Classification of Isolated Volcano-Seismic Events Based on Inductive Transfer Learning

Manuel Titos, Angel Bueno, Luz García, Carmen Benítez, and J. C. Segura

***Abstract*— Domain-specific problems where data collection is an expensive task are often represented by scarce or incomplete data. From a machine learning perspective, this type of problems has been addressed using models trained in different specific domains as the starting point for the final objective-model. The transfer of knowledge between domains, known as transfer learning (TL), helps to speed up training and improve the performance of the models in problems with limited amounts of data. In this letter, we introduce a TL approach to classify isolated volcano-seismic signals at *"Volcán de Fuego"*, Colima (Mexico). Using the well-known convolutional architecture (LeNet) as a feature extractor and a representative data set containing regional earthquakes, volcano-tectonic earthquakes, long-period events, volcanic tremors, explosions, and collapses, our proposal compares the generalization capabilities of the models when we only fine-tune the upper layers and fine-tune overall of them. Compared with the other state-of-the-art techniques, classification systems based on TL approaches provide good generalization capabilities (attaining nearly 94% of events correctly classified) and decreasing computational time resources.**

***Index Terms*— Classification of isolated events, deep learning, transfer learning (TL), volcano-seismic signals.**

I. INTRODUCTION

SEISMIC signals registered using seismometers in volcanic areas can be classified based on the source mechanisms (seismic events) that originated them [1]. This letter analyzes how transfer learning (TL) [2] can be used to speed up training and improve the performance of the models used to classify volcano-seismic signals.

Deployed in vulcanological observatories, volcano-seismic signal recognition (VSR) systems provide several advantages.

1) They reflect the nature and underlying physics of the source processes involved.
2) Analyzing these seismic streams, geophysicists can separate rapidly into classes a large number of events, which is important in the case of eruptive crisis.
3) They provide consistent catalogs of each type of events improving the knowledge that we have about the state of the volcano.
4) The new knowledge obtained can be used to infer new eruptive crisis studying the temporal evolution of the volcano.

Manuscript received November 16, 2018; revised March 14, 2019, June 19, 2019, and July 10, 2019; accepted July 20, 2019. Date of publication August 14, 2019; date of current version April 22, 2020. This work was supported by the Ministerio de Economía y Empresa (MINECO)/Fondo Europeo de Desarrollo Regional (FEDER) under Grant TEC2015-68752-R. *(Corresponding author: Manuel Titos.)*

The authors are with the Teor ía de la señal, Telemática y Comunicaciones, Universidad de Granada, 18071 Granada, Spain (e-mail: mmtitos@ugr.es).

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

Digital Object Identifier 10.1109/LGRS.2019.2931063

Despite the good performance obtained by the existing classification techniques in terms of time consumption and accuracy rate [3]–[9], the accurate recognition of a certain kind of events remains constrained due to the difficulty of creating both, well-labeled and statistically representative data sets [5], [10]. Hence, one of the most challenging objectives in volcano-seismology is the development of robust data pattern extraction mechanisms which is able to characterize properly each event.

Traditionally, feature-engineering approaches based on the knowledge of human experts were used to extract relevant and discriminative information. However, newer approaches are based on deep hierarchical models that do not have to be supplied with such "hand-crafted" features. They can learn representative features from raw data. These new approaches have become the state of the art in many disciplines, improving the traditional ones, at the cost of much greater demand for training material and computational resources [11].

Given the vast amount of data, computation, and time resources required to develop deep hierarchical models, an emerging approach is to exploit what has been learned in one domain (where a lot of labeled training data are available) to improve generalization in another domain where data are scarce. This is what in the related literature is known as TL [2], alluding to the fact of the translation of knowledge acquired in a domain to a different one. Instead of starting the learning process from scratch, the basic idea is to use the parameters of a well-trained model in one domain (original domain) as a pretrained version for a model in a different domain (in which there are much less training data available). After that, pretrained parameters are fine-tuned using domain-specific available data (in the target domain), transferring the knowledge acquired in the original domain to the target domain to be used as a starting point for the training of the models.

