Benchmarking Seismic-Based Feature Groups to Classify the Cotopaxi Volcanic Activity

Noel Pérez, *Member, IEEE*, Pablo Venegas, Diego Benitez, *Senior Member, IEEE*, Felipe Grijalva, *Member, IEEE*, Román Lara, *Senior Member, IEEE*, and Mario Ruiz

***Abstract*— We systematically and comprehensively tested almost 100 feature groups in four commonly used automatic event classifiers to find the combinations that maximize the classification performance for long-period and volcano-tectonic seismic events at Cotopaxi volcano, Ecuador. The feature groups tested fall into the following categories: time, fast-Fourier, wavelet transform, intensity-statistics, shape, and texture. An analysis of the relevance of feature groups highlights and finds that each classifier performs similar to one another when using the best combination of features, as determined by the mean of the area under the curve metric, with no statistically significant differences between them. The feed-forward back-propagation neural network and random forest classifiers achieved scores of 0.96 while the naive Bayes and k-nearest neighbor (kNN) classifiers achieved scores of 0.95. The shape and texture feature groups were the most frequently utilized (three times each) among the different classifiers, and thus, the most appropriate for classifying both types of seismic events. The kNN ($k = 3$) classifier using the shape feature group reached an excellent performance with lower algorithmic complexity. Moreover, nontraditional features like those computed from spectrogram images overcame traditional time, Fourier, and scale features when classifying seismic events.**

***Index Terms*— Feature relevance, feed-forward back-propagation (FFBP) neural network, k-nearest neighbors (kNN), machine learning classifiers (MLCs), naive Bayes (NB), random forest (RF), seismic events classification, volcano seismic events features.**

I. Introduction

THERE has been considerable progress made in recent years on the use of machine learning classifiers (MLCs) for detection and classification of seismic signals. The detection and classification of seismic events is a fundamental tool used to help analysts to estimate the internal status of a volcano, and efficient and successful automatic schemes are highly sought after to decrease analyst workload. With this aim, different feature representations generated from raw seismogram recordings have been proposed to encode characteristic information of the different types of seismic events, including features from time [1]–[5], frequency [1]–[4], [6], and scale domains (i.e., wavelet-based features) [1]–[3], [6]. Other works calculate features from the Mel frequency cepstral coefficients [7], the discrete cosine transform [8], linear predictive coding coefficients [9], and, recently, the intensity statistics, shape, and texture of the spectrogram image [10], Hidden Markov model (HMM) [6], [11], evolutionary algorithms [12], [13], and Gaussian mixture models [1]. Other approaches use unsupervised [14], [15] and semisupervised schemes [2]. Although results are variable in terms of accuracy depending on several factors such as the type of event considered [e.g., long-period (LP), volcano-tectonic (VT)], the data set distribution (e.g., number of events, whether it is balanced or not), and the state of the volcano (e.g., normal, unrest), one key factor that can boost or hinder a classifier's performance is the group of features selected.

In spite of this progress, determining the most relevant feature group or combination of features remains an unsolved problem in the context of volcano seismic events classification [5]. Fourier- and wavelet-based descriptors have consistently demonstrated to be the most popular combination. However, the recent satisfactory results obtained in [10] using a set of nontraditional features motivated us to think about whether the features groups based on time, frequency, and scale are the definitive combinations for tasks such as LP and VT classification. In this context, we explore new spaces that contain sets of nontraditional features that provide information on the classification of these events from a different point of view (e.g., the spectrogram image-processing space).

In this work, we conduct a comparative study of different feature groups to determine the combinations that maximize the area under the curve (AUC) of the receiver operating characteristic (ROC) scores in the classification of LP and VT seismic events using four MLCs with different taxonomies. The goal of this technique goes beyond simple feature selection and instead aims to provide a fairer selection of the best combinations of feature groups from a group space. The group space includes traditional time, frequency, and scale sets in addition to other less-traditional metrics such as intensity, shape, and texture sets computed from the spectrogram images. Throughout the proposed experiment, this research will help to improve the feature catalog consistency for future seismic event classifications. Although explored models are not designed to perform in real time, we avoid unnecessary computation of vast amounts of random features to improve the performance of the models as much as possible.

