

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.

Supplementary Information

A State-of-the-art of classification

047
048
049
050
051
052
053
054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
Table S1 Current status of seismic-volcanic signal classification studies.

Models and presentation year	Volcano – Country	Examples	Spectrogram	Channels	Stations	Classes	Accuracy (%)
Hidden-Markov-Model-Based FSE [1]	Galeras – Colombia	300	No	–	2	3	96
Hidden-Markov-Model-Based FSE [1]	Galeras – Colombia	400	No	–	2	4	85.2
Topic Model (VT, LP, TR) [2]	Galeras – Colombia	300	No	–	1	3	94.1
Topic Model (VT, LP, TR, HB) [2]	Galeras – Colombia	400	No	–	1	4	83.8
Bag-of-Words (VT, LP, TR) [2]	Galeras – Colombia	300	No	–	1	3	91.9
Bag-of-Words (VT, LP, TR, HB) [2]	Galeras – Colombia	400	No	–	1	4	81.8
Support Vector Machines [3]	Ubinas – Peru	109,609	No	–	1	6	93.5±0.5
CNN LeCun [4]	Llaima – Chile	15,895	Yes	–	1	4	97.0
GRU-RNN & LFB [5]	Deception Island – Antarctica	2,193	No	–	1	5	91.6 ¹
LSTM-RNN & LFB [5]	Deception Island – Antarctica	2,193	No	–	1	5	93.8 ¹
Gaussian Mixture Model + uFilter + 5f [6]	Cotopaxi – Ecuador	668	No	–	1	2	96.7
Bayesian Neural Network [7]	St. Helens – U.S.	25,301	No	–	3	3	95.7
Bayesian Neural Network [7]	Bezymianny – Russia	24,916	No	–	3	3	94.5
Bayesian Neural Network [7]	St. Helens – U.S., Bezymianny – Russia	50,217	No	–	6	3	94.6
Bayesian Neural Network [7]	St. Helens – U.S., Bezymianny – Russia	50,217	No	–	6	6	92.1
Support Vector Machines [8]	Ubinas – Peru	3,792	No	3	2	5	83.2
DCNN1 [9]	Cotopaxi – Ecuador	1,108	Yes	–	1	2	95
CNN LeNet [10]	Colima – México	9,332	Yes	–	1	7	88.3
Deep Gaussian Processes – DGP4 [11]	Colima – México	9,332	No	–	–	7	94.8
BIRCH k=2 [12]	Cotopaxi – Ecuador	668	No	–	1	2	88
DNN SeismicNet [13]	Llaima – Chile	3,592	No	–	1	4	95
STuM & MPCA-DR [14]	Ubinas – Peru	5,136	No	3	2	5	91.9 ¹
TCN in b-E3 [15]	Bezymianny – Russia	3,844	Yes	–	1	3	94
RNN-LSTM [16]	Deception Island – Antarctica	2,193	No	3	–	5	84.9
ResNet152 – TL [17]	Lascar – Chile	13,430	Yes	1	10	5	93.2
AlexNet – TL Artificial [17]	Lascar – Chile	67,095	Yes	1	10	5	98.9
AlexNet – TL 10-folds [17]	Lascar – Chile	13,430	Yes	1	10	5	89.1
AlexNet – DL CAE3 Reconstructed Event [17]	Lascar – Chile	13,430	Yes	1	10	5	91.2
AlexNet – TL Real+Artificial 20Hz [18]	Lascar – Chile	13,430	Yes	1	10	5	91.5
AlexNet – TL Artificial 20Hz [18]	Lascar – Chile	67,095	Yes	1	10	5	99.4
AlexNet – TL Artificial 20Hz, Real (Test) [18]	Lascar – Chile	6,145	Yes	1	10	5	99.5
CNN M1-Exp1 [19]	Nevados del Chillán – Chile	313,951	Yes	–	5	8	92.2
UNet & Semantic Segmentation [20]	Nevados del Chillán Volcanic Complex, Laguna del Maule, Villarrica and Puyehue-Cordon Caulle – Chile	24,493	No	1	8	5	91.0 ²

^aCorrect classification rate (CCR).

^bF1 Score.