In this letter, we use the LeNet architecture [12], designed for handwritten and machine-printed character recognition, as a feature extraction algorithm to build a system for automatic classification of volcano-seismic events. Input data, composed by spectrograms, will be processed by the LeNet model resulting in a feature vector that will later be used to train several multilayer perceptrons (MLPs).

The main contribution of this letter is to show the applicability and potential of using hierarchical feature representations obtained by models trained in a different specific problem as efficient information to build a system for automatic classification of volcano-seismic events. Our proposal retrains the model, keeping the spatial and spectral information extracted

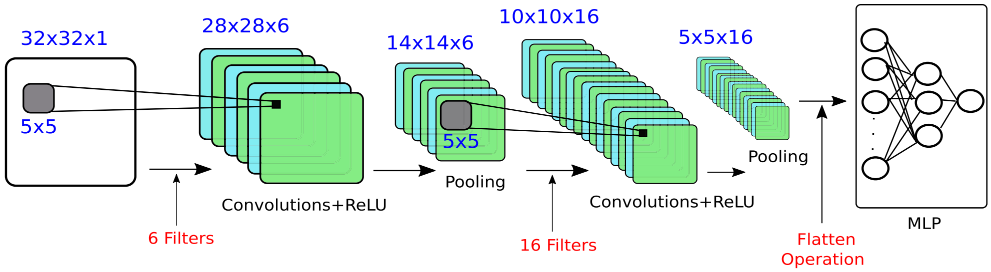

Fig. 1. LeNet architecture.

by the pretrained model in a different domain as the input information.

The rest of this letter is organized as follows: Section II provides a theoretical framework of TL approaches, and how it can be used for discriminative feature modeling of volcano-seismic events. Section III describes from the geophysical point of view, the seismic signals registered at *"Volcán de Fuego"*. Section IV describes the experimental setup and presents the results and discussion. Section V concludes this letter.

II. Theoretical Foundations and Related Works

One of the most promising techniques that could someday increase the capacity of generalization of artificial intelligence is the transfer of knowledge from an environment (domain) to other environment (domain), widely known as TL [2]. As we mentioned above, TL tries to exploit the knowledge that has been learned in one task to improve generalization in a different but related task. Following [2], a domain consists of two components: a feature space and a marginal probability distribution. Given a specific domain, a task (learning problem to be solved) can be defined as a label space and an objective predictive function that will be learned from the training data. Therefore, based on the different relations between domains and tasks, TL can be categorized into inductive TL (ITL), transductive TL (TTL), and unsupervised TL (UTL). Considering the nature of our proposal, we will describe only the inductive approach.

In ITL, the target and the source task are different. The domains of these two tasks may differ. In this case, as the purpose of both classification tasks differs, some labeled data are required in the target domain to induce its particular predictive model. The parameters of previously trained models (source) can be seen as a starting point of a new developing model, where the later layers of the original model are fine-tuned using available domain-specific data. This approach is based on the idea that low-level features (earlier layers) contain generic information (edge detectors, color regions detectors, and so on), while, progressively, the middle and later ones extract shapes and some task-specific features, respectively [13]. Therefore, given computation and time resources required to develop new models from scratch, ITL has become a very useful solution in the areas as computer vision (CV) [14], natural language processing (NLP) [15], or automatic speech recognition (ASR) [16] to speed up training and improve the performance of the models.

Applied to geoscience disciplines, TL has been found to be helpful in the domain-adaption problem as hyperspectral image analysis [17], remote sensing data classification [18], wind-speed prediction [19], and cyclone tracking [20], among others.

III. Data and Methods

This section describes the data set used in the study and the proposed architectures used for the experimental setup.

A. Proposed Architectures

Given that convolutional neural networks (CNNs) [12] have proven great success in fields such as CV, NLP, or ASR, it was decided to incorporate some of the most accurate models as base of our classification system to finally adjust their final layers with our data set.