II. Volcano Seismic Event Data Set

We used a subset of the publicly available *SeisBenchV1* data set from the ESeismic repository [16]. *SeisBenchV1* includes

Manuscript received May 29, 2020; revised August 1, 2020 and September 4, 2020; accepted September 29, 2020. This work was supported in part by the Universidad San Francisco de Quito under Grant 16916 and Grant 12494 and in part by the Universidad de la Fuerzas Armadas under Grant 2016-EXT-038. *(Corresponding author: Diego Benitez.)*

Noel Pérez, Pablo Venegas, and Diego Benitez are with the Colegio de Ciencias e Ingenierías "El Politécnico", Universidad San Francisco de Quito (USFQ), Quito 170157, Ecuador (e-mail: dbenitez@usfq.edu.ec).

Felipe Grijalva is with the Departamento de Electrónica, Telecomunicaciones y Redes de Información, Escuela Politécnica Nacional, Quito 170109, Ecuador.

Román Lara is with the Departamento de Eléctrica, Electrónica y Telecomunicaciones, Universidad de las Fuerzas Armadas–ESPE, Sangolquí 171103, Ecuador.

Mario Ruiz is with the Instituto Geofísico, Escuela Politécnica Nacional, Quito 170109, Ecuador.

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

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

Digital Object Identifier 10.1109/LGRS.2020.3028193

1187 feature vectors computed from signals recorded on two stations (VC1 and BREF) during 2012, January 2013, January 2014, January 2018, and January to March 2019. Since these features depend on the station location [16], we only considered feature vectors from the VC1 station.

Specifically, our data set contains 668 feature vectors distributed between 587 LPs and 81 VTs. Other events were not considered due to their small representation. Each feature vector is a 99-dimensional vector represented by 84 features (originally computed in the *SeisBenchV1* data set), that is, 13 from the time-domain, 21 from the frequency domain, and 50 from the scale domain, and 15 features (proposed and computed in [10]) related to the intensity, shape, and texture from gray-scale spectrogram images.

III. Considered Groups of Features

We considered the six different groups of features of our data set, in which three of them contain traditional features of the time, frequency, and scale domains. In contrast, the others include nontraditional features extracted from the intensity statistics, shape, and texture of the seismic event patterns of the computed spectrogram images. These features were selected based on their acceptance in the volcanology community. The time-, frequency-, and scale-based features have been widely employed to analyze raw signals. As had been demonstrated in state of the art [5], [17] and previously developed methods [1], [3], they consistently provide good classification results. On the other hand, the intensity statistics, shape, and texture features have been less explored, but they demonstrated a generalizable classification performance across several MLCs [10].

The use of these groups allows us to gather information on the importance of traditional versus nontraditional features for the problem being analyzed, and whether traditional features are the ultimate combination to classify LPs and VTs. For more detailed information about the features' calculation refer to [3], [10], [16]. A description of these feature groups is as follows:

A. Time-Based Features (13)

these are focused on statistical calculations (e.g., variance), entropy, ratios (e.g., peak-to-rms), and relevant signal landmarks (e.g., time to reach its maximum amplitude). They try to capture the signal's waveform properties, which are discriminative to separate VT from LP events because of their distinctive shapes (see Fig. 1). However, time features alone are not enough to discriminate shallow VTs with waveforms similar to that of LPs. In contrast, shallow VTs present a distinctive spectrum captured by frequency features.

B. Fourier-Based Features (21)

these are related mostly to peaks in frequency, ratios (e.g., the density of peaks above rms value), entropy, and statistical measures in frequency-domain (e.g., kurtosis). As depicted in Fig. 1, high-frequency spectral content of VTs is highly discriminative when compared to the persistent low-frequency content of LPs.

