

Evaluating data augmentation techniques in the classification of multi-station seismic-volcanic signals

Franz Yupanqui Machaca

frayup@gmail.com

Universidad Católica del Norte

Pablo Salazar Reinoso

Universidad Católica del Norte

Claudio Meneses Villegas

Universidad Católica del Norte

Article

Keywords: Deep Learning, Transfer Learning, Data Augmentation, Machine Learning classification, Seismic-volcanic signals, Multi-Scale Structural Similarity (MS-SSIM)

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11155964/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Evaluating data augmentation techniques in the classification of multi-station seismic-volcanic signals

Franz Yupanqui Machaca^{1,2,5*}, Pablo Salazar Reinoso^{2,3†} and
Claudio Meneses Villegas^{4†}

¹*Doctoral Program in Sustainable Engineering, Universidad Católica
del Norte, Avenida Angamos 0610, Antofagasta, 1270709, Chile.

²Millennium Institute on Volcanic Risk Research - Ckelar Volcanoes,
Avenida Angamos 0610, Antofagasta, 1270709, Chile.

³Department of Geological Sciences, Universidad Católica del Norte,
Avenida Angamos 0610, Antofagasta, 1270709, Chile.

⁴Department of Computing and Systems Engineering, Universidad
Católica del Norte, Avenida Angamos 0610, Antofagasta, 1270709, Chile.

⁵4D Perceptio, c. Los Cedros No.33 Zona Aranjuez, La Paz, Bolivia.

*Corresponding author(s). E-mail(s): frayup@gmail.com;

Contributing authors: pasalaz@ucn.cl; cmeneses@ucn.cl;

[†]These authors contributed equally to this work.

Abstract

Effective volcano monitoring requires rapid identification of seismic events to understand activity levels and provide timely warnings. Traditional methods rely on manual analysis, which may introduce bias and is labor-intensive. This work aims to evaluate data augmentation techniques on the performance of models classifying seismic-volcanic signals by measuring similarity and diversity. Data augmentation aims to balance the dataset and enhance diversity while keeping similarity to the original data low. Techniques are tested on 6,145 seismic events at Lascar Volcano. The datasets are divided into two types: one with real and augmented data and one with only augmented data for performance comparison. The use of similarity and diversity metrics on augmented volcanic seismic data helps link them to model performance. Several datasets containing only augmented data offer superior performance in classification models, such as Drifting, SpecAugment, and Jittering, with test accuracies of 97.2%, 95.4%,

and 93.9%, respectively. Using data augmentation techniques on volcanic seismic signals helps create more generalizable models that can be applied to different volcanoes. The contributions of this work are important for hazard and risk assessment through monitoring the dynamic evolution of volcanic centers, which is essential for understanding the phases of the volcanic eruption cycle.

Keywords: Deep Learning, Transfer Learning, Data Augmentation, Machine Learning classification, Seismic-volcanic signals, Multi-Scale Structural Similarity (MS-SSIM)

1 Introduction

1.1 Problem of recognition

Volcanic and seismic events can be hazardous and may occur unexpectedly, although they are often preceded by detectable signals. These signals are continuously recorded worldwide using seismometers and subsequently classified by expert geophysicists. Seismic-volcanic activity has been studied for many years. It is currently recorded by sensors that capture three-component signals. Seismic-volcanic pattern recognition is typically performed manually, although state-of-the-art studies have implemented semi-automatic or automatic approaches [1–6]. These studies commonly consider one of the three available channels, one of several stations, and a single volcano. In this context, classical machine learning (ML) and deep learning (DL) algorithms for the automatic classification of seismic-volcanic signatures constitute an emerging research field that can address the limitations of manual classification. This study proposes and applies a series of data augmentation (DA) techniques to volcanic seismic events to increase diversity in the event feature space and analyzes their impact on event-classification performance. It considers a database of events collected from Lascar Volcano over an 8-month period and labeled into five classes by CKELAR-VOLCANES experts; see also [1]. Given the nature of the seismic phenomena occurring in and around the volcano, the database is imbalanced and requires a balancing strategy. DA techniques are applied to generate two types of datasets: balanced datasets containing real and augmented data, and balanced datasets composed entirely of augmented data. Similarity and diversity metrics are then applied to characterize the generated data. Classification models are trained using DL and transfer learning (TL) and evaluated using performance metrics.

Each seismic event is represented by a spectrogram. These images are processed using a transfer-learning strategy based on the pretrained AlexNet convolutional neural network. Event-classification models are trained on datasets generated using different DA techniques, and their performance is evaluated.

The results of the experiments are presented in terms of performance metrics, comparing the impact of each data augmentation method on the classification of seismic-volcanic events. The evaluation includes models trained on real and augmented data, as well as models trained on entirely augmented datasets. The influence of similarity and diversity metrics on model generalization is also analyzed.

Several models trained on augmented-only datasets outperform those trained on datasets containing both real and augmented data while also exhibiting greater data diversity. In particular, Drifting, SpecAugment, and Jittering achieve test accuracies of 97.2%, 95.4%, and 93.9%, respectively. Applying various DA techniques to multi-station volcanic seismic signals produces more generalizable classification models.

The main contributions of this study are as follows:

1. A comparative analysis of data augmentation techniques for seismic signals.
2. The introduction of similarity and diversity metrics for evaluating augmented seismic signals.
3. A method for improving model generalization by increasing data diversity while preserving relevant similarities to the original data.

The remainder of this paper is organized as follows. The most recent bibliography on this subject is presented in Section 2. Section 3 presents the methodology. The methods used are described in Section 4. The main results are presented in Section 5. Section 6 discusses the results of this study, and conclusions are presented in Section 7.

2 Literature review

2.1 Summary of methods for classifying volcanic seismic events

Seismic-volcanic signals are nonstationary time series that contain information about physical processes occurring within volcanic systems. Magma movement, gas emissions, and rock fracturing are among the sources that generate these signals. Analyzing these signals is essential for understanding volcanic activity and assessing eruption risk. In the analysis and treatment of seismic-volcanic time signals, transformations are usually applied to represent time-domain observations in the frequency and cepstral domains, as well as using statistical and information theory metrics, among others [3, 6]. Among the methods used for the analysis of this type of signal, the most common are the Hidden Markov model (HMM) [4, 7, 8] and its variants, such as the Hidden Semi-Markov Models (HSMM), the Continuous Wavelet Transform (CWT) [7] in the detection of singularities, Gaussian Mixture Model (GMM) [9], and others, such as the HMM-based generative embedding [4].

Several studies have proposed automatic classification of seismic-volcanic signals using supervised Machine Learning (ML) methods such as Support Vector Machines (SVM) [3, 5, 6, 9–12]; Multilayer Perceptron (MLP) [6, 9, 10]; Random Forest [3, 6, 13]; Linear Discriminant Analysis [6]; Convolutional Neural Networks (CNN) [2, 13, 14]; Temporal Convolutional Neural Networks (TCN) [13, 15]; Bayesian Neural Networks [16]; Tensor Learning Framework with Multilinear Principal Component Analysis (MPCA) and Support Tensor Machine (STM) [17] techniques on multichannel and multistation data; Scalable Variational Inference for Gaussian Process (SVGP) and Deep Gaussian Processes (DGP) [18]; and Bag of Words in seismic document classification as a set of seismic words [19].

139 Unsupervised methods include Kohonen Self-Organizing Map (SOM) [9, 20],
140 Cluster Analysis (CA) [9], and Balanced Iterative Reducing and Clustering using
141 Hierarchies (BIRCH) [21].

142 Transfer Learning (TL) is applied in the classification of volcanic seismic events
143 using CNNs [13, 22].

144 Deep Learning (DL) has been applied to seismic-volcanic signal classification,
145 experimenting with different CNN [22] and Recurrent Neural Network (RNN)
146 architectures [2, 23, 24], and through Deep Gaussian Processes (DGP) [18].

147 RNNs provide real-time tracking capabilities for volcanic activity, even if the
148 seismic sources change over time. They have been applied to seismic-volcanic signal
149 classification and tested using several architectures [2]. In detection and classification,
150 the following RNN architectures have been applied: vanilla-RNN, Long Short-Term
151 Memory (LSTM), and Gated-Recurrent Unit (GRU) [25]; in [15], vanilla-RNN and
152 LSTM were used, in addition to identifying the degree of specialization of the neurons.