In this sense, using the spectrogram images as the parameterization scheme, we proposed to use an LeNet network [12] as the base model.

Basically, an LeNet network is a CNN with seven levels of depth trained with the MNIST data set to classify handwritten and machine-printed character images of 32 × 32 pixels in gray-scale (Fig. 1). The model consists of several convolution layers, each of them followed by a max pooling operation.

1) Each convolutional layer can be understood as a feature extractor taking as inputs and the outputs from its previous layer in the hierarchy. It takes as input a stack of input planes and produces as output some number of output planes known as feature maps. At the same time, each feature map O_k can be understood as an arranged map of responses of a spatially local nonlinear operation, applied identically over the whole input planes. The main building block used to construct the nonlinear transformation is the convolution operation. Hence, each feature map O_k is associated with one kernel and computed as follows:

$$O_k = \sigma(b_k + \sum_r W_{kr} * X_r). \qquad (1)$$

$X_r, W_{kr}, *, b_k, and \sigma$ are the rth input channel, the subkernel for that channel, the convolution operation, the bias term, and the elementwise nonlinearity (sigmoid, hyperbolic tangent or Rectified Linear Unit) applied to the result of the kernel convolution, respectively.

2) The pooling step can be understood as the down-sampling operation along the spatial dimensions (width and height). Thus, max pooling operation consists of substituting each subwindow of size pxp by the

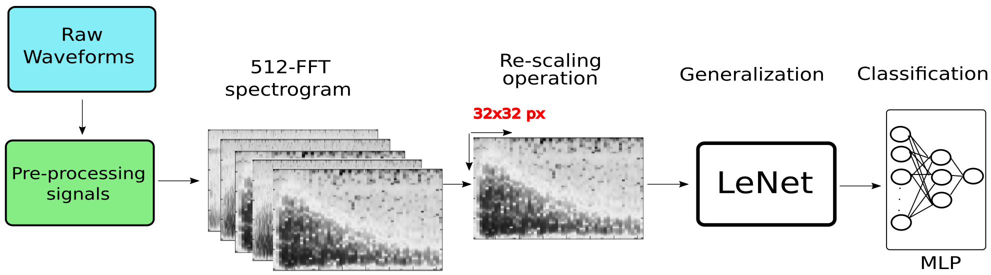

Fig. 2. Overview of the data preprocessing pipeline. First, signals are bandpass-filtered between 1 and 25 Hz. For each signal, we obtain its spectrogram using an FFT of 512 points. Finally, each spectrogram is resized to 32 × 32 and transformed to grayscale.

maximum feature value in it. This procedure can be formalized as follows:

$$H_{k,ij} = \max_{p} \left(O_{k,Si+p,Sj+p}\right) \tag{2}$$

where p and S determine the pooling window size and the stride value that corresponds to the horizontal and vertical increments at which pooling subwindows will be positioned.

3) The final extracted features are flattened and used as an input vector to feed one fully connected layer (or even two fully connected layers) added in the end.

The basic idea behind the LeNet architecture [12] is that the earlier convolutions are able to extract lower features as generic information (edge detectors, color regions detectors, and so on), while later convolutions are specialized on higher level features as specific shapes.

The topological structure in the first step uses a bank of six 5×5 filters with stride 1. This filter design results in six feature maps of 28×28 pixels. The pooling operation (using a filter width of 2 and stride of 2) reduces the dimension by a factor of 2 and ends up with six feature maps of 14×14 pixels. The second step applies another bank of 16 5×5 filters, resulting in 16 feature maps of 10×10 pixels. Again, applying the same pooling operation, 16 feature maps of 5×5 pixels are obtained.

Once the features have been extracted, they are flattened into a 1-D vector to feed one fully connected layer (or even more fully connected layers) with a specific number of nodes. Finally, on top, we add a softmax layer to normalize per-class output probabilities corresponding to each of the available events.

The final number of parameters is approximately 60k, including the parameters associated with filter design and fully connected layers.