C. Scale-Based (Wavelet Transform) Features (50)

the wavelet transform extracts representative frequency and time features at different scales and is thus well-suited for discriminating between VTs and LPs which have different dominant frequency bands. Features are extracted using

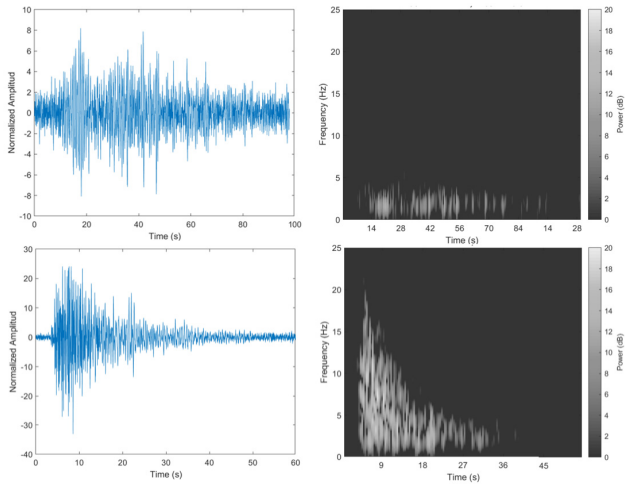

Fig. 1. (Top) Typical LP and (Bottom) VT events and their spectrograms.

Daubechies db10 wavelets up to level 6, and are mainly related to frequency peaks, signal statistics (e.g., mean), ratios (e.g., peak-to-rms), and wavelet component energy distribution.

D. Intensity Statistics-Based Features (5)

these are mainly focused in the pixel intensity distribution of a spectrogram and how skewed, stretched, and dotted this distribution is [10]. The features in this group are excess kurtosis, maximum, mean, skewness, and standard deviation.

E. Shape-Based Features (6)

these are related to an event's geometrical form in a spectrogram, being computed from both the segmentation and bounding box curve where the seismic event pattern is embedded [10]. The main features in this group are: the area, perimeter, circularity, and x and y center of mass.

F. Texture-Based Features (6)

these are computed from the gray-level cooccurrence matrices of the seismic event patterns found on their gray-level spectrogram image [10]. These features constitute the well-known Halarick's descriptors [18] and describe the intensity relationship of a pair of neighbor pixels. The features considered were: angular second moment, contrast, correlation, entropy, and inverse difference moment.

IV. Machine Learning Classifiers

The decision boundary between LP and VT classes can be modeled as a supervised classification problem. After the model's training, it will assign a class to new unseen samples. We used four different supervised MLCs with different taxonomies.

1) The naive Bayes (NB), which is based on probabilistic models with strong (naive) conditional independence assumptions. NB estimates the class priors and the probability distribution of the features; any test sample will follow the decision rule according to the Bayes' theorem that provides the most probable value of the class.
2) The k-nearest neighbors (kNN) classifier, which is a nonparametric technique that assigns a test sample to the class of the majority of its k-neighbors (usually by the Euclidean distance).

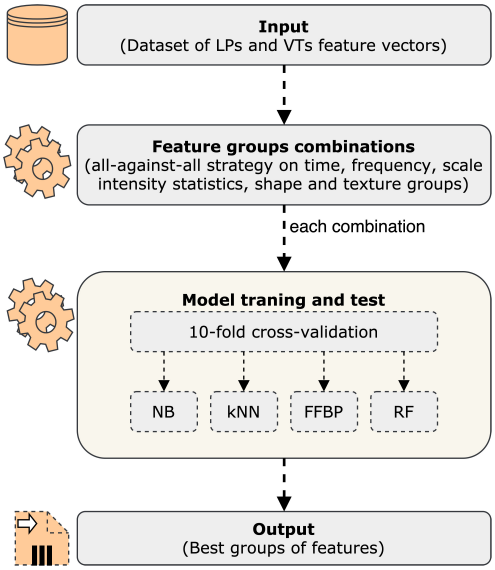

Fig. 2. Experimental workflow.

3) The feed-forward back-propagation (FFBP) neural network, which provides a nonlinear mapping between its inputs and output according to the back-propagation algorithm.
4) The random forest (RF), which randomly divides the inputs to feed several tree predictors, and whose output is the statistical mode of all classification results. For a complete description of these classifiers refer to [10].