B Lascar volcano

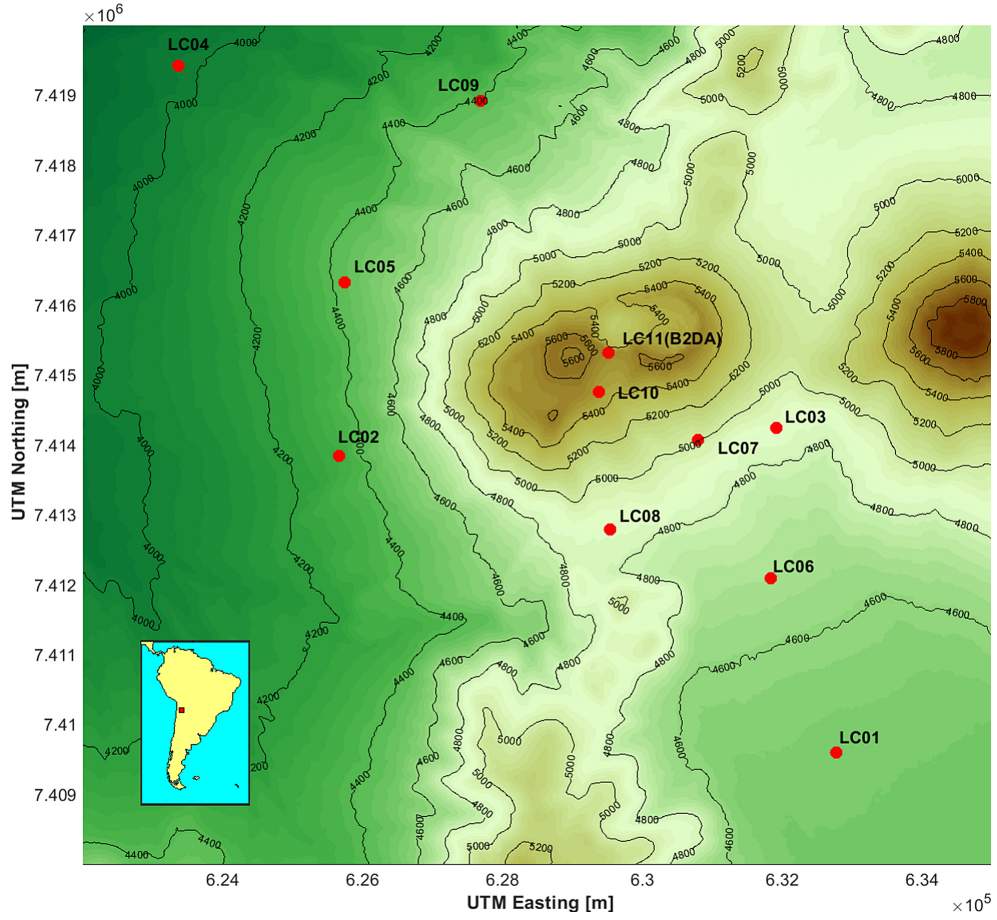


Fig. S1 Location of Lascar volcano. Red dots indicate seismic-station positions. Source: [18]. Map lines delineate study areas and do not necessarily depict accepted national boundaries.

Table S2 Event contributions by station in Lascar database.

Event\Station	LC01	LC02	LC03	LC04	LC05	LC06	LC07	LC08	LC09	LC10	B2DA	Total
HY	0	0	0	7	45	4	0	3	12	136	6	213
LP	11	0	4	2	142	28	5	60	23	195	107	577
TC	9	0	11	97	284	36	157	127	144	1,171	162	2,198
TR	1	0	5	29	21	4	61	13	35	292	10	471
VT	2	0	18	52	84	6	206	14	41	2,249	14	2,686
TOTAL	23	0	38	187	576	78	429	217	255	4,043	299	6,145

Table S3 Event statistics for Lascar volcano.

Events	Quantity (%)	Quantity	Maximum (s)	Minimum (s)	Mean (s)	Median (s)	Mode (s)	Standard Deviation (s)
HY	3%	213	265	10	57.2	59	40	35.3
LP	9%	577	600	9	82.3	67	25	63.5
TC	36%	2,198	670	9	78.5	65	20	56.2
TR	8%	471	216	9	46.2	31	20	33.8
VT	44%	2,686	189	5	19.6	17	10	12.3
TOTAL	100%	6,145	670	5	49.9	28.0	20	50.0

C Seismic events

The seismic events considered are as follows:

- *Long-Period (LP or LF)*: These events are directly associated with the dynamics of fluids (magma, gas, or water) within the volcanic system [3, 8, 21]. They are key indicators for predicting eruptions, as they signal the movement of magmatic or hydrothermal fluids toward the surface [6, 22]. Their energy is concentrated in a narrow band of low frequencies, mostly between 0.5 and 5 Hz [5, 7, 10]. Some sources cite ranges of 0.2–10 Hz or 1–6 Hz [11].
- *Volcano Tectonic (VT, VTE, or HF)*: These represent the brittle response of solid rock to stress changes induced by magma movement or internal pressures [8, 19, 21, 22]. These are fracture earthquakes ("rock breaking"), similar to tectonic earthquakes, but located within or below the volcanic edifice. They usually last less than 30 seconds [6]. They are high-frequency (HF) events. They have a broadband spectrum typically above 5 Hz [6, 22], reaching up to 30 or 40 [5, 11].
- *Hybrid (HY, HB, or MX)*: These events share characteristics of VT and LP events. They are distinguished by starting with a high-frequency signal (similar to a VT, associated with brittle fracture), immediately followed by a low-frequency, long-duration signal (similar to an LP, associated with fluid resonance) [3, 5, 8]. Hybrid events are related to imminent pre-eruptive episodes [3, 5, 22]. Their spectrum is broadband, typically covering energy in the frequency range of 1 to 20 Hz [7].
- *Tectonic (TC or REG)*: These are earthquakes produced by the dynamics of geological faults, not directly related to the active movement of magmatic fluids within the volcanic edifice. They have clear P and S wave arrivals, an impulsive onset, and exponential decay [10, 13]. They can have frequencies of up to 15-40 Hz if they are local, or lower if they are regional due to attenuation [10, 11].
- *Tremor (TR)*: This is a continuous signal indicating sustained activity in the volcanic system. It is defined by a sustained amplitude vibration that can last from minutes to days, months, or even be permanent in active volcanoes [3, 8, 10, 21]. It is associated with the continuous movement of magma or gases, prolonged resonance in the conduits, degassing processes, or the continuous succession of LP events. It is generally restricted to low frequencies, typically in the range of 1 to 5 Hz [5, 10, 22], although it can extend up to 1–9 Hz [3, 21].

