Dilated-RNNs: A Deep Approach for Continuous Volcano-Seismic Events Recognition

Manuel Titos, Joe Carthy, Luz García, Talfan Barnie, and Carmen Benítez

***Abstract*—Monitoring continuous volcano-seismic signals is often performed by systems trained on scarce or incomplete datasets. From a machine learning perspective, these types of systems are typically built by learning from seismic records containing information not only on the volcanic dynamics, but also on the complex inner structure of the volcanic edifice. The dual nature of the information content presents a challenge when it comes to modeling events temporally. Here, we show that while existing recurrent-neural-network-based monitoring systems recognize almost 90% of events annotated in seismic catalogs, the long-range temporal dependencies are still hard to model. We found that dilated recurrent neural networks based on multiresolution dilated recurrent skip connections between layers have the remarkable capability to simultaneously enhance the efficiency of the model, reducing the number of training parameters, while increasing the performance of the model when compared with classical recurrent neural networks in sequence modeling tasks involving very long-term seismic records. Our results offer the potential to increase the reliability of monitoring tools despite the variations in the geophysical properties of the seismic events within the volcano across eruptive periods.**

***Index Terms*—Deep learning, dilated recurrent neural networks (Dilated-RNNs), transfer learning (TL), volcanic monitoring.**

I. Introduction

ALTHOUGH the combination of multiple remote sensing and real-time monitoring disciplines has allowed rapid advances in volcanic hazard assessment and risk mitigation, the study of seismic waves in the volcanic environment offers an understanding of what complex processes are happening inside the volcano. These processes range from the complex interaction between multiphase fluids and their hosting rock to ductile deformation and brittle failure and processes associated with the magma transport. Volcanic activity can be located and described by seismic waves that are characterized (in terms of waveform, energy, duration, and frequency content) based on the internal processes generating them. Therefore, developing operational frameworks that allow for the real-time assessment of the large volumes of data obtained by volcano observatories is an important challenge to address. Particularly, in the field of volcanic seismology, it will help to understand the behavior behind the physical processes of the volcano.

Starting with Ohrnberger [1], the past two decades have borne witness to the development of automatic volcano-seismic monitoring systems. The overwhelming success of data-driven models to solve complex real-world problems [2] positions them as an effective alternative to physics-driven models, which currently form the basis of many of the monitoring systems deployed in volcanological observatories. In addition to their computational cost, physics-driven approaches rely on data to unmask and improve the theoretical explanation behind them. Data-driven approaches allow for improvements in terms of discovering the inner workings of volcanoes, in volcanic hazards assessment approaches, and in risk mitigation protocols [3], [4] by analyzing monitored data through well-defined mathematical rules. As a consequence, to develop improved predictive models, one needs to pursue a balanced approach taking elements from both data and physics-driven approaches, keeping data as a common point. This balanced approach can be achieved by using machine learning (ML) [5].

Focusing on volcanic monitoring through seismic data, ML algorithms, employing both supervised and unsupervised approaches, have played a pivotal role since the groundbreaking works in [6] and [7]. These algorithms aim to uncover meaningful patterns and generate new features for interpreting information collected by seismometers, ultimately enhancing automatic early warning systems [8], [9], [10], [11], [12], [13], [14], [15]. In seismic event recognition, there are two key perspectives: 1) classifying isolated events, where individual events categorized by expert geophysicists are placed into predefined classes [16], [17], [18], [19], [20], [21], and 2) sequence modeling, which involves processing continuous event sequences and requires efficient models capable of capturing the temporal evolution of such sequences for detection, segmentation, and classification [6], [7], [8], [9], [11], [12], [13], [22], [23]. While the isolated classification of volcano-seismic events remains an active area of research, this study specifically emphasizes sequence modeling detection.

Manuscript received 2 February 2024; revised 13 May 2024 and 7 June 2024; accepted 20 June 2024. Date of publication 2 July 2024; date of current version 12 July 2024. This work was supported in part by the IMPROVE under Grant H2020-MSCA-ITN-2019-85809, in part by the Spanish Project PID2022-143083NB-100 funded by MCIN/AEI/10.13039/501100011033 and FEDER (EU) "Una manera de hacer Europa", in part by the Spanish Project PLEC2022-009271 "DigiVolCan," funded by MCIN/AEI/10.13039/501100011033 and EU NextGenerationEU/PRTR, in part by the Spanish Grant TED2021-132178B-I00 funded by MCIN/AEI/10.13039/501100011033 and "European Union NextGenerationEU/PRTR", in part by the Junta de Andalucía-Consejería de Universidad, Investigacion e Innovacion under Grant P21_00051. *(Corresponding author: Manuel Titos.)*

Manuel Titos, Joe Carthy, Luz García, and Carmen Benítez are with the Department of Signal Theory, Telematics and Communications, and the Center for Information and Communication Technologies Research (CITIC), University of Granada, 18014 Granada, Spain (e-mail: mmtitos@ugr.es).

Talfan Barnie is with the Processing and Research Division, Icelandic Meteorological Office, 105 Reykjavík, Iceland.

Digital Object Identifier 10.1109/JSTARS.2024.3421921

Traditional ML techniques, such as Hidden Markov models, have been widely employed in sequence modeling [11], [24]. However, in recent years, neural networks (NNs), including recurrent and temporal convolutional approaches have taken a leading role due to their strong temporal modeling capabilities. By generating a spatiotemporal sequence of hierarchical features, both recurrent NNs (RNNs) [25], [26] and temporal convolutional networks (TCNs) [27] have recently been successfully applied as seismo-volcanic monitoring tools [12], [13], [14], [22], [23] but also in complex and emerging research geosciences fields, such as climate change monitoring [28] and remote sensing [29].

Sequence modeling is a significant challenge in the development of real-time monitoring systems. Learning from continuous seismo-volcanic records can be arduous [25]. There are multiple challenges in training models both due to the requirement to recognize sequences of different durations and due to other considerations including the following:

1) modeling long-term dependencies while maintaining mid and short-term ones;
2) mathematical limitations affecting the training process (vanishing/exploding gradient problems) [26], [30];
3) learning may not be possible to efficiently parallelize;
4) datasets guiding supervised learning are scarce and, the labeling criteria carried out is sometimes not uniform among the experts;
5) algorithms often cannot provide insights into their behavior (being a black-box to users) potentially hiding bias in the model or problems in the training data.