V. Experimental Design

We designed a procedure to determine the best combinations of feature groups, as shown in Fig. 2. First, a data set of LPs and VTs feature vectors is used to create several feature group combinations. These combinations are then used with a tenfold cross-validation method to form independent training and test sets (per fold), which will serve to estimate the predictive power of the classifiers on unseen data. Finally, the four MLCs are fed with the created partitions to determine the best model and, thus, the best combination of feature groups. A short description of the experimental setup is given below.

A. Combination of Feature Groups

We employed an all-against-all strategy to combine groups of features. It is noted that the combinations are based on entire groups instead of individual features. In this manner, the features vectors' dimension varied across combinations. Thus, according to the considered six groups, the total number of combinations is 63.

B. Training and Test Partitions

We build exclusive training and test partitions using the tenfold cross-validation method [19]. Therefore, each MLC has the opportunity to be trained with different training sets, helping it to learn from different input space representations. In the same way, using different test sets improves the variability in the obtained results. Since the experimental data set is formed with an imbalanced representation between the number of LPs and VTs, we use a stratified cross-validation approach to preserve the percentage of samples for each class. This adjustment allows us to avoid biases in the MLCs and, thus, to obtain reliable results.

C. MLC Configuration

The NB classifier is parameterless and therefore, no configuration was necessary. For the kNN classifier, we varied k (size of the neighborhood) from 3 to 11 in increments of 2 units and with a weighted Euclidean distance metric used to measure neighbor contributions. The FFBP neural network used one hidden layer with a total of $n = (\text{attributes} + \text{number of classes})/2$ neurons and one binary output layer associated with the LP or VT event classification. All layers employed the sigmoid activation function. The number of iterations (epochs) was optimized in the interval from 100 to 2000 epochs with increment of 500 units, and the learning rate was set to 0.3 (by default). The RF classifier used an optimal number of tree classifiers optimized in the range from 100 to 2000 with increment of 500 units. Each tree used the $\log_2(X) + 1$ for randomly selected attributes, where X is the total number of attributes.

D. Validation Metrics

We used the mean AUC metric to evaluate MLCs performance. Additionally, the Wilcoxon statistical test [20] was employed once per comparison to assess the importance of the differences between classification models.

E. Selection Criteria

To identify the best predictive model, we must search through four MLCs with different hyperparameter configurations, all applied to a large variety of feature group combinations. To choose the best classification model, we follow three general rules: 1) if a model has a statistically significant high mean AUC score per classifier then that MLC is chosen, else; 2) if two MLCs have a statistically insignificant (tied) rating performance, the classifier with less algorithm complexity is preferred. Finally, to select the best combination of feature groups, we considered; and 3) the participation frequency of each group across the best classification models chosen by the previous rules. The MLCs were implemented on WEKA toolkit [21].

VI. Results and Discussion

A. Feature Groups Relevance

Next we describe the results obtained across the 63 feature group combinations using the tenfold cross-validation method:

1) NB Classifier Performance: Since NB is parameterless, only one architecture was explored. The highest mean AUC achieved was 0.95 using the combination of intensity, shape, and texture groups [Fig. 3(a)]. This effect could be related to the NB classifier nature and these particular groups. The NB is a probabilistic-based classifier, and the spectrogram pixel intensities are homogeneous in most parts of the image, except where the event pattern is located. Also, the relationships among neighboring pixels inside the formed pattern vary depending on the type of event as well as its shape. Therefore, this combination of features provided the data variation needed by this classifier to find the decision boundary of both classes.

2) kNN Classifier Performance: The kNN models provided successful results using the shape group. This classifier reached mean AUC scores of over 0.95 for all cases [Fig. 3(b)]. These results did not show statistical differences among them, which is indicative of excellent performance of the classifier. This situation is linked to the difference in the pattern's shape of both types of classes, that is, the LP events are concentrated on low-frequency values, representing a more elliptical form