153 An Autoencoder (AE) is an unsupervised learning algorithm. The AE is
154 considered a nonlinear compression structure that reduces signal size and
155 dimensionality into a lower-dimensional latent space. In [22], various AE
156 architectures generate encoded and reconstructed seismic events, which are
157 subsequently classified using different DL and TL model architectures. In [15], AEs
158 are applied to obtain the segmentation mask of a seismic event.

159 DA involves generating new valid data from existing data following certain rules
160 and restrictions. Amplitude deformation, frequency deformation, noise addition, and
161 horizontal shifts are applied to the seismic signals [14]. In [1] the importance of DA in
162 avoiding biased models in the classification of seismic-volcanic events is demonstrated.

163 With regard to volcanic seismic data, most studies consider a single station
164 channel for the classification of volcanic seismic events, but other studies consider
165 several channels and stations [5, 6, 23, 26]. However, [1, 13, 23] showed that using a
166 multi-station scheme is more accurate than a single-station approach. This
167 state-of-the-art review in seismic-volcanic signal classification is summarized in
168 Supplementary Table ??.

169

170 **3 Methodology**

171

172 Currently, Machine Learning is used for the recognition of seismic-volcanic signal
173 patterns. In fact, previous research developed [1, 22] models that implemented a new
174 methodology for classifying these types of seismic signals using Deep Learning,
175 Transfer Learning, and spectrograms.

176 Current research on recognizing seismic-volcanic signal patterns considers data
177 from a single Chilean volcano. The [Millennium Institute on Volcanic Risk Research –](#)
178 [Ckelar Volcanoes](#) monitors many of these volcanoes using seismic stations equipped
179 with triaxial sensors.

180

181

182

183

184

3.1 Lascar volcano

Lascar Volcano is located in northern Chile (23°22' S, 67°44' W; 5592 m a.s.l.), 270 km northeast of Antofagasta and 70 km southeast of San Pedro de Atacama. CKELAR-VOLCANES monitors the area using a network of 11 three-component seismic stations. Short-period 2-Hz seismometers continuously record data at a sampling rate of 200 Hz. The signal database was constructed using only the Z component. The stations are distributed around the volcano, as shown in Supplementary Fig. ??.

3.2 Lascar's database

The Lascar database contains 6,145 seismic events classified by expert geophysicists from CKELAR-VOLCANES. The data were recorded between March and October 2018. Supplementary Tables ?? and ?? show the data distribution. The event classes are long-period (LP), volcano-tectonic (VT), hybrid (HY), tectonic (TC), and tremor (TR); these classes are described in Supplementary Section ??.

The Lascar database comprises five event types (see Supplementary Fig. ??): 213 HY events, 577 LP events, 471 TR events, 2,686 VT events, and 2,198 events unrelated to volcanic activity and classified as TC (Supplementary Tables ?? and ??).

4 Methods

A transfer-learning solution is proposed to accelerate training and improve deep-learning model performance [13, 16], particularly given the limited number of examples available for training a model from scratch. The pretrained AlexNet model [27] is selected because it provides shorter training times and performance comparable to that of other evaluated models, including GoogLeNet, ResNet-152, and VGG-19bn [22]. The input data are spectrograms generated from the labeled waveforms for each seismic-event class. The processing sequence consists of five steps:

1. preprocessing the seismic-event time series;
2. generating spectrograms from the labeled events;
3. applying data augmentation techniques to expand the datasets;
4. constructing the prediction model; and
5. evaluating model performance and data similarity.

In the following paragraphs, each step is explained.

1. Preprocessing: The seismometers record data at sampling rates of 100, 200, or 500 Hz. Each event is resampled to 100 Hz, mean-centered, filtered using a 1-Hz Butterworth high-pass filter followed by a 1–20-Hz Butterworth band-pass filter, and normalized. This preprocessing is performed using the ObsPy Python toolbox [28].
2. The spectrograms are calculated by applying the short-time Fourier transform defined as

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

where $x(t)$ and $w(t)$ represent the seismic signal and the short-time Fourier transform window, respectively. The sliding-window length is set to 1 s with 95% overlap. The generated spectrograms are RGB images of 224×224 pixels.

3. Data augmentation techniques are used to address class imbalance and reduce model bias (Supplementary Table ??).
4. The pretrained AlexNet model [27] is fine-tuned on the spectrograms using transfer learning.
5. Model performance is evaluated using the metrics described in the following sections. The metrics are implemented with the [TorchMetrics](#) Python library.

4.1 Data similarity

Kullback–Leibler (KL) divergence and Jensen–Shannon (JS) divergence are used to measure similarity between synthetic and real datasets; these metrics are described in Supplementary Section ??.

4.2 Data augmentation

Several data augmentation techniques for volcanic seismic data are presented below; their algorithms are provided in Supplementary Section ??.

4.2.1 Rotation

This transformation considers the rotation of the spectrogram around the time axis within a given percentage range of the signal length and at a given axis position. The axis position can be at the beginning, middle, or end of a spectrogram. When applying this technique, the combination of the percentage rotation value, location of the rotation axis, and spectrogram of the event to which it will be applied is obtained. When generating new instances, care is taken to ensure that these three values are not repeated, thus ensuring unique instances in the dataset (see Supplementary Fig. ??).

4.2.2 Jittering

Jittering is a data augmentation technique commonly employed in time series analysis [29, 30]. It is considered one of the simplest yet most effective transformation-based methods, as it involves adding noise to a time series.

Formally, an original time series $x = x_1, \dots, x_t, \dots, x_T$ is transformed into $x' = x_1 + \epsilon_1, \dots, x_t + \epsilon_t, \dots, x_T + \epsilon_T$ by introducing noise at each time step t , where ϵ typically represents Gaussian noise distributed as $N(\theta, \sigma^2)$. The standard deviation σ of the added noise is a hyperparameter that must be defined in advance [29]. This approach assumes that the time series patterns of a dataset are inherently noisy. Such cases may include data derived from sensors, audio signals, or electroencephalograms (EEG). For instance, jittering has been applied to wearable sensor data for Parkinson’s disease monitoring (see Supplementary Fig. ??).

4.2.3 Drifting

One challenge in both online and offline settings is data drift, defined as a change in the data distribution over time. A random walk is often used to simulate such drift. In mathematics, a random walk is classified as a stochastic process that represents a path composed of successive random steps in a mathematical space, such as the integers or the reals. A simple example is the integer number line $\mathbb{Z}$, where the walk begins at θ and moves either $+1$ or -1 with equal probability at each step. More generally, a random move of $+a$ or $-a$ with equal probability can occur in any vector space, such as the real space, where $a \in \mathbb{R}$ denotes the step magnitude [31]. Let d define a path starting at position $d_1 = \phi(\tau)$. A random walk can then be modeled as:

$$d_t = d_{t-1} + \phi(\tau) \quad (2)$$

where ϕ is the random variable describing the probability distribution for the next step, and τ is the time interval between consecutive steps. Consequently, an original time series $x = x_1, \dots, x_t, \dots, x_T$ can be transformed into $x' = x_1 + d_1, \dots, x_t + d_t, \dots, x_T + d_T$. Data drift simulation can be used to generate more robust training datasets, thereby enhancing model resilience and reducing sensitivity to drift when it arises in real-world data (see Supplementary Fig. ??).

4.2.4 Genetic Algorithms

Genetic Algorithms (GA) belong to the broader class of Evolutionary Algorithms (EA) and are inspired by the principle of natural selection. Their primary objective is to generate high-quality solutions to complex problems using bio-inspired operators such as selection, crossover, and mutation [32]. In [33], GAs have been applied to data augmentation of time series derived from centers of pressure for the detection of neuropathies. Building on GA operators, a seismogram data augmentation technique is developed with the following procedure:

- *Selection*: Real seismograms of the same class are selected, without repetition, to serve as parents.
- *Crossover*: Genetic information from two or more parent seismograms is combined to generate a new instance. This is achieved by exchanging time series segments between the parents' chromosomes.
- *Mutation*: Small modifications are introduced in a limited portion of the population. In this case, a smoothing operation is applied to the scales of the exchanged segments.
- *Validation*: KL and JS divergences are used to measure similarity between datasets. Values closer to zero for both KL and JS divergences indicate greater similarity in the frequency-domain power distribution between the original and augmented seismic signals.