Each of these challenges may hinder the development of high-performance models.

This work introduces dilated recurrent NNs (Dilated-RNNs) [31] as an architecture to build a new seismo-volcanic monitoring system. Based on multiresolution dilated recurrent skip connections between layers, the Dilated-RNN possesses enhanced efficiency reducing the training parameters (compared with standard stacked RNN cells) while matching state-of-the-art performance in sequence modeling tasks involving very long-term dependencies. Our objective is to show the usefulness of Dilated-RNN simultaneously tackling the first, second, and third challenges referred to above. To address the fifth challenge (enhancing learning explainability), this work adopts the method outlined in [22] to assess neural activation dynamics in the most informative network layers. This is achieved by pruning the least influential neurons, with the objective of 1) simplifying the architecture of the network and 2) studying the ability of the neurons to specialize in the recognition of a specific physical event by mapping their activation levels.

The rest of the article is organized as follows: Section II offers a theoretical overview of dilated architectures (Dilated-RNN). Section III details the database and experimental framework. In Section IV, we present the results and engage in discussions. Finally, Section V concludes this article.

II. Dilated-RNNs

Dilation operations in NNs allow for inserting *holes* between consecutive information elements in each network layer during the learning/prediction process. This creates a trainable mechanism for selectively emphasizing or disregarding individual features based on their relevance to the prediction task. This dynamic feature selection allows the network to effectively analyze temporal series, adjusting to diverse temporal structures and sequence durations. This adaptive behavior ensures alignment with the specific requirements of the task at hand.

The introduction of dilations in NNs was initially proposed in [32], leading to TCNs. Processing of temporal series has not been successfully attained with classic Convolutional NNs (CNNs) because their fixed input data size is unable to identify longer-term patterns for events of higher durations. When dealing with sequential training, CNNs have primarily been used as a preliminary feature-extractor, e.g., [31], [33], [34]. TCNs proposed a compact approach to learning temporal structures. They combined very deep networks (augmented with residual layers) with dilated convolutions to efficiently increase the size of the receptive field employed. This increase in the effective receptive field allows TCN models to better capture short-to-medium sequences in the data. However, the capability of TCN models to capture longer temporal dependencies is still constrained by their kernel size while RNNs, using autoregressive modeling, can theoretically capture longer term of temporal dependencies with a modest number of parameters [31].

Despite issues related to extended training durations and limited parallelization due to their gradient propagation through time training approach, RNNs and more specifically their memory-gate versions such as long short-term memory (LSTM) [25], [35] and gated recurrent unit (GRU) [36], [37] have provided interesting advances in parallel to those of TCNs in effective sequential training in several applications including, speech recognition and processing [38], [39], or volcano-seismic events recognition [22], [23].

While RNN-LSTMs and RNN-GRUs achieve high-performance, long-sequence learning remains a challenging problem often confronted with vanishing and exploding gradients (impeding learning) and time-consuming sequential training periods. The usage of dilations offers potential improvements in regard to both of these issues, which has led to the approach named Dilated-RNNs [31], that are theoretically capable of capturing longer-term dependencies with a small number of parameters and parallel computation if desired. Dilated-RNNs are multilayer and contain cell-independent neural connection architectures, analogous to TCNs, but under a recurrent setting (see Fig. 1). The dilated recurrent skip connections organized into multiple dilated recurrent layers with hierarchical dilations alleviate the impeding learning problems and extend the range of temporal dependencies with fewer parameters and lower computational cost permitting parallel computation.

According to the work in [31], dilated recurrent skip connections can be mathematically described using a general formula for forward propagation at each time step. Let c_t^l denote any RNN cell, such as a Vanilla RNN cell, LSTM, GRU, etc., in layer l at time t. Assuming an input at time t and layer l denoted by x_t^l, the dilated skip connection is represented as follows:

$$c_t^l = f\left(W^l * x_t^l + U^l * c_{t-s^l}^l\right). \tag{1}$$

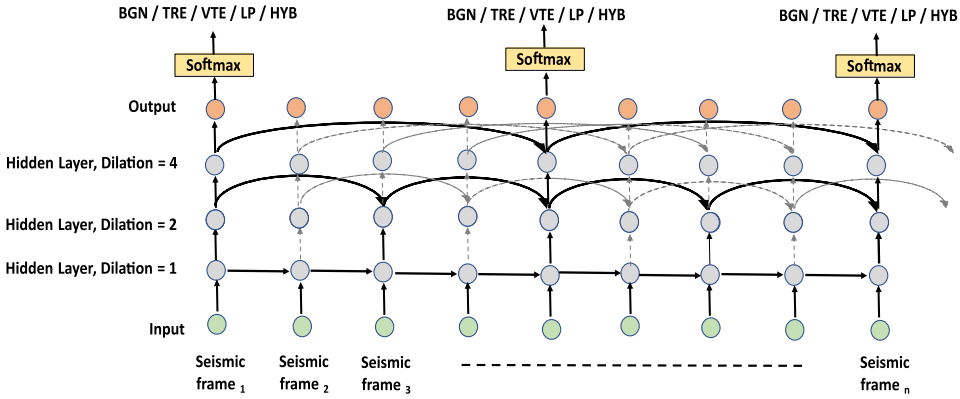

Fig. 1. Dilated-RNN architecture studied in this investigation. Each gray point represents an LSTM cell. A classical LSTM configuration is essentially equivalent when there is just one layer, and there are no dilated recurrent skip connections.

Here, s^l is referred to as the skip length or dilation of layer l, W^l is the weight matrix at layer l for the current input x_t, U^l is the weight matrix at layer l for the dilated connection from the output at time $t - s^l$, and f encompasses any activation function. This formula embodies the central concept of incorporating dilated connections to capture temporal patterns by introducing a connection from the output at a previous time $t - s$. The dilation s determines the time distance of this connection. An essential distinction between dilated and regular skip connection approaches lies in the fact that the former eliminates reliance on c^l_{t-1}, in contrast to the regular skip connections [31]. This reduction in dependence results in a decrease in the number of parameters that need to be tuned. It is important to note that this equation serves as a general example, and the mathematical notation may vary depending on the specific architecture of the Dilated-RNN in use. The architecture of a Dilated-RNN may include multiple layers and structures. For a more in-depth understanding, we encourage readers to explore the work in [31].

