Using CNN to classify spectrograms of seismic events from Llaima volcano (Chile)

Millaray Curilem
Department of Electrical Engineering
Universidad de La Frontera
Francisco Salazar 01145.
Temuco – Chile.
Email: millaray.curilem@ufrontera.cl

João Paulo Canário
Department of Computer Science
Federal University of Bahia
Adhemar de Barros. s/n. Ondina.
Salvador – Brasil.
Email: jopacanario@gmail.com

Luis Franco
Observatorio Vulcanológico de los Andes Sur
Rudecindo Ortega 03850.
Temuco – Chile.
Email: luis.franco@sernageomin.cl

Ricardo A. Rios
Department of Computer science
Federal University of Bahia
Adhemar de Barros. s/n. Ondina.
Salvador – Brasil.
Email: ricardoar@ufba.br

***Abstract*—The monitoring of volcanoes is mainly performed by sensors installed on their structures, aiming at recording seismic activities and reporting them to observatories to be later analyzed by specialists. However, due to the high volume of data continuously collected, the use of automatic techniques is an important requirement to support real time analyses. In this sense, a basic but challenging task is the classification of seismic activities to identify signals yielded by different sources as, for instance, the movement of magmatic fluids. Although there exists several approaches proposed to perform such task, they were mainly designed to deal with raw signals. In this paper, we present a 2D approach developed considering two main steps. Firstly, spectrograms for every collected signal are calculated by using Fourier Transform. Secondly, we set a deep neural network to discriminate seismic activities by analyzing the spectrogram shapes. As a consequence, our classifier provided outstanding results with accuracy rates greater than 95%.**

***Keywords*—volcano monitoring, spectrogram analysis, supervised learning, deep neural network**

I. Introduction

Volcanic activity involves physical processes related to fluid and energy transportation. A fraction of this energy is released as discrete seismic oscillations, also referred to as volcano seismic events, which have specific patterns [1]. The recognition of these patterns provides useful information to remotely monitor volcanic activities. However, such task is not trivial once general factors as, for instance, source geometry, soil characteristics, and station site effects may affect the seismic events, modifying their phase, introducing delays, amplifying or attenuating their amplitude in particular frequencies, and so on.

Aiming at addressing this pattern recognition problem, several approaches have been proposed, basically, considering two steps: i) extracting features from collected signals; and ii) adjusting classification models on features. The feature extraction in volcanic seismicity considers amplitude, spectral and phase parameters. Amplitude statistics (mean, standard deviation, skewness and kurtosis) [2], LTA/STA, autocorrelation functions, obtained using Fast Fourier Transform [3], wavelet transform [4], cepstral coefficients [5], Linear predictor coding [6] are some of the features extracted from seismic signals. Afterwards, these features are used to train classification models using, for instance, Support Vector Machines [7], Self Organizing Maps [8], Hidden Markov Models [9], Multi-Layer Perceptron [10] and Gaussian Mixture Models [11].

Besides these aforementioned features directly extracted from signals, spectrograms have also been widely applied in seismology [8]. However, they are commonly used by specialists to perform manual analyses on seismic events. In general, classifiers are not modeled by using the spectrograms indeed, but using features extracted from them as well [12][13][14].

In this work, we evaluated the advantage of using spectrograms to classify seismic events by using techniques that take into account their 2D representation as images. Our hypothesis relies on the idea that the 2D approach retrieves important information to learn and recognize the behavior of frequency components along time. We state that the 2D local structures (bandwidth and amplitude) formed by the time-frequency dependencies can be used to discriminate volcano events.

Our experiments were performed using Convolutional Neural Network (CNN) [15], due to its ability to identify local structures in the first layers of the network. Moreover, it is well-known that when receptive fields of CNN filters are properly adjusted, they can learn spatial and temporal features extracted from images or time-domain waveforms. This is why CNNs have been successfully applied in object recognition tasks [16], document classification [17], and face recognition [18]. In summary, the CNN performance is based on the ability of automatically learning basic features in the first layers and more complex ones in deeper hierarchies [19], [20]. This is the reason why the CNN models are usually very deep, ranging the number of layers from tens to hundreds [21]. As

978-1-5090-6014-6/18/$31.00 ©2018 IEEE

a consequence, such deep structure requires a high number of examples to train and validate the learning process, making the use of CNN a challenge, specially for new application domains.

In such situation, the lack of enough labeled data to support training leads to the need for data augmentation techniques, which can increase the amount of data keeping their main characteristics. In the context of volcano monitoring, even if the volume of data is huge, the labels are not always available and some events are very sparse in time. Thus, we also performed a scrutiny to assess whether our results were influenced by number of signals in our dataset. Results emphasize our approach was able to found differences among the 2D local structures (bandwidth and amplitude) formed by the time-frequency dependencies in every event type. Such differences allowed us to obtain outstanding classification results, even when the signal was influenced by noise.

This paper is organized as follows: Section II provides details about the Llaima Volcano in Chile; Section III provides the necessary background on CNN; The experimental setup and results are presented and discussed in Sections IV and V, respectively; Finally, section VI draws the main conclusions of the study.

II. Seismic Events of the Llaima Volcano, Chile.

The Llaima volcano is considered one of the most dangerous volcanoes in South America. It is located at the Araucania region of Chile (S $38°41'$ - W $71°44'$), on the western edge of Andes. Several researchers have studied its seismicity to design automatic classifiers for its events [22], [23], [24]. The Observatorio Vulcanológico de los Andes Sur (OVDAS) is the state agency responsible for its monitoring. OVDAS uses the criteria defined in [25] to identify and label the different kind of volcanic events, as discussed in the following section.

