Autoencoders as a Characterization Technique and Aid in the Classification of Volcanic Earthquakes

Paula Andrea Montenegro C., Oscar Ernesto Cadena I., and Anna Diva P. Lotufo, *Senior Member, IEEE*

***Abstract*—Volcanic seismicity is one of the most relevant parameters for the evaluation of volcanic activity and consequently the prognosis of eruptions. Earthquakes of volcanic origin are of different classes, directly related to the physical process that generates them. The distribution of the data between classes of seismic-volcanic signals generally presents an unbalanced profile (imbalanced datasets), which can hinder the performance of the classification in machine learning models. Therefore, this research presents a characterization technique (feature extract) that, in addition to reducing the dimension of each seismic record, allows a representation of the signals with the most relevant and significant information. This work proposes the use of a dual autoencoder feature, which is compared with conventional characterization techniques, such as linear prediction coefficients and principal component analysis. The training of the model was performed with a dataset containing volcano-tectonic (VT) earthquakes, long period events, and Tornillo-type events of the Galeras volcano, one of the most active volcanoes in Colombia. The classification results reach 99% of the classification of the mentioned classes.**

***Index Terms*—Characterization techniques, classification, dual autoencoder, lower dimensional representation, unbalanced dataset, volcano-seismic signals.**

I. Introduction

SINCE the last two decades, automatic recognition algorithms have emerged as a reliable answer to classify seismic data. Among the automatic systems for recognizing and classifying these signals, artificial neural networks (ANN) [1], [2], [3], [4], [5], [6], [7], and hidden Markov models (HMM) [8], [9], [10], have been the most widely applied [8]. Orozco-Alzate et al. [11] indicated that HMMs may sometimes fail to meet the theoretical assumptions made in certain practical situations, due to the nontemporal nature of seismic signals. It is also worth mentioning the extensive processing time and computational cost that continues to be the limiting factor for HMMs [11].

The performance of automatic volcanic earthquake signals detection and classification models presents several challenges, the main one being the availability of a dataset of sufficient size to perform training and validation of the classification model. Although large amounts of seismic data are continuously collected, many of these data remain underutilized because labeling them is an expensive task. Another common problem in volcanic earthquake log data is their unbalanced profile, meaning that one of the classes has a higher number of examples, this is because volcanoes produce signal types that may be predominant, depending on the volcanic activity present. Therefore, classification models must be robust enough to address the class imbalance problem in the dataset, to avoid the tendency to learn the majority or dominant class and fails to predict the remaining classes [12], [13], [14]; this problem is accentuated when the datasets do not have a sufficient number of instances (seismic records in this case).

For the period of volcanic activity selected for this study in the Galeras volcano, the majority class corresponds to vulcano-tectonic (VT) earthquakes of which 1736 have been selected, followed by the long period (LP) class with a selection of 402, and finally the *Tornillo*-type signals (TOR) with only 67 samples.

Usually, in literature, it is chosen to balance the dataset by discarding data from the majority classes until a relatively balanced distribution to the number of examples of the minority class is obtained, further reducing the size of the dataset. Many classification studies of volcanic earthquakes have datasets of sizes smaller than 500 data [1], [7], [15], [16], [17]. However, the application of these techniques generates a loss or deterioration of information in the data [14], [18], [19], [20]. In contrast to class balancing, there is an approach based on feature selection [14], which is also known as characterization or representation [20], [21], in which an unbalanced dataset is allowed to be presented to the classification model without affecting its performance.

Another important challenge presented by seismic data of volcanic origin is its high dimensionality, which can make the classification model computationally wasteful, for this reason feature selection is fundamental in the performance of classification algorithms, which significantly reduces the size of each data. Moreno [12] made reference in this sense by mentioning that the amount of samples needed to obtain statistically significant results grows exponentially with dimensionality. Therefore, the generalization capability of the classification model declines as smaller the quotient between the number of samples and the dimension of the feature space [22], [23].

This article proposes a dual autoencoder feature (DAF) as a novel characterization technique, capable of obtaining a lower

Manuscript received 20 March 2023; revised 10 May 2023; accepted 17 May 2023. Date of publication 29 May 2023; date of current version 18 September 2023. This work was supported in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and in part by the Coordinação de Aperfeiçoamento de Pessoal de nível Superior (CAPES) under Grant Finance Code 001. *(Corresponding author: Paula Andrea Montenegro C.)*

Paula Andrea Montenegro C. and Anna Diva P. Lotufo are with the Department of Electrical Engineering, São Paulo State University - UNESP, Ilha Solteira 01049-010, Brazil (e-mail: paula.andrea@unesp.br; annadiva@ieee.org).

Oscar Ernesto Cadena I. is with the The Colombian Geological Survey, Volcanological and Seismological Observatory of Pasto, Pasto 111321, Colombia (e-mail: ocadena@sgc.gov.co).

Digital Object Identifier 10.1109/JSTARS.2023.3280416

dimensional representation of volcanic earthquake signals and simultaneously addressing the problem for unbalanced datasets.

The proposed feature extraction representation learning model is based on two stacked autoencoders that use different activation functions since the *sigmoid* function is less sensitive to changes, providing a more robust representation, while the *tanh* activation function is more sensitive, offering more detailed information about the data [20], in order to obtain a feature vector that efficiently represents the seismic signals.

Thus, the main contributions of this work include addressing the problem of unbalanced volcanic seismic signal datasets, especially when it is not possible to have a dataset of sufficient size, a common problem in volcanic seismic signals. In addition to offering a new, more robust and efficient characterization technique, capable of defining and differentiating the classes of volcanic seismic signals, facilitating the classification process. This proposed characterization technique is compared with conventional techniques, such as the linear prediction coefficients (LPC) technique [24] and the principal component analysis (PCA) technique [25].

The rest of this article is organized as follows. Section II describes from a geophysical point of view the seismic signals recorded at the Galeras volcano. Section III provides a general theoretical framework on the conventional characterization techniques that are usually applied in literature. Section IV deals with autoencoders and the proposed technique of this work is presented: DAF. Section V describes the experimental setup and Section VI presents the experimental results that serve as a basis for the discussion in Section VII. Finally, Section VIII concludes this article.

II. Seismicity of Galeras Volcano

This study applies to seismic records generated in the Galeras volcano, which is located in the department of Nariño, approximately 9 km west of Pasto city, at coordinates 1°13'43.8" north latitude and 77°2"33.0" west longitude and with a height of 4276 m above sea level. It is considered one of the most active volcanoes in Colombia.

Volcanoes are surface manifestations of dynamic processes that occur in the interior of the Earth, coupling physical and chemical processes of great complexity, which gives rise to a great variety of seismic signals that can be recorded [26], [27]. Although, a uniform global classification scheme for these signals has not yet been established [11], [15], it is generally observed that volcanoes produce earthquakes with comparable characteristics that can be associated to different volcanic sources [9], [11]. For the case study, the main seismo-volcanic events can be grouped as follows (Fig. 1).

1) Vulcano-Tectonic (VT) Earthquakes: They are impulsive signals that originate in a range of depths between 2–10 km. They are produced by stresses caused by a brittle fracture, generating seismic waves in which primae (P) and secundae (S) phases can be identified. The spectral content of this signal is wide, reaching up to 30 Hz.