Two strategies are considered for the crossover operation:

- *Without time adjustment:* The segments exchanged between parents may differ in length, resulting in a child seismogram with a duration that can vary significantly from that of the original seismograms (see Supplementary Fig. ??).
- *With time adjustment:* The exchanged segments are rescaled to equal length by applying a time-scaling transformation, ensuring that the generated seismogram maintains a duration comparable to the original signals (see Supplementary Fig. ??).

4.2.5 SpecAugment

SpecAugment is a data augmentation technique designed for Automatic Speech Recognition (ASR). Its primary purpose is to help the network learn useful features that are robust to time warping, the partial loss of frequency information, and the partial loss of small signal segments. It is a simple and computationally inexpensive method, as it acts directly on the log-mel spectrogram as if it were an image, allowing for online application during training [34]. SpecAugment policies consist of three main components: frequency masking, time masking, and time warping [35, 36]. In this work, SpecAugment is applied to the spectrograms using frequency masking, time masking, and time warping policies (see Supplementary Fig. ??).

4.2.6 Interpolation AEs

Autoencoders can interpolate in certain situations by decoding the convex combination of the latent codes of two data points, thereby producing an output that semantically blends the features of the data points [37]. First, an input $x \in \mathbb{R}^{d_x}$ is passed through an encoder $z = f_\theta(x)$, parameterized by θ , to obtain a latent code $z \in \mathbb{R}^{d_z}$. The latent code is then passed through a decoder $\hat{x} = g_\phi(z)$, parameterized by ϕ , to produce an approximate reconstruction $\hat{x} \in \mathbb{R}^{d_x}$ of the input x . The case in which both f_θ and g_ϕ are implemented as multilayer neural networks is considered. The encoder and decoder are trained jointly with respect to θ and ϕ to minimize a distance measure between the input x and the output $\hat{x}$, such as the squared L_2 distance $|x - \hat{x}|^2$. Interpolation with an autoencoder refers to the process of using the decoder g_ϕ to reconstruct a mixture of two latent codes. Typically, the latent codes are combined via a convex combination, so that interpolation amounts to computing $x_\alpha = g_\phi(\alpha z_1 + (1 - \alpha)z_2)$ for some $\alpha \in [0, 1]$, where $z_1 = f_\theta(x_1)$ and $z_2 = f_\theta(x_2)$ are the latent codes corresponding to the data points x_1 and x_2 (see Supplementary Fig. ??).

4.3 Model performance - Metrics

Machine learning metrics make it possible to quantify the performance of a trained model and to compare it with other generated models or with the current baseline. Evaluating the learning algorithm is a crucial part of any project. While classification accuracy is most commonly used to assess model performance, it is insufficient on its own to provide a complete evaluation. Therefore, various complementary classification metrics are taken into account for evaluation, including the following: Recall (Sensitivity), Precision, F_β , Accuracy, Specificity, Cohen's kappa; these are described in the Supplementary Section ??.

4.4 Evaluation metrics for data augmentation

Data augmentation techniques have become a widely adopted strategy for improving the generalization capacity of deep neural networks. An important consideration in this context is the balance between similarity and diversity in the data [38]. Similarity reflects the degree of resemblance between the original and augmented data, whereas diversity captures the variation in complexity introduced through augmentation. Therefore, several data similarity and diversity metrics are considered: Fréchet Inception Distance (FID), Kernel Inception Distance (KID), and Multi-Scale Structural Similarity (MS-SSIM). These metrics are described in Supplementary Section ???. Since it is not recommended to apply the IS, KID, or FID metrics to datasets other than ImageNet, and given that the spectrogram dataset is not part of the ImageNet dataset, the MS-SSIM score is used. These metrics are calculated using the [TorchMetrics](#) Python library from Lightning AI.

4.5 AI-assisted manuscript preparation

While preparing this manuscript, the author used Google Gemini 3.6 Flash to improve the English in the paragraphs after they were written, and OpenAI Codex, powered by GPT-5, was used to verify compliance with the author guidelines. After using these tools or services, the author reviewed and edited the content as necessary and assumes full responsibility for the content of the publication.

5 Results

The data distribution is imbalanced, comprising 3% HY, 9% LP, 36% TC, 8% TR, and 44% VT events, as shown in Supplementary Table ???. To address this imbalance, new data are generated using various augmentation techniques. Each data augmentation technique is applied to balance the existing dataset, resulting in a dataset containing 13,430 events (2,686 examples per class). Furthermore, balanced datasets composed solely of augmented data are created, ensuring an equal number of events, as detailed in Supplementary Table ??.

Multiple experiments are designed using the datasets generated by the respective DA methods (see Table 1). For each balanced dataset, a deep-learning model is trained using a transfer-learning strategy based on pretrained AlexNet. Each dataset is divided into 80% training data and 20% test data. Training uses 20 epochs, a mini-batch size of 8, an initial learning rate of 0.00005, and the Adam optimization algorithm. The training and validation performance scores for selected experiments are illustrated in Supplementary Fig. ??, and confusion matrices in Supplementary Fig. ?. The trained models are subsequently evaluated using the original imbalanced dataset of 6,145 events, with the results presented in Table 2.

5.1 Augmented datasets

The original imbalanced dataset of 6,145 events is balanced using each data augmentation technique. This approach increases the number of examples in the underrepresented classes, resulting in a balanced dataset of 13,430 events. In addition,

415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460

Table 1 Performance of AlexNet transfer-learning experiments.

Dataset	Data augmentation	Accuracy (%)	F1 (%)	Recall (%)	Precision (%)	Cohen Kappa (%)	Class Accuracy (%)				
							HY	LP	TC	TR	VT
i	No	62.8	63.9	62.8	66.7	67.6	38.9	74.8	73.5	33.3	93.4
ii	Rotation(5%-25%)	91.5	91.4	91.5	91.4	89.2	98.4	92.4	83.5	93.3	89.6
iii	Rotation(5%-25%)	94.2	94.0	94.2	94.0	92.6	99.8	96.3	83.1	96.1	95.8
iv	Jittering(0.2)	90.2	90.0	90.2	90.2	87.5	94.0	93.6	77.9	95.9	89.4
v	Jittering(0.2)	91.1	91.1	91.1	91.2	88.8	97.1	95.3	85.2	93.3	84.3
vi	Drifting Drift(0.01-0.1)	84.1	83.5	84.1	83.7	79.6	94.4	90.7	62.1	82.4	90.8
vii	Drifting Drift(0.01-0.1)	87.7	87.6	87.7	87.9	84.6	98.7	88.5	78.1	91.3	81.7
viii	Drifting Drift(0.001-0.01)	90.8	90.6	90.8	90.6	88.3	96.9	95.7	77.7	90.2	93.6
ix	Drifting Drift(0.001-0.01)	93.5	93.4	93.5	93.5	91.8	98.9	95.9	83.0	97.0	92.8
x	Drifting Drift(0.0001-0.001)	93.8	93.5	93.8	93.7	92.0	99.8	100.0	81.0	98.0	90.0
xi	Drifting Drift(0.0001-0.001)	97.6	97.6	97.6	97.6	97.0	100.0	99.0	93.4	100.0	95.6
xii	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	87.8	87.6	87.8	87.7	84.6	95.8	91.5	73.6	91.7	86.7
xiii	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	86.4	86.2	86.4	86.3	82.8	92.6	90.7	74.6	85.6	88.6
xiv	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	86.4	86.1	86.4	86.2	82.7	94.6	88.5	72.0	86.0	91.0
xv	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	85.8	85.5	85.8	85.9	81.9	90.4	94.4	75.1	77.3	91.8
xvi	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	84.9	84.7	84.9	84.7	80.8	88.1	91.5	74.1	80.0	90.8
xvii	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	83.0	82.9	83.0	83.0	78.6	89.2	83.1	77.2	73.9	91.8
xviii	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	83.7	83.4	83.7	83.8	79.3	94.0	84.3	75.0	73.2	91.8
xix	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	83.2	83.2	83.2	83.3	78.8	89.5	86.0	75.3	76.9	88.2
xx	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Marks(2)	90.5	90.2	90.5	90.2	87.9	98.2	95.5	77.9	90.2	90.4
xxi	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Marks(2)	91.7	91.5	91.7	91.5	89.4	97.5	94.8	83.5	89.5	93.2
xxii	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Marks(2)	88.6	88.6	88.6	88.5	85.7	94.9	86.8	80.5	89.5	91.4
xxiii	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Marks(2)	86.6	86.4	86.6	86.4	83.1	96.0	85.6	75.8	87.2	88.2
xxiv	Interpolation AE Signal percent(40%-60%)	87.5	87.4	87.5	87.6	84.1	89.3	90.9	76.9	89.3	91.0
xxv	Interpolation AE Signal percent(40%-60%)	93.1	93.1	93.1	93.3	91.3	97.6	96.5	83.7	96.5	91.4

Table 2 Testing trained models on real dataset of 6,145 events.