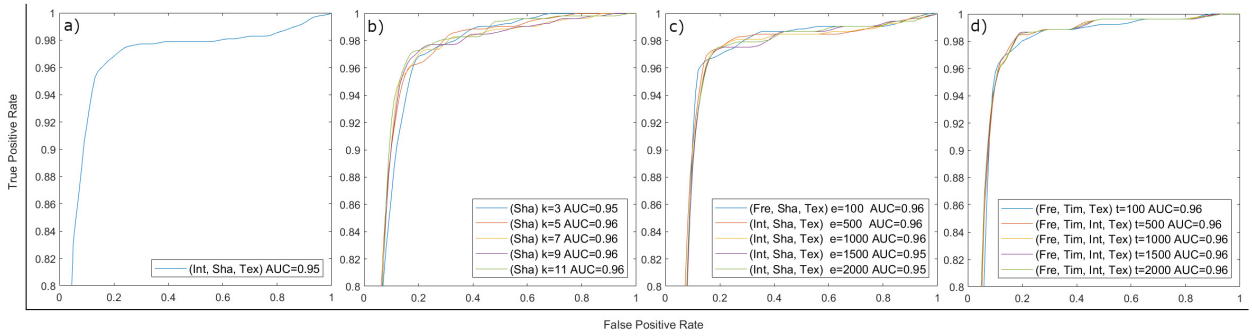

Fig. 3. Receiver operating characteristic (ROC) curve of the classification performance result for (a) NB, (b) kNN, (c) FFBP, and (d) RF classifiers.

(the small diameter along the y-axis), while VT events are accentuated on both the low and high frequencies, depicting a more triangle form. Thus, there is a clear morphological difference between both classes, which was appreciated by the kNN classifier. The classifier using $k = 3$ was selected as the best model by applying the logic of lower algorithm complexity (second selection criterion) form above.

3) FFBP Neural Network Classifier Performance: This classifier provided excellent results independently of the employed configurations. For all cases, it reached a mean of AUC scores over the 0.95 threshold, as shown in Fig. 3(c). While the epochs were changing, the best combination of features changed too. At 100 epochs, it provided the best result using the frequency, shape, and texture groups. The next best model results were obtained using the intensity, shape, and texture groups. The variation in terms of group combinations is likely linked to the learning ability of this classifier: as the iterations were increasing, the classifier was balancing the learning capability. This explains why the latter models consistently used the same combination of feature groups to reach the optimal classification performance. In terms of feature relevance, this classifier has the ability to determine the decision boundary between classes, no matter the degree of complexity in the feature space. Since there is no statistical difference in AUC score among these models, the one using 100 epochs was selected as the best model in this architecture by applying the second selection criterion.

4) RF Classifier Performance: All the models in this classifier achieved the same mean AUC score of 0.96, which is considered to be an outstanding classification performance [see Fig. 3(d)]. The combination of frequency, time, and texture groups was most successful for forests with 100 trees. Subsequently, as the number of trees was increased, the number of combinations increased by adding a new group of features (intensity) to the former set of features. Trees are information entropy-based classifiers. Thus, as the number of trees increases, more data are needed. That could explain the variation of the best combination of features starting from 100 epochs. According to the second selection criterion, the RF with 100 trees is the best model from this classifier.

B. Comparison of Feature Groups

We considered the best model per classifier to carry out the comparison among employed groups of features, as shown in Table I. From this table, the ordered sequence of the feature groups by its participation is shape and texture (equal importance), frequency, time, and intensity (equal importance), and scale. Concerning the classification performance on these groups or combinations of them, except for the RF model, the other MLCs considered the shape group (three times) as an essential set that provides excellent results. This situation is explained because RF models are classifiers dependent on the information variation in the data set. Thus, high variations in the attribute information or feature space makes entropy-based indexes (like in tree classifiers) get more involved. Time and frequency groups are composed of 13 and 20 features, respectively. On the other hand, the shape group included only six features. Thus, finding the decision boundary by each tree in the forest becomes more satisfactory when incorporating more features in the classification space. But, the tree induction could go more in-depth as well, which is a drawback.

