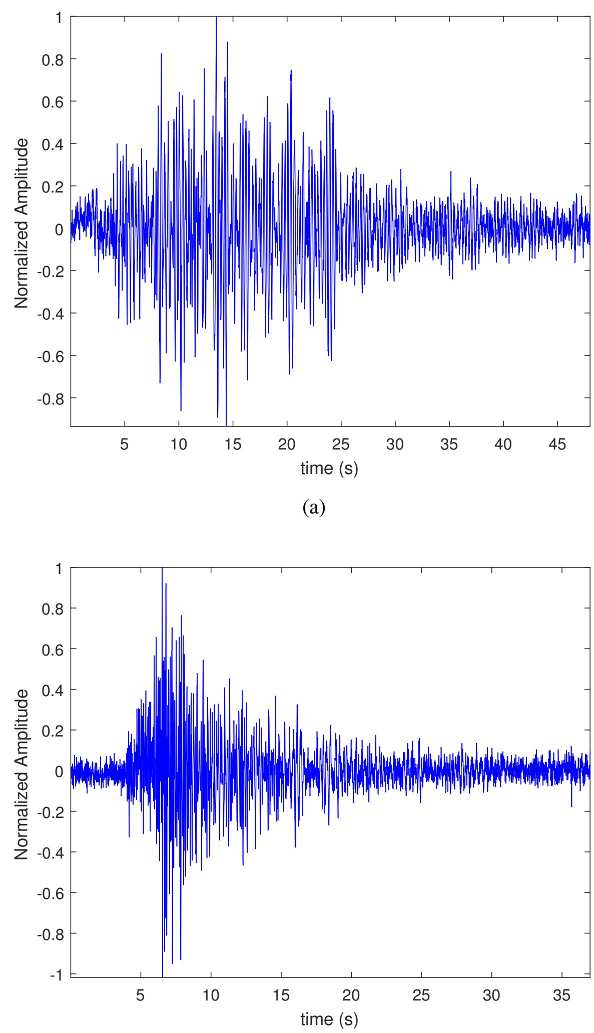
Pablo Venegas, *Student Member, IEEE,* Noel Pérez, *Member, IEEE,* Diego Benítez, *Senior Member, IEEE,* Román Lara-Cueva *Member, IEEE,* and Mario Ruiz

***Abstract*—This paper proposes an exhaustive evaluation of five different filter-based feature selection methods in combination with a Gaussian Mixture Model classifier for almost real time classification of Long-Period and Volcano-Tectonic seismic events recorded at Cotopaxi volcano in Ecuador between 2009 and 2010. 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 using a balanced seismic event dataset formed by 80 feature vectors per class, each composed by 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 Long-Period seismic event classification was obtained by using the Information Gain method with 10 features, achieving 91.75%, 91.04%, 92.75% and 0.0027s, respectively. While for Volcano-Tectonic seismic event classification, the CHI2 discretization method with 10 features reached the scores of 90.81%, 91.12%, 90.63% and 0.0025s, respectively. For the classification of both seismic events together, the Information Gain method with 10 features, yields 91.88%, 91.94%, 91.86% and 0.0027s, respectively. According to the Wilcoxon statistical test, these classification schemes demonstrated to provide competitive seismic event classification while reducing the processing time.**

***Index Terms*—Feature Selection Methods, GMM Classifier, Redundancy Analysis, Seismic Events Classification.**

I. Introduction

Nowadays, developing technologies and systems capable of helping scientist for forewarning to the population in the case of volcanic eruptions is of great importance for saving human lives and minimizing possible consequences. In such sense, the analysis and identification of seismic signals is an essential activity for studying the volcanic dynamic process and its intrinsic structure. Currently, volcano observatories dispose of a great amount of monitoring sensor networks (sismometers or geophones) capable of detecting seismic signals of low intensity or microseisms, where each seismogram may contain various 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. Given the high sensibility of the recording instruments, in some cases, signals of non-volcanic origin such as lightnings (LGH) or tectonic earthquakes, can also be recorded by the seismograph. Examples of LP and VT seismic signals are illustrated in Fig.1.



(a)

(b)

Fig. 1. Examples of seismic signals from Cotopaxi volcano: (a) LP event (b) VT event.

P. Venegas, N. Pérez and D. Benítez are with Colegio de Ciencias e Ingenierías "El Politécnico", Universidad San Francisco de Quito USFQ, Campus Cumbayá, Casilla Postal 17-1200-841, Quito, Ecuador.

R. Lara-Cueva is with Grupo de Investigación en Sistemas Inteligentes (WiCOM-Energy), and Centro de Investigaciones de Redes Ad-Hoc (CIRAD), Departamento de Eléctrica y Electrónica, Universidad de las Fuerzas Armadas ESPE, Sangolquí-Ecuador, 171-5-231B.

M. Ruiz is with Instituto Geofísico, Escuela Politécnica Nacional, P.O. Box 17-01-2759, Quito-Ecuador.

Manuscript received XX, 2018; revised XX, 2018.