2) Long Period (LP) Seismic Events: The origin of these signals is related to the fluid dynamics inside the volcanic edifice, where as a consequence of the movement of fluids in cracks or conduits, waves are produced within the fluids involved that cross the crust and reach the receivers. The existence of pressure transients within the fluid–gas mixture also gives rise to resonance phenomena. The sources of such events can be located at depths of more than 2 km, and their frequency content is generally restricted to a range between 0.5–15 Hz.

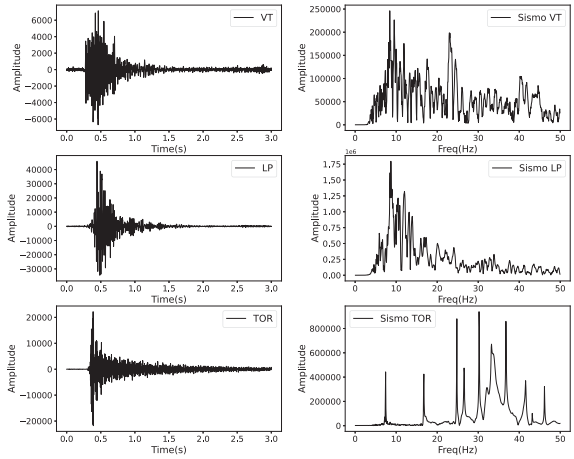

Fig. 1. Types of volcanic seismic signals registered in Galeras volcano. Seismograms (left). Spectra (right).

3) Tornillo *Type Earthquakes (TOR):* These events present characteristic waveforms, where the frequency distribution is homogeneous, with a slow decay coda that can last several minutes and a small amplitude compared to the duration. It is considered that this type of earthquake is associated with resonance processes, where the loss of energy in the resonant cavity is slow. Generally, this type of earthquakes have been precursors of some eruptions in the Galeras volcano [28], [29].

B. Galeras Volcano Dataset

The dataset provided by the Volcanological and Seismological Observatory of Pasto (SGC-OVSP), includes information between July 1, 2004 and December 31, 2010, corresponding to the Cufiño station. Ibarra [1] made a complete description of the sensor systems and signal acquisition mechanism for the Galeras volcano. The data were labeled by seismologists of the observatory, based on professional knowledge and specific experience of the volcano activity. From the total number of earthquakes, the most representative of each class were selected. As a result, a total of 2205 events were obtained with the following distribution: 1736 VT, 402 LP, and 67 TOR.

The histogram in Fig. 2 shows that for LP signals the seismic events have a homogeneous duration of 30 s. However, for the TOR seismic signals, the distribution of the duration is very varied, in addition to being in a smaller class (with a reduced number of earthquakes). Finally, the signals corresponding to VT seismic events, where most of the earthquakes have a duration between 10 and 30 s.

III. Characterization Techniques

These are techniques that allow dimensionality reduction through feature extraction, removing redundant information

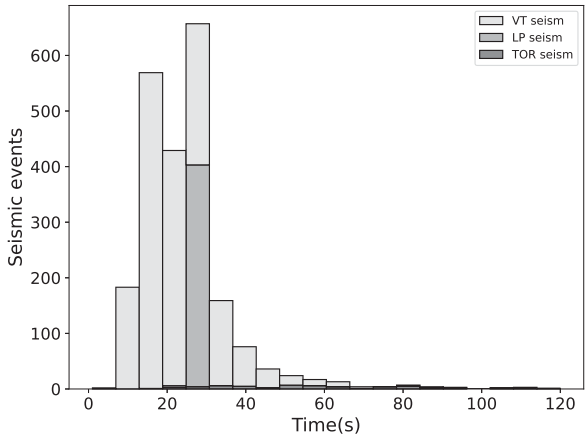

Fig. 2. Histogram summarizes the distribution of the duration of seismic events recorded during 2004–2010 at the Galeras volcano. It is also possible to identify the distribution of instances by seismic event class.

present in the data, and thus, generating a subset of values more suitable for the classification task [30].

Due to the high dimensionality of seismic signal data, it is necessary to address issues, such as scattering, numerical instability, and overfitting. This is a common challenge for high-dimensional data, known in literature as the curse of dimensionality [31]. Techniques that can solve the dimensionality problem, also facilitate the interpretation of the data, which allows for greater generalizability in data analysis models, since they reduce the adaptation of the learning model to the trained data (overfitting) and, finally increase the computational capacity when they reduce the vector dimension, favoring data visualization [17].

In general, algorithms for dimension reduction are classified into two types: 1) resource (or feature) selection, and 2) resource transformation.

A. Feature Selection

The goal is to form a set of the most efficient and relevant features. Among the most commonly used parameters to represent seismic signals are features of geophysical character [21], [32], [33], resources used by specialists in supervised manual classification (impulsivity of the wave onset, the shape of the envelope, energy, periodicity, and harmonics, etc). There are also features based on statistical properties of time series data in the time domain and also in frequency (standard deviation, mean, median, kurtosis and skewness, the mean spectrum and the energy of the spectrum, etc.) [34].

B. Resource Transformation

It is about transforming the signal domain to a space where the information is less sparse. This approach has been popular in literature in the last two decades, for the characterization of volcanic earthquake signals as part of the preprocessing of the data. Among the most common are the Fourier transform, the Hilbert transform, the wavelet transform, the logarithmic frequency cepstral coefficients (LFCC), and the wavelet transform.

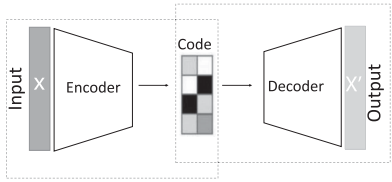

Fig. 3. Autoencoder architecture diagram.

It is about transforming the signal domain to a space where the information is less sparse. This approach has been popular in literature in the last two decades for the characterization of volcanic earthquake signals as part of the preprocessing of the data, among the most common we can mention the Fourier transform, Hilbert transform, wavelet transform, the logarithmic frequency cepstral coefficients (LFCC) [6], [8], [35], [36], [37], Mel-scale frequency cepstral coefficients (MFCC) [38], LPC [2], [15], [24], [39], [40], [41], PCA [16], [38], [42], [43], and among other nonlinear variations derived from the previous ones.

This work will address two of the conventional techniques, LPC and PCA. They will act as a basis of comparison for the proposed characterization technique in the preprocessing of seismic volcanic signals using autoencoders, to transform the signals to a feature subspace as a support in the classification, by reducing the dimension of the data, eliminating redundant information, and as a solution when the dataset presents an unbalanced profile.

IV. Autoencoders

Autoencoders are ANNs that are generally designed to encode the input into a more compact and meaningful representation, which is called code, as shown in Fig. 3. These architectures were first introduced by [44]. It is said that these architectures do not correspond to supervised learning, but to self-supervised learning, because there is no need to label the training dataset. In literature, this type of architectures have been used as a resource extraction technique in the field of remote sensing for the classification of images with super resolution achieving good results [45], [46]. In the field of seismology, its application has been mainly for the elimination of noise in the data [47], [48], [49], the reduction of its dimension [50], [51], as well as the object of this work.