TABLE I
SUMMARY OF THE RECURRENCE OF THE APPEARANCE OF THE FEATURE GROUPS WITHIN DIFFERENT MLC SCHEMES

Best model	Groups of features						AUC*
	Tim.$^{+}$	Fre.$^{+}$	Sca.$^{+}$	Int.$^{\pm}$	Sha.$^{\pm}$	Tex.$^{\pm}$	
NB				x	x	x	0.95
kNN with $k = 3$					x	x	0.95
FFBP with $e = 100$		x			x	x	0.96
RF with $t = 100$	x	x				x	0.96
Appearance rate	1	2	0	1	3	3	

Tim-time; Fre-Frequency; Sca-scale; Int-intensity; Sha-shape; Text-texture; $^{+}$computed from raw signals; $^{\pm}$computed from spectrograms; *rounded to two decimal places

Likewise, the texture group was relevant to three of the four MLCs explored. The information regarding the gray-level correlation between pixel values was beneficial for deciding between both types of seismic events. The VTs represent a closer relationship between pixels when compared to LPs (see Fig. 1, right plot). But, the kNN classifier ($k = 3$) model decided only for the shape-based features rather than the texture or other group. This result was expected since there is a comprehensible difference between the formed event pattern on the spectrogram images. Thus, for a distance-based classifier like the kNN, the pattern difference constitutes an optimal separation margin. The frequency group was significant for the FFBP and RF classifiers. The FFBP model combined this group with others, such as shape and texture, to provide the best classification results. Meanwhile, the RF model considered the combination of time and texture with this group to reach the best performance. These results were expected; the FFBP model is a robust nonlinear classifier that, when well-trained, can identify the decision boundaries between event classes with more complex feature spaces. On the other hand, the RF model is constrained by the necessity of information. Thus, it is prevalent to use it on high dimensional feature spaces. That explains why this classifier added another group of features when increasing the number of trees in the forest [see Fig. 3(d)].

The time and intensity groups were the least-used by the implemented MLCs with only one participation each. The RF classifier used the time group, while the NB classifier used the intensity group. Since the RF classifier is based on random trees, it is reasonable to demand increased attribute variety as the number of trees increases. On the other hand, the NB classifier took advantage of the homogeneous pixel intensity in most parts of the spectrogram images to differentiate between seismic event patterns using the posterior likelihood calculation. The scale group was considered the worst. It was irrelevant for the LPs and VTs classification. This situation is probably linked to the redundant information among these features. Since the scale descriptors are computationally related to the time-domain features, the level of correlated information between both groups could be higher than expected. Therefore, after being optimized, most of the chosen MLCs decided by the smaller of the two sets of features (time group). Finally, it should be noted that feature groups computed from the spectrogram images overcame the groups calculated from the raw signal with a participation frequency of 7 against 3. This result might be explained by the fact that because the spectrogram is a time-frequency representation, features from this space efficiently encode highly relevant features from time and frequency domains at the same time.

VII. Conclusion and Future Work

The analysis of feature group relevance highlights the fact that the four employed classifiers performed successfully with respect to the mean of the AUC metric with no statistically significant differences between them. The FFBP and RF classifiers obtained scores of 0.96; meanwhile, the NB and kNN classifiers acquired scores of 0.95. The shape and texture groups were the most frequently used (three times each) from the classifiers' perspective and, thus, the most appropriate to classify both types of seismic events. On the other hand, the kNN with $k = 3$ model using the shape group was the best classification scheme in terms of providing good performance with low overall algorithmic complexity. Moreover, nontraditional features like those computed from spectrogram images outperformed traditional time, Fourier, and scale features in the context of seismic event classification.

As future work, we plan to include more feature sets and additional MLCs to extend the results found in this study. Moreover, we plan to compare features automatically learned from deep learning schemes. Finally, a study on the effects of the path and the site should be carried out to eliminate its impact on the feature sets. This would greatly simplify the classification and feature selection problem while allowing the scheme to be transferable to other stations and volcanoes.

Acknowledgment

The authors thank the Applied Machine Learning Research Group of USFQ for the NVidia DGX workstation to run our experiments. The authors thank S. Hernandez for proofreading the final document.

References

[1] P. Venegas, N. Pérez, D. Benítez, R. Lara-Cueva, and M. Ruiz, "Combining filter-based feature selection methods and Gaussian mixture model for the classification of seismic events from Cotopaxi Volcano," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1991–2003, 2019.