In this sense, several systems for automated recognition of microseisms have been developed and proposed in the literature, for both detection [1], [2] and classification stages [3], [4]. For the latter, several approaches for analyzing the seismic signals in time, frequency, and scale domains have also been proposed, in which well-known features have been used by the classification algorithms. Some previous works reported in the literature have employed feature selection techniques to determine the most relevant characteristics in seismic events of volcanic origin. In Alvarez et al. [5], for example, discriminative feature selection technique (DFS) was used to select the features with more discriminatory information between the different types of events for the Colima Volcano, in Mexico. In Cárdenas et al. [6], for classifying the seismic events from the Nevado del Ruiz, in Colombia, a feature selection strategy based on a relevance measure of time-variant features and k-nearest neighbors, was proposed. Meanwhile for the Cotopaxi volcano, in Ecuador, several previous works used mixed feature selection techniques i.e. wrapper and embedded methods in [3], and filter and embedded methods in [7]. However, determining the most appropriate subset of features for rapid and accurate seismic event classification remains as 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 optimal feature subset related to a classification problem is always relative to a certain evaluation function [8]. That means, the optimal feature subset chosen, using one evaluation function may not be the same as those used by another evaluation function. Generally, an evaluation function tries to measure the discriminating ability of a feature or a subset, in order to distinguish the different class labels. Therefore, the use of different evaluation functions may provide important information about the nature of each feature or subset (respect to the class) in the feature space.

Feature selection techniques are represented by three different paradigms, according to the type of features searching algorithm inside the classification model: wrapper and embedded methods, which incorporate machine learning classifiers for deciding the features merit, and filter methods, which use the data characteristics as main heuristics rather than machine learning classifiers to assess the features importance [9]. Thus, filter methods are less complex and much faster than wrapper or embedded methods [9].

This work aims to study the behavior of five different filter based feature selection methods: Information Gain (IG) [10], One Rule [11], RELIEF [12], Chi2 Discretization [13] and uFilter [14], in combination with a Gaussian Mixture Model (GMM) [15] for classifying LP and VT seismic events in almost real time on a dataset of seismic signals based features (volcanic origin collected from the Cotopaxi Volcano, in Ecuador).

The remainder of the paper is organized as follows: Section II "Materials and methods" describes the employed seismic events dataset, the five filter based feature selection methods, the GMM classifier and the experimental methodology; Section III "Results and discussions" covers and discusses the experimental results obtained both from the selection of the best classification scheme for each scenario (before and after the redundancy analysis) and from the global comparison against others developed methods. Finally, in the "Conclusions" section, we outline the principal achievements of this work and our future work.

II. Materials and methods

A. Dataset

The dataset used in this work correspond to several seismic events records from 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. As illustrated in Fig. 2, a network of seismometers has been installed around the Cotopaxi volcano, which comprises: six short period (SP) seismological stations (PITA, NAS2, VC1, REFU, CAMI, and TAMB) with a frequency response range of 1–50 Hz, four of these stations have vertical-axis sensors and two of them have three-axis sensors, six broadband (BB) stations (BREF, BVC2, BTAM, BNAS, BMOR, and VCES), with a frequency response range of 0.1–50 Hz [16].

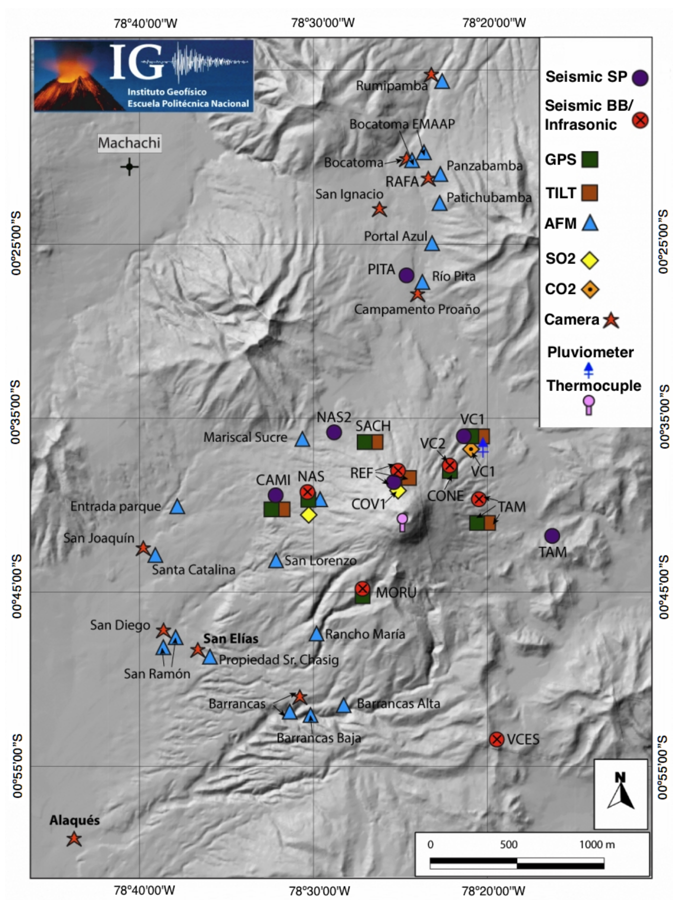


Fig. 2. Network of seismological stations deployed at Cotopaxi Volcano, the dataset has been taken from the BVC2 broad band seismic station. Image provided by IGEPN and taken from [3].

The dataset consisted of $N = 914$ seismic records (759 LP, 116 VT, 30 HYB, and 9 TRE). The seismic records were recorded during 2009 and 2010 at the BVC2 station, located at 3 km from the Cotopaxi summit, by a three-axial seismometer (CMG-40T Guralp) with a sensitivity of 1600 V/ms^{-1}. Seismic records were sampled at 100 Hz using a 24-bit analog-to-digital converter (Geotech Smart 24D digitizer). The acquisition system uses the STA/LTA algorithm [17] to detect events and then stores them in files of 12.000 s of duration. Each seismic record contains one microseism, from each independent microseism record, a total of 84 parameters computed from the signal, were used as features: 13 in time domain, 21 in the frequency domain, and 50 in scale domain using wavelets. See [3], [18] and references therein for details.