As for its operation, as shown in Fig. 3, for a sample $x^{(i)}$ in the training dataset $D = \{x^{(1)}, x^{(2)}, \ldots, x^{(N)}\}$, where i and N, denote the n-dimensional input vector of the ith sample and N the number of samples, respectively . The coding layer is defined by

$$f(x) = s_e(\mathbf{W}_e x + b_e) \tag{1}$$

where $\mathbf{W}_e$, b_e, and $s_e()$ denote the weight matrix, the threshold (bias), and the activation function of the coding layer, respectively. Similarly, the decoding layer is defined as

$$g(x) = s_d(\mathbf{W}_d x + b_d). \tag{2}$$

Therefore, the autoencoder output is defined as follows:

$$y = g(f(x)). \tag{3}$$

The purpose of the autoencoder is to learn a representation from the most significant features in the coding layer, so that at the output it is possible to reconstruct the inputs. Therefore, the autoencoder learning problem is to find a set of parameters, $\theta = \mathbf{W}_e, b_e, \mathbf{W}_d, b_d,$, that minimize the reconstruction error between the inputs and outputs of the sutoencoder. Therefore the formal definition is $A : \mathbb{R}^n \rightarrow \mathbb{R}^p$(encoder) and $B : \mathbb{R}^p \rightarrow \mathbb{R}^n$(decoder), satisfying

$$\arg\min_\theta \sum_{i=1}^{N} L\left(x^{(i)}, g\left(f\left(x^{(i)}\right)\right)\right) \tag{4}$$

where

$$L(x, y) = \|y - x\|_2. \tag{5}$$

To avoid overfitting, the regularization penalty ℓ_2 is added in the loss function L to constrain the magnitudes of the weights [52], [53]. Therefore, the optimization problem is constrained as

$$\arg\min_\theta \sum_{i=1}^{N} L\left(x^{(i)}, g\left(f\left(x^{(i)}\right)\right)\right) + \frac{1}{2}\lambda\|\mathbf{W}\|_2. \tag{6}$$

Being

$$\ell_2 = \frac{1}{2}\lambda\|\mathbf{W}\|_2 \tag{7}$$

where λ and $\mathbf{W}$ denote the regularization parameter and the matrix consisting of two weight matrixes $\mathbf{W}_e$ and $\mathbf{W}_d$, respectively. In general, the parameter θ is optimized using the backpropagation method in the same way as a standard multilayer ANN, with the exception that for an autoencoder the outputs match the inputs.

For a special case of autoencoder (nonstacked), where the activation function is linear and for a number of hidden neurons M less than the number of neurons of the input n, the problem is reduced to a PCA [54].

In ANN, the activation functions are generally nonlinear. When $M < n$, the autoencoder performs feature reduction, similar to the PCA process. Conversely, when $M > n$, the autoencoder learns the identity function. In practice, the autoencoder explores statistical regularities in the dataset and learns useful features [40], where instead of finding a lower dimensional hyperplane in which the data lies, it is able to learn a nonlinear variety.

The representation capability of an autoencoder with a single hidden layer is limited [54], [55]. A deeper architecture can be formed by stacking several autoencoders to improve the representation capability of features learned from the input data, obtaining deeper and more abstract features [20]. In stacked autoencoders, the outputs of the encoding layer of the first autoencoder correspond to the inputs of the next autoencoder [56]. In this sense, only the first autoencoder uses the original data as input.

A. Dual Autoencoder Features (DAF)

The DAF combines the functions learned by two stacked autoencoders using the sigmoid and tanh functions as trigger functions. Fig. 4 shows the general DAF scheme.

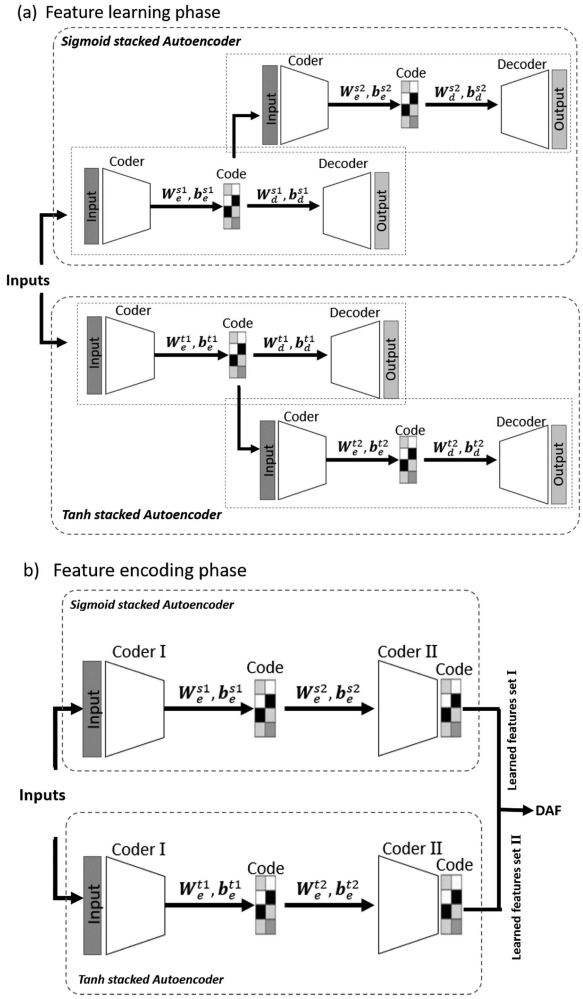

Fig. 4. Dual autoencoder features-DAF. (a) Learning procedures. (b) Feature encoding procedures.

The DAF uses stacked autoencoders that allow extracting the most representative features from the volcanic seismic signals. The DAF does not limit the number of stacked autoencoders, however, a single layer autoencoder may not be robust enough to learn useful features [54], [57], [58]. On the other hand, more layers can produce a good representation, albeit at a high computational cost of training. For this work, the DAF trains two autoencoders independently using two activation functions.

As shown in Fig. 4(a), the first autoencoder encodes the input data, then the second autoencoder uses the outputs of the encoding layer of the first autoencoder as input, encoding the information once again, and a set of learned features is obtained.

Ng et al. [20] used two stacked autoencoders, each with two different activation functions, thus obtaining feature learning from two different perspectives, which are concatenated to form the DAF.

B. Activation Functions

Autoencoders with different activation functions produce different responses to inputs and therefore learn different features, generating more robust representations. The sigmoid and tanh

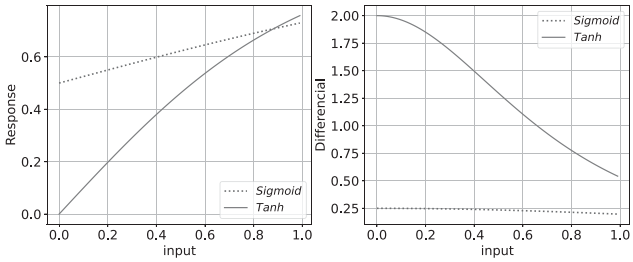

Fig. 5. Activation functions used in DAF for the range [0, 1]. (a) Response of *Sigmoid* and *Tanh* activation functions. (b) Respective differentials of these functions.

activation functions are widely used in ANN. The activation responses and their corresponding differentials can be seen in Fig. 5.

