Journal of Posthumanism
2026
Volume: 6, No: 6, pp. 264–272
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v6i6.4259

Evaluating the Generalization Capacity of Volcanic Seismic Classification via Transfer Learning

John Jehu Meneses Muñoz[1], Oscar Ernesto Cadena Ibarra[2], Manuel Bolañoz Gonzales[3]

Abstract

Effective volcanic monitoring depends on the timely detection and classification of seismic signals; however, the automation of these processes faces the challenge of site specificity, requiring extensive labeled databases for each individual volcano. To address this limitation, this study evaluates the generalization capability of deep learning models in seismic environments not included in the training phase. A transfer learning approach via fine-tuning was employed using EfficientNetB0, ResNet50, and DenseNet121 architectures. The training process utilized data from a single station, while generalization capacity was validated using data from two external volcanoes. The results demonstrate that although EfficientNetB0 possesses superior generalization capability, the levels of precision and specificity obtained at external stations remain insufficient for autonomous implementation. These findings suggest that fine-tuning pre-trained models is a promising strategy; however, it is concluded that models trained under a single-station scheme do not achieve the operational reliability necessary for deployment in contexts other than their origin.

Keywords: *Volcanic monitoring, Seismic signal classification, Transfer Learning, Fine-tuning, EfficientNetB0, Model generalization.*

Introduction

Volcanic activity monitoring is a fundamental pillar for risk mitigation and the protection of populations living near active volcanic systems. Among the various monitoring techniques, volcanic seismology has become one of the most useful tools for eruption forecasting and monitoring [1]. The timely detection and classification of events—such as Volcano-Tectonic (VT) earthquakes, Long-Period (LP) events, and volcanic tremors—allow observatories to issue early warnings and understand the state of volcanic unrest.

Traditionally, the analysis of these signals has relied on the manual labor of expert seismologists. However, the exponential increase in data generated by modern high-density monitoring networks has made manual processing unsustainable, leading to significant delays in information analysis. To address this, the scientific community has turned toward automated systems based on Machine Learning [2], [3], [4], [5] and, more recently, Deep Learning [6]. Convolutional Neural Networks (CNNs) [7], [8], [9], [10] have shown remarkable success in seismic signal processing due to their ability to automatically extract complex features from spectrograms or

[1] Ingeniero de Sistemas, Mg(c) Maestría en Ingeniería de Sistemas y Computación, Universidad de Nariño, Pasto, Colombia. jjmenesesm@gmail.com , https://orcid.org/ 0009-0006-4040-4028
[2] Físico, PhD en Geociencias, Servicio Geológico Colombiano, Pasto, Colombia. ocadena@sgc.gov.co , https://orcid.org/ 0000-0002-4351-2263
[3] Ingeniero de sistemas, PhD en Ciencias de la electrónica, Universidad de Nariño, Pasto, Colombia, mbolanos@udenar.edu.co , https://orcid.org/ 0000-0002-3077-415X

raw waveforms.

Despite these advances, site specificity remains a critical barrier. Most deep learning models are developed using datasets from a single volcano or station. However, due to variations in local geology, sensor coupling, and background noise, these models often exhibit limited generalization capacity, restricting their efficacy to the original training environment. Developing models for every station is practically unfeasible, as many observatories lack both exhaustive labeled catalogs and the computational infrastructure required for large-scale training.

This research addresses this limitation by investigating the generalization capability of deep learning models, their ability to maintain high performance on data from volcanic environments not seen during training. We employ a transfer learning [11] strategy combined with fine-tuning [12], leveraging architectures originally designed for large-scale image recognition, such as EfficientNetB0[13], ResNet50[14], and DenseNet121[15]. By adapting these pre-trained models to seismic-volcanic data, we aim to determine which architecture provides the most robust performance across different contexts.

The results of this study contribute to the development of more versatile monitoring tools, demonstrating that the EfficientNetB0 architecture yielded the highest generalization capacity. These findings are promising for the development of models capable of operating in contexts that lack the labeled data or resources necessary for site-specific training from scratch.