III. Database Description and Experimental Framework

A. Database Description

Our dataset consists of seismological signals from Deception Island (62° 59' S, 60° 41' W), a prominent active volcanic island in the South Shetland archipelago and the Antarctic Peninsula [40], [41], [42]. These signals were gathered during three Antarctic seismic surveys conducted in the austral seasons of 1994–1995, 1995–1996, and 2001–2002. A thorough account of the sensor and acquisition systems can be accessed in [43]. For a more extensive understanding of the dataset and the seismic events themselves, one can refer to the work in [12]. Data labeling was conducted by experienced geophysicists knowledgeable about the volcano's dynamics. All volcano-seismic events have been selected as the most representative of each class. This yielded 512 continuous data streams, comprising 2193 events, distributed among the classes as follows: 1222 background noise (BGN), 77 tremors (TRE), 765 long-period events (LP), 75 volcano-tectonic earthquakes (VTE) and 54 hybrid events (HYB). Fig. 2 illustrates the volcano-seismic events related to the Deception Island dataset examined in this article, providing a comprehensive depiction of their temporal organization across both time and frequency domains. The fluctuations in amplitude and frequency observed throughout the event stem directly from the distinctive source mechanism that characterizes this occurrence. Consequently, these variations can be utilized as distinctive markers for identification purposes.

B. Experimental Framework and Recognition Results

To facilitate a comprehensive comparison and evaluate the efficacy of these models in recognizing volcano seismic signals, two architectures have been implemented in this study: 1) RNN-LSTM and 2) Dilated-RNN with LSTM cells (Dilated-LSTM).

Rather than directly employing the waveform, the raw streaming records underwent parameterization. This involved applying the log filter bank parameterization scheme, incorporating temporal context information (δ, $\delta\delta$), as outlined in [22], to establish the baseline systems. Such feature extraction workflow follows the process outlined in the following.

1) Raw data undergo windowing using 4-s Hamming windows with a 3.5-s overlap.
2) For each window, a 512-point fast Fourier transform is computed. The magnitude of the resulting spectrum serves as input for a bank of 16 triangular filters, evenly distributed on a logarithmic frequency scale with 50% overlap between adjacent filters. Subsequently, we compute the logarithm of the filter-bank energies, resulting in a 16-component feature vector for each window.
3) The feature vector is finally augmented with details about the temporal context. This includes the incorporation of the first and second-order temporal derivatives (δ, $\delta\delta$) for each 16-component feature vector [44]. Consequently, the size of the feature vector triples in comparison to the initial one.

The RNN-LSTM network topologies utilized in this study are based on the architectures described in [12]. Specifically, we explored a range of configurations by adjusting the number of hidden units for the LSTM layer from 10 to 300 and the

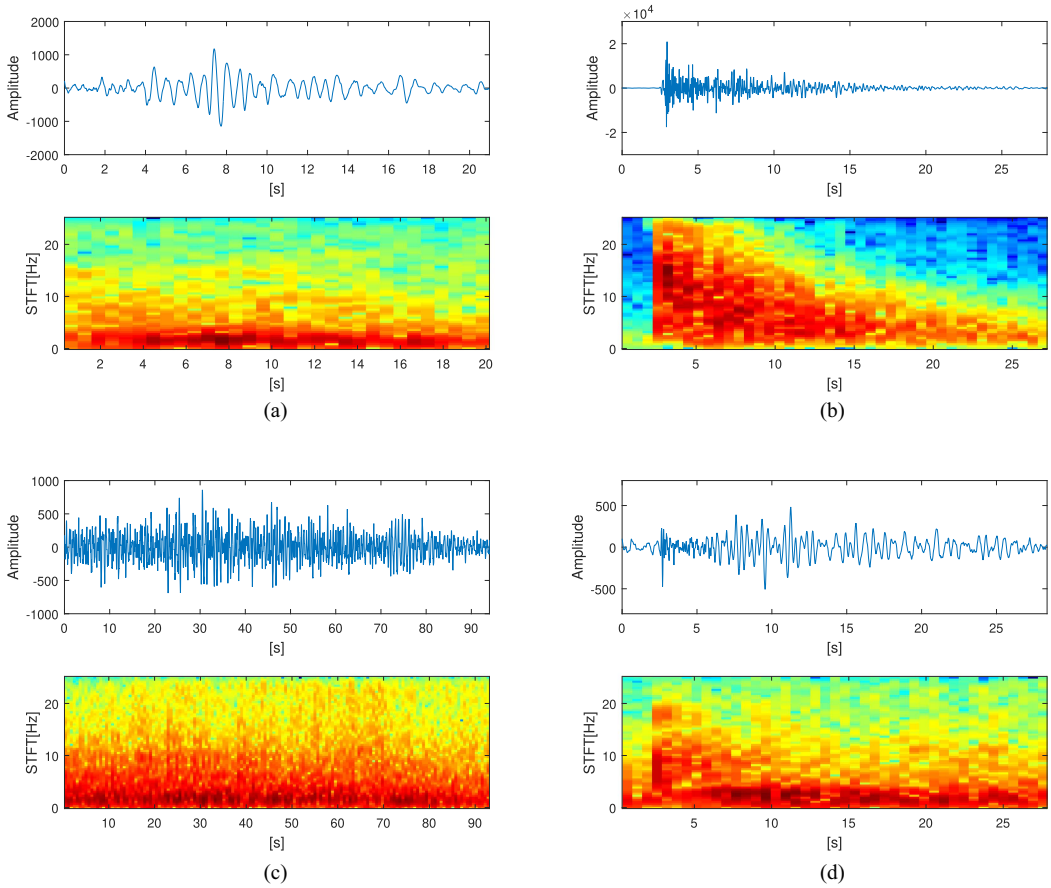

Fig. 2. Amplitude and spectrograms of the four types of volcano-seismic events recorded at *Deception Island* volcano during three seismic surveys: 1994–1995, 1995–1996, and 2001–2002. (a) LP. (b) Volcano-tectonic earthquake (VTE). (c) TRE. (d) HYB.

learning rate from 0.004 to 0.01. To mitigate overfitting, we employed early stopping and L2 regularization techniques. The optimal RNN-LSTM model, which featured a single hidden layer without dilations, demonstrated superior performance with 210 hidden units (including $\delta, \delta\delta$) and 130 hidden units (excluding $\delta, \delta\delta$). For the dilated version, we varied the number of hidden units per layer from 10 to 150, incorporating dilation factors of 2, 4, and 8 (considering the length of the input signals). This parameter also determined the depth of the system, ranging from 1 to 3 layers. The most effective configuration for the Dilated-LSTM model included three hidden layers, each with 50 hidden units, and 2–4 dilated recurrent skip connections per layer. System performance was assessed using accuracy-based validation metrics. The experimental configuration is defined as follows.