The differentials of the *sigmoid* and *tanh* functions are defined as follows:

$$\text{sigmoid}'(x) = \frac{e^{-x}}{(1+e^{-x})^2} \tag{8}$$

$$\tanh'(x) = \frac{4}{e^{2x}+e^{-2x}}. \tag{9}$$

From Fig. 5(a) it can be identified that the *tanh* function is more sensitive than the *sigmoid* function. The output of the *tanh* function varies whenever there is a small change in the inputs. Likewise, in Fig. 5(b), it is possible to recognize that the *tanh* function is more sensitive by generating larger differentials, thus having a wider range for the activation values, compared to the *sigmoid* function. Therefore, it is possible to capture more detailed information to represent the volcanic seismic signals with the *tanh* activation function.

In contrast, a classifier with low sensitivity produces higher robustness when it ignores detailed changes in the inputs [59]. In this context, the *sigmoid* function is more resistant to noise and provides a more global representation of the data. In this way, it is possible to explore both detailed local information and robust global representations to aid the automatic classification process.

V. Experimental Setup

This work proposes a structure composed of two main stages: data preprocessing in which three characterization techniques were used: 1) LPC, 2) PCA, and 3) DAF; the latter being the proposed technique, followed by the learning and classification stage (Fig. 6).

A. Preprocessing

First, a rectangular window was applied to standardize the signal duration, a duration of 3 s was established for all seismic records to avoid discarding significant information from the seismic records. It is worth mentioning that, for signals with a duration shorter than this value, it was decided to fill with zeros starting from the last recorded sample of the signal.

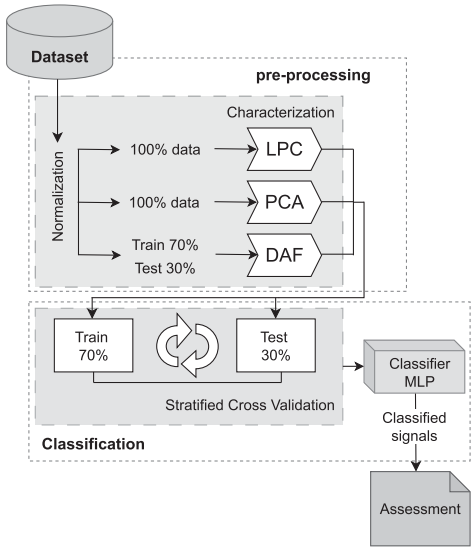

Fig. 6. Experimental setup scheme.

The data were then normalized to the range [0, 1], a process necessary to apply the characterization techniques. The normalization takes into account the maximum and minimum values of the signals, values that are subtracted from each sample to then form part of the denominator as

$$X' = \frac{X - \min|X|}{(\max|X| - \min|X|)}. \tag{10}$$

The normalized data are characterized using the techniques: LPC, PCA, and DAF, obtaining vectors with representations of 12, 50, and 400 features, respectively. The DAF technique generates 200 features corresponding to the coding using the *sigmoid* activation function and 200 features using the *tanh* activation function.

For the LPC technique, the all-pole filter model of order 12 was used, following the methodology used by [4]. For the PCA technique, 50 principal components were calculated taking into account that one of the classes required 49 principal components to conserve 95% of the variance of the data. Finally, the configuration of neurons per DAF layer was $3000 \times 2000 \times 1000 \times 1000 \times 800 \times 1000 \times 1000 \times 2000 \times 3000$ for the autoencoder receiving the original input dice, and $800 \times 400 \times 200 \times 400 \times 400 \times 800$ of the stacked autoencoder.

B. Classification

Once the respective representations of the seismic signals (of lower dimension) are obtained, these are presented as the inputs to the multilayer perceptron (MLP) classifier [60], [61]. The configuration of neurons per layer of the MLP was: $400 \times 500 \times 100 \times 3$. The classifier is trained separately, with one dataset for each type of representation. The learning and classification process was carried out using the stratified cross-validation method, in order to test the success rate of the classification model. The dataset is divided into six stratified groups to maintain the proportion of seismic events per class (having an unbalanced dataset) and then divide the set into training and test groups.

Finally, the confusion matrix of the classification results is obtained. The values of TP, FN, FP, and TN are defined as follows.

1) *True Positive (TP)*: refers to the correct prediction of the class of interest.
2) *False Negative (FN)*: refers to the prediction of the class of interest as if it were a different class.
3) *False Positive (FP)*: refers to the incorrect prediction of a different class as if it were the class of interest.
4) *True negative (TN)*: refers to the correct prediction of classes other than the class of interest.

C. Assessment

The model is evaluated in order to objectively infer the behavior and performance of the classifier under different conditions. The classification model is compared using the data characterized with the proposed technique (DAF) with conventional characterization techniques (LPC and PCA). The metrics used are sensitivity (true positive rate, or recall or TPR) (11) [62], *f1* score (12) and (13) [63], receiver operator characteristic curve (ROC) [64], [65], and its area under the curve (AUC) [66]. It should be noted here that the metrics are calculated as if it were a binary problem from a multiclass problem comparing one class with the rest [One versus Rest (OvsR)].

$$\text{TPR} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$f1 = \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{13}$$

VI. Results

A. Preprocessing

Fig. 7 identifies the representation obtained for VT, LP, and TOR seismic signals, applying the characterization techniques. As can be seen in Fig. 7(a), the LPC technique is the lowest dimension representation and allows defining very well the class of TOR type earthquakes, by obtaining a representation that keeps much similarity between two randomly chosen signals for the same signal type, indicating that this technique manages to extract characteristics that represent well this class. It is worth mentioning that TOR earthquakes have well-defined characteristics both in the time and frequency domain (see Fig. 1). On the contrary, this representation technique does not define well the VT and LP classes, as can be identified in Fig. 7(a), by presenting variations in the representation for two signals of the same type.

On the other hand, the representations obtained by the PCA technique define well the VT class of earthquakes, as shown in Fig. 7(b); however, the LP and TOR classes present variation in the representation of two random signals for the same type of earthquake, this variation is more significant in the TOR class signals. Finally, in Fig. 7(c), it is observed that the representations obtained with the DAF define each of the classes quite well, not finding variation between two random signals

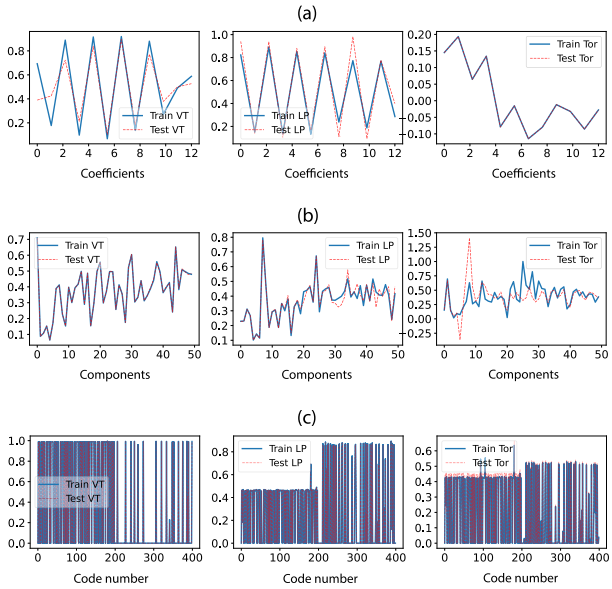

Fig. 7. Characterization of two random signals (from the training and test set) of volcanic earthquakes. (a) LPC technique. (b) PCA technique. (c) DAF technique.

of the same type, thus demonstrating that this technique is able to extract the most representative characteristics in each type of earthquake. This quality can favor the classification process, even though the dataset is unbalanced. Furthermore, it reduces the dimension of the data from 3000 signal samples to a vector with 400 characteristics.