Data

This study focuses on the classification of two types of seismic signals: volcano-tectonic (VT) and long-period (LP) events. These categories allow for the differentiation of distinct physical processes within the volcanic edifice: whereas VT earthquakes are associated with rock fracturing, LP events are linked to fluid dynamics (magma, gases, or water) [16].

Data for this study were obtained from two primary sources: the Colombian Geological Survey (SGC), via the Pasto Volcanological and Seismological Observatory [17], and the ESeismic repository [18]. Seismic data from Galeras and from Chiles-Cerro Negro volcanic complex, were provided by the SGC. These records consist of signals from broadband velocity sensors with a sampling rate of 100 samples per second (100 Hz). From the ESeismic repository, the MicSigV1 dataset corresponding to the Cotopaxi volcano was used. The length of the data segments used is variable, depending on the specific duration of each seismic event. To rigorously evaluate the models' generalization capability, data from the Galeras and Cotopaxi volcanoes were entirely excluded from the training phase and reserved exclusively for external validation testing.

Table 1 details the distribution of the dataset by station, providing a breakdown of the samples used during the training, local validation, and testing phases.

Table1. distribution of the dataset.

Volcano	Station	Type	Total	Training	Validation	Testing
CHILES	CHIP	VT	12000	7680	1920	2400

CHILES	CHIP	LP	12000	7680	1920	2400
GALERAS	ANGP	VT	1087	0	0	1807
GALERAS	ANGP	LP	1087	0	0	1807
COTOPAXI	VC1	VT	1825	0	0	1825
COTOPAXI	VC1	LP	2164	0	0	2164

Methodology

The methodology was developed in three main phases: data preparation, model configuration, and robustness evaluation.

For data preprocessing and representation, seismic time-series signals were transformed using the spectrogram technique. Given that the selected CNN architectures are optimized for image processing, this conversion is ideal, as the spectrogram enables the visualization of the dynamic evolution of frequency content over time. This representation is crucial for capturing the distinctive features that differentiate VT from LP events.

Regarding model configuration, three benchmark architectures were evaluated: ResNet50, EfficientNetB0, and DenseNet121, all pre-trained on the ImageNet dataset [19]. The training process followed a transfer learning strategy structured into two experimental modalities:

- Feature Extraction: The model keeps all layers of the base architecture frozen, optimizing only the weights of the added classification layer (Top Layer) tailored to the seismic problem.
- Fine-tuning: Experiments were conducted by selectively unfreezing the last 1, 2, 5, 10, 15, 20, 25, and 30 layers of each architecture. This allows the high-level filters of the networks to specialize in detecting morphological patterns specific to seismic signals.

The selection of the ResNet50, EfficientNetB0, and DenseNet121 architectures stems from the need to evaluate different approaches to connectivity and efficiency in seismic signal processing. ResNet50 was chosen for its ability to mitigate the vanishing gradient problem through residual connections, facilitating the training of deep networks [14]. DenseNet121 was included due to its dense connectivity, which promotes feature reuse and reduces the number of parameters compared to traditional architectures [15]. Finally, EfficientNetB0 was integrated as it represents the state-of-the-art in computational efficiency, utilizing uniform scaling of depth, width, and resolution to optimize performance in monitoring environments with limited resources [13].

The robustness evaluation of the system was structured into three critical scenarios to ensure the validity of both detection and classification. Training was conducted exclusively using data from the CHIP station. Subsequently, an internal test was performed with unseen data from the same station to establish a performance baseline. Finally, a generalization test (external validation) was executed using a cross-volcano scheme, incorporating data from the Galeras (ANGP station) and Cotopaxi (ESeismic repository) volcanoes. This final scenario is essential for measuring the model's robustness against geological environments and noise levels distinct from those in the training phase, thereby assessing its applicability in unseen seismic contexts. Across all three test scenarios, performance was quantified using accuracy, precision, recall, and the F1-score. These metrics enabled a comparative evaluation of each architecture's performance and served to determine the impact of progressive fine-tuning on the models' generalization capacity.