Dataset	Data augmentation	Accuracy (%)	F1 (%)	Recall (%)	Precision (%)	Cohen Kappa (%)	Class Accuracy (%)				
							HY	LP	TC	TR	VT
i	No	-	-	-	-	-	-	-	-	-	-
ii	Rotation(5%-25%)	96.6	96.2	96.6	95.9	95.7	97.7	95.3	97.0	94.9	98.1
iii	Rotation(5%-25%)	93.7	88.9	93.7	85.5	86.9	100.0	96.0	82.5	93.8	96.0
iv	Jittering(0.2)	96.2	96.0	96.2	95.8	95.0	94.8	96.7	95.8	96.2	97.5
v	Jittering(0.2)	93.9	87.3	93.9	82.9	86.2	99.5	97.1	86.0	95.8	91.3
vi	Drifting Drift(0.01-0.1)	91.5	89.4	91.5	88.1	90.0	93.0	95.5	90.0	81.5	97.6
vii	Drifting Drift(0.01-0.1)	66.6	53.5	66.6	55.2	48.3	69.5	96.5	42.0	56.5	68.4
viii	Drifting Drift(0.001-0.01)	95.0	95.4	95.0	95.8	94.7	93.0	97.1	94.9	91.3	98.8
ix	Drifting Drift(0.001-0.01)	93.7	86.7	93.7	82.4	85.5	98.1	97.4	83.4	97.7	91.8
x	Drifting Drift(0.0001-0.001)	98.3	96.7	98.3	95.3	96.0	99.5	100.0	95.5	98.5	97.8
xi	Drifting Drift(0.0001-0.001)	97.2	93.8	97.2	91.0	92.9	100.0	98.3	93.3	99.6	94.9
xii	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	95.9	93.7	95.9	91.8	93.8	96.7	97.6	94.1	94.3	97.0
xiii	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	82.7	73.6	82.7	70.0	71.8	86.9	93.4	65.2	78.3	89.8
xiv	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	95.2	93.6	95.2	92.2	93.6	95.3	96.7	93.2	92.4	98.3
xv	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	75.2	67.3	75.2	64.9	64.3	77.5	91.2	57.3	61.4	88.8
xvi	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	93.3	94.0	93.3	94.9	93.8	90.6	98.1	95.5	84.1	98.3
xvii	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	81.4	74.2	81.4	70.7	72.4	85.4	83.9	67.9	78.6	91.1
xviii	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	94.0	94.0	94.0	94.2	93.8	95.3	94.8	95.1	86.4	98.5
xix	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	80.1	68.1	80.1	64.9	65.8	88.7	92.5	56.9	76.6	85.9
xx	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)	96.7	96.4	96.7	96.1	95.4	98.6	98.6	95.8	92.4	98.2
xxi	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)	95.4	91.1	95.4	87.8	89.8	100.0	97.2	88.1	96.6	95.1
xxii	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Masks(2)	95.5	95.7	95.5	95.9	94.9	97.7	90.5	96.3	94.9	98.4
xxiii	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Masks(2)	93.6	87.5	93.6	83.7	85.2	100.0	95.1	82.8	97.5	92.4
xxiv	Interpolation AE Signal percent(40%-60%)	90.4	91.8	90.4	93.4	92.0	82.2	90.5	94.7	86.4	98.1
xxv	Interpolation AE Signal percent(40%-60%)	80.7	71.0	80.7	67.5	67.8	88.7	85.8	60.4	80.7	88.1

a second balanced dataset consisting exclusively of 13,430 augmented events—2,686 instances per class—is generated for comparison. The original imbalanced dataset is retained as a baseline to assess the effect of class imbalance on model accuracy. The generation parameters for each data augmentation technique are:

- *Rotation*: Random value $r \in [0.05, 0.25]$ with random rotation axes (*start*, *center*, *end*) without repetition.
- *Jittering*: $\sigma = 0.2$ for Gaussian noise distributed as $N(0, \sigma)$.
- *Drifting*: Random value $d \in [a, b]$. Ranges considered $[0.01, 0.1]$, $[0.001, 0.01]$, $[0.0001, 0.001]$.
- *Genetic Algorithm & Without Time Adjustment*: 3 and 5 crossover segments, each lasting 5 seconds.
- *Genetic Algorithm & With Time Adjustment*: 40% and 45% of the signal for crossover, divided into 10 crossover segments.
- *SpecAugment*: Two time and frequency masks, each covering 10% and 15% of the signal.
- *Interpolation AEs*: Random value $x \in [0.4, 0.6]$. Convolutional autoencoder used is illustrated in Supplementary Fig. ??.

5.2 Model performance

The experimental evidence underscores the importance of data augmentation techniques in improving deep-learning model performance. Models trained without data augmentation exhibited low precision, as shown in Table 1. To compare the generated models with the state-of-the-art, performance metrics such as accuracy, precision, recall, specificity, F1 score and Cohen’s Kappa have been taken into account. In tables 1 and 2, it is evident that the Drifting data augmentation technique achieved the best results across almost all trained models. This success may be attributed to the Drifting technique’s ability to effectively cover the feature space. The most notable result is from Drifting (Augmented)(xi), which, with a drift range of 0.0001–0.001, achieved an accuracy of 97.6% over an entirely augmented dataset. An additional noteworthy result is the Rotation (Augmented)(iii) technique, which, with a range of 5%–25%, achieved an accuracy of 94.2% on another augmented dataset. In both cases, the F1 score metrics closely correspond to their respective accuracies. Although both models are derived from entirely augmented data, they also perform well in real data tests, achieving accuracies of 97.2% for Drifting (Augmented)(xi) and 93.7% for Rotation (Augmented)(iii) (see Table 2). However, in real-data tests, the Drifting (Real+Augmented)(x) model achieved the highest accuracy (98.3%), using both real and augmented data. Furthermore, this dataset (Real+Augmented)(x), generated by Drifting in the range of 0.0001–0.001, achieved an accuracy of 93.8% during model training. Next, the dataset of augmented data (ix) generated by Drifting in the range 0.001–0.01, reached an accuracy of 93.5%. Then, the augmented data dataset (xxv) generated by Autoencoder Interpolation in the range 40%–60% obtained an accuracy of 93.1% (see Table 1). These models also show good performance in tests with real data, with accuracies of 98.3% for Drifting (Real+Augmented)(x), 93.7% for Drifting (Augmented)(ix) and

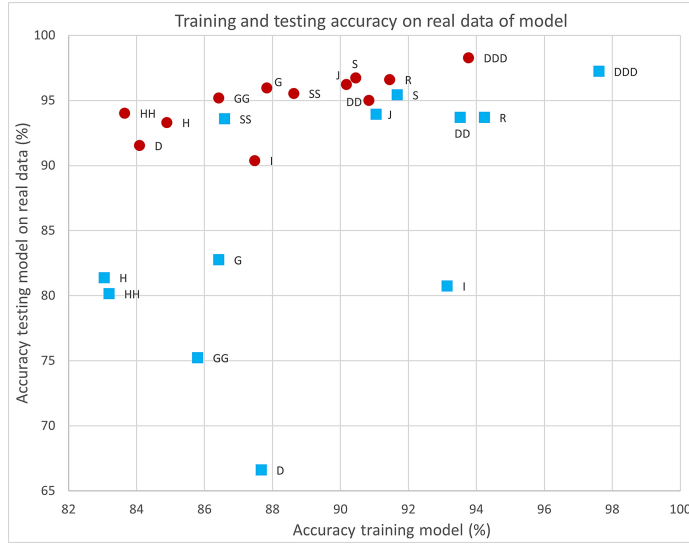


Fig. 1 Model training versus real-data testing accuracy: green shape represents real dataset with 62.8% accuracy (not visible because segment with accuracy $\geq$ 82% is displayed); red circles indicate datasets with real and augmented data; turquoise squares represent fully augmented dataset. Dataset codes appear in Table 3.