B. Classification

The learning performance, once the classifier was trained and validated (stratified cross-validation), can be seen in Fig. 8, which indicates the behavior of the loss function of the model for each dataset corresponding to the characterization technique used; the percentage of success achieved is also shown in the same figure.

In Fig. 8, it is possible to identify that the MLP classifier presents the highest value for the error minimization in the loss function compared to the error minimization achieved by the model when using a dataset from the PCA or DAF characterization techniques. The accuracy achieved by the classifier does not exceed 82% [Fig. 8(a)], which maintained this performance throughout the stratified cross-validation process. On the other hand, when the datasets corresponding to the PCA [Fig. 8(b)] and DAF [Fig. 8(c)] techniques were used, the error was minimized. The accuracy achieved by the MLP classifier using the dataset when applying PCA is 96% and when applying the DAF technique it reaches 99%. However, it is necessary to continue analyzing the results exhaustively, since this type of behavior can be the result of the correct classification of the majority class only, obtaining erroneous classification results for the minority classes when the dataset is unbalanced [18], [21], [64].

Fig. 9 shows the confusion matrixes resulting from the classification using the different datasets resulting from the characterization techniques. With the LPC technique, the MLP classifier

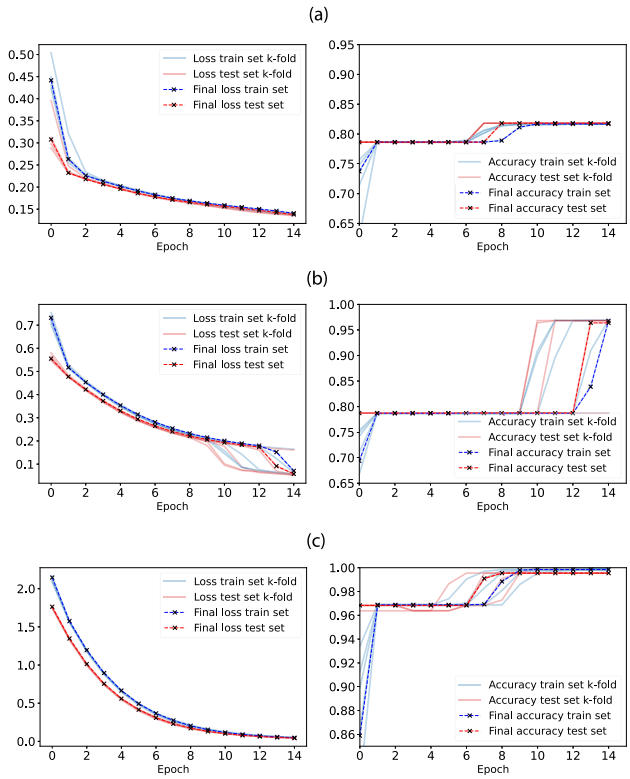

Fig. 8. Behavior of the loss function and the hit achieved in the MLP classifier for each dataset represented. (a) LPC technique. (b) PCA technique. (c) DAF technique.

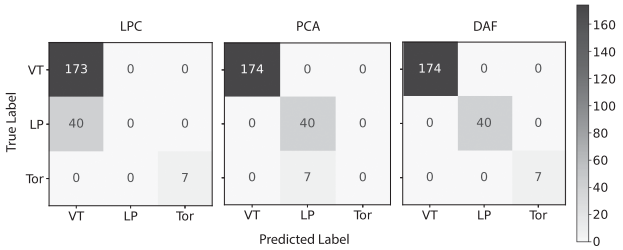

Fig. 9. Confusion matrix for the MLP classifier using the datasets after applying the LPC, PCA, and DAF characterization techniques.

model cannot learn the LP earthquake class, as can be seen in the confusion matrix, this type of signals are confused with those of type VT. On the other hand, it is important to mention that, although the TOR seismic signals correspond to the minority class, the LPC technique allows to classify and correctly label this type of seismic signal.

On the other hand, the classifier results using the dataset from the PCA technique are good for VT and LP earthquake signals. However, the model does not learn to recognize the class of TOR-type earthquakes, thus inferring that this representation technique does not support learning for an unbalanced dataset. Subsequently, the confusion matrix resulting from the classification using the proposed technique for the representation of seismic signals (DAF) indicates that the model is able to correctly learn and predict the signals despite the unbalanced dataset.

TABLE I
RANKING REPORT BY EVALUATION METRICS

Characterization technique		Seism type	Metric Sensibility	Precision	f1-score
	LPC	VT	1.00	0.81	0.89
		LP	0.00	0.00	0.00
		TOR	1.00	1.00	1.00
	PCA	VT	0.99	1.00	0.99
		LP	1.00	0.83	0.91
		TOR	0.00	0.00	0.00
	DAF	VT	1.00	1.00	1.00
		LP	1.00	1.00	1.00
		TOR	1.00	1.00	1.00

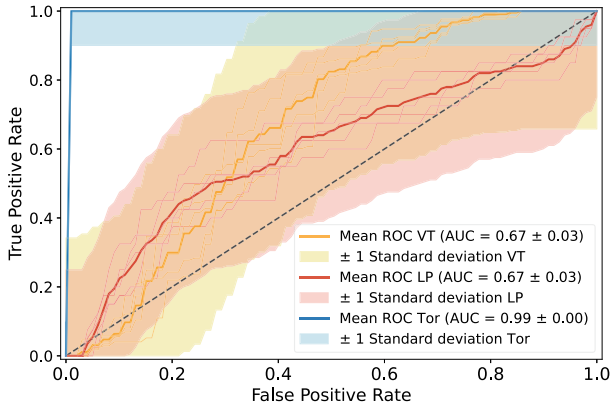

Fig. 10. ROC OvsR curve for classification using the LPC technique.

Finally, to evaluate the classifier and know its performance, it is necessary to treat the problem as a binary classification problem of type OvsR, and thus, implement the performance metrics mentioned in the previous section.

Table I shows the analysis report of the model performance evaluation metrics. It can be observed that when using the dataset resulting from the LPC technique, the TOR type earthquake class is correctly classified reaching 100% for all metrics. The opposite is the case for the LP class, which is confused with the VT type.

On the other hand, when analyzing the performance metrics of the classifier using the dataset resulting from applying the PCA technique, it can be observed that this technique allows an excellent definition of the VT and LP classes; however, the accuracy in the classification of the LP earthquakes only reaches 83.3%. The minority class TOR is totally detrimental by obtaining 0% as a result in all metrics for this dataset. The dataset resulting from applying the proposed DAF technique allowed a remarkable performance in the classification of all earthquakes, regardless of the unbalanced distribution of earthquakes by class that this dataset presents, obtaining 100% in all metrics.