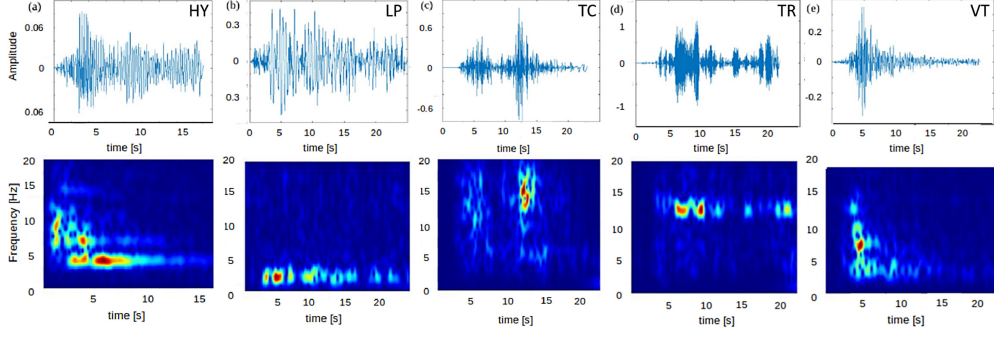


Fig. S2 Time series (top) and spectrogram examples (bottom) for seismic-event classes in Lascar database: (a) hybrid (HY), (b) long-period (LP), (c) tectonic (TC), (d) tremor (TR), and (e) volcano-tectonic (VT).

D Data Similarity

Kullback–Leibler divergence (KL divergence) is used to measure similarities between synthetic and real datasets, and is defined as

$$D_{KL}(P||Q) = \sum_i P(x_i) \cdot \log \left(\frac{P(x_i)}{Q(x_i)} \right) \quad (1)$$

where P and Q are the probability distributions whose distance is calculated over the samples x_i of the distribution [23]. Kullback–Leibler divergence (KL divergence) is not a symmetric distance, so it can be symmetrized to give rise to the so-called Jensen–Shannon divergence (JS divergence), defined as

$$JS(P||Q) = D_{KL}(P||(P+Q)/2) + D_{KL}(Q||(P+Q)/2) \quad (2)$$

This measure quantifies the distance between time-series distributions. It is based on the calculation of the Wasserstein distance between time-series data. This metric is defined by measuring the Wasserstein distance of the energy between frequencies. The probability distributions have a Wasserstein–Fourier distance, which is calculated as follows [23]:

$$WF([x], [y]) = W_2(s_x, s_y) \quad (3)$$

where s_x and s_y are the normalized power spectral densities of the distributions.

E Data Augmentation Pseudocodes

Algorithm 1 Rotation algorithm

```

1: function ROTATION( $tTrace, fPercent, sAxisLoc$ )
2:   if  $sAxisLoc$  in [ $I', F', A'$ ] then
3:     for all  $tr$  in  $tTrace$  do
4:       if  $sAxisLoc = I'$  then ▷ start
5:          $tr.data \leftarrow [0] * \text{int}(\text{len}(tr.data) * fPercent) + tr.data$ 
6:       end if
7:       if  $sAxisLoc = F'$  then ▷ end
8:          $tr.data \leftarrow tr.data + [0] * \text{int}(\text{len}(tr.data) * fPercent)$ 
9:       end if
10:      if  $sAxisLoc = A'$  then ▷ center
11:         $tr.data \leftarrow [0] * (\text{len}(tr.data) * fPercent / 2) + tr.data +$ 
12:           $[0] * (\text{len}(tr.data) * fPercent / 2)$ 
13:      end if
14:    end for
15:  end if
16:  return  $tTrace$ 
17: end function

```

Algorithm 2 Jittering algorithm

```

1: function JITTERING( $tTrace, fSigma = 0.2$ )
2:   for all  $tr$  in  $tTrace$  do
3:      $noise \leftarrow$  Gaussian noise array with distribution  $\mathcal{N}(\mu = 0, \sigma = fSigma)$ 
4:      $tr.data \leftarrow tr.data + noise$ 
5:   end for
6:   return  $tTrace$ 
7: end function

```

Algorithm 3 Drifting algorithm

```

1: function DRIFTING( $tTrace, fDrift = 0.1$ )
2:    $vStep \leftarrow []$ 
3:    $vStep.append(fDrift * (\text{random} < 0.5 ? -1 : 1))$ 
4:   for  $i = 1$  to  $\text{len}(tTrace.data) - 1$  do
5:      $value \leftarrow vStep[i - 1] + (fDrift * (\text{random} < 0.5 ? -1 : 1))$ 
6:      $vStep.append(value)$ 
7:   end for
8:   for all  $tr$  in  $tTrace$  do
9:      $tr.data \leftarrow tr.data + vStep$ 
10:  end for
11:  return  $tTrace$ 
12: end function

```

```

323 Algorithm 6 SpecAugment algorithm
324 1: function SPECAUGMENT(trTrace, iSpectrogramSize, fFrequencyPercent = 0.1,
325                      iFrequencyCant = 2, fTimePercent = 0.1, iTimeCant = 0)
326 2:   imgSpec  $\leftarrow$  GenerateSpectrogram(trTrace)
327 3:   fDuration  $\leftarrow$  trTrace.duration()
328 4:   lFrequencyMask  $\leftarrow$  GenerateMask(20.0, fFrequencyPercent, iFrequencyCant)
329 5:   lTimeMask  $\leftarrow$  GenerateMask(fDuration, fFrequencyPercent, iFrequencyCant)
330 6:   for all (y1, y2) in lFrequencyMask do
331 7:     AddImageMark(imgSpec, Rectangle((0, y1), fDuration, y2 - y1),
332 8:                 color = WhiteNoise)
333 9:   end for
334 10:  for all (x1, x2) in lTimeMask do
335 11:    AddImageMark(imgSpec, Rectangle((x1, 0), x2 - x1, 20.0), color = WhiteNoise)
336 12:  end for
337 13:  return imgSpec
338 14: end function