Results and discussion

The evaluation on the local dataset (figure 1), comprising data from the CHIP station not seen during training, yielded exceptional results across all architectures. For ResNet50, the baseline Transfer Learning (TL) model already achieved an accuracy of 0.9996, reaching a perfect score (1.0000) in all metrics after unfreezing 10, 20, and 25 layers. EfficientNetB0 followed a similar trend, showing its peak performance with perfect metrics at 2, 5, and 20 unfrozen layers. DenseNet121 demonstrated the most consistent improvement through progressive fine-tuning; while its initial performance (FT_1) was slightly lower than the others (Accuracy: 0.9981), it achieved and maintained perfect performance (1.0000) consistently from 10 unfrozen layers onwards.

These results indicate that for the local station, the features learned during the training phase are highly representative, allowing the models to classify events with almost no margin of error. The high Precision and Recall values (frequently 1.0000) confirm that the models are equally robust at avoiding false positives and classifying all seismic events within their own volcanic environment.

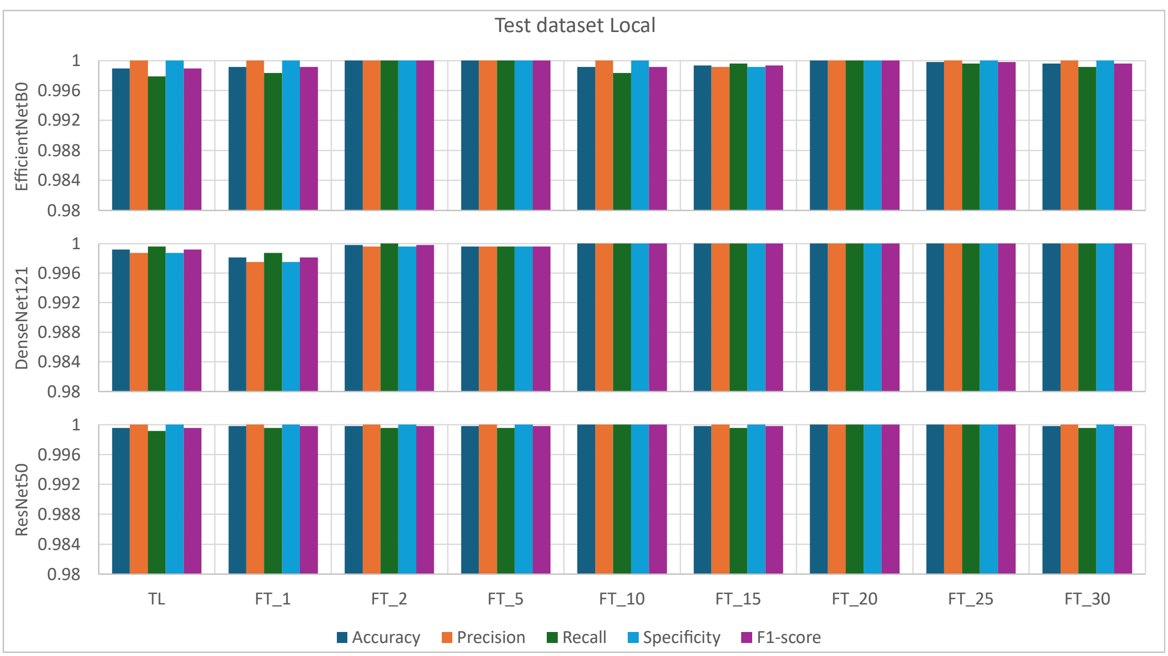

Figure 1: Results in the local test environment (CHIP station).

The evaluation of models on the ANGP station dataset (figure 2) reveals a significant drop in performance compared to the local test, highlighting the challenges of cross-volcano generalization. While all models maintained high recall values, frequently exceeding 0.99, there was a marked increase in false positives (FP), which drastically reduced the specificity and precision across all architectures. This suggests that while the models are highly capable of detecting seismic events (low False Negatives), they tend to over-classify background noise or local site effects as seismic signals when moved to a new volcanic environment.