A. Database of seismic signals

Llaima is monitored by seven seismic stations as shown in Figure 1. As recommended by OVDAS experts, the events selected for this study were recorded by the Z-vertical component from the LAV station (highlighted in red in the figure), during the period between 2010 and 2016. Then, the signals were sampled at 100 Hz and filtered with a 10th order Butterworth bandpass filter between 1 and 10 Hz, preserving the bandwidth that contains the main interest frequency range for Llaima events. The signals were also normalized by their maximum value and organized in 4 classes as shown in Table I. It is important to emphasize that the events were individually reviewed, segmented and labelled by experts from OVDAS.

TABLE I. Number of events per class

LP	TR	VT	TC	Total
1310	490	304	1488	**3592**

Volcano-tectonic (VT) is a class of events associated with rock fracturing in the conduits located inside the volcanic structures. The VT events present a frequency pattern with a broadband spectrum (frequency range that concentrate most of energy), that may reach the 10 Hz. Their spectrograms are very characteristic, presenting an impulsive beginning and an exponential decay.

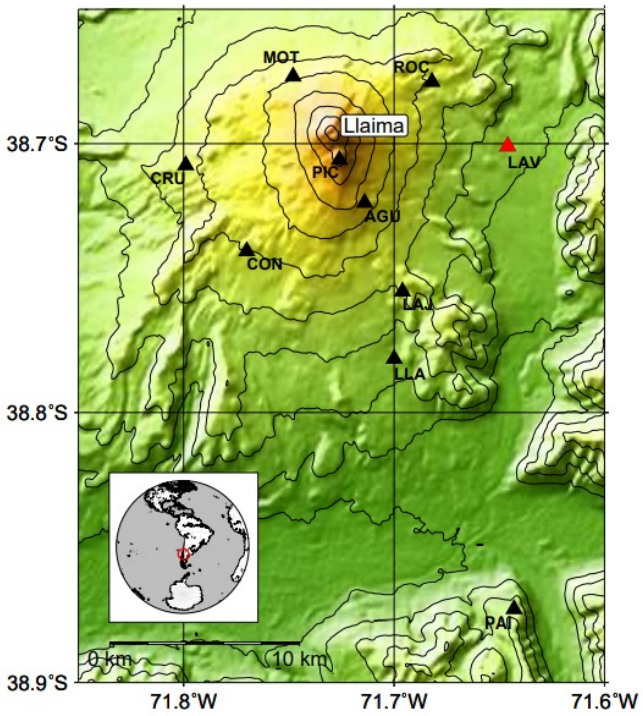

Fig. 1. Seismic Stations of the Llaima Volcano (△)

The long period (LP) events are related to the transit of magmatic and hydrothermal fluids inside the volcanic conduits [26]. The spectral pattern of LP is narrower than VT, bounded mainly in a range between 0.5 and 5Hz. Their spectrograms have often a non impulsive beginning and a slow decay.

The tremor (TR) events generally present a duration longer than LP. The source of the tremor is understood as a sustained pressure disturbance over the magmatic and hydrothermal fluid, which may be continuous or a sequence of transient signals similar to those ones generated by LP. As a consequence, their spectrograms are characterized by a very narrow, regular, and long shape. Their broadband spectrum is typically between 0.5 and 3.0 Hz, being slowly attenuated at the end of the event.

In addition to these events, the stations record other events, not related to volcanic activity. One of the most common event is called tectonic (TC), which is produced by the dynamic of the geological faults [26]. These events can be local, regional or distant according to the location of their epicentre. The further the event is, the lower the frequency pattern is. Moreover, depending on the proximity between the epicentre and the station, TC can be confused with the LP or VT events. Similarly to VT, the spectral content of TC is characterized by an impulsive beginning and an exponential decay. Meanwhile, TC signals carry more energy.

B. Data Augmentation technique

The aim of this technique is to increase the number of examples in the training and test datasets. The controlled

manipulation of the input data may contribute to create more robust classifiers, once artificial (synthetic) examples may cover regions on the input space not considered by the lack of original data.

The main challenge of data augmentation techniques is to transform inputs without changing their labels [21]. In image recognition, the augmentation is performed by creating artificial images from the original ones, using techniques such as translation, mirroring, rotation, scaling and elastic deformations, among others [27]. Although spectrograms can be transformed by these techniques [28], in this work, the transformation was applied on signals to get closer to real-world conditions. In summary, the most common techniques used to transform signals are pitch shift, time-stretch [29], mixing signals, additive background noise [30], and rotational distortions [21].

In this study, the seismic signal transformations were carried out following specialists' recommendation, who suggested us to keep the signal to noise ratios and the frequency ranges for each event class. The augmentation process was as random as possible, but within bounded ranges of variation to preserve the nature of the events. The following modifications were performed:

1) Amplitude deformation: a random number of maxima of the signal were randomly amplified or attenuated. The number of affected maxima is random, but less than 60% of the total. The amount of amplification or attenuation is also random, but inferior to 60% of the absolute amplitude of the maxima.

2) Frequency deformation: the signal is randomly re-sampled in a variable proportion whose maximum is 20% of variation from the original frequency.

3) Noise addition: white noise with a 0 mean and random standard deviation, inferior to 40% of the maximum amplitude, was added to the signals.

4) Horizontal shift: the signal is right-shifted with a random percentage of displacement inferior to 15%.