For the purpose of this study, an experimental balanced dataset, formed by a randomly sample selection of 80 feature vectors per class from the previously described dataset, and belonging to the LP and VT seismic event types, was created. Other types of seismic events were not taken into consideration because of the insufficient number of samples (less than 30 instances per class), that could affect the minimum number of samples required for further statistical analysis. Table I shows a brief description of some of the used features (random selection). Detailed information about the features extracted from the seismic signals can be found in [3], [7], [18].

TABLE I: Summary of features.

ID	Name	ID	Name	ID	Name
f3	Time variance	f29	FFT energy	f49	WAV D3 FFT average freq.
f4	Entropy time	f30	FFT density of number of peaks on rms	f50	WAV D2 FFT VMAX
f5	Kurtosis time	f31	FFT 2nd highest peak	f52	WAV D2 FFT average freq.
f7	Time max. peak	f32	Freq. 2nd peak in FFT	f58	Percentage of energy D4
f9	Peak to peak time	f33	FFT 3th highest peak	f60	Percentage of energy D3
f15	Max. Frequency	f34	Freq. 3th peak in FFT	f61	Percentage of energy D1
f17	Average frequency	f35	WAV A6 FFT VMAX	f63	WAV A6 RMS
f23	FFT value threshold 10Hz - 20 Hz	f37	WAV A6 FFT average freq.	f76	WAV D3 difference max-min values
f27	FFT RMS	f38	WAV D6 FFT VMAX	f78	WAV D2 RMS
f28	FFT peaks RMS	f40	WAV D6 FFT average freq.	f79	WAV D2 difference max-min values

B. Feature Selection Methods

This work explored five different features selection methods belonging to the filter paradigm. Thus, their algorithm complexity are lower and the performance are faster when compared to wrapper and embedded methods [9]. These advantages are considerable when developing applications in almost real time. A brief description of selected methods are presented here:

1) Information Gain: The IG measurement, normalized with the symmetrical uncertainty coefficient [10], 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 information gain bias toward attributes with more values.

2) One Rule: This method estimates the predictive accuracy of individual feature building rules based on a single feature (it can be thought of as single level decision tree) [11]. As it is used for training and testing datasets, it is possible to calculate a classification accuracy for each rule, and hence each feature. Then, from the classification scores, a ranked list of features is obtained. In previous studies, experiments choosing a selected number of the highest ranked features, and using them with common machine learning algorithms, have showed that, on average, by just using the top three or more features, the results obtained are as accurate as using the original feature set. However, this approach is unusual, due to the fact that no search is conducted.

3) RELIEF: This method uses instance based learning to assign a relevance weight to each feature [12]. Each feature weight reflects its ability to distinguish among the class values. The 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 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), while for numeric features, the difference is the actual difference normalized in the interval [0..1].

4) Chi2 Discretization: This method consists on a justified heuristic for supervised discretization [13]. 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 non-parametric Mann-Whitney U-test [14]; 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, providing therefore 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. Gaussian Mixture Model Classifier

The Gaussian Mixture Model (GMM) is a probabilistic method used for representing multimodal data distributions. It determines a normal distribution for each Gaussian $k-$component inside the model, which are further employed for computing the weighted sum of the Gaussian density function (output).
The GMM can be univariate or multivariate regarding to the number of variables existing in the model, and the most important parameters include: the weight (ϕ_k), mean ($\vec{\mu}_k$) and covariance matrix (Σ_k) of each $k-$component [15].

These parameters are estimated by using the Expectation-Maximization (EM) method, which is an iterative algorithm to estimate the maximum likelihood involving two steps: (1) the calculation of the parameters expectation for each $k-$component and (2) the maximization of computed expectations. Hereafter, the model parameters are updated with the output of the maximization step; this process is repeated until the algorithm converges, providing the maximum likelihood estimation [19]. The formulation of the Gaussian density function of a component is given by:

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

where $\vec{x}$ is the set of data vectors with D dimension (number of variables).
Besides, the GMM is calculated as:

$$p(\vec{x}) = \sum_{i=1}^{K} \phi_i \mathcal{N}(\vec{x} \mid \vec{\mu}_i, \Sigma_i) \tag{2}$$

with the normalization constrain:

$$\sum_{i=1}^{K} \phi_i = 1 \tag{3}$$

where ϕ_i is the weight of each k-component.

D. Experimentation

This section outlines the experimental evaluation of the selected filter based feature selection methods in combination with a GMM classifier on an experimental balanced dataset, for classifying first LP events, then VT events, and finally both seismic events together. The experimentation was conducted on three training-test split scenarios. The overall procedure involved the following steps:

1) Normalizing the values of the experimental dataset using the min-max normalization method [20], in order to bring all values into the range [0,1] (See Fig. 3 step 1).
2) Applying five filter based feature selection methods: Information Gain [10], Relief [12], One Rule [11], CHI2 [13] and uFilter [14] on the experimental dataset to produce five different features rankings (See Fig. 3 step 2).
3) Creating several ranked subset 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 Fig. 3 step 3).
4) Applying a ten times $k-$fold cross-validation (CV) method [21] on each ranked feature subset to create three different training-test split scenarios (depending of the k value selected): i.e. 25%-75%, 50%-50% and 75%-25% with $k = 4$, $k = 2$, and $k = 4$, respectively. Therefore, samples will not appear simultaneously in the training and test data set, thus guaranteeing disjoint test partitions and preventing the classifier from overfitting [21] (See Fig. 3 step 4).
5) Classifying the generated ranked subset of features using the GMM classifier over the three formed scenarios (See Fig. 3 step 5). Only two types of events (LP and VT) will be considered by the GMM configuration, thus the $k-$component was set to two and, the parameters ϕ_k,$\vec{\mu}_k$ and Σ_k were estimated by the EM method [19].
6) Establishing a comparative analysis based on metrics such as: accuracy, precision, recall, and processing time scores, for selecting the best classification schemes for LP, VT, and both seismic event together (See Fig. 3 step 6). All comparisons were made using the Wilcoxon statistical test [22], [23] in order to evaluate the statistical difference between classification schemes. In case of any tie (according to the statistical test) among the classification schemes, the criteria used for selecting the most appropriate scheme was to select the one who had the smaller number of features, and subsequently the smaller value of processing time. It should be pointed out that the accuracy, precision, and recall metrics were individually computed depending on the events type: for LP or VT events (as disjoint units), it was used the confusion matrix output, while for both events together (as a binary classification problem), the use of the weighted average technique [24] based on the confusion matrix was mandatory.

In the last step of the experiment, we also performed a redundancy analysis over the most appropriated classification schemes using the Pearson correlation [25], in order to separate the redundant features from relevant ones, and thus to produce the final optimal subset of features.

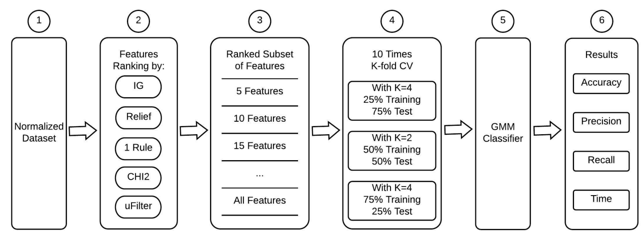


Fig. 3. Applied experimental workflow; CV means cross-validation.

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 over 100 runs highlighted interesting results for classifying LP, VT, and both seismic event together on three different training-test split scenarios. For better interpretation of the obtained results, we only only included the set of classification schemes that did not represent statistical difference in terms of performance (considering the assessment metrics) among them.

A. Long-Period Seismic Event Classification

The highest metrics scores in the 25-75% training-test split scenario was obtained by using the GMM classifier in combination with the uFilter feature selection method, with 15 features. As it can be seen in Table II, this result was statistically superior than the majority of the remaining combinations of the GMM classifier with other filter selection schemes. However, there were also others combinations with similar statistically performances at $p = 0.05$. According to these results, it is possible to select the GMM classifier and the IG method with 10 features as the most suitable classification scheme in this scenario, since, it reached a similar performance but using less number of features.

TABLE II: LP event classification performance results for the GMM classifier with different feature selection filters for the 25-75% training-test split scenario.

Best Scheme					Other Schemes				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)	Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)
UF+15f	91,6042	90,9243	92,4583	0,0025	Relief+20f	91,4375	90,5523	92,5833	0,0026
					IG+10f	**91,3750**	**90,5734**	**92,4167**	**0,0028**
					Relief+15f	91,1875	90,6676	91,8750	0,0027
					UF+20f	91,1875	90,5813	91,9583	0,0027

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

In the 50-50% training-test split scenario, the highest metrics scores were obtained by the GMM classifier using the IG method with 10 features (See Table III). Although, there was another classification scheme with similar statistically performance at $p = 0.05$, the one with the best metrics scores was selected as the most appropriate classification scheme for this scenario, since it uses the less number of features.

TABLE III: LP event classification performance results for the GMM classifier with different feature selection filters for the 50-50% training-test split scenario.

Best Scheme					Other Schemes				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)	Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)
IG+10f	**91,7500**	**91,0448**	**92,7500**	**0,0027**	CHI2+35f	91,2500	89,6776	93,5000	0,0067

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

In the last scenario (75-25% training-test split), the highest metrics scores were achieved by the combination of the GMM classifier with the Relief method using 20 features (See Table IV). This result exhibited a statistical superiority when compared to the majority of other possible GMM and filter selection schemes. But, it did not show statistical difference ($p = 0.05$) in terms of performance against a small set of classification schemes. Therefore, it is also possible to select the GMM classifier in combination with the IG method with 10 features as the most adequate classification scheme in this scenario, since it required less number of features for obtaining similar performance.

TABLE IV: LP event classification performance results for the GMM classifier with different feature selection filters for the 75-25% training-test split scenario.

Best Scheme					Other Schemes				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)	Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)
Relief+20f	91,9375	92,2162	91,9375	0,0038	UF+15f	91,8750	92,0590	91,8750	0,0025
					Relief+15f	91,8125	92,1308	91,8125	0,0041
					IG+30f	91,6250	91,9020	91,6250	0,0049
					IG+10f	**91,3125**	**91,5175**	**91,3125**	**0,0030**

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

Regarding the LP seismic events classification, it seems that the GMM classifier in combination with the IG method with 10 features reached the best performance for all scenarios. However, the classification scheme presented in the 50-50% training-test split scenario (see Table III) was slightly better in terms of processing time, ensuring it, as the best selection.

B. Volcano-Tectonic Seismic Event Classification

The highest metrics scores in the the 25-75% training-test split scenario were obtained by the GMM classifier in combination with the Relief method with 20 features (See Table V). This result was statistically superior than the majority of remaining classification schemes. However, there were also others schemes with similar performances at $p = 0.05$. From these results, the GMM classifier using the Relief method with 5 features was the best classification scheme in this scenario. It performed statistically similar while requiring less number of features.