```

```

341 Algorithm 7 Interpolation with AutoEncoder algorithm
342 1: function INTERPOLATIONAE(trTrace1, trTrace2, modelAE, fPercent = 0.5)
343 2:   imgSpec1  $\leftarrow$  GenerateSpectrogram(trTrace1)
344 3:   imgSpec2  $\leftarrow$  GenerateSpectrogram(trTrace2)
345 4:   vEncoded1  $\leftarrow$  modelAE.encoder(imgSpec1)
346 5:   vEncoded2  $\leftarrow$  modelAE.encoder(imgSpec2)
347 6:   vResult  $\leftarrow$  vEncoded1  $\cdot$  (1 - fPercent) + vEncoded2  $\cdot$  fPercent
348 7:   imgResult  $\leftarrow$  modelAE.decoder(vResult)
349 8:   return imgResult
350 9: end function

```

```

351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368

```

F Data Augmentation

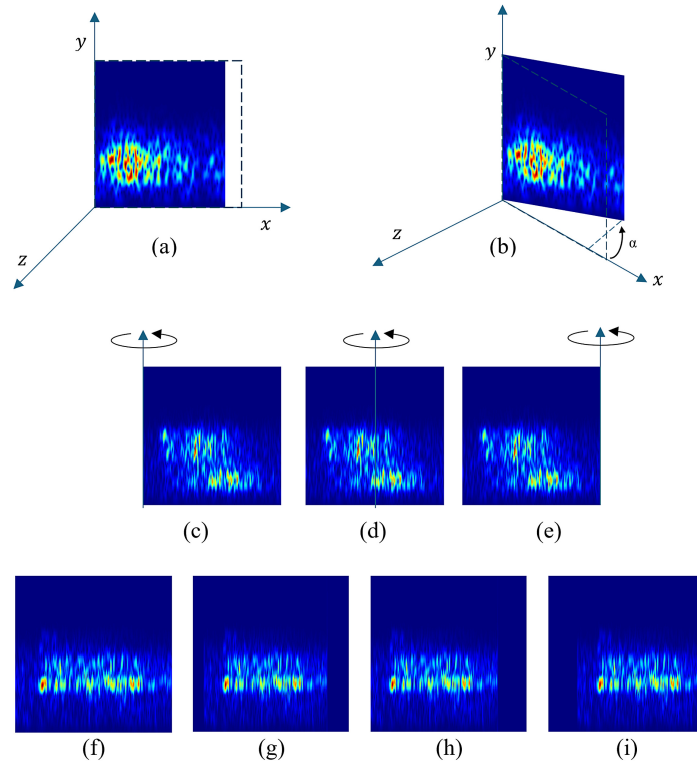


Fig. S3 Rotation transformation or time stretching (a) produced by applying a rotation angle (b). Possible rotation axes are located at the beginning (c), center (d), and end (e) of the spectrogram. Examples include (f) the original spectrogram, (g) a 19% rotation around the central axis, (h) a 23% rotation around the initial axis, and (i) a 24% rotation around the final axis.

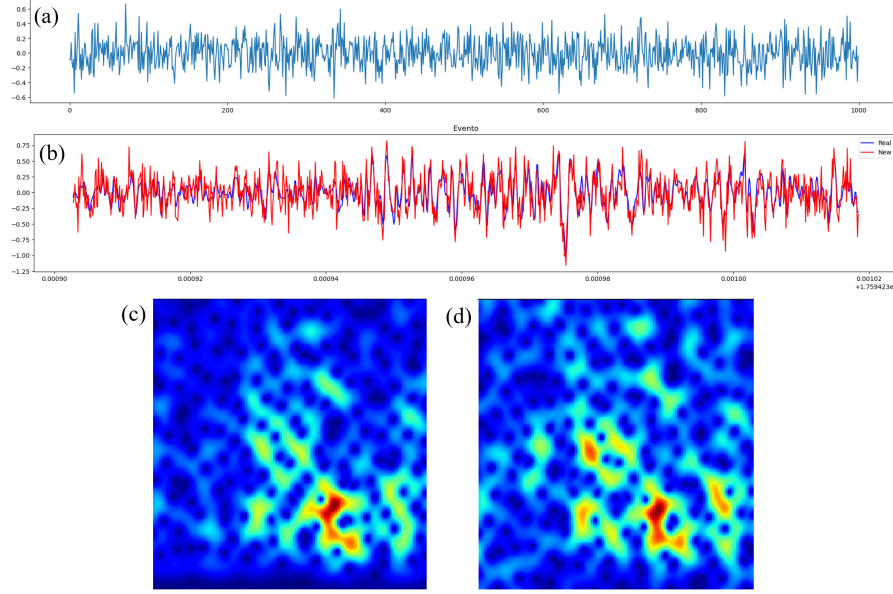


Fig. S4 Jittering-based signal generation. (a) Gaussian noise with distribution $N(\mu = 0, \sigma = 0.2)$. (b) Original (blue) and generated (red) signals. (c) Original-signal spectrogram. (d) Jittered-signal spectrogram.