To limit the amount of data created, the artificial data did not exceed 5 times the amount of the real data per class. All the new artificial signals were normalized and filtered with the same filter used for the original signals. An example of the transformation for an LP event is presented in Figure 2.

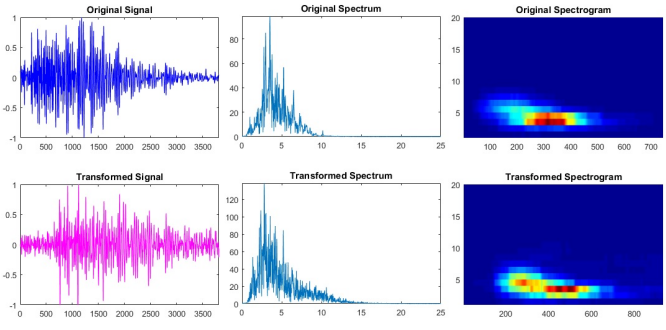

Fig. 2. Creation of a new signal from an original one.

The artificial signals were classified by a classifier proposed in a previous work [24] and only the events correctly classified were considered. This step discarded approximately 30% of the augmented events. The new database formed by the artificial signals is presented in Table II.

TABLE II. NUMBER OF EVENTS PER CLASS FOR THE ARTIFICIAL DATABASE

LP	TR	VT	TC	Total
6835	4726	2471	1863	**15895**

III. CONVOLUTIONAL NEURAL NETWORKS

A. Overview

Convolutional neural networks (CNNs) are a multi-layer neural networks variant capable of recognizing patterns with extreme variability, distortions, and geometric transformations [15]. As other neural networks, CNNs are made up of neurons with learnable weights and biases. In summary, each neuron receives several inputs, takes a weighted sum over them, applies an activation function, and produces an output. It is also important to highlight, such neurons are spread over multiple layers, such as: an input layer, a set of hidden layers, and an output layer [15]. Also, the hidden layers are usually composed by four different types of layers [16]:

1) Convolutional layers: Compute the output of neurons that are connected to local regions in the input. Each layer calculates a dot product between the weights and a small region where they are connected to the input volume.

2) Activation layers: Apply an activation function on outputs to avoid learning problems as, for instance, getting stuck near zero or indefinitely growing up.

3) Pooling layers: Perform a downsampling operation along with the spatial dimensions.

4) Fully connected layers: Compute the class scores. Simillary to ordinary neural networks, each neuron in a layer is connected to all the neurons in the previous one.

In the following section, we present the experimental setup designed to classify volcano events from Llaima, including details on the CNN architecture.

IV. EXPERIMENTAL SETUP

A. Spectrograms

The signals were recorded with a sampling frequency of 100Hz. The spectrograms were obtained by applying the Short-Time Fourier Transform (STFT), described in

$$SP[x(t)](n,k) = |\sum_{m=0}^{N-1} x(m) \cdot w(m-n) \cdot e^{i2\pi mk}|^2, \quad (1)$$

where $x(t)$, and $w(t)$ represent the seismic signal and the STFT sliding window, respectively. The proposed work uses expert's specifications that recommend to define the sliding window size to 1 second with 95% of overlapping. The signals were all set to 6000 samples (60 seconds) filling with zeros when necessary. Although the signals were filtered from 1 to 10Hz, the frequency interval considered to calculate the spectrograms

was from 1 to 20Hz, to show a better visualization of their shapes. Other intervals can be considered in future work. With these specifications, all the spectrograms had the same dimensions, in time and frequency. The frequency components of the spectrograms were smoothed by using a moving average filter to obtain their envelopes. The moving average was calculated by

$$SmoothedSP[x(t)](n,k) = \frac{1}{Sf}\sum_{j=0}^{Sf-1} SP(n,k+j), \quad (2)$$

where Sf is the number of the averaged samples (Smooth factor). Different values for Sf were tested and the best results were obtained with 150 samples. Figure 3 shows one example for each class of the event and their normal and smoothed spectrograms.

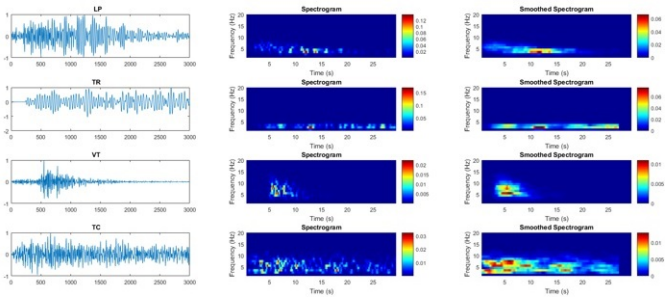

Fig. 3. Seismic events and their spectrograms.

All the spectrograms were transformed to RGB images of 20 x 20 pixels. Figure 4 shows the shape of the spectrograms for the different classes.

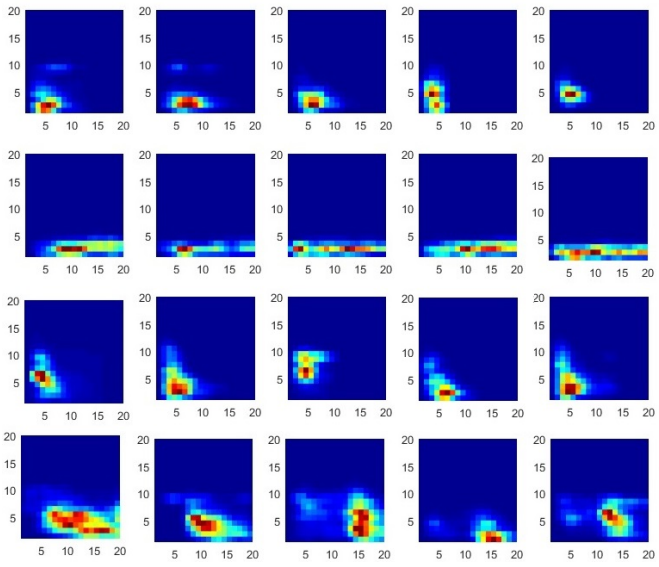

Fig. 4. Smoothed spectrograms of the events: each line presents five examples of the smoothed spectrograms for the LP, TR, VT and TC classes respectively.