1) All experiments were conducted within a conda environment. The setup of this environment includes CUDA version 9.2, Python 3.7, PyTorch 1.2.0, CUDNN 7602, and requisite libraries. The seismic data were divided into two segments for model training and testing purposes. The training set was further divided into two subsets: 1) the training dataset and 2) the validation dataset, constituting 80% and 20% of the data, respectively.
2) Data Normalization: We use a standard deviation normalization across the training, validation, and blind testing datasets, defined as

$$y_{\text{i-normalized}} = \frac{y_i - \mu_i}{\sigma_i}$$

where μ_i and σ_i represent the mean and standard deviation of the ith feature for all the seismic instances in the training set, respectively, and y_i denotes the ith feature of the input vector.
3) Optimizer: SGD with a fixed learning rate ranging from 0.004 to 0.01.

It is important to acknowledge the inherent limitations of the dataset, stemming from its scarcity and imbalance. The PyTorch implementation of both architectures' source codes, along with preparameterized databases, is accessible to users through the provided link.

Table I summarizes per-frame recognition results obtained for RNN-LSTM and Dilated-LSTM models when applying a leave-one-out cross-validation approach. All the experiments were carried out using fivefold, resulting in 80% and 20% of the data for the training, validation, and test sets, respectively. A closer look at the table reveals that the dilated model outperforms the classical model, both with and without contextual information, while also reducing the parameter count. As emphasized in Section II, the parameter reduction in the dilated architecture is evident not just in a reduction of hidden units

TABLE I
RECOGNITION CAPABILITIES ANALYSIS: RECOGNITION PERFORMANCE OBTAINED BY THE PROPOSED ARCHITECTURES USING DATA BELONGING TO THE 1994–1995, 1995–1996, AND 2001–2002 SPANISH ANTARCTIC CAMPAIGNS

acc. (%)	LSTM $\delta, \delta\delta$	Dilated-LSTM $\delta, \delta\delta$	LSTM	Dilated-LSTM
Test 1	86.61	90.22	85.02	86.60
Test 2	80.13	82.13	78.11	79.01
Test 3	82.05	83.04	79.20	83.26
Test 4	89.09	89.16	87.23	86.84
Avg.	84.47	**86.13**	82.39	83.92
No. Param	21 9455	61 055	77 615	**54 655**
Train times (s)	**4200**	6300	3995	6000

The bold values correspond to the best results.

per layer but also in the elimination of reliance on c^l_{t-1}. This represents a fundamental distinction between dilated and regular skip connection approaches. Interestingly, when contextual information is introduced, the RNN-LSTM yields results similar to the Dilated-LSTM without context. These results suggest that the use of dilations, effectively expanding the receptive field, imparts knowledge akin to that obtained from contextual information. Since training RNNs often exhibit oscillatory behavior due to the variation in training segment sizes and, consequently, error propagation at different time steps, we did not utilize an early stopping-based regulator. However, we have observed that dilated models converge earlier than their classical counterparts in terms of the number of iterations. The training times are associated with a fixed training duration of 300 iterations.

IV. ANALYZING ARCHITECTURES: DISCUSSION

As discussed, our goal is to assess Dilated-LSTM effectiveness in seismo-volcanic monitoring systems, specifically in offline recognition tasks. We will examine how multiresolution dilation operations impact result robustness and reliability. The results obtained for RNN-LSTM models in [22] and [12] are considered as baselines and three aspects are analyzed as follows.

1) We examine a sparse version of Dilated-LSTM, where selected hidden units are pruned to reduce complexity while operating in high-dimensional feature spaces.
2) We explore the completeness of the system's learning mechanism, focusing on the idea of abstraction through the specialization of sub-structures, neurons, or units.
3) We assess the system's robustness by characterizing data from seismic campaigns or periods with varying volcanic dynamics.

A. Posttraining Sparsification Analysis

Understanding the impact of sparsity on NNs remains a challenge, but there is growing interest in this area. Deep learning models are typically dense and overparameterized, leading to overfitting and memorization of random patterns, ultimately impairing their generalization abilities [44]. Sparsified networks, on which some hidden units or neurons are selectively pruned, consistently match and often outperform their initially dense versions [45], [46]. That is, they reduce the representational complexity of the systems by zeroing out subsets of parameters while they continue to operate in high-dimensional feature spaces [45].

Although sparsification approaches in the field of deep learning are very diverse, this work focuses on a data-driven selection scheme where the network is pruned after training, by removing a certain amount of components with the lowest heat values to match the required sparsity level [45]. Following the work in [22], instead of relying on an absolute weight magnitude-based pruning method (data-free selection scheme), our posttraining sparsification approach is grounded in the neurons' excitation behavior. Rather than pruning connections solely based on their magnitude (as smaller absolute magnitude weights are assumed to have less impact), we prune hidden units using an absolute-heat-value-based threshold. The heat value is determined as the average activation value for each hidden unit when recognizing events within the same category. Additionally, sparsification, alongside having an impact on a network's generalization and robustness, can aid in assessing the saliency of each layer as well as the high-heat-value neurons.