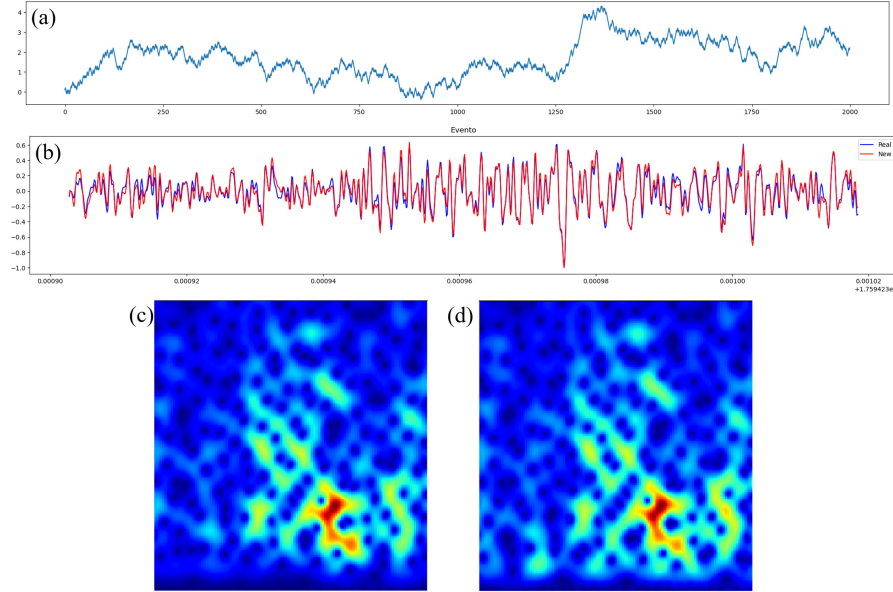


Fig. S5 Drifting-based signal generation. (a) Drifting signal with $\tau = 0.2$. (b) Original (blue) and generated (red) signals. (c) Original-signal spectrogram. (d) Drifted-signal spectrogram.

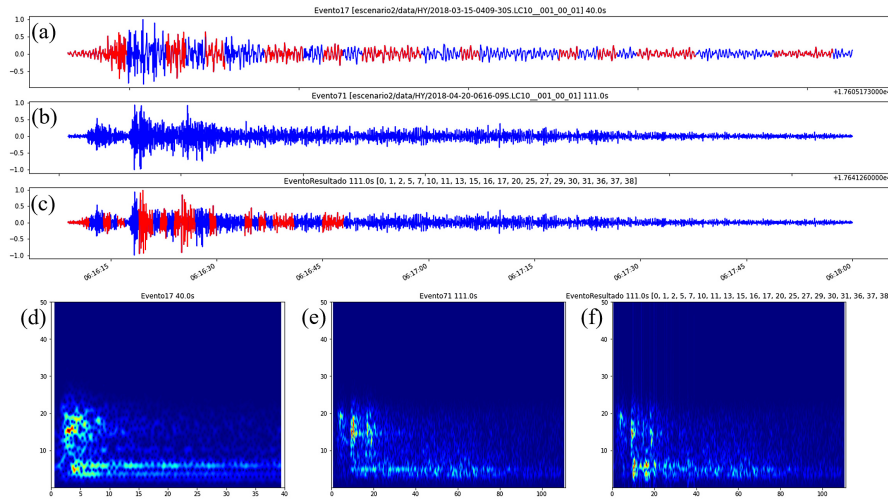


Fig. S6 Signal generation using a Genetic Algorithm without time adjustment. (a,b) Parent signals, with red highlighting exchanged segments. (c) Generated signal. (d,e) Parent-signal spectrograms. (f) Generated-signal spectrogram.

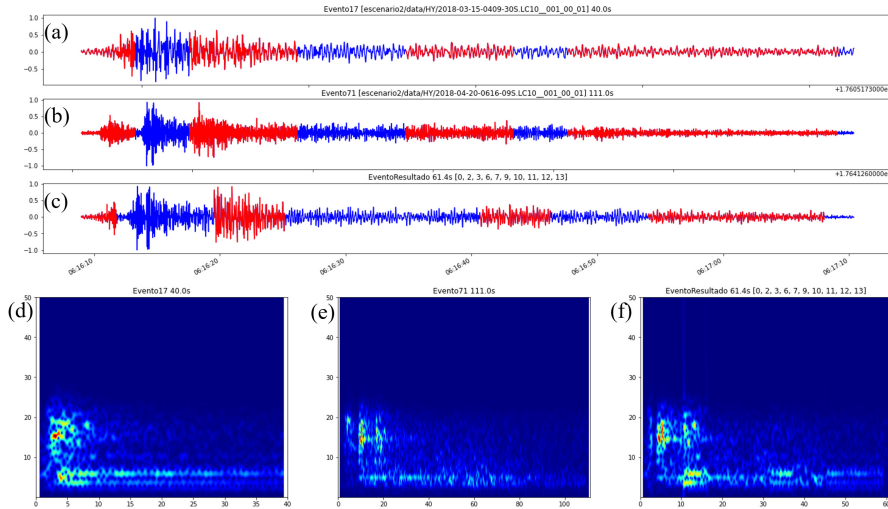


Fig. S7 Signal generation using a Genetic Algorithm with time adjustment. (a,b) Parent signals, with red highlighting exchanged segments. (c) Generated signal. (d,e) Parent-signal spectrograms. (f) Generated-signal spectrogram.

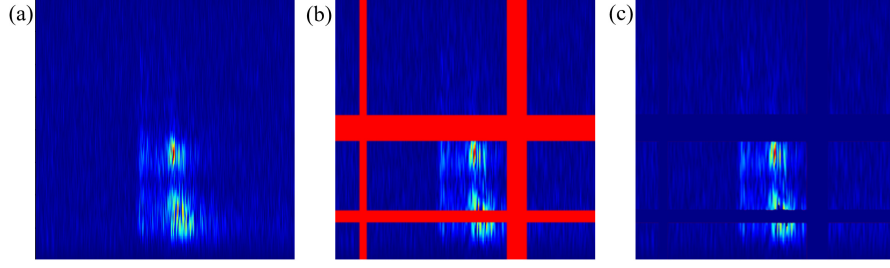


Fig. S8 Signal generation using SpecAugment. (a) Original spectrogram. (b) Spectrogram with two time-frequency distortions highlighted in red. (c) Spectrogram with two time-frequency distortions with white noise.