Figs. 10–12 graphically summarize the performance of the MLP classifier for each dataset through the ROC curve analysis. Fig. 10 shows the performance of the model during the validation process when the LPC technique was applied to the dataset, indicating a rather low performance in the prediction of the VT and LP classes, with an area under the curve of 0.64 for both

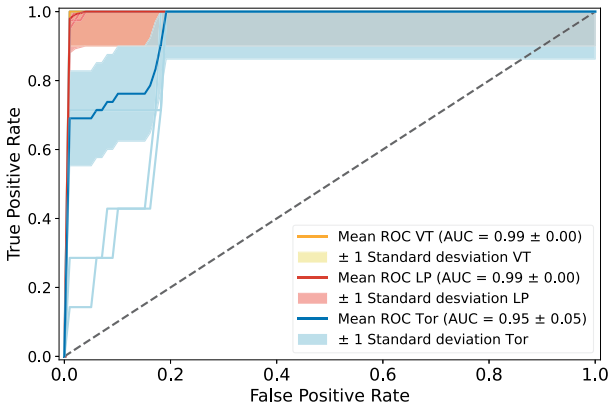

Fig. 11. ROC OvsR curve for classification using the PCA technique.

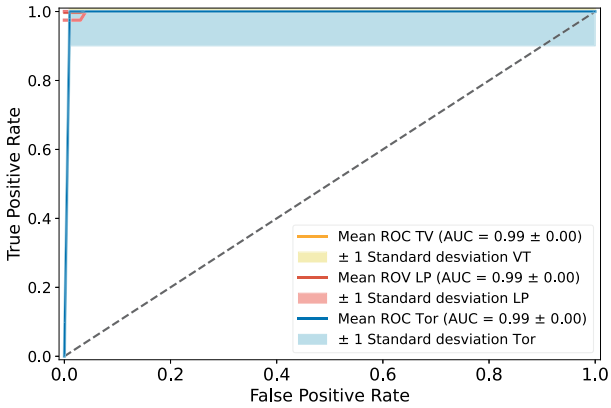

Fig. 12. ROC OvsR curve for classification using the DAF technique.

classes. Although the VT class has achieved good sensitivity results, the AUC obtained indicates that the model confuses the LP class as being VT, resulting in high false positives.

The ROC curve from the classification of the dataset when applying the PCA technique is shown in Fig. 11. The curve indicates a better performance of the model by obtaining an AUC of 0.99 for the classification of VT and LP type earthquakes, although the prediction of the TOR class presents an AUC of 0.95. This is due to the presence of false positives during the model prediction.

Finally, Fig. 12 allows identifying a high performance of the classification model when the dataset was applied using the DAF characterization technique, achieving an AUC of 0.99 in all classes.

VII. Discussion

The raw seismic signals correspond to the simplest representation that can be provided to the classifier model. However, this option requires having a sufficiently large and balanced dataset to train the model and obtain satisfactory results [18], [19], [20]. To overcome this problem, it is necessary to reduce the dimension in the data, to have a technique to discard the redundant information present in the seismic records, and also to reduce the dimensionality, thus favoring the learning and prediction during the classification process, mainly when dealing with datasets with a low number of data, and also with an unbalanced profile of classes. Preprocessing techniques for seismic log data are advantageous for obtaining higher classification performance results. As is known, there are large amounts of seismic log data acquired by in situ sensors; however, that requires labeling in order to fulfill their purpose of long-term monitoring and interpretation of internal volcanic activity. The results showed that conventional preprocessing techniques applied on volcanic earthquake signals could be improved (LPC and PCA) [2], [16], [24], [38], [39], [40], [41], [42], [43]. The representation of data by means of resource transformation using methods that are usually successfully applied to signals that are similar to seismic signals, such as speech signals, are not always compatible [7], [15], as is the case with widely used algorithms, such as MFCC, LFCC, LPC, and PCA, among others. Volcanic earthquake signals rarely exceed 30 Hz, thus not following the Mel scale, e.g., [7]; human speech signals are around 20 kHz. Although many works in literature adopt the LPC technique for the characterization of seismic signals, the results do not show success in all cases [2], [40], [67]. For this particular study, the LPC technique did not achieve separability between VT and LP classes, which was reflected in the classification of these signals, obtaining 0% accuracy, sensitivity and f1-score for LP earthquakes, and an AUC of only 0.64. On the other hand, the PCA technique is able to generate a new subset from the orthogonal transformation of the data. However, it is possible to identify that it is not always possible to obtain good results when applied to volcanic earthquake signals because this method has limitations when assuming linearity and Gaussian distribution in the variables [17], [26]. In this study, this method did not allow obtaining good results when the TOR class was characterized by principal components, affecting its classification, obtaining 0% in sensitivity, precision and f1-score, and an AUC of 0.95−0.07 for this class. Satisfactory results were achieved by applying the autoencoder architecture. The representation of the signal by learning the intrinsic patterns achieved in the nonlinear transformation of the signal resources allowed a good separation between the classes reaching 100% sensitivity, accuracy and f1-score for all, and an AUC of 0.99 in the same way. Regarding the MLP classifier model used, good results are obtained without using a complex architecture that implies a high computational cost and memory usage. For this case, the model required 26.9 s and 3734.25 MiB, using the Google Colab environment, with a Tesla T4 GPU unit with a capacity of 16 GB. Titos et al. [15] used more complex architectures (LSTM) with higher computational cost with a training time of more than 27 000 s, obtaining 94% accuracy using the LPC characterization technique. The Bayesian neural network [6], which is used in the PICOSS platform [36] achieves a performance of 92%, in which the LFCC technique is used to characterize the seismic signals. In the same way, this performance is achieved using the LPC, LFCC, PCA, and MFCC signal characterization techniques in [2], [4], [6], [10], [15], [16], and [38]. This work demonstrates that the use of a DAF is a good technique for characterizing volcanic earthquake signals, thus assisting in the classification of seismic records without the need to address complex architectures that require higher computational cost and achieving satisfactory results.

VIII. Conclusion

The dimension reduction obtained with the LPC, PCA, and DAF techniques allows an efficient compression of information in each of the data, avoiding significant loss of the general characteristics of the signals. However, the LPC technique does not achieve a good definition between VT and LP classes. On the other hand, the PCA technique achieves a good definition between the VT and LP classes, however, the TOR class cannot be well defined by this technique. Finally, the DAF technique obtains well-defined representations for all classes: VT, LP, and TOR.

The performance of the MLP classifier model improves when a dataset is presented by applying characterization techniques, even if an unbalanced distribution of earthquakes per class is present. The preprocessing of the data using the LPC technique allows the correct classification of the minority class, reaching an AUC of 0.99. However, the classifier does not recognize the other classes (VT and LP). The use of the PCA technique in the characterization of the dataset allows the classification between the VT and LP classes reaching an AUC of 0.99 for both, although the TOR class is not recognized. Finally, the DAF technique used in the dataset preprocessing allows the correct classification for all classes.

Future work is expected to apply the signal characterization technique to datasets of volcanic earthquakes belonging to new eruptive scenarios, as well as to datasets collected from different active volcanoes, in order to validate the performance of this technique by exporting it to different conditions.