[2] C. Brusil, F. Grijalva, R. Lara-Cueva, M. Ruiz, and B. Acuna, "A semi-supervised approach for microseisms classification from Cotopaxi volcano," in *Proc. IEEE Latin Amer. Conf. Comput. Intell. (LA-CCI)*, Nov. 2019, pp. 1–6.

[3] R. A. Lara-Cueva, D. S. Benitez, E. V. Carrera, M. Ruiz, and J. L. Rojo-Alvarez, "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.

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

[5] M. Malfante, M. D. Mura, J.-P. Metaxian, J. I. Mars, O. Macedo, and A. Inza, "Machine learning for volcano-seismic signals: Challenges and perspectives," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 20–30, Mar. 2018.

[6] P. Alasonati, J. Wassermann, and M. Ohrnberger, "Signal classification by wavelet-based hidden Markov models: Application to seismic signals of volcanic origin," in *Statistics in Volcanology*. London, U.K.: Geological Society London, 2006.

[7] M. Orozco-Alzate, J. M. Londoño-Bonilla, V. Nale, and M. Bicego, "Towards better volcanic risk-assessment systems by applying ensemble classification methods to triaxial seismic-volcanic signals," *Ecol. Informat.*, vol. 51, pp. 177–184, May 2019.

[8] J. M. Ibáñez, C. Benítez, L. A. Gutiérrez, G. Cortés, A. García-Yeguas, and G. Alguacil, "The classification of seismo-volcanic signals using hidden Markov models as applied to the Stromboli and Etna volcanoes," *J. Volcanol. Geothermal Res.*, vol. 187, nos. 3–4, pp. 218–226, Nov. 2009.

[9] B. Apolloni *et al.*, "Support vector machines and MLP for automatic classification of seismic signals at Stromboli volcano," in *Proc. 19th Italian Workshop Neural Nets*, vol. 204. Vietri sul Mare, Italy: IOS Press, May 2009, p. 116.

[10] N. Pérez, P. Venegas, D. Benitez, R. Lara-Cueva, and M. Ruiz, "A new volcanic seismic signal descriptor and its application to a data set from the Cotopaxi volcano," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6493–6503, Sep. 2020.

[11] S. M. Bhatti *et al.*, "Automatic detection of volcano-seismic events by modeling state and event duration in hidden Markov models," *J. Volcanol. Geothermal Res.*, vol. 324, pp. 134–143, Sep. 2016.

[12] J. Liu, H. A. Abbass, and K. C. Tan, "Evolutionary computation," in *Evolutionary Computation and Complex Networks*. Cham, Switzerland: Springer, 2019, pp. 3–22.

[13] K. M. Asim, A. Idris, T. Iqbal, and F. Martínez-Álvarez, "Seismic indicators based earthquake predictor system using genetic programming and AdaBoost classification," *Soil Dyn. Earthq. Eng.*, vol. 111, pp. 1–7, Aug. 2018.

[14] A. Köhler, M. Ohrnberger, and F. Scherbaum, "Unsupervised pattern recognition in continuous seismic wavefield records using self-organizing maps," *Geophys. J. Int.*, vol. 182, no. 3, pp. 1619–1630, Sep. 2010.

[15] A. Duque *et al.*, "Exploring the unsupervised classification of seismic events of Cotopaxi volcano," *J. Volcanol. Geothermal Res.*, vol. 403, 2020, Art. no. 107009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377027320302821, doi: 10.1016/j.jvolgeores.2020.107009.

[16] N. Pérez, D. Benítez, F. Grijalva, R. Lara-Cueva, M. Ruiz, and J. Aguilar, "Eseismic: Towards an Ecuadorian volcano seismic repository," *J. Volcanol. Geothermal Res.*, vol. 396, May 2020, Art. no. 106855.

[17] M. S. Khan, M. Curilem, F. Huenupan, M. F. Khan, and N. B. Yoma, "A signal processing perspective of monitoring active volcanoes [applications corner]," *IEEE Signal Process. Mag.*, vol. 36, no. 6, pp. 125–163, Nov. 2019.

[18] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[19] 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, Mar. 2006.

[20] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.

[21] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explor. Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.