80.7% for Interpolation (Augmented)(xxv) (see Table 2). Only the last Interpolation model (Augmented)(xxv) exhibits a large gap between training and test accuracy on real data. Figure 1 shows that Drifting, Rotation, Interpolation, SpecAugment and Jittering have slightly better accuracy values than the other techniques.

Table 3 Relationship between Data Augmentation technique and Dataset.

Code	Data augmentation technique	Datasets	
		Real + Augmented	Augmented
N	Real dataset	i	
R	Rotation (5%-25%)	ii	iii
J	Jittering (0.2)	iv	v
D	Drifting (0.01-0.1)	vi	vii
DD	Drifting (0.001-0.01)	viii	ix
DDD	Drifting (0.0001-0.001)	x	xi
G	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	xii	xiii
GG	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	xiv	xv
H	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	xvi	xvii
HH	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	xviii	xix
S	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)	xx	xxi
SS	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Masks(2)	xxii	xxiii
I	Interpolation AE Signal percent(40%-60%)	xxiv	xv

5.3 Diversity and Similarity Metrics

The diversity and similarity metrics are calculated by comparing two datasets and the same event type.

599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644

Table 4 Multiscale Structural Similarity Index (MS-SSIM) as a similarity and diversity metric for experimental datasets.

Dataset	Data augmentation	Internal						External					
		HY	LP	TC	TR	VT	MS	HY	LP	TC	TR	VT	MS
i	No	0.34	0.41	0.33	0.29	0.34	0.34	1.00	1.00	1.00	1.00	1.00	1.00
ii	Rotation(5%-25%)	0.38	0.42	0.36	0.29		0.36	0.37	0.41	0.33	0.32	1.00	0.53
iii	Rotation(5%-25%)	0.36	0.42	0.36	0.33	0.35	0.37	0.39	0.41	0.34	0.32	0.37	0.37
iv	Jittering(0.2)	0.28	0.36	0.27	0.26		0.29	0.26	0.37	0.32	0.25	1.00	0.49
v	Jittering(0.2)	0.26	0.33	0.26	0.21	0.23	0.26	0.30	0.34	0.27	0.24	0.26	0.28
vi	Drifting Drift(0.01-0.1)	0.40	0.46	0.42	0.35		0.41	0.44	0.45	0.34	0.33	1.00	0.55
vii	Drifting Drift(0.01-0.1)	0.80	0.81	0.81	0.76	0.71	0.78	0.44	0.47	0.39	0.34	0.38	0.40
viii	Drifting Drift(0.001-0.01)	0.37	0.42	0.38	0.31		0.37	0.38	0.42	0.33	0.30	1.00	0.53
ix	Drifting Drift(0.001-0.01)	0.42	0.48	0.43	0.37	0.36	0.41	0.36	0.43	0.35	0.31	0.33	0.35
x	Drifting Drift(0.0001-0.001)	0.35	0.40	0.33	0.29		0.34	0.38	0.41	0.33	0.29	1.00	0.53
xi	Drifting Drift(0.0001-0.001)	0.35	0.40	0.32	0.31	0.34	0.34	0.32	0.40	0.33	0.30	0.35	0.34
xii	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	0.38	0.40	0.35	0.30		0.36	0.43	0.41	0.33	0.30	1.00	0.54
xiii	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	0.35	0.40	0.37	0.30	0.36	0.36	0.37	0.40	0.32	0.30	0.35	0.35
xiv	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	0.35	0.41	0.35	0.30		0.35	0.36	0.40	0.33	0.30	1.00	0.52
xv	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	0.35	0.40	0.37	0.33	0.40	0.37	0.40	0.40	0.34	0.28	0.38	0.36
xvi	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	0.36	0.41	0.35	0.30		0.35	0.35	0.40	0.32	0.30	1.00	0.52
xvii	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	0.34	0.41	0.32	0.29	0.34	0.34	0.37	0.40	0.33	0.30	0.34	0.34
xviii	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	0.35	0.41	0.33	0.30		0.35	0.36	0.40	0.33	0.29	1.00	0.52
xix	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	0.31	0.41	0.35	0.30	0.34	0.34	0.32	0.41	0.33	0.30	0.35	0.34
xx	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)	0.37	0.41	0.33	0.30		0.35	0.39	0.40	0.34	0.30	1.00	0.53
xxi	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)	0.34	0.40	0.35	0.31	0.34	0.35	0.38	0.40	0.33	0.29	0.35	0.35
xxii	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Masks(2)	0.35	0.41	0.33	0.31		0.35	0.35	0.41	0.33	0.31	1.00	0.53
xxiii	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Masks(2)	0.35	0.40	0.36	0.32	0.39	0.37	0.36	0.40	0.34	0.31	0.37	0.36
xxiv	Interpolation AE Signal percent(40%-60%)	0.38	0.44	0.36	0.33		0.38	0.37	0.44	0.34	0.32	1.00	0.54
xxv	Interpolation AE Signal percent(40%-60%)	0.43	0.48	0.41	0.37	0.41	0.42	0.37	0.44	0.35	0.33	0.39	0.38

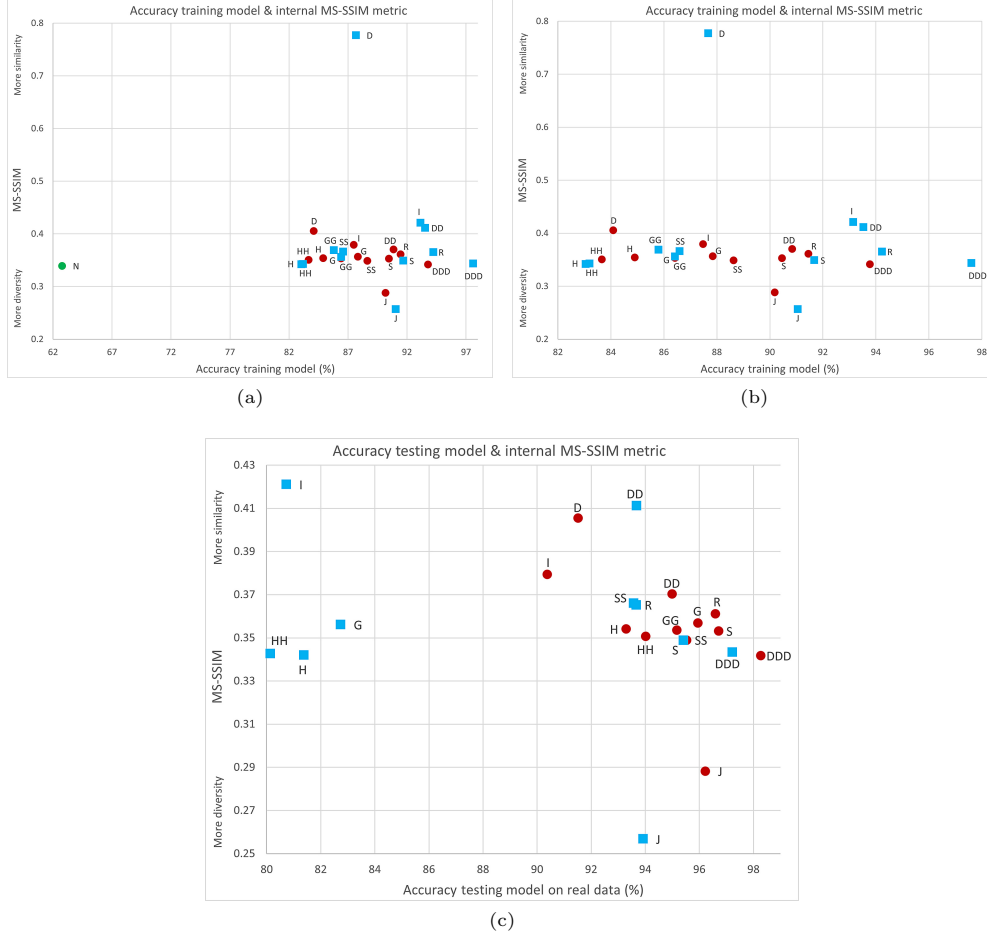


Fig. 2 (a) Model-training accuracy versus internal MS-SSIM. (b) Detailed view of model-training accuracy versus internal MS-SSIM. Real dataset with 62.8% accuracy (not visible because segment with $\text{accuracy} \geq 82\%$ is displayed). (c) Real-data testing model accuracy versus internal MS-SSIM. Green shape represents real dataset; red circles represent datasets combining real and augmented data; turquoise squares represent fully augmented datasets. Dataset codes appear in Table 3.