B. Data Set

The database used to test the TL architectures proposed consists of 9332 volcano-seismic signals distributed per class as follows: 1738 volcano-tectonic earthquakes (VTEs), 2699 long-period events (LPEs), 1170 volcanic tremors (TREs), 455 regional earthquakes (REGs), 1406 collapses (COL), 278 explosions (EXPs), and 1586 volcanic noise (NOISE). Following [5] and [21], each type of event can be described according to its properties (source mechanism, length, and frequency content).

Volcano-Tectonic Earthquakes: VTE events are originated by seismic stress when a solid fracture takes place producing a seismic wave; it is possible to identify the P wave (pressure) and S wave (shear) arrivals. Spectral content could reach up to 30 Hz.

Regional Earthquakes: These earthquakes occur outside the volcanic structure and are related to tectonic stresses and to fault fractures. They can have larger duration and magnitude than VTE, but similar spectral content. P- and S-waves arrivals are generally clear.

Long-Period Events: Their source models are generally associated with the resonance of fluid-filled cavities such as cracks or magmatic conduits, in the shallow part of the volcano. Their spectra usually present one or several dominant peaks below 5 Hz.

Volcanic Tremor: Its spectral content is below 5 Hz, and the duration is highly variable, lasting from a few minutes to months. Volcanic tremor is a sign of high activity inside the volcano. Some theories suggest that it is caused by the movement of magma or gas, being almost identical to long-period events, except for the duration.

Explosions: They are characterized by variable duration (from second to tens of minutes) and a distinctive spectrogram with a narrow energy peak around 20 Hz. Explosions are naturally related to sonic boost waves, produced when the expanding gas is accelerated within the volcano structure.

Lava Flow: Volcanic debris processes located at the volcano surface exhibiting a frequency content of above 5 Hz.

Environmental Noise: Mainly introduced by nearby populations, human activities will interfere the frequency range where most of the volcanic spectral content is located.

IV. Experiment and Discussion

This section illustrates the performance of the proposed method. We compare the results obtained with other methods and parameterization schemes in terms of classification accuracy.

A. Model Training

Following [5], the feature extraction process is summarized in Fig. 2. The input of the model is the full data set of 9332 seismic signals (belonging to the station EZ5V4) in the time domain, sampled at 50 Hz and preprocessed using a bandpass filter between 1 and 25 Hz.

TABLE I

CLASSIFICATION RESULTS OBTAINED BY DIFFERENT ARCHITECTURES. 1 FC CORRESPONDS TO ONE FULLY CONNECTED LAYER. L AND A CORRESPOND TO MODELS WHERE THE LAST AND ALL LAYERS WERE FINE-TUNED. CNN-128 AND CNN-512 CORRESPOND TO MODELS WITH A SIMILAR TOPOLOGY TO LENET BUT WITHOUT TL STAGE

#Model	#Parameterization	#Topology	#Acc(%)	# Speed up(%)
SVM	LPC+Statistical information	RBF Kernel	91.55	-
SVM	LPC+Statistical information	Lineal Kernel	92.32	-
RF	LPC+Statistical information	120 estimators	92.80	-
MLP	LPC+Statistical information	500 hidden units	93.57	-
sDA-2H	LPC+Statistical information	260-385	94.32	-
CNN-LeNet 128 L	**512 FFT Spectrogram**	**LeNet + 1 FC layer (128 units)**	**88.3**	56.5
CNN-LeNet 512 L	**512 FFT Spectrogram**	**LeNet + 1 FC layer (512 units)**	**89.3**	36.4
CNN-LeNet 128 A	**512 FFT Spectrogram**	**LeNet + 1 FC layer (128 units)**	**93.4**	11.3
CNN-LeNet 512 A	**512 FFT Spectrogram**	**LeNet + 1 FC layer (512 units)**	**94.1**	32.3
CNN-128	**512 FFT Spectrogram**	**LeNet + 1 FC layer (128 units)**	**92.5**	-
CNN-512	**512 FFT Spectrogram**	**LeNet + 1 FC layer (512 units)**	**93.0**	-

After the preprocessing stage, a data set of 9332 spectrogram images using short-time fast Fourier transform (FFT) of 512 points is obtained, with their associated labels. To extract representative features using an LeNet architecture, we need to adapt the input dimensions of the images at 32 × 32 px. For that, spectrogram images are rescaled. Finally, the flattened features from the LeNet architecture will be used as training data for the classifier. As we are working with different streams of data, a direct consequence of normalizing all images to the same dimensions is that the longer signals lose more information. However, as we shall see later, the large size of the data set used to train the LeNet architecture minimizes the impact of this fact.