Fig. 3(a) depicts the average posttraining sparsification results for the Dilated-LSTM (with three hidden layers). Average posttraining sparsification results related to RNN-LSTM are described in [22]. Each line represents recognition accuracy (%) at various sparsification levels, considering all possible pruning combinations among layers. *"All-layers"* illustrates performance while varying activation percentile and pruning neurons in all layers. *"Layers-1-2," "Layers-1-3,"* and *"Layers-2-3"* show the same analysis but keep all the neurons at layer 3 and only pruning neuron in layers 1 and 2 for the first; keeping all the neurons at layer 2 and only pruning neuron in layers 1 and 3 for the second and keeping all the neurons at layer 1 and only pruning neurons in layers 2 and 3 for the latest one respectively. Finally, *"Layer-1," "Layer-2,"* and *"Layer-3"* show the results when pruning only layer 1, 2, or 3, respectively. The Y-axis represents the percentile of neuron activity (relative to the most active neuron) used for testing the architectures. For example, the 100th percentile on the Y-axis uses all neurons for testing while the 10th percentile uses only the top 10% of the most active neurons (those with 90% of the maximum activation value). The X-axis represents classification results in terms of accuracy percentage. Fig. 3(b) illustrates the number of neurons pruned per layer during posttraining sparsification for Dilated-LSTM when keeping other layers unpruned.

The results exhibit varying trends when considering different layers during pruning. Fig. 3(a) demonstrates that, when evaluating layer 3, performance deteriorates earlier, even with a relatively low activation percentile. This highlights both the greater specialization of layer 3 compared to layers 1 and 2 and the robustness of the extracted features. Even with a substantial number of neurons pruned from layers 1 and 2, performance remains acceptable. The specialization of layer 3 is further evident in Fig. 3(b). Assuming specialization when only a subset of units within a layer are highly activated, layer 3 is highly specialized. The average number of pruned neurons, even at low activation percentiles, is significantly higher in layer 3 compared to layers 1 and 2. Approximately 40% of neurons in layer 3, on average,

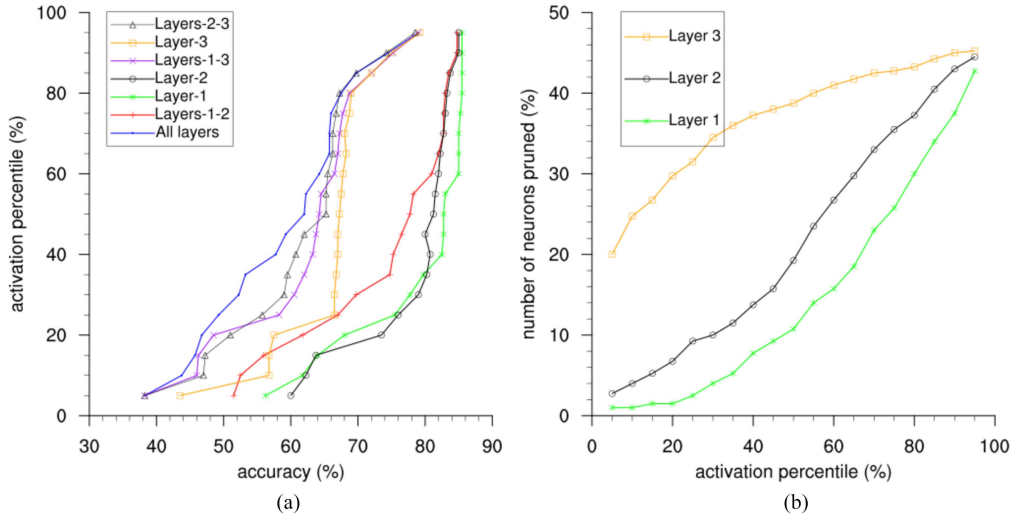

Fig. 3. Posttraining sparsification analysis. (a) Recognition accuracy (%). (b) Number of neurons pruned per layer.

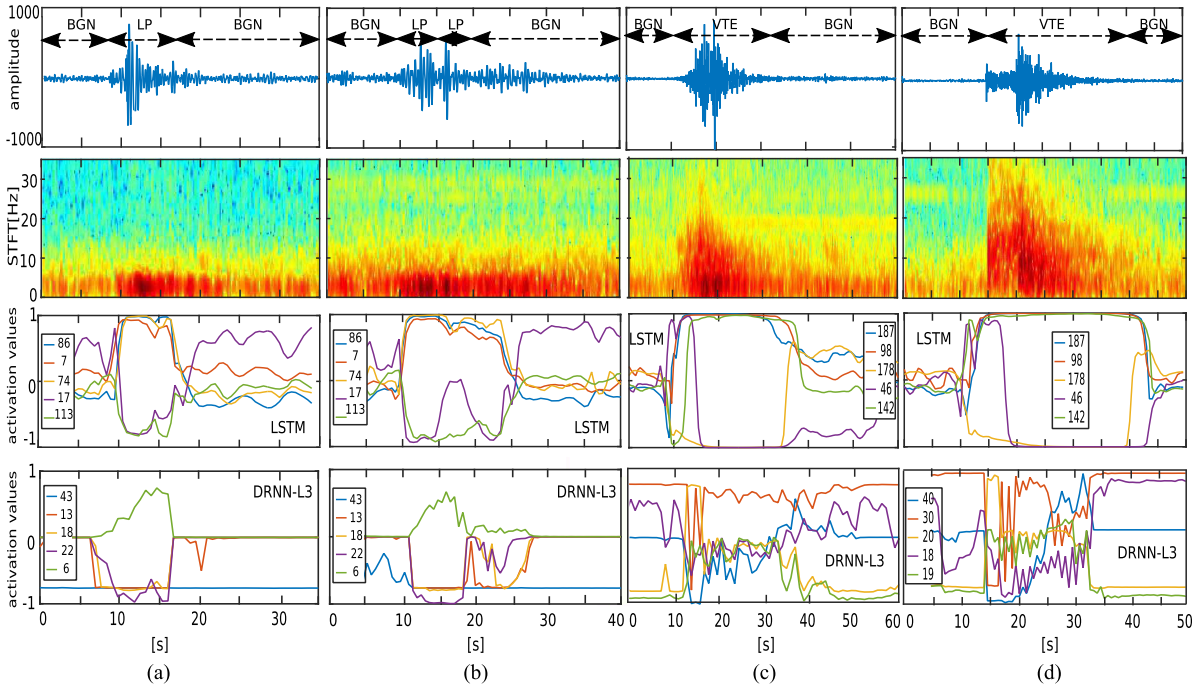

Fig. 4. Waveforms (first row), spectrograms (second row) and time-evolution profiles of the top five specialized hidden units for RNN-LSTM (third row) and Dilated-LSTM (fourth row) architectures when recognizing different LPs (a and b) and VTEs (c and d). L3 in Dilated-LSTM corresponds to the final layer depth.

exhibit activation levels below 5% of the highest one. While 40%–50% of neurons in layers 1 and 2 can be pruned having a little effect on the performance, when layer 3 is included, the number of pruned neurons grows hugely, losing elemental information and dropping out the performance quickly.