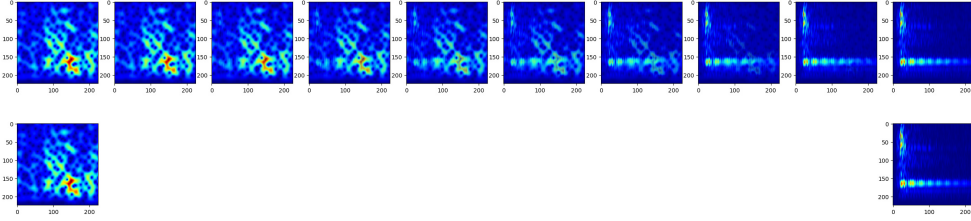


Fig. S9 Interpolating events (HY) at different intensity levels yields greater diversity.

G Model performance – Metrics

The following performance metrics are available:

- *Confusion Matrix*: A tabular representation of true labels versus model predictions. Although not a performance metric in itself, it serves as the foundation for deriving many other evaluation measures (see Table S4).

Table S4 Confusion matrix.

		Predicted	
		Positive (P)	Negative (N)
Actual	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

Where TP denotes true positives, TN true negatives, FP false positives, and FN false negatives.

- *Recall (Sensitivity)*: Measures the proportion of actual positive cases that the model correctly identifies.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

- *Precision*: Measures the proportion of positive predictions that are correct, indicating the model’s reliability in classification tasks.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

- F_β : Combines precision and recall into a single value, facilitating comparison across models. The parameter $\beta > 0$ adjusts the relative importance of recall versus precision: values $\beta > 1$ favor recall, while $\beta < 1$ favor precision. This metric is particularly useful for imbalanced datasets.

$$F_\beta = (1 + \beta^2) \cdot \frac{precision \cdot recall}{\beta^2 \cdot precision + recall} \quad (6)$$

- *Accuracy*: Measures the overall proportion of correctly classified cases. However, it can be misleading when dealing with imbalanced datasets.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

- *Specificity*: Measures the proportion of actual negative cases that the model correctly identifies.

$$Specificity = \frac{TN}{TN + FP} \quad (8)$$

- *Cohen’s kappa*: Measures the agreement between the model and human evaluators, taking randomness into account. It is particularly useful for imbalanced datasets.

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \quad (9)$$

where p_0 is the empirical probability of agreement and p_e is the expected agreement when both annotators assign labels randomly.

H Similarity and Diversity metrics

H.1 Fréchet Inception Distance (FID)

The Fréchet Inception Distance (FID) is a widely used metric for evaluating the quality of images generated by Generative Adversarial Networks (GANs) [24]. Its primary objective is to quantify the similarity between generated and real images [25]. Compared to earlier metrics such as the Inception Score (IS), the FID offers an improvement because it leverages statistics from real-world samples for comparison, whereas the IS does not [26]. The core idea of the FID is to embed both real and generated images into a vision-relevant feature space and then compute the distance between the distributions of these embeddings. In practice, this feature space is typically the penultimate layer (pool3, with 2048 features) of an Inception-V3 classifier pre-trained on ImageNet [26].

$$FID(\mu_r, \Sigma_r, \mu_g, \Sigma_g) = \|\mu_r - \mu_g\|_2^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}}) \quad (10)$$

where (μ_r, Σ_r) and (μ_g, Σ_g) denote the sample mean and covariance of the embeddings of the real and generated data, respectively, and $Tr(\cdot)$ indicates the matrix trace [26]. The theoretical range of the FID is $[0, +\infty)$, as it is a distance measure. Lower values correspond to higher-quality images [27]. In theory, an FID of 0 indicates that the feature distributions of the generated and real images are identical in terms of mean and covariance under the Gaussian approximation [25]. Empirical studies have shown that FID correlates well with human judgments of image fidelity [25].

H.2 Kernel Inception Distance (KID)

The squared Maximum Mean Discrepancy (MMD) is used by the Kernel Inception Distance (KID) to measure the difference between the feature representations of real and generated images. These features are extracted from the Inception network. Unlike the Fréchet Inception Distance (FID) [25], KID employs an unbiased estimator, which makes it more reliable, particularly when the number of test images is smaller than the dimensionality of the Inception features. A lower KID score indicates greater visual similarity between real and generated images [28]. KID uses a polynomial kernel $k(x, y) = (\frac{1}{d}x^T \cdot y + 1)^3$ where d is the dimension of the Inception representation vector, and x and y are two feature vectors. The kernel $K(x, y) = k(\phi(x), \phi(y))$ can then be applied as an MMD for input images, where images are mapped into Inception representations through ϕ [29]. Since KID is a distance metric, a value of 0 indicates a perfect match between the distributions of real and generated features. Theoretically, if the two distributions are identical, the distance equals zero. A key advantage of KID lies in its simple and unbiased estimator, which distinguishes it from FID [29].

H.3 Multi-Scale Structural Similarity (MS-SSIM)

MS-SSIM is a multi-scale improvement of SSIM that aims to eliminate aspects of an image that are not crucial for human perception [30, 31]. The calculation of SSIM, which serves as the basis for MS-SSIM, compares three principal components: luminance (intensity), contrast, and structure. It employs basic statistical moments, such as the mean and standard deviation, within local windows to derive a similarity score.

$$MS-SSIM(x, y) = [l_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} \cdot [s_j(x, y)]^{\gamma_j} \quad (11)$$

where $l_M(x, y)$ represents the luminance component between images x and y , $c_j(x, y)$ denotes the contrast component at scale j , and $s_j(x, y)$ reflects the structural similarity in scale j . The parameters α_M , β_j , and γ_j define the relative importance of the luminance, contrast, and structure components, respectively. Lastly, M signifies the number of scales employed in the MS-SSIM calculation [32]. The range of MS-SSIM values is between 0 and 1, and images with higher MS-SSIM values are perceived as more similar [33], with 1 indicating perfect similarity. Lower MS-SSIM scores indicate higher diversity [27, 34]. Unlike metrics such as FID (Fréchet Inception Distance) or KID (Kernel Inception Distance), MS-SSIM does not require a pre-trained Inception network [35]. MS-SSIM is applied to a fingerprint dataset [35].