TABLE V: VT event classification performance results for the GMM classifier with different feature selection filters for the 25-75% training-test split scenario.

Best Scheme					Other Schemes				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)	Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)
Relief+20f	91,4375	92,4697	90,2917	0,0026	Relief+10f	89,8125	88,0092	92,2917	0,0026
					Relief+5f	**89,6250**	**86,9856**	**93,2500**	**0,0030**

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

Similarly, for the 50-50% training-test split scenario, the highest metrics scores were obtained by the GMM classifier using the CHI2 discretization method with 10 features (See Table VI). This result appeared as the absolute winner in this scenario ($p = 0.05$).

TABLE VI: VT event classification performance results for the GMM classifier with different feature selection filters for the 50-50% training-test split scenario.

Best Scheme				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)
CHI2+10f	**90,8125**	**92,4697**	**91,1163**	**0,0025**

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

In the case of the 75-25% training-test split scenario, the GMM classifier using the Relief method with 20 features obtained the highest performance among others classification schemes (See Table VII). However, the CHI2 discretization method with 10 features was considered the most suitable classification scheme in this scenario, since it required less number of features and took less processing time for reaching a similar performance at $p = 0.05$.

TABLE VII: VT event classification performance results for the GMM classifier with different feature selection filters for the 75-25% training-test split scenario

Best Scheme					Other Schemes				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)	Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)
Relief+20f	91,9375	93,4937	90,3750	0,0038	UF+15f	91,8750	92,4971	91,2500	0,0025
					Relief+15f	91,8125	92,8142	91,0000	0,0041
					Relief+25f	91,5000	92,8753	90,1250	0,0042
					IG+10f	91,3125	92,0725	90,6250	0,0030
					UF+20f	91,0625	91,6328	90,5000	0,0028
					CHI2+10f	**91,0000**	**91,3535**	**90,7500**	**0,0028**

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

Regarding VT seismic events classification, the good performance obtained by the GMM classifier using the CHI2 discretization method with 10 features was predominant for the majority of scenarios. However, the combination in the 50-50% training-test split scenario (see Table VI) overcame the renaming classification schemes for all metrics. Therefore, it was selected as the best method for classifying VT seismic events.

C. Joint Long-Period and Volcano-Tectonic Seismic Event Classification

The results shown in Table VIII, indicate that the GMM classifier using the uFilter method with 15 features reached the highest metrics scores in the 25-75% training-test split scenario. But, there were others classification schemes that demonstrated to be statistically similar in terms of performance ($p = 0.05$). Therefore, the GMM classifier that uses the IG method with 10 characteristics (the least number of characteristics used) appeared as the most appropriate classification scheme in this scenario.

TABLE VIII: Joint LP and VT event classification performance results for the GMM classifier with different feature selection filters for the 25-75% training-test split scenario.

Best Scheme					Other Schemes				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)	Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)
UF+15f	91,6042	91,6436	91,6042	0,0025	Relief+20f	91,4375	91,5110	91,4375	0,0026
					IG+10f	**91,3750**	**91,4344**	**91,3750**	**0,0028**
					Relief+15f	91,1875	91,2479	91,1875	0,0027
					UF+20f	91,1875	91,2459	91,1875	0,0027
					Relief+25f	90,9583	91,0250	90,9583	0,0026
					IG+30f	90,9167	91,0523	90,9167	0,0029
					CHI2+35f	90,8333	91,0127	90,8333	0,0029
					OR+35f	90,1250	90,4700	90,1250	0,0030

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

Similarly, in the 50-50% training-test split scenario, the highest metrics scores were obtained by the GMM classifier using the uFilter method with 15 features (See Table IX). This result was statistically superior when compared to others classification schemes ($p = 0.05$). Although, the IG method with 10 features achieved similar classification results while using less number of features.

TABLE IX: Joint LP and VT event classification performance results for the GMM classifier with different feature selection filters for the 50-50% training-test split scenario.

Best Scheme					Other Schemes				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)	Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)
UF+15f	91,8750	91,9405	91,8750	0,0027	**IG+10F**	**91,7500**	**91,8798**	**91,7500**	**0,0027**
					Relief+20f	91,6875	91,7679	91,6875	0,0040
					IG+25f	91,3750	91,5754	91,3750	0,0057
					UF+20f	91,3750	91,4290	91,3750	0,0035
					CHI2+35f	91,2500	91,5138	91,2500	0,0068
					IG+30f	91,2500	91,4366	91,2500	0,0040
					Relief+25f	91,2500	91,3302	91,2500	0,0028
					Relief+15f	91,2500	91,3178	91,2500	0,0034

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

The last scenario (75-25% training-test split) was lead by the GMM classifier and the Relief method with 20 features. This combination obtained the highest metrics scores and statistically overcame the majority of the classification schemes (See Table VIII). But, the GMM classifier using the uFilter method with 15 features was considered as the most suitable classification scheme in this scenario, since, it performed statistically similar and used less number of features.

TABLE X: Joint LP and VT event classification performance results for the GMM classifier with different feature selection filters for the 75-25% training-test split scenario.

Best Scheme					Other Schemes				
Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)	Method	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)
Relief+20f	91,9375	92,2162	91,9375	0,0038	**UF+15f**	**91,8750**	**92,0590**	**91,8750**	**0,0025**
					Relief+15f	91,8125	92,1308	91,8125	0,0041
					IG+30f	91,6250	91,9020	91,6250	0,0049
					CHI2+35f	91,6250	91,8254	91,6250	0,0066
					IG+35f	91,5000	91,7851	91,5000	0,0066
					Relief+25f	91,5000	91,7539	91,5000	0,0042
					IG+25f	91,5000	91,7129	91,5000	0,0058

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

Regarding the classification of both seismic events together, the GMM classifier in combination with the IG method with 10 features reached the best performance results over most scenarios. But, the selected classification scheme in the 50-50% training-test split scenario demonstrated to be satisfactory while requiring less number of features (see Table IX). Thus, it was ratified as the best selection.