B. Neural Activation: Analyzing the Knowledge Acquired

Analyzing the system's output is essential for detecting training biases and ensuring its correct functioning. We aim to address the challenge of delivering accurate yet interpretable insights into the system's behavior (completeness of interpretation) [47]. To examine the completeness of the system's interpretation, we highlight the concept of abstraction through the specialization of substructures or hidden units. Given that data are parameterized according to frequency domain information and grouped by bands, the objective is to incorporate visual cues for gaining "potential insights." These insights aid in comprehending the system's behavior and the reasons behind specific input–output relationships.

Fig. 4 depicts the temporal neural activity of the top five highly specialized hidden units in the architectures while recognizing

different LP and VTE events. The most specialized units are identified by first grouping events by type and calculating the average absolute activation values of neurons over time during event recognition. Subsequently, the heat value for each hidden unit is determined. This heat value represents the average neural activation of each hidden unit when recognizing all events grouped by type. Finally, the five neurons with the highest heat values for each event type are selected.

Upon closer examination of neural activation patterns, we observe a clearer class-selectivity among neurons, which is defined as the ability to enhance event recognition or detection by reducing global neural activity concentrated in a few highly sensitive neurons with high activation values [48], [49]. While the RNN-LSTM demonstrates a comprehensible behavior in detecting and delineating both types of events, Dilated-LSTM exhibits a slightly less intuitive behavior when detecting VTEs: The most specialized neurons reduce their activity when a VTE is detected and become active again when the energy levels drop. In the case of LPs, the most specialized neurons maintain a state of minimal activity, activating efficiently to detect incoming events when there is a change in energy levels. Compared to the RNN-LSTM, the Dilated-LSTM displays a relative quasi-plain behavior before and after the event occurrence. In terms of event detection, both architectures prove effective.

C. Generalization Capability: Reviewing Recent Dataset

One of the most important challenges in automatic volcano-seismic recognition systems is to build robust computational models that can easily readapt themselves to the highly dynamical internal seismic sources inside volcanoes. Over time, volcanic dynamics may alter specific seismic signal characteristics, potentially leading to significant changes in event characterization. Consequently, in the long term, these alterations can result in a recognition system's failure to detect certain events, even if the dynamic changes have not given rise to entirely new seismic event types. Given the highly dynamic nature of the data under examination, we assess whether a system trained on data from campaigns in 1994–1995, 1995–1996, and 2001–2002 can effectively monitor the XXX Spanish Antarctic Campaign data at the *Deception Island* volcano.

We employed a 3.5-h seismic record from the 2016–2017 Antarctic survey as our testing data. To maintain consistency with the training dataset's frequency range (see Section III), all signals were filtered between 1 and 20 Hz. This new dataset was derived from various unprocessed seismic records, featuring unknown events and lacking prior human oversight. The inclusion of such events in this dataset might impact the predictions, as they may differ from the meticulously selected prototype events used to train the system. According to the work in [12], the accuracy of recognition and generalization capabilities for all recurrent models proposed could be influenced by the pronounced attenuation and source effects observed in the volcanic environment of Deception Island.

Table II presents our test results, reported in terms of *accuracy* and *correct* detection. *Correct* reflects the percentage of recognized events within the existing seismic catalog while *accuracy* also accounts for insertion errors (recognized events not found in the seismic catalog). Notably, the table reveals that recognition performance (Cor%) is robust, but a significant decline is observed when insertion errors are considered. When comparing architectures, it is noteworthy that Dilated-LSTM consistently introduces a 5% insertion error rate in contrast to the LSTM. These errors predominantly stem from the inclusion of TRE events in noisy seismic records, and the inclusion of BGN events in previously labeled TRE records. These observed phenomena could be related to attenuation effects involving the degradation in peak-to-peak amplitude of the recorded signals in comparison to those included in the master (training) dataset. Although spectrograms reveal a consistent frequency pattern (indicative of a common source mechanism), the discernible variations in their energy levels point to significant attenuation effects. The reason the expanded version encounters this 5% insertion error may be attributed to its larger receptive field, detecting longer-term changes or transients with fluctuating energy levels along the seismic trace.

TABLE II
GENERALIZATION CAPABILITIES ANALYSIS: RECOGNITION PERFORMANCE OBTAINED BY THE PROPOSED ARCHITECTURES USING DATA BELONGING TO XXX SPANISH ANTARCTIC CAMPAIGN, 2017

	Acc (%)	Cor (%)	Insertion error (%)
RNN-LSTM+ $\delta, \delta\delta$	64.9	78.8	13.9
Dilated-LSTM+ $\delta, \delta\delta$	54.96	72.84	17.88
RNN-LSTM	69.20	77.48	8.28
Dilated-LSTM	53.97	68.87	14.9

Regarding the parameterization scheme, the incorporation of contextual temporal information ($\delta, \delta\delta$) results in an insertion error rate ranging from 4% to 5%. These errors primarily stem from the identification of transient events in traces initially labeled as TRE. The inclusion of $\delta, \delta\delta$ makes the models more responsive to subtle variations in seismic traces, leading to the recognition of events that were initially overlooked or considered part of broader occurrences. As highlighted in [12] and [50], seismic tomographies in velocity and attenuation at Deception Island substantiate the presence of aquifers and hot materials near the surface. Interactions between water and hot rocks (source effects) induce a sudden phase change at depth, accompanied by a pressure step and the emission of high-frequency seismic waves. The intricate fault systems in the area contribute to the generation of low-frequency seismic wave swarms that produce overlapping signals. The acquisition of strongly supervised data, including fully ground-truthed catalogs, is impeded by the considerable costs associated with the data-labeling process. Consequently, inherent limitations, stemming in part from the methodologies used to obtain these catalogs, have significant implications for the testing of future monitoring systems. Therefore, these insertion errors may be associated with source effects arising from the completeness of seismic catalogs, which are notably influenced by human decision-making biases. An expert review may be necessary to discern some of these insertions from a geophysical perspective, and certain insertions may not be considered errors after further examination.