Lower FID values indicate superior image quality, and lower MS-SSIM scores reflect increased diversity [34]. Smaller FID and KID values correlate with more realistic images in image translation [36]. Therefore, the use of IS, KID, or FID is not recommended into datasets outside ImageNet, as these metrics inherently utilize weights from the Inception network, which are validated for images similar to those in the ImageNet dataset.

I Training models

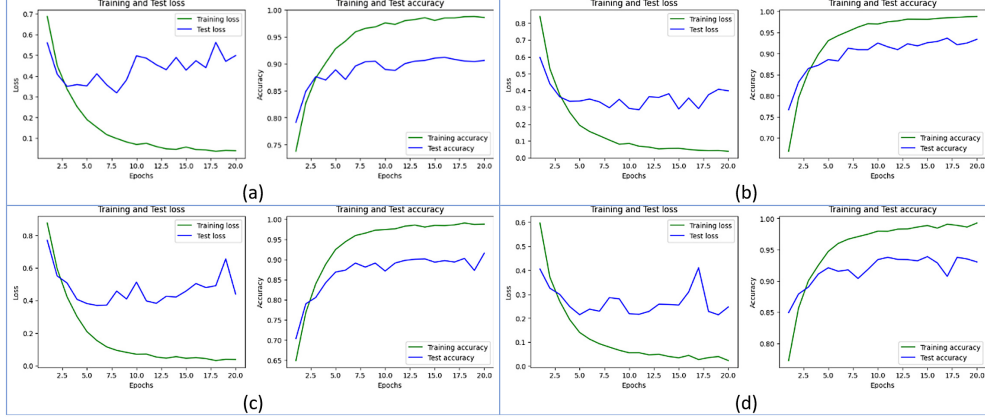


Fig. S10 Model Loss and Accuracy graph for Training and Validation using transfer learning with AlexNet on dataset: (a) Drifting Drift(0.001-0.01)(Real+Augmented)(viii), (b) Drifting Drift(0.001-0.01)(Augmented)(ix), (c) SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)(Augmented)(xxi), and (d) Interpolation AE Signal Percent(40%-60%)(Augmented)(xxv).

Table S5 Transfer-learning settings for the AlexNet pretrained model.

Dataset	Data type	Data Augmentation	Balanced dataset	Epochs	Total data	Train data	Test data	Training time (min.)
i	Real	No	NO	10	6,145	4,916	1,229	4.2
ii	Real + Augmented	Rotation(5%-25%)	YES	20	13,430	10,744	2,686	18.7
iii	Augmented	Rotation(5%-25%)	YES	20	13,430	10,744	2,686	20.9
iv	Real + Augmented	Jittering(0.2)	YES	20	13,430	10,744	2,686	16.8
v	Augmented	Jittering(0.2)	YES	20	13,430	10,744	2,686	17.2
vi	Real + Augmented	Drifting Drift(0.01-0.1)	YES	20	13,430	10,744	2,686	17.9
vii	Augmented	Drifting Drift(0.01-0.1)	YES	20	13,430	10,744	2,686	20.3
viii	Real + Augmented	Drifting Drift(0.001-0.01)	YES	20	13,430	10,744	2,686	21.7
ix	Augmented	Drifting Drift(0.001-0.01)	YES	20	13,430	10,744	2,686	21.0
x	Real + Augmented	Drifting Drift(0.0001-0.001)	YES	20	13,430	10,744	2,686	18.1
xi	Augmented	Drifting Drift(0.0001-0.001)	YES	20	13,430	10,744	2,686	20.9
xii	Real + Augmented	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	YES	20	13,430	10,744	2,686	16.7
xiii	Augmented	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	YES	20	13,430	10,744	2,686	17.0
xiv	Real + Augmented	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	YES	20	13,430	10,744	2,686	16.6
xv	Augmented	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	YES	20	13,430	10,744	2,686	16.9
xvi	Real + Augmented	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	YES	20	13,430	10,744	2,686	16.6
xvii	Augmented	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	YES	20	13,430	10,744	2,686	17.3
xviii	Real + Augmented	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	YES	20	13,430	10,744	2,686	16.5
xix	Augmented	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	YES	20	13,430	10,744	2,686	17.1
xx	Real + Augmented	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)	YES	20	13,430	10,744	2,686	16.5
xxi	Augmented	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)	YES	20	13,430	10,744	2,686	17.0
xxii	Real + Augmented	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Masks(2)	YES	20	13,430	10,744	2,686	16.6
xxiii	Augmented	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Masks(2)	YES	20	13,430	10,744	2,686	16.9
xxiv	Real + Augmented	Interpolation AE Signal percent(40%-60%)	YES	20	13,430	10,744	2,686	16.6
xxv	Augmented	Interpolation AE Signal percent(40%-60%)	YES	20	13,430	10,744	2,686	17.9