According to the results obtained, the performance of the GMM classifier was better when using the 50-50% split training-test scenario. This effect could be associated to the intrinsic statistical nature of the classifier and the learning algorithm used in the training phase (EM based algorithm). Since the GMM classifier uses mixed Gaussian distributions (statistical based objective function) to approximate new incoming samples, it is constrained by the statistical assumption of having the minimum samples in the training phase. This explains the unsatisfactory performance in the 25-75% training-test split scenario. Similarly, training it with the 75% of the data produced satisfactory results, but, very often introduced instability in the leaning model because of the over-training. 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.

The best feature selection methods were: CHI2 discretization for VT seismic events, and Information Gain for LP and both seismic events together. These two methods share 9 out of 10 features selected as the most relevant to the GMM classifier (f52, f78, f79, f50, f23, f49, f30, f28 and f76). This behavior is explained by the fact that the Information Gain (which is an entropy-based feature evaluation), and CHI2 discretization (which used the chi-square statistical test as the main evaluation function) methods assume a given distribution (Gaussian distribution in most case) from which the data have been collected [14]. Thus, they matched perfectly with the GMM classifier. Furthermore, the GMM relies on the Bayes theorem (entropy based) [26], that explained the superior performance obtained when combined with the IG method.

D. Features Subset Validation

Filter methods for features selection are efficient when using high-dimensional data due to their linear time complexity in terms of the dimensionality N (total of features). But, they are unable to remove redundant features because they are individual evaluators (the weight is assigned according to the degree of relevance [27]), and whenever the features are considered relevant for the class, all features will be selected without taking into account the possibility of being highly correlated with each other (redundant). Therefore, feature subsets validation through a redundancy analysis is very appropriate; the main idea is to reduce the features subset size while keeping the prediction accuracy.

In this sense, a two-step procedure was considered that involved: (1) selecting the best subset of features for each scenario (LP, VT and both seismic events together), and (2) performing the redundancy analysis using the Pearson correlation [25], which allowed us to separate the redundant features from relevant ones, and thus to produce the final optimal subset of features.

For better results interpretation, we used two degrees of features relevance: strong and weak (redundant and nonredundant) [28]. Strong relevance points out the feature importance in such way that it can not be removed without losing prediction accuracy. On the other hand, weak relevance makes reference to the feature that contributes in a lower degree to the prediction accuracy. Besides, irrelevant features can never contribute to the prediction accuracy. Thus, a features subset is relevant if it contains only strong and weak features.

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 the starting point for the redundancy analysis step, which used an heuristic value of correlation as the acceptance threshold (c-Pearson values greater than 0.75 on both positive and negative direction).

TABLE XI: Summary of the redundancy feature analysis.

Event Type	Method	Best Features Subset	Redundant Features	c-Pearson	Weakly Relevant	Strongly Relevant
LP	IG+10f	f52, f78, f79, f50, f23, f49, f30, f28, f76, f60	f52= f78, f50, f79 f49= f76 f60= f23	0.98, 0.93, 0.91 0.92 0.83	f52, f49, f60	f30, f28
VT	CHI2+10f	f78, f52, f79, f50, f49, f76, f23, f30, f75, f28	f78=f52, f50, f79 f75= f49, f76	0.98, 0.95, 0.92 0.96, 0.96	f78, f75	f23, f30, f28
LP+VT	IG+10f	f52, f78, f79, f50, f23, f49, f30, f28, f76, f60	f52= f78, f50, f79 f76= f49 f23= f60	0.98, 0.93, 0.91 0.92 0.83	f52, f76, f23	f30, f28

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

From Table XI, 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 importance in the group (redundant). Therefore, only one of the features in the correlated pair is selected to form the subset of weakly relevant features, while the non-correlated features were included in the strongly relevant subset.

According with our results, for the classification of LP, VT and both seismic events together (LP and VT), features f28, f30 and f23 (only for VT events) were selected as the main strongly relevant features, where feature f23 corresponds to the amplitude and frequency of the maximum peak in the band 10–20 Hz, feature f28 are the amplitudes and frequencies of the second and third peaks of the FFT of the signal, while feature f30 is the number of peaks in Power Spectral Density (PSD) of the signal above its RMS value.

These features were consider as strongly relevant features because they are uncorrelated and their absence in the final subset significantly decreased the accuracy, precision and recall scores: for the classification of LP seismic events, from 91.7500 to 85.3750 ($p < 0.01$), 91.0448 to 78.3948 ($p < 0.01$) and 92.7500 to 87.8750 ($p < 0.01$), respectively. Similarly, for VT seismic events classification, the scores were reduced from 90.8125 to 79.5000 ($p < 0.01$), 91.1163 to 88.8637 ($p < 0.01$) and 90.6250 to 67.5000 ($p < 0.01$), respectively. While for the classification of both seismic events together, the scores were reduced from 91.8750 to 85.1875 ($p < 0.01$), 91.9405 to 87.1086 ($p < 0.01$) and 91.8750 to 85.1875 ($p < 0.01$), respectively. These results were expected, since features f28, f30 and f23 are features taken from the frequency domain with higher discriminative power, i.e. the mean value of the f23 feature for LP seismic events is 0.1179 and for VT seismic events is 0.6233, that represents a value difference of 0.5054. Something similar occurred for the remaining f28 and f30 features. These significant differences can be seen as a clear boundary separation between classes when contrasting the features of each others in the features space (see Fig 4 top row) and when the GMM classifier is used, as illustrated in Fig 4 bottom row).