Furthermore, irrespective of the selected parameterization scheme, both architectures exhibit decreased performance in terms of accurately detecting events. This decline can be attributed to the classification of all earthquakes with negative magnitudes or those exhibiting significant attenuation due to propagation or side effects as LPs.

Finally, it is worth noting that while the LSTM architectures, both with and without contextual information, yield quite similar results in terms of events correctly recognized, the dilated version shows a significant discrepancy. This difference becomes more pronounced, reaching 4%, when contextual information is not included. In both scenarios, the drop in Dilated-LSTM performance primarily results from misidentifying sequences of events within the same seismic category. Where the LSTM is capable of separately recognizing each of the events forming the sequence, the Dilated-LSTM, mainly due to its broader receptive field, recognizes consecutive events (smaller than the model's own receptive field) as a single event, leading to a higher number of omitted events and a decrease in performance.

V. Conclusion

In this work, we introduce Dilated-RNNs as a foundational framework for robust and efficient volcano-seismic monitoring systems. Our results demonstrate that the incorporation of multiresolution dilated recurrent skip connections between layers not only enhances efficiency by reducing training parameters but also surpasses classical RNN architectures in sequence modeling tasks involving seismic records. These networks enable hierarchical representations that can be effectively leveraged in dynamic and streaming data environments. Moreover, while their performance with recent seismic campaign data may be influenced by catalog incompleteness, their learning/adaptability capabilities have the potential to outperform traditional approaches.

References

[1] M. Ohrnberger, "Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia," Ph.D. dissertation, Fac. Math. Natural Sci., Potsdam, Univ., Postdam, Germany, 2001.

[2] P. P. Shinde and S. Shah, "A review of machine learning and deep learning applications," in *Proc. 4th Int. Conf. Comput. Commun. Control Automat.*, 2018, pp. 1–6.

[3] J. Palmer, "The new science of volcanoes harnesses AI, satellites and gas sensors to forecast eruptions," *Nature*, vol. 581, no. 7808, pp. 256–260, 2020.

[4] A. Witze, "How AI and satellites could help predict volcanic eruptions," *Nature*, vol. 567, no. 7747, pp. 156–158, 2019.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[6] P. Alasonati, J. Wassermann, M. Ohrnberger, H. Mader, C. Connor, and S. Coles, "Signal classification by wavelet-based hidden Markov models: Application to seismic signals of volcanic origin," *Statist. Volcanol.*, vol. 1, pp. 161–174, 2006.

[7] 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. 2006.

[8] L. Gutiérrez et al., "Volcano-seismic signal detection and classification processing using hidden Markov models. application to San Cristóbal volcano, Nicaragua," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2009, pp. IV-522–IV-525.

[9] 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.*, 2009, pp. II-1012–II-1015.

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

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

[12] 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. 1936–1940, Apr. 2019.

[13] A. Bueno, M. Titos, C. Benítez, and J. M. Ibáñez, "Continuous active learning for seismo-volcanic monitoring," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, Art. no. 7505405, doi: 10.1109/LGRS.2021.3121611.

[14] A. B. Rodríguez et al., "Recurrent scattering network detects metastable behavior in polyphonic seismo-volcanic signals for volcano eruption forecasting," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–23, 2021.

[15] G. Cortés, R. Carniel, P. Lesage, M. Á. Mendoza, and I. D. Lucia, "Practical volcano-independent recognition of seismic events: VULCAN.ears project," *Front. Earth Sci.*, vol. 8, 2021, Art. no. 616676.

[16] C. Hibert, F. Provost, J.-P. 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," *J. Volcanol. Geothermal Res.*, vol. 340, pp. 130–142, 2017.

[17] M. Titos, A. Bueno, L. Garcia, and C. Benitez, "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.

[18] A. Bueno, M. Titos, L. García, I. Álvarez, J. Ibañez, and C. Benítez, "Classification of volcano-seismic signals with Bayesian neural networks," in *Proc. 26th Eur. Signal Process. Conf.*, 2018, pp. 2295–2299.

[19] M. Malfante, M. D. Mura, J. I. Mars, J.-P. Métaxian, O. Macedo, and A. Inza, "Automatic classification of volcano seismic signatures," *J. Geophysical Res.: Solid Earth*, vol. 123, no. 12, pp. 10–645, 2018.

[20] J. P. Canario, R. Mello, M. Curilem, F. Huenupan, and R. Rios, "In-depth comparison of deep artificial neural network architectures on seismic events classification," *J. Volcanol. Geothermal Res.*, vol. 401, 2020, Art. no. 106881.

[21] M. Bicego, A. Rossetto, M. Olivieri, J. M. Londoño-Bonilla, and M. Orozco-Alzate, "Advanced KNN approaches for explainable seismic-volcanic signal classification," *Math. Geosci.*, vol. 55, pp. 59–80, 2022.

[22] M. Titos et al., "Toward knowledge extraction in classification of volcano-seismic events: Visualizing hidden states in recurrent neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2311–2325, 2022.

[23] M. Titos, L. Gutiérrez, C. Benítez, P. R. Devesa, I. Koulakov, and J. M. Ibáñez, "Multi-station volcano tectonic earthquake monitoring based on transfer learning," *Front. Earth Sci.*, vol. 11, 2023, Art. no. 1204832.

[24] G. Cortes et al., "Parallel system architecture (PSA): An efficient approach for automatic recognition of volcano-seismic events," *J. Volcanol. Geothermal Res.*, vol. 271, pp. 1–10, 2014.

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[26] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, 2015.

[27] C. Lea, M. D. Flybbn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1003–1012.

[28] J. Yan, L. Mu, L. Wang, R. Ranjan, and A. Y. Zomaya, "Temporal convolutional networks for the advance prediction of ENSO," *Nature Sci. Rep.*, vol. 10, 2020, Art. no. 8055.

[29] M. Račič, K. Oštir, D. Peressutti, A. Zupanc, and L. Č. Zajc, "Application of temporal convolutional neural network for the classification of crops on sentinel-2 time series," *Int. Arch. Photogrammetry, Remote Sens., Spatial Inf. Sci.*, vol. XLIII-B2-2020, pp. 1337–1342, 2020.