B. CNN design

By considering CNNs are not commonly used on spectrograms of seismic activity, we assessed them using standard architectures. In this sense, we used the same architecture proposed by LeCun *et al*. (1998) [15], empirically, analyzing the following characteristics:

1) Changing the number of fully-connected layers or convolutional layers by adding or removing them;
2) Adding a dropout layer following the convolutional layers;
3) Using dropout layers following the fully-connected layers to avoid overfitting;
4) Changing the optimization algorithm from stochastic gradient descent to Adam [31].
5) Changing the hyperparameters of some layers, such as the number of filters on convolutional layers, dropout rate, number of neurons on fully-connected layers;

After evaluating all these CNN characteristics, the final architecture used in our experiments presents: i) one input layer that represents the seismic event and is denoted by an input image $I(h,w,c)$, where h is the height, w is the width and c is the color channels; ii) two convolutional layers designed by $C(K,F,S,P)$, where K is the number of filters, F is their spatial extent, S is the stride and P is the ammount of zero padding; iii) two max pooling layers denoted as $P(F,S)$, where F is the size of the square pooling regions and S is the stride; iv) two fully connected layers expressed as $F(u)$, where u is the number of units or neurons in layer; v) in the output of each convolutional layer and the first fully-connected layer, the ReLU activation function was applied considering $f(x) = max(0,x)$, where x is the output of each neuron from previous layer; vi) two dropout layers defined by $D(dr)$, where dr is the dropout rate; and, finally, vii) the softmax activation function was applied on the output of proposed architecture. The CNN architecture and its adjusted hyperparameters are represented by Figure 5.

C. Classification Models

In our first experiment, four CNN models were adjusted, using a 10-folds cross validation strategy, to assess the proposed methodology. Firstly, we only used real-world data. Secondly, artificial data was used to train the classification model, which was latter validated by using real and artificial data, separately. Next, a combined model was designed using the database formed by real and artificial data. Finally, to evaluate whether the learning was influenced by the artificial database, a last model was implemented using the combined database in the training step and the real one to test the generalization performance.

In a second experiment, a fifth class was included. This class is a contrast group, referred to as "other" (OT), which is formed by events that do not belong to any previous class. This class groups random events recorded from the volcano as, for instance, ice breaking, weather events (wind, rain, etc.), and high-energy noise. Thus, a total of 171 real and 552 artificial OT data were added to the real and artificial databases. The shape of the spectrograms for this class is more variable than the other ones, as shown in Figure 6, imposing a new challenge for the classification models. The same tests of the first experiment were performed to evaluate the classification performance for this class.

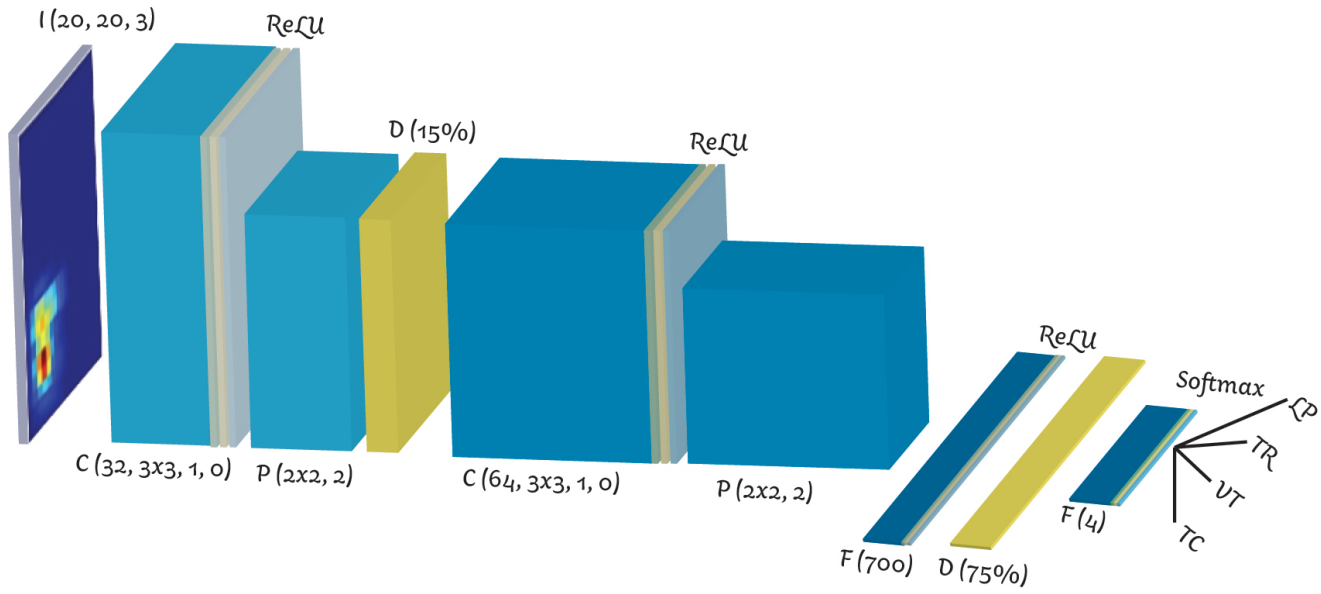

Fig. 5. CNN summary: conv. 2D, ReLU, max pool., dropout, conv. 2D, ReLU , max pool., fully-connected, ReLU, dropout, fully-connected and output layer with softmax activation function.