Considering the inductive nature of our proposal, we compare the result obtained by the models training all the layers and the later fully connected one. To do that, we use a sigmoid function as the activation function and an Adam optimizer [22] to optimize the loss function (negative log-likelihood). The overfitting scenarios are controlled during training using a validation set and early stopping criteria with patience of ten iterations. The data set was divided into training (75%) and test (25%) sets. This yields a training set of 7000 training instances and 2332 test instances. Furthermore, we used 50% of the test set (1166 instances) as validation data [5]. Other techniques as dropout or batch normalization did not offer any improvement. All the experiments were carried out using cross-validation with four partitions of the original database.

B. Classification of Isolated Events

The basis of the comparative study was derived from [5] where volcano-seismic signals were classified by several classical and deep architectures. As classical approaches, this letter used MLP [23], random forest (RF) [24], and support vector machines (SVMs) [25] with both linear and radial kernels. Regarding the deep ones, the architectures tested were deep belief networks (DBNs) [26] and stacked denoising autoencoders (sDAs) [27]. The parameterization scheme was based on linear prediction coefficients (LPCs) and statistical information associated with the impulsivity of the signals.

The best results obtained for different architectures and parameterization schemes are summarized in Table I. In order to prove the efficiency of the TL-based models against the version trained from scratch, we measure the relation between the runtime of both training algorithms according to

$$\text{speed up} = \frac{\mathrm{TL}_{\text{line}} - \text{base}_{\text{line}}}{\text{base}_{\text{line}}} \tag{3}$$

where $\text{base}_{\text{line}}$ is the runtime spent without TL version and $\mathrm{TL}_{\text{line}}$ is the total runtime achieved with TL version. The reported metrics are based on accuracy. Only the best results obtained after testing several configurations varying different units at first and second hidden layers have been reported.

Compared with the TL approach, several conclusions can be drawn from these results.

1) By applying the trained model as feature extraction, we note that ITL provides useful features for the discriminative stage, outperforming handcrafted ones applied to shallow classical classifiers (SVM, MLP, and RF).
2) Although the results obtained do not improve those obtained by deep networks using a specific handcrafted parameterization [5], they are really promising, most especially considering the vast amount of data, computation, and time resources required to develop deep hierarchical models from scratch.
3) Compared with specific features based on signal processing approaches, the ones extracted from spectrogram images have proven to be very useful for the classification of isolated seismo-volcanic events.
4) Given the large number of images of very different quality used to train the LeNet architecture, the rescaling operation of spectrogram does not have an undesired effect. The hierarchical features obtained from resized images focus mainly on the contours and shapes of the spectrograms proving to be sufficiently discriminative.

Moreover, the following should be pointed out. First, the dimension and resolution of the spectrogram images often have a big influence on the performance of the systems. Therefore, changing both, or even the number of input channels, models could obtain better discriminative information, improving the performance in learning and classification tasks. However, the use of higher resolution, even if the models use the same filter design, will result in larger feature maps increasing the number of parameters to be tuned and, therefore, affecting the size of the data set necessary to guide the optimization process. Second, given the size of the data set used in this letter, we noted that the inclusion of more than one hidden layer between the flattened and softmax layers degrades the performance. This degradation may be due to an overfitting problem. The convolutional features extracted are very representative. Thus, the inclusion of new nonlinear transformations leads the system to model noise and memorize rather than to generalize data.