Acknowledgment

The authors would like to thank the Colombian Geological Survey for providing the waveform data of the volcanic seismic events of the Galeras volcano.[1]

References

[1] O. E. C. Ibarra, "Detección y clasificación automática de registros sísmicos en el observatorio vulcanológico y simológico de Pasto, utilizando redes neuronales artificiales," M.S. thesis, Departamento de Geociencias, Universidad Nacional de Colombia, Bogota, Colombia, 2011.

[2] E. D. Pezzo, A. Esposito, F. Giudicepietro, M. Marinaro, M. Martini, and S. Scarpetta, "Discrimination of earthquakes and underwater explosions using neural networks," *Bull. Seismological Soc. Amer.*, vol. 93, no. 1, pp. 215–223, 2003.

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

[4] A. M. Esposito, L. D'Auria, F. Giudicepietro, R. Peluso, and M. Martini, "Automatic recognition of landslides based on neural network analysis of seismic signals: An application to the monitoring of Stromboli volcano (Southern Italy)," *Pure Appl. Geophys.*, vol. 170, pp. 1821–1832, 2013.

[5] Z. Geng and Y. Wang, "Automated design of a convolutional neural network with multi-scale filters for cost-efficient seismic data classification," *Nature Commun.*, vol. 11, no. 1, 2020, Art. no. 3311.

[6] A. Bueno, S. D. A. C. Benitez, A. D. Moreno, and J. M. Ibánez, "Volcano-seismic transfer learning and uncertainty quantification with Bayesian neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 892–902, Feb. 2020.

[7] 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.

[8] M. C. Benítez 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.

[9] J. M. Ibáñez et al., "The classification of seismo-volcanic signals using hidden Markov models as applied to the Stromboli and Etna volcanoes," *J. Volcanol. Geothermal Res.*, vol. 187, no. 3/4, pp. 218–226, 2009.

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

[11] M. Orozco-Alzate, C. Acosta-Muñoz, and J. M. Londoño-Bonilla, "The automated identification of volcanic earthquakes: Concepts, applications and challenges," *Earthq. Res. Anal.- Seismol., Seismotectonic Earthq. Geol.*, vol. 51, pp. 345–370, 2012.

[12] G. C. Moreno, "Reconocimiento de señales sismo-volcánicas mediante canales específicos basados en modelos ocultos de Markov," Ph.D. dissertation, Departamento de Teoría de la senal, telemática y comunicaciones, Univ. de Granada, Granada, Spain, 2015.

[13] J. Bi and C. Zhang, "An empirical comparison on state-of-the-art multi-class imbalance learning algorithms and a new diversified ensemble learning scheme," *Knowl.-Based Syst.*, vol. 158, pp. 81–93, 2018.

[14] M. Wasikowski and X. wen Chen, "Combating the small sample class imbalance problem using feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1388–1400, Oct. 2010.

[15] M. Titos, A. Bueno, L. García, M. C. Benítez, and J. Ibañez, "Detection and classification of continuous volcano-seismic signals with recurrent neural networks," *IEEE Trans. Geosci. Remote Sens.*, no. 4, pp. 1936–1948, Apr. 2019.

[16] C. Brusil, E. Grijalva, R. Lara-Cueva, M. Ruiz, and B. Acuña, "A semi-supervised approach for microseisms classification from Cotopaxi volcano," in *Proc. IEEE Latin Amer. Conf. Comput. Intell.*, 2019, pp. 1–6.

[17] G. Cortéss, M. C. Benítez, L. García, I. Álvarez, and J. M. Ibanez, "A comparative study of dimensionality reduction algorithms applied to volcano-seismic signals," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 1, pp. 253–263, Jan. 2016.

[18] H. Ali, M. N. M. Salleh, R. Saedudin, K. Hussain, and M. F. Mushtaq, "Imbalance class problems in data mining: A review," *Indonesian J. Elect. Eng. Comput. Sci.*, vol. 14, no. 3, pp. 1560–1571, 2019.

[19] V. Garcia, J. S. Sánchez, and R. A. Mollineda, "On the effectiveness of preprocessing methods when dealing with different levels of class imbalance," *Knowl.-Based Syst.*, vol. 25, no. 1, pp. 13–21, 2012.

[20] W. Ng, G. Zeng, J. Zhang, D. S. Yeung, and W. Pedrycz, "Dual autoencoders features for imbalance classification problem," *Pattern Recognit.*, vol. 60, pp. 875–889, 2016.

[21] G. Cortés, L. Gracia, I. Álvarez, C. Benítez, Ángel De la Torre, and J. Ibáñez, "Parallel system architecture (PSA): An efficient approach for automatic recognition of volcano-seismic events," *J. Volcanol. Geothermal Res.*, vol. 271, pp. 81–10, 2014.

[22] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.

[23] R. Corpotatio and R. Bellman, *Adoptive Control Processes: A Guided Tour*. Oxford, U.K.: Oxford Univ. Press, 1961.

[24] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[25] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933.

[26] B. Chouet, "Dynamics of a fluid-driven crack in three dimensions by the finite difference method," *J. Geophys. Res.: Solid Earth*, vol. 91, no. B14, pp. 413967–13992, 1986.

[27] J. M. Ibáñez et al., "The 1998–1999 seismic series at Deception Island volcano, Antarctica," *J. Volcanol. Geothermal Res.*, vol. 128, no. 1/3, pp. 65–88, 2003.

[28] C. R. A. Torres, G. M. D. M, and M. L. Narváez, "Unusual seismic signals associated with the activity at Galeras volcano, Colombia, from July 1992 to September 1994," *Ann. Geophys.*, vol. 39, no. 2, pp. 299–310, 1996. Online. Available: http://hdl.handle.net/2122/1661.

[29] L. Narváez, R. A. Torres, D. M. Gómez, G. P. Cortés, H. Cepeda, and J. Stix, "Tornillo-type seismic signals at Galeras volcano, Colombia, 1992–1993," *J. Volcanol. Geothermal Res.*, vol. 77, no. 1/4, pp. 159–171, 1997.

[30] F. Z. Chelali and A. Djeradi, "Face recognition system using neural network with Gabor and discrete wavelet transform parameterization," in *Proc. Int. Conf. Soft Comput. Pattern Recognit.*, 2014, pp. 17–24.

[31] M. Verleysen and D. François, "The curse of dimensionality in data mining and time series prediction," in *Proc. Int. Work- Conf. Artif. Neural Netw.*, 2005, pp. 758–770.