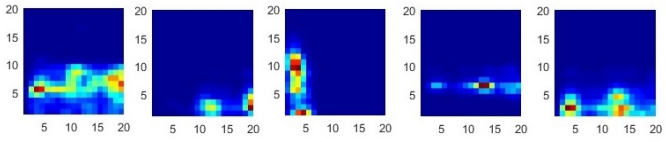

Fig. 6. Smoothed spectrograms: five examples for the OT class.

D. Performance measures

To evaluate the performance of different classifying models, we considered the following measures [32]: i) accuracy (Equation 3); ii) error (Equation 4); iii) specificity (Equation 5); and F1-score (Equation 6).

$$Accuracy = \frac{TP + TN}{n}, \tag{3}$$

$$Error = \frac{FP + FN}{n}, \tag{4}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{5}$$

$$F1\text{-}score = 2 \times \frac{Recall \times Precision}{Recall + Precision}. \tag{6}$$

In such measures, n is the total number of events, TP is the number of true positive classifications, TN is the number of true negative classifications, FP is the number of false positive classifications, and FN is the number false negative classifications. The $Recall$ and $Precision$ are obtained by

$$Recall = \frac{TP}{TP + FN}, \quad Precision = \frac{TP}{TP + FP}. \tag{7}$$

The Accuracy (ACC) and the F1-score (FS) indices were calculated for each class and each fold, and their mean is used to evaluate the overall modeling performance. Accuracy, Error, Recall, and Specificity were also used to calculate the performance per class.

In addition to the previous indices, the Kappa coefficient [32] was used to measure the general agreement between the classification provided by CNN and the experts. Kappa values near to one show excellent levels of agreement. This coefficient is based on the difference between the "observed" agreement (P_o) and the agreement that would be expected to occur by chance, a so-called "expected" agreement (P_e), according to

$$P_o = \sum_{i=1}^{C} p_{ii}, \quad P_e = \sum_{i=1}^{C} p_{i.} p_{.i}, \tag{8}$$

where p_{ii} is the joint proportion of the agreement (diagonal), $p_{.i}$ and $p_{i.}$ are the sum of the joint proportions of the classifier (column) and the expert (rows), for each class and C is the number of classes. Finally, the Kappa coefficient (K) is obtained by

$$K = \frac{P_o - P_e}{1 - P_e}. \tag{9}$$

V. Results

The first set of experiment was performed on all real and artificial seismic signals, but the ones from OT. As a consequence, Table III shows the obtained accuracy and F1-score, whereas Table IV presents the corresponding confusion matrices and the Kappa values.The tables show that the mean accuracy is high (superior to 97%) and presents a slightly increase for the combined model. The F1-score, that measures the ability of recognizing the true positive events of each

TABLE III. ACCURACIES (ACC) AND F1-SCORES (FS) OF 10-FOLD CROSS VALIDATION WITHOUT OT CLASS

	Real		Artificial		Combined (real + artificial)		Combined (train) + Real (test)	
Fold	**ACC**	**FS**	**ACC**	**FS**	**ACC**	**FS**	**ACC**	**FS**
1:	96.39%	94.52%	99.75%	99.69%	99.08%	98.94%	96.39%	94.81%
2:	97.22%	94.56%	99.75%	99.71%	99.13%	98.92%	98.06%	96.72%
3:	98.33%	97.30%	99.69%	99.60%	98.97%	98.80%	98.33%	97.94%
4:	96.66%	96.06%	99.25%	99.14%	99.08%	98.90%	96.66%	96.10%
5:	96.94%	96.26%	99.62%	99.50%	99.18%	98.97%	97.21%	96.85%
6:	96.94%	95.71%	99.50%	99.38%	99.08%	98.94%	96.10%	94.11%
7:	98.05%	97.32%	99.87%	99.81%	99.08%	98.99%	96.94%	95.27%
8:	97.77%	97.67%	99.87%	99.89%	99.08%	98.86%	97.77%	97.11%
9:	96.94%	95.29%	99.87%	99.83%	99.64%	99.57%	96.66%	95.22%
10:	95.54%	93.69%	99.43%	99.31%	99.08%	98.87%	96.66%	95.85%
Mean:	**97.08%**	**95.84%**	**99.66%**	**99.59%**	**99.14%**	**98.98%**	**97.08%**	**96.00%**
Std:	**0.78%**	**1.27%**	**0.20%**	**0.23%**	**0.17%**	**0.20%**	**0.71%**	**1.11%**

TABLE IV. CONFUSION MATRIX OBTAINED ON EVALUATION OF CNN WITHOUT OT CLASS

		Real				Artificial				Combined				Combined (train) + Real (test)			
		Predicted				**Predicted**				**Predicted**				**Predicted**			
		LP	**TR**	**VT**	**TC**	**LP**	**TR**	**VT**	**TC**	**LP**	**TR**	**VT**	**TC**	**LP**	**TR**	**VT**	**TC**
True	**LP**	**1282**	3	11	14	**6815**	3	15	2	**8101**	7	27	10	**1287**	5	13	5
	TR	4	**473**	0	13	**TR**	1	**4718**	0	7	**TR**	9	**5187**	0	20	**TR**	4	**483**	0	3
	VT	13	0	**279**	12	**VT**	13	0	**2457**	1	**VT**	23	0	**2746**	6	**VT**	9	0	**293**	2
	TC	9	12	14	**1453**	**TC**	3	9	0	**1851**	**TC**	16	30	20	**3285**	**TC**	13	28	23	**1424**
		Kappa: 0.956				**Kappa: 0.995**				**Kappa: 0.988**				**Kappa: 0.957**			

class, was superior to 95% for all experiments. The standard deviation of the performance for the 10-fold cross-validation is low (inferior to 0.78%). The Kappa values for all experiments are superior to 0.95, presenting a very good performance.