5.3.1 Internal MS-SSIM

This metric is called an *internal* diversity and similarity metric if only the internal data of the dataset are considered for the comparison (see Table 4). Depending on the dataset type, the following cases apply:

- *Real + augmented:* Since the VT event has sufficient examples and does not require augmented data to balance the dataset, this metric is not calculated for this specific event due to the lack of comparison data. For the other events, the comparison is made between real and augmented data.

- *Augmented*: Because all data are augmented, the dataset is divided into two distinct subsets, which are then compared with each other.

Most *internal* MS-SSIM values fall between 0.30 and 0.42, with three exceptions (see Fig. 2a). Jittering produces greater diversity in the Augmented (v) and Real+Augmented (iv) datasets, with MS-SSIM values of 0.26 and 0.29, respectively. The dataset with the highest similarity and lowest diversity is Drifting (0.01–0.1) (Augmented) (vii), with an MS-SSIM of 0.78. It is followed by Interpolation (Augmented) (xxv), and Drifting (Real+Augmented) (vi), with MS-SSIM values of 0.42 and 0.41, respectively. Figure 2a also shows the relationship between internal MS-SSIM and model-training accuracy. Two points are clearly separated from the main cluster: the real-only dataset (i), which has low accuracy because of class imbalance, and the Drifting dataset (Augmented) (vii), which exhibits substantially higher similarity. The detailed view in Fig. 2b shows that almost all datasets, including the real dataset, have similar diversity levels; however, the augmented Jittering dataset (v) is more diverse. When augmented-only datasets are compared with real-plus-augmented datasets generated using the same DA method, the former generally exhibit equal or greater diversity and superior training performance. This pattern is especially evident for Jittering (Augmented) (v); SpecAugment (Augmented) (xxi); and Genetic Algorithm (Augmented) (xvii, xix). In Figure 2c, certain augmented and real datasets demonstrate better diversity and testing accuracy values compared to their counterparts within the same DA method. Notable examples include Rotation (ii, iii), Drifting (vi, vii), Drifting (viii, ix), Drifting (x, xi), Genetic Algorithm (xiv, xv), SpecAugment (xxii, xxiii), and Interpolation AE (xxiv, xxv).

5.3.2 External MS-SSIM

The metric is termed *external* when a study dataset is compared with the real dataset. Table 4 shows that the real dataset (i) exhibits perfect similarity (see Fig. 3a), as expected when a dataset is compared with itself. Figure 3b shows external MS-SSIM as a function of model-training accuracy. The real dataset (i) is clearly distinguishable because it combines perfect similarity with low accuracy caused by class imbalance. Two additional groups can be identified (see Fig. 3c):

- The first group comprises datasets that combine real and augmented data, with MS-SSIM values ranging from 0.50 to 0.55. Two cases lie outside this range: Drifting (Real+Augmented) (vi), has the greatest similarity, with an MS-SSIM of 0.552, whereas Jittering (Real+Augmented) (iv), has the greatest diversity in this group, with a value of 0.49.
- The second group comprises fully augmented datasets, which are more diverse than those in the first group. Their MS-SSIM values range from 0.30 to 0.40. Two cases lie outside this range: Drifting (Augmented) (vii), has an MS-SSIM of 0.401 and is the least diverse dataset in the group, whereas Jittering (Augmented) (v), has an MS-SSIM of 0.28 and is the most diverse dataset both within the group and overall.

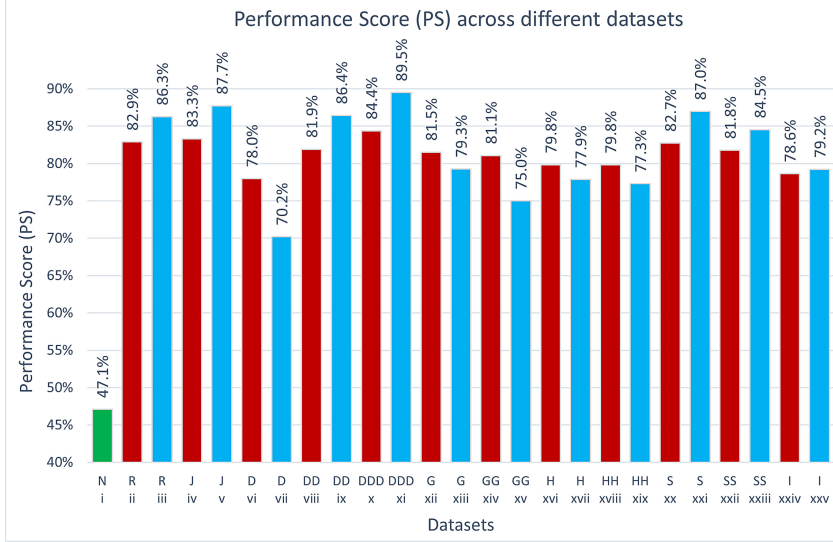


Fig. 4 Performance Score (PS) across datasets; higher values indicate better performance. Green bar represents real dataset; red bars represent datasets combining real and augmented data; turquoise bars represent fully augmented datasets. Dataset codes appear in Table 3.

the range $[0, 1]$. The test accuracy (A_{te}) is assigned twice the weight of the other components because of its direct relationship with model generalization. The *external* MS-SSIM metric is used because its computation considers the entire real dataset. Figure 4 shows that models trained on the Drifting (Augmented)(xi), Jittering (Augmented)(v), SpecAugment (Augmented)(xxi), and Drifting (Augmented)(ix) fully augmented datasets exhibit the highest PS values, making them the most relevant in terms of the trade-off between model accuracy and dataset diversity. In contrast, models with low PS values, such as the Drifting fully augmented dataset (Augmented)(vii) and the real event dataset (i), are negatively impacted by their lower diversity and imbalanced nature, respectively.

6 Discussion

The experiments conducted in this study highlight the limitations of models trained on imbalanced datasets. Specifically, the real dataset (i) produced a model biased toward the LP, TC, and VT event classes. The results indicate that balanced datasets yield better-performing models because they increase the number of examples within the event feature space. When subsequently evaluated on real data, most models achieve competitive classification performance. Transfer learning also reduces the time required to develop the deep-learning models. Several models trained entirely on augmented data achieve higher accuracy than models trained on datasets combining real and augmented data. This result confirms that the data augmentation techniques evaluated increase data diversity. Data from several stations are integrated to construct a generalized multi-station classification model.

Furthermore, models trained on fully augmented datasets achieve high classification accuracy on real data in most cases.

Regarding the similarity and diversity metrics, most *internal* MS-SSIM values are similar, and no specific group of DA techniques stands out under the selected generation parameters.

The *external* MS-SSIM results reveal two groups, one of which is more dispersed than the other. The more compact group corresponds to datasets containing real and augmented data, which exhibit less diversity than the fully augmented datasets. A statistical t-test comparing the mean MS-SSIM values of the two groups yields a p -value of approximately $5.03 \cdot 10^{-15}$, indicating a statistically significant difference.