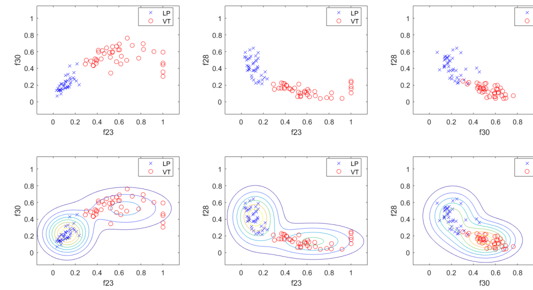


Fig. 4. Strongly relevant features distribution plot: in the natural features space (top-row) and using the GMM classifier with the EM convergence (bottom-row).

Following the definition of "relevant and irrelevant features" stated in [28], we joint together both weakly and strongly relevant features to form the optimal subset of features. These subsets were used for feeding the GMM classifier in order to establish a performance statistical comparison between the best selected subset of features and its corresponding optimal feature subset after the redundancy analysis. Table XII summarizes the performance comparison.

TABLE XII: Statistical comparison of feature subsets based on the employed metrics between the best and optimal feature subset.

Event Type	Best Subset	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)	Optimal Subset	Acc. (%)	Pre. (%)	Rec. (%)	Time (s)	Wilcoxon ($\alpha = 0.05$)
LP	f52, f78, f79, f50, f23, f49, f30, f28, f76, f60	91.7500	91.0448	92.7500	0.0027	f52, f49, f30, f28, f60	93.6875	92.7632	94.875	0.0026	$p < 0.05$
VT	f78, f52, f79, f50, f49, f76, f23, f30, f75, f28	90.8125	91.1163	90.6250	0.0025	f78, f23, f30, f75, f28	92.8125	93.1934	92.5000	0.0023	$p < 0.05$
LP+VT	f52, f78, f79, f50, f23, f49, f30, f28, f76, f60	91.8750	91.9405	91.8750	0.0027	f52, f76, f23, f30, f28	93.4375	93.5272	93.4375	0.0021	$p < 0.05$

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

As shown in Table XII, the optimal feature subset increased the performance of the GMM classifier while decreasing the number of features and the processing time in the classification of all types of seismic events. Despite the satisfactory classification results obtained when using the best subset of characteristics (subset of features before the redundancy analysis), the limited performance of the GMM classifier under the presence of correlated characteristics is remarkable [29].

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, while previous studies conducted in [30] and [3] were focus only in the classification of a single event (LP). By classifying LP seismic events, those studies reached better performance results, i.e. the scores obtained in [3], using a linear Support Vector Machine (SVM) classifier (linear kernel) with fifteen features (obtained using a wrapper based feature selection method) as input, seems to be the best for detecting this type of seismic event, the classifier required 50 milliseconds to fulfill the task. On the other hand, for the approach proposed in [30], the Decision Tree (DT) classifier with five features (feature selection obtained based on a combination of wrapper and embedded methods) was the fastest, consuming a processing time of 31 milliseconds.

Despite the good processing time obtained by these approaches, the method proposed in this paper achieves its best classification results in 2.6 milliseconds, guarantying the shortest processing time among them. This difference is mainly justified by the fact that previous approaches employed classifier dependant feature selection methods (as wrapper or embedded) 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 the method previously developed in [4], which use a multi-class SVM classifier with four features (extracted from the scale domain) for classifying LP and VT seismic events as disjoint units, highlighted the proposed method as the best classification scheme in almost all metrics for both types of seismic events (see Table XIII).

Regarding the relevance of feature types, the best selected subset used in [3] was formed by features from the time domain (one), frequency domain (five), and scale domain (nine); in [30], the best features (five) were all extracted from the frequency domain, while the proposed method considered two features from the frequency domain and three from the scale domain. It is important to notice that feature f23 (computed from the frequency domain, as it is shown in Table I) was selected as one of the most relevant features by the method proposed in [30] and also by the method proposed here. This behaviour corroborated the importance of using frequency domain features for seismic events classification. In general, these results provided clearly experimental evidence that the proposed method produces good classification results in almost real-time.

TABLE XIII: Comparison based on the accuracy, precision, recall and time between previous works and the proposed method

Method	Event	Acc. (%)	Pre. (%)	Reca. (%)	Time (s)
Artificial Neural Network [30]	LP	97	100	93	0,230
DT [30]	LP	96	98	93	0,031
Multi-Class SVM [4]	LP/VT	95/91	86/88	95/75	-
Linear SVM [3]	LP	97	97	96	0.05
Proposed Method	LP/VT	94*/93*	93*/93*	95*/93*	0,0026

Acc. - Accuracy; Pre. - Precision; Rec. - Recall; f - Features; *Rounded Values

IV. Conclusions and Future Work

In this paper, different filter based feature selection methods have been 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 IG method showed the best performance, while for VT seismic event classification, the CHI2 discretization method outperformed the others. Similarly, regarding the classification of both seismic events together, the IG method reached the best performance results over most scenarios. In all cases, the classification schemes presented in the 50-50% training-test split scenario performed slightly better in terms of processing time, ensuring it, as the best selection.