The performance obtained on the real data shows the network learned to discriminate the classes from the spectrograms. It is interesting to note that the models estimated on the artificial and combined databases obtained very high performances as well. This is mainly due to the fact that the artificial events were pre-classified, as described in Section II-B. Another observation is that the standard deviations decreased among the cross-validation folds, following the trend that states higher number of examples decreases the variability of the results. However, by comparing the performance for the real dataset with the combined (train) + Real (test) model, we noticed the accuracy, F1-score and standard deviations remain almost the same, showing that the augmented data did not have a significant impact on learning. The results presented in Table V show the performance indices obtained per class by the real and the Combined (train) + Real (test) models. According to the table, although the accuracy remains the same, there is a slight improvement on the ability of the last model to recognize the true positives (Recall).

Finally, Table VI shows the accuracy and F1-score measures obtained after including the OT signals. Table VII presents the contingency matrices and the Kappa values. As expected, the OT class was responsible for decreasing the overall modeling performance. In particular, F1-score was the most affected index. In general, the models adjusted on artificial and combined data presented better results by comparing the F1-score measures. The Kappa values for all models are around 0.90, which is a good performance when compared with literature, specially by considering a previous work on Llaima volcano [22] where a structure using station-dependant classifiers reached a kappa of 0.94, but without considering the TC class.

TABLE V. PERFORMANCE INDICES FOR THE REAL AND THE COMBINED (TRAIN) + REAL (TEST) DATABASES, WITHOUT OT CLASS (%)

Real					
	LP	**TR**	**VT**	**TC**	**Mean**
Accuracy	98.5	99.11	98.61	97.94	98.54
Error	1.5	0.89	1.39	2.06	1.46
Recall	97.86	96.53	91.78	97.65	95.95
Specificity	98.86	99.52	99.24	98.15	98.94
Combined (train) + Real (test)					
	LP	**TR**	**VT**	**TC**	**Mean**
Accuracy	98.64	98.89	98.69	97.94	98.54
Error	1.36	1.11	1.31	2.06	1.46
Recall	98.24	98.57	96.38	95.7	97.22
Specificity	98.86	98.94	98.91	99.52	99.06

Similarly to the first experiment without the OT class, a new model was also adjusted using the Combined (train) + Real (test) strategy. When comparing this model with the model trained only with the real data, we noticed that the use of the artificial database improved the performance of the classifiers. However, in this case, the standard deviation between the cross-validation folds has increased.

Table VIII show the performance indices for the models obtained on real and Combined (train) + Real (test) data. The table show improvements on the recall and specificity mean. However, the recall for the OT class is very low (inferior to 38%), which demonstrates that the network could not discriminate the OT events. This result was expected once the difficulty of recognizing OT events from their spectrograms is a well-known problem.

VI. FINAL REMARKS

The present work applied CNN to recognize seismic events produced by different sources and registered in a seismic station of the Llaima volcano in Chile. The major contribution of the work is the use of the spectrograms to classify different

TABLE VI. ACCURACIES (ACC) AND F1-SCORES (FS) OF 10-FOLD CROSS VALIDATION WITH OT CLASS

	Real		Artificial		Combined		Combined (train) + Real (test)	
Fold	**ACC**	**FS**	**ACC**	**FS**	**ACC**	**FS**	**ACC**	**FS**
1:	93.63%	75.35%	98.97%	96.78%	97.87%	94.09%	93.37%	82.23%
2:	93.37%	83.28%	98.48%	95.19%	97.23%	92.74%	93.63%	82.79%
3:	92.84%	81.74%	98.72%	95.50%	97.43%	93.42%	95.76%	91.51%
4:	93.35%	79.58%	98.42%	95.91%	97.97%	94.43%	92.82%	81.92%
5:	92.29%	82.18%	98.78%	95.75%	97.48%	93.79%	93.35%	83.96%
6:	93.88%	85.16%	98.12%	94.67%	97.53%	93.81%	93.88%	88.33%
7:	92.82%	78.58%	98.36%	94.61%	97.58%	94.40%	92.29%	79.13%
8:	90.96%	81.40%	98.78%	96.63%	97.58%	91.92%	91.76%	84.55%
9:	94.41%	80.64%	98.91%	96.06%	97.62%	93.92%	95.21%	85.71%
10:	92.29%	79.24%	98.60%	96.12%	97.43%	93.70%	91.49%	80.61%
Mean:	**92.98%**	**80.71%**	**98.61%**	**95.72%**	**97.57%**	**93.62%**	**93.36%**	**84.07%**
Std:	**0.93%**	**2.58%**	**0.25%**	**0.70%**	**0.21%**	**0.73%**	**1.31%**	**3.50%**