V. Conclusion

In this letter, we present the use of ITL as a knowledge base from which to build reliable and efficient volcano-seismic classification systems. Based on the results obtained, we conclude that the use of previously adjusted CNN and, more specifically, the hierarchical learning representation that they implement can be efficiently exploited in the classification of isolated seismo-volcanic signals, taking advantage of the invariance and locality characteristics that convolution operations offer.

Acknowledgment

The data used in this letter have been obtained by the Seismological Network of Colima State (RESCO). They have been provided to the authors with the agreement of Gabriel Reyes Dávila and Raúl Arámbula.

References

[1] B. Chouet, "Volcano seismology," *Pure Appl. Geophys.*, vol. 160, nos. 3–4, pp. 739–788, 2003.

[2] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[3] M. Masotti, S. Falsaperla, H. Langer, S. Spampinato, and R. Campanini, "Application of Support Vector Machineto the classification of volcanic tremor at Etna, Italy," *Geophys. Res.Lett.*, vol. 33, 2006, Art. no. L20304.

[4] A. M. Esposito, F. Giudicepietro, S. Scarpetta, and S. Khilnani, "A neural approach for hybrid events discrimination at Stromboli volcano," in *Multidisciplinary Approaches to Neural Computing*. Cham, Switzerland: Springer, 2018, pp. 11–21.

[5] M. Titos, A. Bueno, L. García, and 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.

[6] 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 atRaoul Island, New Zealand Kermadec Arc," *Bollettino di Geofisica Teorica ed Applicata*, vol. 54, no. 1, pp. 39–52, 2013.

[7] A. D. Jolly *et al.*, "Seismo-acoustic evidence for an avalanche driven phreatic eruption through a beheaded hydrothermal system: An example from the 2012 Tongariro eruption," *J. Volcanol. Geothermal Res.*, vol. 286, pp. 331–347, Oct. 2014.

[8] M. Beyreuther and J. Wassermann, "Hidden semi-Markov model based earthquake classification system using weighted finite-state transducers," *Nonlinear Processes Geophys.*, vol. 18, no. 1, pp. 81–89, 2011.

[9] 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.*, vol. 57, no. 4, pp. 1936–1948, Apr. 2019.

[10] G. Cortés *et al.*, "Evaluating robustness of a HMM-based classification system of volcano-seismic events at colima and popocatepetl volcanoes," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 2, Jul. 2009, pp. II–1012–II-1015.

[11] T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: A survey," *J. Mach. Learn. Res.*, vol. 20, no. 55, pp. 1–21, 2019.

[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[13] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.

[14] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[15] J.-T. Huang, J. Li, D. Yu, L. Deng, and Y. Gong, "Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 7304–7308.

[16] J. Deng, Z. Zhang, F. Eyben, and B. Schuller, "Autoencoder-based unsupervised domain adaptation for speech emotion recognition," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1068–1072, Sep. 2014.

[17] H. L. Yang and M. M. Crawford, "Domain adaptation with preservation of manifold geometry for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 543–555, Feb. 2016.

[18] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, Jun. 2016.

[19] Q. Hu, R. Zhang, and Y. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renew. Energy*, vol. 85, pp. 83–95, Jan. 2016.

[20] S.-S. Ho and A. Talukder, "Automated cyclone tracking using multiple remote satellite data via knowledge transfer," in *Proc. Aerosp. Conf.*, Mar. 2009, pp. 1–7.

[21] R. Arámbula-Mendoza, P. Lesage, C. Valdés-González, N. R. Varley, G. Reyes-Dávila, and C. Navarro, "Seismic activity that accompanied the effusive and explosive eruptions during the 2004–2005 period at Volcán de Colima, Mexico," *J. Volcanol. Geothermal Res.*, vol. 205, nos. 1–2, pp. 30–46, 2011.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[23] I. Goodfellow, Y. Bengio, A. Courville, and F. Bach, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[24] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[25] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[26] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[27] P. Vincent, H. Larochelle, I. Lajoie, Y. 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, Dec. 2010.