The Performance Score (PS) quantifies how model accuracy and dataset diversity jointly contribute to overall performance. Applying DA techniques to volcanic seismic events diversifies the training data and encourages models to learn invariant, fundamental event characteristics rather than memorize specific details. In this sense, DA acts as a form of regularization that reduces overfitting and improves model robustness. The initial PS formulation (Eq. 3) uses fixed weights to facilitate model comparison. However, these weights could be treated as hyperparameters in future applications.

To improve model performance, it is essential to quantify the relationship between model accuracy and the diversity generated by the data augmentation method. Specifically, this requires determining the gain in model accuracy relative to the variation in dataset diversity.

Integrating these data augmentation strategies with other machine learning approaches, such as ensemble learning or reinforcement learning, could lead to the development of more robust and accurate models for seismic-volcanic signal recognition.

Finally, a large dataset could be generated using various data augmentation techniques from different channels, stations, and volcanoes.

7 Conclusion

This study applies various data augmentation techniques to diversify the training data used to classify seismic-volcanic signals, resulting in more robust models with greater generalization ability. These techniques are first used to balance datasets containing real observations and are then used to generate fully augmented datasets. Deep-learning classification models are trained on these datasets through transfer learning with pretrained AlexNet. The performance of the trained classification models is evaluated, revealing that data augmentation techniques enhance model performance. The trained classification models are competitive with the baselines presented in previous research and the state of the art. Additionally, the following results are achieved:

- The integration of data from multiple stations on the Lascar volcano facilitated the construction of more generalizable probabilistic models, thereby mitigating biases associated with individual stations through various data augmentation techniques.

- Data augmentation techniques improve the performance metrics of deep learning models, as they enable the generation of additional examples within the feature space of volcanic seismic events.
- Similarity and diversity metrics are used to analyze these characteristics in datasets generated by various data augmentation techniques, as well as their effect on model performance.
- The evaluated data augmentation techniques generate more diverse data and, in several cases, better-performing models, particularly when Drifting, Jittering, and SpecAugment are used.
- Transfer learning accelerated the generation of deep learning models compared to training from scratch.

The findings of this work can be used to improve the performance of deep learning models in seismic-volcanic signal recognition, considering the data augmentation techniques presented. Furthermore, the results indicate that models trained exclusively on augmented datasets show excellent performance when tested with real data, in addition to exhibiting diversity. This finding represents a substantial contribution to seismic-volcanic signal recognition, as well as to the development of more generalizable models.

8 Future Work

Future research should optimize the parameters of the existing data augmentation methods to improve model performance. Evaluating these techniques on different types of seismic events and under varying noise levels would also provide deeper insight into their applicability.

Future work should explore a methodology to formulate conditions that determine whether a data augmentation technique increases dataset diversity to safe levels without compromising class separation.

It would be extremely valuable to build a large dataset consisting entirely of augmented data generated using data augmentation techniques, train an event classification model on that dataset, and then test the trained model on real data.

The possibility remains open to retrain the classification model with each new dataset from different volcanoes, or to train a new model using data from multiple volcanoes.

Acknowledgements

We thank the CKELAR-VOLCANES team for its support.

Author contributions

Franz Yupanqui Machaca: Investigation, Writing – original draft, Writing – review & editing, Software, Formal analysis, Visualization. Pablo Salazar Reinoso: Conceptualization, Methodology, Supervision, Resources, Formal analysis, Funding

acquisition, Validation. Claudio Meneses Villegas: Conceptualization, Methodology,	921
Supervision, Resources, Formal analysis, Funding acquisition, Validation.	922
	923
Funding	924
	925
This research was supported by the Millennium Institute on Volcanic Risk	926
Research–Ckelar Volcanoes (ANID/MILENIO) and Universidad Católica del Norte	927
(UCN). Franz Yupanqui Machaca was funded by the Millennium Institute on	928
Volcanic Risk Research–Ckelar Volcanoes (ANID/MILENIO; grant no.	929
ICN2021_038) and a doctoral scholarship from UCN, as part of the Doctoral	930
Program in Sustainable Engineering.	931
	932
Competing interests	933
	934
The authors declare that they have no known competing financial interests or personal	935
relationships that could have appeared to influence the work reported in this paper.	936
	937
Data availability	938
	939
The data are available from Zenodo at https://doi.org/10.5281/Zenodo.6001869 [39].	940
	941
Code availability	942
	943
Name of the code: daSeismicSignals.	944
Contact: frayup@gmail.com , franz.yupanqui@alumnos.ucn.cl	945
Hardware requirements: Win 11, NVIDIA GeForce GTX 1060 GPU.	946
Programming language: Python.	947
Software required: Python ≥ 3.9 , CUDA	948
The code has been open-sourced on GitHub and is available at:	949
https://github.com/franznet/daseismicsignals .	950
	951
References	952
	953
[1] Salazar, P., Yupanqui, F., Meneses, C., Layana, S. & Yáñez, G. Multi-station	954
automatic classification of seismic signatures from the lascar volcano database.	955
<i>Natural Hazards and Earth System Sciences Discussions</i> 2022 , 1–27 (2022).	956
	957
[2] Salazar, A., Arroyo, R., Pérez, N. & Benítez, D. Deep-learning for volcanic seismic	958
events classification. In <i>2020 IEEE Colombian Conference on Applications of</i>	959
<i>Computational Intelligence (IEEE ColCACI 2020)</i> , 1–6 (IEEE, 2020).	960
	961
[3] Malfante, M. <i>et al.</i> Automatic classification of volcano seismic signatures. <i>Journal</i>	962
<i>of Geophysical Research: Solid Earth</i> 123 , 10–645 (2018).	963
	964
	965
	966

- 967 [4] Bicego, M., Acosta-Munoz, C. & Orozco-Alzate, M. Classification of seismic
968 volcanic signals using hidden-markov-model-based generative embeddings. *IEEE*
969 *transactions on geoscience and remote sensing* **51**, 3400–3409 (2012).
970
- 971 [5] Curilem, M. *et al.* Pattern recognition applied to seismic signals of llaima volcano
972 (chile): An evaluation of station-dependent classifiers. *Journal of Volcanology and*
973 *Geothermal Research* **315**, 15–27 (2016).
974
- 975 [6] Lara, P. E. E. *et al.* Automatic multichannel volcano-seismic classification using
976 machine learning and emd. *IEEE Journal of Selected Topics in Applied Earth*
977 *Observations and Remote Sensing* **13**, 1322–1331 (2020).
978
- 979 [7] Beyreuther, M., Hammer, C., Wassermann, J., Ohrnberger, M. & Megies, T.
980 Constructing a hidden markov model based earthquake detector: application to
981 induced seismicity. *Geophysical Journal International* **189**, 602–610 (2012). URL
982 <https://doi.org/10.1111/j.1365-246X.2012.05361.x>.
983
- 984 [8] Carniel, R. Characterization of volcanic regimes and identification of significant
985 transitions using geophysical data: a review. *Bulletin of Volcanology* **76**, 1–22
986 (2014).
987
- 988 [9] Langer, H. *et al.* Synopsis of supervised and unsupervised pattern classification
989 techniques applied to volcanic tremor data at mt etna, italy. *Geophysical Journal*
990 *International* **178**, 1132–1144 (2009). URL [https://doi.org/10.1111/j.1365-246X.](https://doi.org/10.1111/j.1365-246X.2009.04179.x)
991 [2009.04179.x](https://doi.org/10.1111/j.1365-246X.2009.04179.x).
992
- 993 [10] Giacco, F., Esposito, A. M., Scarpetta, S., Giudicepietro, F. & Marinaro, M.
994 Support vector machines and mlp for automatic classification of seismic signals
995 at stromboli volcano In *Neural Nets WIRN09* 116–123 (IOS Press, 2009).
996
- 997 [11] Curilem, M. *et al.* Improving the classification of volcanic seismic events extract-
998 ing new seismic and speech features. In *Progress in Pattern Recognition, Image*
999 *Analysis, Computer Vision, and Applications: 22nd Iberoamerican Congress,*
1000 *CIARP 2017, Valparaíso, Chile, November 7–10, 2017, Proceedings 22*, 177–185
1001 (Springer, Cham, 2018).
1002
- 1003 [12] Malfante, M. *et al.* Machine learning for volcano-seismic signals: Challenges and
1004 perspectives. *IEEE Signal Processing Magazine* **35**, 20–30 (2018).
1005
- 1006 [13] Titos, M., Bueno, A., García, L., Benítez, C. & Segura, J. C. Classification
1007 of isolated volcano-seismic events based on inductive transfer learning. *IEEE*
Geoscience and Remote Sensing Letters **17**, 869–873 (2019).
1008
- 1009 [14] Curilem, M., Canário, J. P., Franco, L. & Rios, R. A. Using cnn to classify spec-
1010 trograms of seismic events from llaima volcano (chile). In *2018 International*
1011 *Joint Conference on Neural Networks (IJCNN)*, 1–8 (2018).
1012