TABLE VII. CONFUSION MATRIX OBTAINED ON EVALUATION OF CNN USING REAL AND ARTIFICIAL DATA WITH OT CLASS

		Real					Artificial						Combined						Combined (train) + Real (test)					
		Predicted					Predicted						Predicted						Predicted					
		LP	**TR**	**VT**	**OT**	**TC**		**LP**	**TR**	**VT**	**OT**	**TC**		**LP**	**TR**	**VT**	**OT**	**TC**		**LP**	**TR**	**VT**	**OT**	**TC**
	LP	**1271**	2	12	14	11	**LP**	**6789**	2	23	18	3	**LP**	**8065**	5	28	37	10	**LP**	**1280**	4	12	11	3
	TR	5	**473**	0	2	10	**TR**	2	**4713**	0	2	9	**TR**	8	**5178**	0	4	26	**TR**	3	**484**	0	2	1
True	**VT**	14	0	**278**	3	9	**VT**	13	0	**2444**	13	1	**VT**	20	0	**2734**	11	10	**VT**	10	0	**290**	2	2
	OT	30	1	32	**37**	71	**OT**	23	3	17	**450**	59	**OT**	45	3	40	**527**	108	**OT**	34	3	31	**64**	39
	TC	9	11	12	16	**1440**	**TC**	1	2	0	37	**1823**	**TC**	13	29	22	72	**3215**	**TC**	11	25	23	34	**1395**
		Kappa: 0.898					**Kappa: 0.980**						**Kappa: 0.966**						**Kappa: 0.905**					

TABLE VIII. PERFORMANCE INDICES FOR THE REAL AND THE COMBINED (TRAIN) + REAL (TEST) DATABASES WITH OT CLASS (%)

Real						
	LP	**TR**	**VT**	**OT**	**TC**	**Mean**
Accuracy	97.42	99.18	97.82	95.51	96.04	97.19
Error	2.58	0.82	2.18	4.49	3.96	2.81
Recall	97.02	96.53	91.45	21.64	96.77	80.68
Specificity	97.64	99.57	98.38	99.03	95.56	98.03
Combined (train) + Real (test)						
	LP	**TR**	**VT**	**OT**	**TC**	**Mean**
Accuracy	97.66	98.99	97.87	95.85	96.33	97.34
Error	2.34	1.01	2.13	4.15	3.67	2.66
Recall	97.71	98.78	95.39	37.43	93.75	84.61
Specificity	97.64	99.02	98.09	98.64	98.02	98.28

type of seismic events. In this way, the classification was based on the recognition of the spectrogram 2D local structures, removing the need for a high number of features extracted from signals. To obtain the spectrograms, the signals of the seismic events were manually segmented normalized by the OVDAS analysts. The signals spectrograms were smoothed in order to highlight the general shape of their frequency components over time. Thus, it should be noted that no sophisticated pre-processing was performed on the data and that the spectrogram was directly placed at the input of the network.

In summary, the network was composed of two convolutional layers, followed by a ReLU, dropout – only in the first layer – and max pooling layers; one fully connected layer, followed by a ReLU and dropout layer; and one fully connected layer as output of the CNN with a softmax activation function. A data augmentation process was also carried out to improve the generalization capacity of our CNN and to ensure the obtained results were not affected by over-fitting.

In a first set of experiments, only four classes were considered. The results obtained only with the real data were very good and the contribution of the artificial data was not significant, although there was a slight improvement in the results. In a second experiment, a contrast OT class was added. This new class is difficult to classify from the spectrograms, since they have a great variability. As expected, the performance of the network decreased, although there was a significant improvement of the model performance when artificial data was included. The low performances for the OT class happen due to the fact that its spectrogram is not representative, but also because this group had a smaller number of examples compared to the other classes. Since OVDAS is still collecting more OT examples, a deeper analysis will be carried out in a future work. However, it is worth emphasizing the classification performance of the volcanic seismicity (LP, TR and VT) was not affected after including this class, confirming the robustness of our proposal. An alternative widely adopted in real-world analyses is to consider information from multiple stations to remove the events that are not originated from volcanoes like TC and OT. Therefore, in a real-world scenario, our approach can support OVDAS by classifying LP, TR, and VT with high accuracy.

Results have confirmed our hypothesis that the 2D representation of seismic signals contains important information to perform the discrimination of the classes. The fact that the OT group was not well discriminated also supports this hypothesis. This was a preliminary study, thus subsequent studies must be carried out to evaluate the proposal with a larger real database and in different volcanoes. Also, the data augmentation technique can be improved to generate more relevant new data to significantly impact the performance of CNN classifiers.

ACKNOWLEDGMENT

The authors would like to thank the FONDEF IDeA IT15I10027 project for financing part of the study, the PNPD/CAPES for supporting the postdoctoral position in the

dept of Computer Science at Federal University of Bahia and the OVDAS for providing the datasets.

REFERENCES

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

[2] V. Joevivek, N. Chandrasekar, and Y. Srinivas, "Improving seismic monitoring system for small to intermediate earthquake detection," *International Journal of Computer Science and Security*, vol. 4, no. 3, pp. 308–315, 2010.

[3] H. Langer, S. Falsaperla, T. Powell, and G. Thompson, "Automatic classification and a-posteriori analysis of seismic event identification at soufrière hills volcano, montserrat," *Journal of Volcanology and Geothermal Research*, vol. 153, no. 1-2 SPEC. ISS., pp. 1–10, 2006.

[4] G. Erlebacher and D. A. Yuen, "A wavelet toolkit for visualization and analysis of large data sets in earthquake research," *Pure and Applied Geophysics*, vol. 161, no. 11-12, pp. 2215–2229, 2004.