[30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learn.*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[31] S. Chang et al., "Dilated recurrent neural networks," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2017.

[32] C. Lea et al., "Temporal convolutional networks: A unified approach to action segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 47–54.

[33] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2015.
[34] J. Bradbury, S. Merity, C. Xiong, and R. Socher, "Quasi-recurrent neural networks," in *Proc. Conf. Neural Inf. Process. Syst.*, 2017.
[35] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[36] J. Chung, C. Gülçehre, K. H. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. NIPS Workshop Deep Learn.*, 2014.
[37] K. Cho, B. V. Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," in *Proc. 8th Workshop Syntax Semantics Struct. Statist. Transl.*, 2017, pp. 103–111.
[38] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," 2014, *arXiv:1402.1128*.
[39] Q. Chen, X. Zhu, Z. Ling, S. Wei, J. Hui, and D. Inkpen, "Enhanced LSTM for natural language inference," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 103–111.
[40] J. L. Smellie, "Recent observations on the volcanic history of Deception Island, South Shetland islands," *Brit. Antarctic Surv. Bull.*, vol. 81, pp. 83–85, 1988.
[41] E. Carmona, J. Almendros, I. Serrano, D. Stich, and J. M. Ibáñez, "Results of seismic monitoring surveys of Deception Island volcano, Antarctica, from 1999–2011," *Antarctic Sci.*, vol. 24, no. 5, pp. 485–499, 2012.
[42] J. Martí, A. Geyer, and G. Aguirre-Diaz, "Origin and evolution of the Deception Island Caldera (South Shetland Islands, Antarctica)," *Bull. Volcanol.*, vol. 75, pp. 1–18, 2013.
[43] J. M. Ibáñez et al., "Seismovolcanic signals at Deception Island volcano, Antarctica: Wave field analysis and source modeling," *J. Geophysical Res. Solid Earth*, vol. 105, no. B6, pp. 13905–13931, 2000.
[44] S. Zhang et al., "Cambricon-X: An accelerator for sparse neural networks," in *Proc. 49th Annu. IEEE/ACM Int. Symp. Microarchitecture*, 2016, pp. 1–12.
[45] T. Hoefler, D. Alistarh, T. Ben-Nun, N. Dryden, and A. Peste, "Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks," *J. Mach. Learn. Res.*, vol. 22, no. 241, pp. 1–124, 2021.
[46] L. Timpl et al., "Understanding the effect of sparsity on neural networks robustness," in *Proc. ICML Workshop Overparameterization: Pitfalls Opportunities*, 2021.
[47] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: An overview of interpretability of machine learning," in *Proc. IEEE 5th Int. Conf. Data Sci. Adv. Analytics*, 2018, pp. 80–89.
[48] M. Leavitt, "Selectivity considered harmful: Evaluating the causal impact of class sensitivity in DNNs," in *Proc. Int. Conf. Learn. Representations*, 2021.
[49] B. Zhou et al., "Revisiting the importance of individual units in CNNs via ablation," 2018, *arXiv:1806.02891*.
[50] J. Prudencio et al., "The 3D attenuation structure of Deception Island (Antarctica)," *Surv. Geophys.*, vol. 36, pp. 371–390, 2015.

Manuel Titos received the M.Sc. degree in computer engineering, the master's degree in computer and networks engineering, and the Ph.D. degree in information and communication technologies from the University of Granada, Granada, Spain, in 2012, 2013, and 2018, respectively.

He is currently working on developing quantitative probabilistic volcanic hazard assessment methods to evaluate the impact of ash-forming eruptions on air traffic, as well as advancing Artificial Intelligence and signal processing techniques to describe and characterize volcano-seismic signals. His main research interests include deep learning techniques and computational intelligence for remote sensing signals, particularly in short- and long-term volcanic hazard assessments, also focusing on satellite image processing for fire detection and in modeling techniques to predict air-quality conditions.

Joe Carthy received the master's degree in electronic and computer engineering from University College Dublin, Dublin, Ireland, in 2021. He is currently working toward the Ph.D. degree in information and communication technologies with the University of Granada, Granada, Spain.

He is currently an early-stage Researcher with the University of Granada. He is currently a member of a European early training network, IMPROVE, with a role focused on signal processing, and machine learning applied to volcano-seismic signals. These align with his primary research interests of time-series analysis, and deep learning for finding different representations to better understand data.

Luz García received the M.Sc. degree in telecommunication engineering from the Polytechnic University of Madrid, Madrid, Spain, in 2000, and the Ph.D. degree in signals, systems, and radiocommunications from the University of Granada, Granada, Spain, in 2008.

After working as a Support Engineer for communication networks with Ericsson-Spain for five years, she joined a European research project with the University of Granada. Since 2005, she has been with the Department of Signal Theory, Telematics and Communications, University of Granada, having worked first as an Assistant Professor and currently as an Associate Professor. Her research interests include signal processing, pattern recognition, and machine learning in the fields of speech and geophysics.

Talfan Barnie received the undergraduate degree in geological sciences from the University of Cambridge, Cambridge, U.K., the master's degree in remote sensing and geographical information systems from the University of Wales, Aberystwyth, U.K., and the Ph.D. degree in remote sensing of volcanic eruptions from the University of Cambridge.

He is currently a specialist in the remote sensing of volcanoes with Icelandic Meteorological Office, Reykjavík, Iceland. His research interests include remote sensing, Bayesian modeling of volcanic processes, and using neural networks for volcano monitoring and model emulation.

Carmen Benítez received the M.Sc. and Ph.D. degrees in physics from the University of Granada, Granada, Spain, in 1991 and 1998, respectively.

She was a Visiting Researcher with International Computer Science Institute, Berkeley, CA, USA, and United States Geological Survey (USGS), Menlo Park, CA, USA. From 1990 to 2004, she was with the Department of Electronics and Computer Sciences, Faculty of Science, University of Granada, where she has been with the Department of Signal Theory, Telematics and Communications, Escuela Técnica Superior of Computer and Telecommunication Engineering, since 2004. From 2003 to 2018, she was an Associate Professor with the University of Granada and, from 2015 to 2019, the Head of the Department of Signal Theory, Telematics and Communications, where she is currently a Full Professor. Her research interests include signal processing, computational geophysics, speech recognition, machine learning, and pattern recognition.