- [15] Rodriguez, A. B., Benitez, C., Zuccarello, L., De Angelis, S. & Ibáñez, J. M. Bayesian monitoring of seismo-volcanic dynamics. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–14 (2021). 1013
1014
1015
1016
- [16] Bueno, A., Benitez, C., De Angelis, S., Moreno, A. D. & Ibáñez, J. M. Volcano-seismic transfer learning and uncertainty quantification with bayesian neural networks. *IEEE transactions on geoscience and remote sensing* **58**, 892–902 (2019). 1017
1018
1019
1020
1021
- [17] Peixoto, A. A. T. *et al.* Tensor-based learning framework for automatic multi-channel volcano-seismic classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **14**, 4517–4529 (2021). 1022
1023
1024
1025
- [18] López-Pérez, M., García, L., Benítez, C. & Molina, R. A contribution to deep learning approaches for automatic classification of volcano-seismic events: Deep gaussian processes. *IEEE Transactions on Geoscience and Remote Sensing* **59**, 3875–3890 (2020). 1026
1027
1028
1029
- [19] Bicego, M., Londoño-Bonilla, J. M. & Orozco-Alzate, M. Volcano-seismic events classification using document classification strategies. In *Image Analysis and Processing—ICIAP 2015: 18th International Conference, Genoa, Italy, September 7–11, 2015, Proceedings, Part I* **18**, Vol. 9279, 119–129 (Springer, 2015). 1030
1031
1032
1033
1034
- [20] Esposito, A. M. *et al.* Unsupervised neural analysis of very-long-period events at stromboli volcano using the self-organizing maps. *Bulletin of the Seismological Society of America* **98**, 2449–2459 (2008). 1035
1036
1037
1038
- [21] Duque, A. *et al.* Exploring the unsupervised classification of seismic events of cotopaxi volcano. *Journal of Volcanology and Geothermal Research* **403**, 107009 (2020). URL <https://www.sciencedirect.com/science/article/pii/S0377027320302821>. 1039
1040
1041
1042
1043
- [22] Yupanqui Machaca, F. *Clasificación de señales sísmico volcánicas mediante técnicas de Transferencia de Aprendizaje y Aprendizaje Profundo*. Master’s thesis, Universidad Católica del Norte, Avenida Angamos 0610, Antofagasta, Chile (2021). Available at: https://ucn.primo.exlibrisgroup.com/discovery/fulldisplay?docid=alma990000595470107936&context=U&vid=56UCN_INST:UCN&lang=es. 1044
1045
1046
1047
1048
1049
1050
- [23] Ferreira, A., Curilem, M., Gomez, W. & Rios, R. Deep learning and multi-station classification of volcano-seismic events of the nevados del chillán volcanic complex (chile). *Neural Computing and Applications* **35**, 24859–24876 (2023). 1051
1052
1053
1054
- [24] Canario, J. P., Mello, R., Curilem, M., Huenupan, F. & Rios, R. In-depth comparison of deep artificial neural network architectures on seismic events classification. *Journal of Volcanology and Geothermal Research* **401**, 106881 (2020). 1055
1056
1057
1058
URL <https://www.sciencedirect.com/science/article/pii/S0377027319306171>.

- 1059 [25] Titos, M., Bueno, A., García, L., Benítez, M. C. & Ibañez, J. Detection and
1060 classification of continuous volcano-seismic signals with recurrent neural networks.
1061 *IEEE Transactions on Geoscience and Remote Sensing* **57**, 1936–1948 (2018).
1062
- 1063 [26] Maggi, A. *et al.* Implementation of a multistation approach for automated event
1064 classification at piton de la fournaise volcano. *Seismological Research Letters* **88**,
1065 878–891 (2017).
1066
- 1067 [27] Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with
1068 deep convolutional neural networks. *Advances in neural information processing*
1069 *systems* **25**, 1097–1105 (2012).
1070
- 1071 [28] Beyreuther, M. *et al.* Obspy: A python toolbox for seismology. *Seismological*
1072 *Research Letters* **81**, 530–533 (2010).
1073
- 1074 [29] Iwana, B. K. & Uchida, S. An empirical survey of data augmentation for time
1075 series classification with neural networks. *Plos one* **16**, 1–32 (2021). URL <https://doi.org/10.1371/journal.pone.0254841>.
1076
- 1077 [30] Iglesias, G., Talavera, E., González-Prieto, Á., Mozo, A. & Gómez-Canaval, S.
1078 Data augmentation techniques in time series domain: a survey and taxonomy.
1079 *Neural Computing and Applications* **35**, 10123–10145 (2023). URL <http://dx.doi.org/10.1007/s00521-023-08459-3>.
1080
- 1081 [31] Fields, T., Hsieh, G. & Chenou, J. Mitigating drift in time series data with noise
1082 augmentation. In *2019 International Conference on Computational Science and*
1083 *Computational Intelligence (CSCI)*, 227–230 (IEEE, 2019).
1084
- 1085 [32] Nethravathi, P. & Karibasappa, K. Augmentation of the customer’s profile
1086 dataset using genetic algorithm. *International Journal of Research and Scientific*
1087 *Innovation (IJRSI), IV* 33–39 (2017).
1088
- 1089 [33] Villegas, C. M., Curinao, J. L., Aqueveque, D. C., Guerrero-Henríquez, J. &
1090 Matamala, M. V. Data augmentation and hierarchical classification to support
1091 the diagnosis of neuropathies based on time series analysis. *Biomedical Signal*
1092 *Processing and Control* **95**, 106302 (2024). URL <https://www.sciencedirect.com/science/article/pii/S1746809424003604>.
1093
- 1094 [34] Park, D. S. *et al.* Specaugment: A simple data augmentation method for auto-
1095 matic speech recognition. *arXiv preprint arXiv:1904.08779* 2613–2617 (2019).
1096 URL <http://dx.doi.org/10.21437/Interspeech.2019-2680>.
1097
- 1098 [35] Park, D. S. *et al.* Specaugment on large scale datasets. In *ICASSP 2020-*
1099 *2020 IEEE International Conference on Acoustics, Speech and Signal Processing*
1100 *(ICASSP)*, 6879–6883 (IEEE, 2020). URL <https://arxiv.org/abs/1912.05533>.
1101 [arXiv:1912.05533](https://arxiv.org/abs/1912.05533).
1102
1103
1104

- [36] Soni, M., Panda, A. & Kopparapu, S. K. Generalized specaugment: Robust online augmentation technique for end-to-end automatic speech recognition. In *2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 1–5 (IEEE, 2024).
- [37] Berthelot, D., Raffel, C., Roy, A. & Goodfellow, I. Understanding and improving interpolation in autoencoders via an adversarial regularizer. *arXiv preprint arXiv:1807.07543* (2018). URL <https://arxiv.org/abs/1807.07543>.
- [38] Yang, S., Guo, S., Zhao, J. & Shen, F. Investigating the effectiveness of data augmentation from similarity and diversity: An empirical study. *Pattern Recognition* **148**, 110204 (2024). URL <https://www.sciencedirect.com/science/article/pii/S0031320323009019>.
- [39] Salazar, P., Yupanqui, F., Meneses, C., Layana, S. & Yáñez, G. Data lacar experiment (2018.03 - 2018.10) (2022). URL <https://doi.org/10.5281/zenodo.6001870>.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [snsupplementary.pdf](#)