[5] 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," *Journal of Volcanology and Geothermal Research*, vol. 187, no. 3-4, pp. 218–226, 2009.

[6] S. Scarpetta, F. Giudicepietro, E. C. Ezin, S. Petrosino, E. Del Pezzo, M. Martini, and M. Marinaro, "Automatic classification of seismic signals at mt. vesuvius volcano, italy, using neural networks," *Bulletin of the Seismological Society of America*, vol. 95, no. 1, pp. 185–196, 2005.

[7] M. Masotti, S. Falsaperla, H. Langer, S. Spampinato, and R. Campanini, "Application of support vector machine to the classification of volcanic tremor at etna, italy," *Geophysical Research Letters*, vol. 33, no. 20, 2006.

[8] R. Carniel, L. Barbui, and A. D. Jolly, "Detecting dynamical regimes by self-organizing map (som) analysis: An example from the march 2006 phreatic eruption at raoul island, new zealand kermadec arc," *Bollettino di Geofisica Teorica ed Applicata*, vol. 54, no. 1, pp. 39–52, 2013.

[9] C. Cassisi, M. Prestifilippo, A. Cannata, P. Montalto, D. Patanè, and E. Privitera, "Probabilistic reasoning over seismic time series: Volcano monitoring by hidden markov models at mt. etna," *Pure and Applied Geophysics*, vol. 173, no. 7, pp. 2365–2386, 2016.

[10] A. M. Esposito, L. D'Auria, F. Giudicepietro, T. Caputo, and M. Martini, "Neural analysis of seismic data: Applications to the monitoring of mt. vesuvius," *Annals of Geophysics*, vol. 56, no. 4, 2013.

[11] G. Cortés, L. García, I. Álvarez, C. Benítez, T. de la Torre, and J. Ibáñez, "Parallel system architecture (psa): An efficient approach for automatic recognition of volcano-seismic events," *Journal of Volcanology and Geothermal Research*, vol. 271, pp. 1–10, 2014.

[12] M. Ibs-von Seht, "Detection and identification of seismic signals recorded at krakatau volcano (indonesia) using artificial neural networks," *Journal of Volcanology and Geothermal Research*, vol. 176, no. 4, pp. 448–456, 2008.

[13] B. Sick, M. Guggenmos, and M. Joswig, "Chances and limits of single-station seismic event clustering by unsupervised pattern recognition," *Geophysical Journal International*, vol. 201, no. 3, pp. 1801–1813, 2015.

[14] C. Hibert, F. Provost, J. Malet, A. Maggi, A. Stumpf, and V. Ferrazzini, "Automatic identification of rockfalls and volcano-tectonic earthquakes at the piton de la fournaise volcano using a random forest algorithm," *Journal of Volcanology and Geothermal Research*, vol. 340, pp. 130–142, 2017.

[15] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 2, 2012, pp. 1097–1105.

[17] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, vol. 2003-January, 2003, pp. 958–963.

[18] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701–1708.

[19] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, 2015, pp. 1–9.

[21] M. M. Krell and S. K. Kim, "Rotational data augmentation for electroencephalographic data," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2017, pp. 471–474.

[22] M. Curilem, F. Huenupan, D. Beltrán, C. San Martin, G. Fuentealba, L. Franco, C. Cardona, G. Acuña, M. Chacó, M. S. Khan, and N. Becerra Yoma, "Pattern recognition applied to seismic signals of llaima volcano (chile): An evaluation of station-dependent classifiers," *Journal of Volcanology and Geothermal Research*, vol. 315, pp. 15–27, 2016.

[23] S. M. Bhatti, M. S. Khan, J. Wuth, F. Huenupan, M. Curilem, L. Franco, and N. B. Yoma, "Automatic detection of volcano-seismic events by modeling state and event duration in hidden markov models," *Journal of Volcanology and Geothermal Research*, vol. 324, pp. 134–143, 2016.

[24] M. Curilem, C. Soto, F. Huenupan, C. San Martin, G. Fuentealba, C. Cardona, and L. Franco, *Improving the classification of volcanic seismic events extracting new seismic and speech features*, ser. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, vol. LNCS.

[25] J. Lahr, B. Chouet, C. Stephens, J. Power, and R. Page, "Earthquake classification, location, and error analysis in a volcanic environment: implications for the magmatic system of the 1989–1990 eruptions at redoubt volcano, alaska," *Journal of Volcanology and Geothermal Research*, vol. 62, no. 1, pp. 137 – 151, 1994, the 1989-1990 Eruptions of Redoubt Volcano, Alaska. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0377027394900310

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

[27] N. McLaughlin, J. M. Del Rincon, and P. Miller, "Data-augmentation for reducing dataset bias in person re-identification," in *AVSS 2015 - 12th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2015.

[28] N. Jaitly and G. E. Hinton, "Vocal tract length perturbation (VTLP) improves speech recognition," in *Proc. ICML Workshop on Deep Learning for Audio, Speech and Language*, vol. 117, 2013.

[29] B. McFee, E. Humphrey, and J. Bello, "A software framework for musical data augmentation," in *16th International Society for Music Information Retrieval Conference*, ser. ISMIR, 2015.

[30] E. L. Ferguson, R. Ramakrishnan, S. B. Williams, and C. T. Jin, "Convolutional neural networks for passive monitoring of a shallow water environment using a single sensor," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2017, pp. 2657–2661.

[31] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[32] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2016, pp. 1–621.