Among the evaluated architectures, EfficientNetB0 demonstrated the most robust generalization capability; however, its performance reveals a critical limitation in class discrimination. While the EfficientNetB0 FT_25 model achieved the highest Accuracy (0.736) and F1-score (0.803), these results are offset by a poor specificity of 0.435, this low specificity indicates a significant failure in correctly identifying LP events, which are frequently misclassified as VT. In contrast, ResNet50 and DenseNet121 exhibited significant difficulties regarding site-specificity.

ResNet50 stayed within an Accuracy range of 0.55 to 0.59, regardless of the level of fine-tuning. DenseNet121 showed a concerning trend in several fine-tuning stages (FT_10 through FT_25), where specificity dropped to near zero, indicating that the model began classifying almost all incoming data as positive events. These results confirm that EfficientNetB0, through progressive fine-tuning, is the most suitable architecture for adapting to new volcanic contexts where labeled data is unavailable.

Figure 2. Results for the ANGP test set.

The results obtained with the ESeismic (MicSigV1) dataset expose the generalization constraints of the evaluated architectures in volcanic environments different from those used during training. While EfficientNetB0 stands out as the architecture with the highest relative generalization capacity, its performance does not reach the standards required for real-world operational applications. In its optimal configuration (FT_2), the model achieves an F1-score of 0.602 and a precision of merely 0.439; these values reveal that, despite an outstanding sensitivity (Recall) of 0.960, the network is unable to rigorously discriminate against the positive class.

Performance further degrades when unfreezing intermediate layers (FT_5 to FT_20), resulting in a collapse of the F1-score from 0.602 to 0.322. Rather than abstracting universal seismic patterns, the model overfits to the characteristics of the source station (CHIP). Consequently, the superiority of the baseline model (TL) over deeper fine-tuning versions should not be interpreted as a definitive solution, but rather as evidence that weight adjustment degrades the robustness of generic feature abstraction, leading to poor performance in unseen environments. Ultimately, although EfficientNetB0 demonstrates superior generalization compared to ResNet50 and DenseNet121, its implementation in autonomous monitoring systems remains hindered by its low selectivity.

Conversely, ResNet50 and DenseNet121 demonstrated a notable lack of adaptability to the ESeismic dataset. ResNet50 consistently underperformed, with accuracy levels stagnating below 0.380 regardless of the fine-tuning configuration, while DenseNet121 exhibited extreme instability, characterized by a total collapse in specificity beyond the FT_10 stage. This decline to zero indicates that DenseNet121 ceased to function as a discriminator, instead defaulting to a majority-class prediction. These findings consolidate the conclusion that EfficientNetB0 possesses the most robust capacity for extracting transferable features, maintaining its structural integrity where deeper or more densely connected architectures succumb to site-specific noise.