Further redundancy analysis allowed to obtain an optimal set of features 50% smaller than those obtained by the IG and CHI2 discretization methods before the redundancy analysis. All the features determined as most significant (i.e. strongly relevant), have the resemblance of being in the frequency domain, while those considered to be weakly relevant are in the scale and time domains, providing less relevant information to the classifier. According to the Wilcoxon statistical test, the proposed method demonstrated that it can provide competitive schemes for seismic event classification while reducing the processing time required.

Future work will focus on extending the GMM classifier scheme for use in multi-class classification problems (several types of seismic events).

Acknowledgment

The authors gratefully acknowledge the financial support during the development of this project of Universidad San Francisco de Quito (USFQ) under the Poli Grants Program (Grants no. 10100 and XXXX). This work was also partially supported by Universidad de las Fuerzas Armadas-ESPE under Grants 2013-PIT-014 and 2016-EXT-038. The seismic data used in this study was provided by Instituto Geofísico, EPN.

References

[1] R. A. Lara-Cueva, A. S. Moreno, J. C. Larco, and D. S. Benítez, “Real-time seismic event detection using voice activity detection techniques,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 12, pp. 5533–5542, Dec 2016.

[2] R. V. Allen, “Automatic earthquake recognition and timing from single traces,” *Bulletin of the Seismological Society of America*, vol. 68, no. 5, p. 1521, 1978. [Online]. Available: http://dx.doi.org/

[3] 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 Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5247–5257, Sept 2016.

[4] 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 *2017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, Nov 2017, pp. 1–6.

[5] I. Alvarez, L. Garcia, G. Cortes, C. Benítez, and A. D. la Torre, “Discriminative feature selection for automatic classification of volcano-seismic signals,” *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 2, pp. 151–155, March 2012.

[6] D. Cárdenas-Peña, M. Orozco-Alzate, and G. Castellanos-Dominguez, “Selection of time-variant features for earthquake classification at the nevado-del-ruiz volcano,” *Computers & Geosciences*, vol. 51, pp. 293 – 304, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0098300412002919

[7] 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,” vol. 316, pp. 34–49, 03 2016.

[8] M. Dash and H. Liu, “Feature selection for classification,” *Intelligent data analysis*, vol. 1, no. 3, pp. 131–156, 1997.

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

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

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

[12] K. Kira and L. A. Rendell, “A practical approach to feature selection,” in *Machine Learning Proceedings 1992*, D. Sleeman and P. Edwards, Eds. San Francisco (CA): Morgan Kaufmann, 1992, pp. 249 – 256. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781558602472500371

[13] H. Liu and R. Setiono, “Chi2: feature selection and discretization of numeric attributes,” in *Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence*, Nov 1995, pp. 388–391.

[14] N. P. Pérez, M. A. G. López, A. Silva, and I. Ramos, “Improving the mannwhitney statistical test for feature selection: An approach in breast cancer diagnosis on mammography,” *Artificial Intelligence in Medicine*, vol. 63, no. 1, pp. 19 – 31, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0933365714001419

[15] D. Reynolds, “Gaussian mixture models,” *Encyclopedia of biometrics*, pp. 827–832, 2015.

[16] H. D. Ortiz Erazo, “Study of the site effects for the construction of a seismic activity index at the cotopaxi volcano - estudio de los efectos de sitio para la construcción de un índice de actividad sísmica en el volcán cotopaxi,” *Masters thesis, Escuela Politécnica Nacional*, 2013.

[17] Y. Vaezi and M. Van der Baan, “Comparison of the sta/lta and power spectral density methods for microseismic event detection,” *Geophysical Journal International*, vol. 203, no. 3, pp. 1896–1908, 2015. [Online]. Available: http://dx.doi.org/10.1093/gji/ggv419

[18] 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 *2015 Asia-Pacific Conference on Computer Aided System Engineering*, July 2015, pp. 129–134.

[19] T. K. Moon, “The expectation-maximization algorithm,” *IEEE Signal processing magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[20] Y. K. Jain and S. K. Bhandare, “Min max normalization based data perturbation method for privacy protection,” *International Journal of Computer & Communication Technology*, vol. 2, no. 8, pp. 45–50, 2011.

[21] 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,” *European Journal of Operational Research*, vol. 169, no. 2, pp. 477 – 489, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221704005491

[22] 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, 03 1947. [Online]. Available: https://doi.org/10.1214/aoms/1177730491

[23] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *Journal of Machine learning research*, vol. 7, no. Jan, pp. 1–30, 2006.

[24] 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

[25] J. D. Gibbons and S. Chakraborti, “Nonparametric statistical inference,” in *International encyclopedia of statistical science*. Springer, 2011, pp. 977–979.

[26] E. Jaynes, “The relation of bayesian and maximum entropy methods,” in *Maximum-entropy and Bayesian methods in science and engineering*. Springer, 1988, pp. 25–29.

[27] 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

[28] G. H. John, R. Kohavi, and K. Pfleger, “Irrelevant features and the subset selection problem,” in *Machine Learning Proceedings 1994*, W. W. Cohen and H. Hirsh, Eds. San Francisco (CA): Morgan Kaufmann, 1994, pp. 121 – 129. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781558603356500234

[29] N. R. Chaganty, R. Sabo, and Y. Deng, “Alternatives to mixture model analysis of correlated binomial data,” *ISRN Probability and Statistics*, vol. 2012, 2012.

[30] R. Lara-Cueva, E. V. Carrera, J. F. Morejon, and D. Benítez, “Comparative analysis of automated classifiers applied to volcano event identification,” in *2016 IEEE Colombian Conference on Communications and Computing (COLCOM)*, April 2016, pp. 1–6.