[1] https://www.sgc.gov.co/

[32] I. Álvarez et al., "Improving feature extraction in the automatic classification of seismic events. Application to Colima and Arenal volcanoes," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2009, pp. 526–529.
[33] I. Álvarez, L. Garcia, G. Cortes, C. Benitez, and Á. 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.
[34] G. Curilem, J. Vergara, G. Fuentealba, G. Acuña, and M. Chacón, "Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms," *J. Volcanol. Geothermal Res.*, vol. 180, no. 1, pp. 1–8, 2009.
[35] G. Cortés et al., "Analysis of Colima, Popocatépetl and Arenal volcanic seismicity using an automatic Continuous Hidden Markov Models-based recognition system," in *The VOLUME Project: VOLcanoes: Understanding Subsurface Mass moveMEnt*. Brussels, Belgium: CORDIS EU, 2009, pp. 150–160.
[36] A. Bueno et al., "Picoss: Python interface for the classification of seismic signals," *Comput. Geosci.*, vol. 142, 2020, Art. no. 104531.
[37] A. Bueno, M. Titos, M. C. Benítez, and J. Ibáñez, "Continuous active learning for seismo-volcanic monitoring," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 7505405.
[38] P. E. E. Lara et al., "Automatic multichannel volcano-seismic classification using machine learning and EMD," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1322–1331, 2020.
[39] M. Titos, A. Bueno, L. García, and M. C. Benítez, "A deep neural networks approach to automatic recognition systems for volcano-seismic events," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1533–1544, May 2018.
[40] F. Giacco, A. M. Esposito, S. Scarpetta, F. Giudicepietro, and M. Marinaro, "Support Vector Machines and MLP for Automatic Classification of Seismic Signals at Stromboli Volcano," in *Proc. Conf. Neural Nets: Proc. 19th Italian Workshop Neural Nets*, 2009, pp. 116–123.
[41] S. Dineva, D. Eaton, and R. Mereu, "Seismicity of the southern Great Lakes: Revised earthquake hypocenters and possible tectonic controls," *Bull. Seismological Soc. Amer.*, vol. 94, no. 5, pp. 1902–1918, 2004.
[42] K. Unglert, V. Radic, and A. M. Jellinek, "Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra," *J. Volcanol. Geothermal Res.*, vol. 320, no. 5, pp. 58–74, 2016.
[43] K. Unglert and A. M. Jellinek, "Principal component analysis for pattern recognition in volcano seismic spectra," in *Proc. EGU Gen. Assem. Conf. Abstr.*, 2016, pp. 526–529.
[44] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
[45] K. Unglert and A. M. Jellinek, "Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution," in *Proc. Comput. Vis. 16th Eur. Conf.*, 2020, pp. 526–529.
[46] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.
[47] S. Mandelli, F. Borra, V. Lipari, P. Bestaginim, A. Sarti, and S. Tubaro, "Seismic data interpolation through convolutional autoencoder," in *Proc. SEG Int. Expo. Annu. Meeting*, 2018, pp. 4101–4105.
[48] J. Jiang, H. Ren, and M. Zhang, "A convolutional autoencoder method for simultaneous seismic data reconstruction and denoising," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, no. 5, 2021, Art. no. 7503405.
[49] O. M. Saad and Y. Chen, "Deep denoising autoencoder for seismic random noise attenuation," *Geophysics*, vol. 85, no. 4, pp. V367–V376, 2020.
[50] Z. Gao, C. Li, N. Liu, Z. Pan, J. Gao, and Z. Xu, "Large-dimensional seismic inversion using global optimization with autoencoder-based model dimensionality reduction," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1718–1732, Feb. 2021.
[51] M. Liu and D. Grana, "Ensemble-based seismic history matching with data reparameterization using convolutional autoencoder," in *SEG Technical Program Expanded Abstracts*, Houston, TX, USA: Soc. Exploration Geophysicists, 2018, pp. 3156–3160.
[52] A. Krogh and J. Hertz, "A simple weight decay can improve generalization," *Adv. Neural Inf. Process. Syst.*, vol. 4, no. 2, 1991, pp. 950–957.
[53] D. Bank and N. Koenigstein, "Autoencoders," in *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, Berlin, Germany: Springer, 2023, pp. 353–374.
[54] J. Bengio, "Learning deep architectures for AI," *Found. trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
[55] J. Bengio, A. Couville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[56] P. Vincent, H. Larochelle, I. Lajoie, J. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, 2010.
[57] S. Bengio and O. Delalleau, "On the expressive power of deep architectures," in *Proc. Algorithmic Learn. Theory: 22nd Int. Conf.*, 2011, pp. 3156–3160.
[58] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proc. Workshop Unsupervised Transfer Learn.*, 2012, pp. 3156–3160.
[59] D. Yeung, J.-C. Li, W. W. Y. Ng, and P. P. K. Chan, "MLPNN training via a multiobjective optimization of training error and stochastic sensitivity," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 5, pp. 978–992, May 2016.
[60] W. S. Mcculloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, pp. 115–133, 1943.
[61] P. J. Werbos, "Generalization of backpropagation with application to a recurrent gas market model," *Neural Netw.*, vol. 1, no. 4, pp. 339–356, 1988.
[62] R. Parikh, A. Mathai, S. Parikh, G. C. Sekhar, and R. Thomas, "Understanding and using sensitivity, specificity and predictive values," *Indian J. Ophthalmol.*, vol. 56, no. 1, pp. 45–50, 2008.
[63] R. Yacouby and D. Axman, "Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models," in *Proc. 1st Workshop Eval. Comparison NLP Syst.*, 2020, pp. 79–91.
[64] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
[65] J. N. Mandrekar, "Receiver operating characteristic curve in diagnostic test assessment," *J. Thoracic Oncol.*, vol. 5, no. 9, pp. 1315–1316, 2010.
[66] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 299–310, Mar. 2005.
[67] B. Rouet-Leduc, C. Hulbert, I. W. McBrearty, and P. A. Johnson, "Probing slow earthquakes with deep learning," *Geophys. Res. Lett.*, vol. 47, no. 4, 2020, Art. no. e2019GL085870.

Paula Andrea Montenegro C. received the graduate degree in environmental engineering from Universidad Mariana, Pasto, Colombia, in 2007, and the M.Sc. degree in civil engineering (water resources and environmental technologies), in 2017, from UNESP Campus of Ilha Solteira, Ilha Solteira, SP, Brazil, where she is currently working toward the doctoral degree in electrical engineering.

She is currently with the Intelligent Systems Research Laboratory - SINTEL at Electrical Engineering Department from UNESP, Campus of Ilha Solteira. Her research interests include artificial intelligence, fuzzy logic, and artificial neural networks.

Oscar Ernesto Cadena I. received the graduate degree in physics from the Universidad de Nariño, Pasto, Colombia, in 2005, and the M.Sc. degree in geophysics and the Ph.D. degree in geosciences from the Universidad Nacional de Colombia, Bogotá, Colombia, in 2011 and 2020, respectively.

Since 2008, he has been with the Volcanological and Seismological Observatory of Pasto, Colombian Geological Survey, Bogota, Colombia. His research interests include the modeling of LP earthquake sources, and the application of machine learning to the detection and classification of volcanic earthquakes.

Anna Diva P. Lotufo (Senior Member, IEEE) received the graduate degree in mathematics from Faculdade de Filosofia Ciências e Letras Imaculada Conceição, Santa Maria, RS, Brazil, in 1977, the graduate degree in electrical engineering from Universidade Federal de Santa Maria, Santa Maria, RS, Brazil, in 1978, the M.Sc. degree in electrical engineering from Universidade Federal de Santa Catarina, Florianopolis, SC, Brazil, in 1982, and the Ph.D. in electrical engineering from UNESP, Campus of Ilha Solteira, Ilha Solteira, SP, Brazil, in 2004.

Since 1984, she has been with the Electrical Engineering Department, UNESP, Campus of Ilha Solteira- SP, where she is currently an Associate Professor. She has experience with transient stability analysis, load forecasting, fuzzy logic, and artificial neural networks.