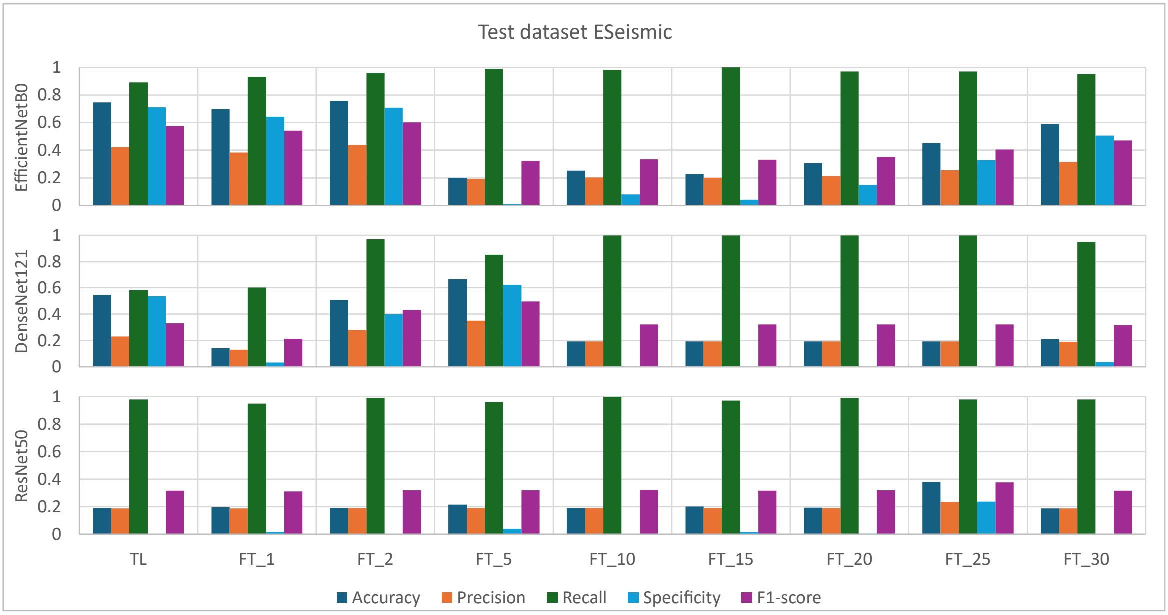

Figure 3. Results in the ESeismic test environment.

The superiority of EfficientNetB0 in generalization tasks across the ANGP and ESeismic datasets is rooted in its architectural optimization. Unlike ResNet50 and DenseNet121, whose depth requires a considerable volume of parameters, EfficientNetB0 implements compound scaling that harmonizes depth, width, and resolution. This structural efficiency not only reduces the computational load but also results in a contracted parametric space during the fine-tuning process.

In the context of volcanic seismology, characterized by a scarcity of labeled catalogs, this parametric simplicity acts as an intrinsic regularizer. This mitigates the risk of the model over-specializing in ambient noise or the site-specific particularities of the CHIP station. Consequently, EfficientNetB0 succeeds in extracting 'universal' seismic features, facilitating a more robust knowledge transfer toward external volcanic systems.

Furthermore, the fact that a lower degree of layer unfreezing favored generalization reinforces the hypothesis that a conservative parametric adjustment preserves the essential morphology of the seismic source, filtering out local particularities that lack transferability.

Conclusions

EfficientNetB0 established itself as the most robust architecture for automated volcanic monitoring among the evaluated models. Unlike ResNet50 and DenseNet121, whose performance collapsed in unseen environments, this architecture demonstrated a superior ability to retain generalizable spectral features, achieving the highest accuracy levels in both regional (ANGP: 0.736) and external (ESeismic: 0.757) scenarios. However, the results also expose the inherent vulnerabilities of knowledge transfer in this domain. Despite its relative robustness, the model's low selectivity (evidenced by a precision of 0.439) and its susceptibility to overfitting during deep fine-tuning underscore that structural efficiency alone does not guarantee full operational autonomy. These findings validate EfficientNetB0 as the most promising baseline for future developments, suggesting that its success hinges on conservative parametric adjustment and increased data diversity to mitigate source-station bias.

The experiments demonstrate that site-specificity constitutes the primary barrier to the universal implementation of Deep Learning models for seismic classification. While the models exhibited outstanding sensitivity within the training environment, the ability to discriminate between VT and LP events was compromised in unseen environments. This suggests that the models tend to overfit to the specific characteristics of the training station.

Future Work

As future lines of research, the implementation of multi-volcano and multi-station training is proposed. Integrating data from diverse volcanic structures would allow the models to converge toward universal seismological features, effectively mitigating the bias induced by site effects.

Additionally, the use of advanced data augmentation techniques, specifically through synthetic noise injection and phase variations, is recommended. This approach aims to simulate greater heterogeneity in signal reception conditions, forcing the network to extract intrinsic source characteristics. Such strategies are essential to overcome the current precision limitations and move toward monitoring systems with genuine autonomous deployment capabilities in unseen volcanic environments.

References

C. Scarpati, "The encyclopedia of volcanoes (second edition)," *Geologos*, vol. 25, no. 2, 2019, doi: 10.2478/logos-2019-0018.

A. Falcin *et al.*, "A machine-learning approach for automatic classification of volcanic seismicity at La Soufrière Volcano, Guadeloupe," *Journal of Volcanology and Geothermal Research*, vol. 411, 2021, doi: 10.1016/j.jvolgeores.2020.107151.

N. Trujillo-Castrillón, C. M. Valdés-González, R. Arámbula-Mendoza, and C. C. Santacoloma-Salguero, "Initial processing of volcanic seismic signals using Hidden Markov Models: Nevado del Huila, Colombia," *Journal of Volcanology and Geothermal Research*, vol. 364, 2018, doi: 10.1016/j.jvolgeores.2018.09.008.

R. Soto, F. Huenupan, P. Meza, M. Curilem, and L. Franco, "Spectro-temporal features applied to the automatic classification of volcanic seismic events," *Journal of Volcanology and Geothermal Research*, vol. 358, 2018, doi: 10.1016/j.jvolgeores.2018.04.025.

O. Ernesto and C. Ibarra, "Detección y clasificación automática de registros sísmicos en el Observatorio Vulcanológico y Sismológico de Pasto utilizando redes neuronales artificiales," *Universidad Nacional de Colombia Facultad de Geociencias, postgrado Geofísica Bogotá, Colombia*, 2011.

V. L. Martínez *et al.*, "Advanced signal recognition methods applied to seismo-volcanic events from Planchon Peteroa Volcanic Complex: Deep Neural Network classifier," *J. South Am. Earth Sci.*, vol. 107, 2021, doi: 10.1016/j.jsames.2020.103115.

F. Lara, R. Lara-Cueva, J. C. Larco, E. V. Carrera, and R. León, "A deep learning approach for automatic recognition of seismo-volcanic events at the Cotopaxi volcano," *Journal of Volcanology and Geothermal Research*, vol. 409, 2021, doi: 10.1016/j.jvolgeores.2020.107142.

G. F. Manley *et al.*, "A Deep Active Learning Approach to the Automatic Classification of Volcano-Seismic Events," *Front. Earth Sci. (Lausanne).*, vol. 10, 2022, doi: 10.3389/feart.2022.807926.

M. Nakano and D. Sugiyama, "Discriminating seismic events using 1D and 2D CNNs: applications to volcanic and tectonic datasets," *Earth, Planets and Space*, vol. 74, no. 1, 2022, doi: 10.1186/s40623-022-01696-1.

D. Rincon-Yanez, E. De Lauro, S. Petrosino, S. Senatore, and M. Falanga, "Identifying the Fingerprint of a Volcano in the Background Seismic Noise from Machine Learning-Based Approach," *Applied Sciences (Switzerland)*, vol. 12, no. 14, 2022, doi: 10.3390/app12146835.

F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," 2021. doi: 10.1109/JPROC.2020.3004555.

J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems*, 2014.

M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *36th International Conference on Machine Learning, ICML 2019*, 2019.

K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in

Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016. doi: 10.1109/CVPR.2016.90.

G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017. doi: 10.1109/CVPR.2017.243.

D. Gómez *et al.*, *Catálogo de señales sísmicas volcánicas de Colombia*, 1st ed., vol. 1. Pasto: INGEOMINAS, 2004.

D. M. G. Martínez, C. M. López Vélez, M. L. M. Bustamante, A. A. Restrepo, G. P. C. Jiménez, and M. L. C. Velasco, “Active volcanism in Colombia and the role of the Servicio Geológico Colombiano,” *Volcanica*, vol. 4, 2021, doi: 10.30909/vol.04.S1.113139.

N. Pérez, D. Benítez, F. Grijalva, R. Lara-Cueva, M. Ruiz, and J. Aguilar, “ESeismic: Towards an Ecuadorian volcano seismic repository,” *Journal of Volcanology and Geothermal Research*, vol. 396, 2020, doi: 10.1016/j.jvolgeores.2020.106855.

J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” 2010. doi: 10.1109/cvpr.2009.5206848.

.