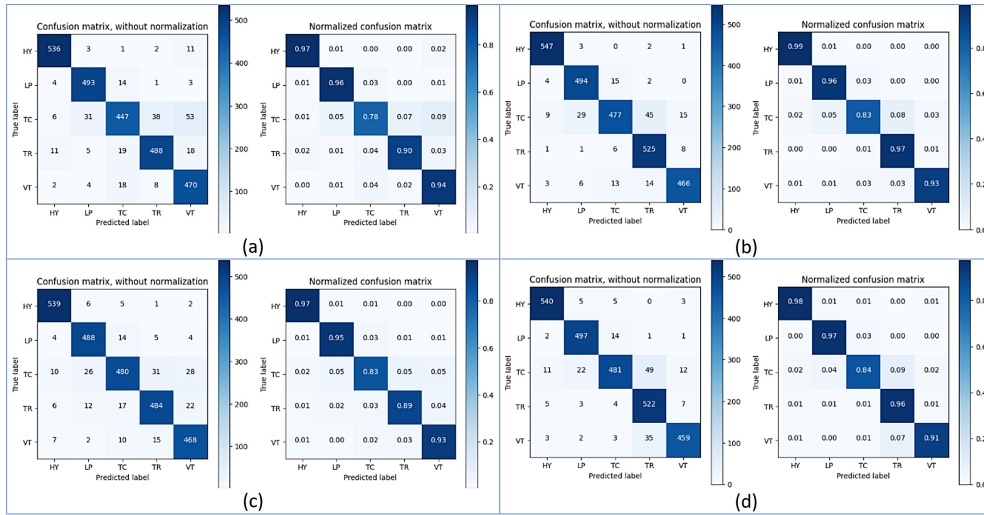


Fig. S11 Normalized and non-normalized confusion matrix of model training using transfer learning with AlexNet on dataset: (a) Drifting Drift(0.001-0.01)(Real+Augmented)(viii), (b) Drifting Drift(0.001-0.01)(Augmented)(ix), (c) SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)(Augmented)(xxi), and (d) Interpolation AE Signal Percent(40%-60%)(Augmented)(xxv).

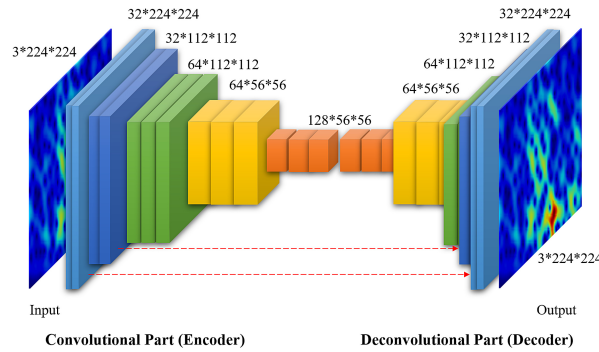


Fig. S12 Convolutional AutoEncoder of spectrograms.

J Similarity and Diversity metrics values

References

- [1] Bicego, M., Acosta-Munoz, C. & Orozco-Alzate, M. Classification of seismic volcanic signals using hidden-markov-model-based generative embeddings. *IEEE transactions on geoscience and remote sensing* **51**, 3400–3409 (2012).

Table S6 Similarity and Diversity internal metrics for experimental datasets.

Dataset	Data augmentation	Fr�chet Inception Distance (FID)					Kernel Inception Distance (KID)					Multiscale Structural Similarity Index (MS-SSIM)							
		HY	LP	TC	TR	VT	FID	HY	LP	TC	TR	VT	KID	HY	LP	TC	TR	VT	MS
i	No	0.16	0.10	0.02	0.05	0.02	0.03	0.14	-0.75	3.78	0.31	-0.23	1.21	0.34	0.41	0.33	0.29	0.34	0.34
ii	Rotation(5%-25%)	0.70	0.61	0.50	0.45		0.56	4.81	-1.40	4.41	2.56		2.59	0.38	0.42	0.36	0.29	0.36	0.36
iii	Rotation(5%-25%)	0.03	0.01	0.04	0.02	0.03	0.02	1.62	0.15	1.16	-1.03	2.34	0.85	0.36	0.42	0.36	0.33	0.35	0.37
iv	Jittering(0.2)	10.09	5.46	7.65	8.88		8.02	56.99	18.80	40.98	48.72		41.37	0.28	0.36	0.27	0.26	0.29	0.29
v	Jittering(0.2)	0.01	0.01	0.02	0.02	0.03	0.02	-2.50	-2.35	3.38	0.24	-2.77	-0.80	0.26	0.33	0.26	0.21	0.23	0.26
vi	Drifting Drift(0.01-0.1)	34.80	30.48	35.93	30.01		32.81	102.53	71.83	89.02	83.60		86.75	0.40	0.46	0.42	0.35	0.41	0.41
vii	Drifting Drift(0.01-0.1)	0.01	0.02	0.01	0.01	0.01	0.01	-0.06	-0.10	0.29	0.51	-0.11	0.11	0.80	0.81	0.81	0.76	0.71	0.78
viii	Drifting Drift(0.001-0.01)	1.50	1.23	2.69	0.84		1.56	3.25	3.61	3.84	1.01		2.93	0.37	0.42	0.38	0.31	0.37	0.37
ix	Drifting Drift(0.001-0.01)	0.03	0.01	0.02	0.03	0.01	0.02	1.53	-1.02	0.30	1.14	-0.80	0.23	0.42	0.48	0.43	0.37	0.36	0.41
x	Drifting Drift(0.0001-0.001)	0.02	0.02	0.03	0.02	0.01	0.02	-4.42	4.68	-3.96	0.79	1.74	-0.73	0.35	0.40	0.32	0.31	0.34	0.34
xi	Drifting Drift(0.0001-0.001)	0.01	0.01	0.02	0.01	0.01	0.01	-5.60	-1.65	2.42	3.00	1.74	-0.02	0.35	0.40	0.32	0.31	0.34	0.34
xii	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	0.89	0.41	0.39	0.28		0.49	2.36	2.98	-1.94	2.55		1.49	0.38	0.40	0.35	0.30	0.36	0.36
xiii	Genetic Algorithm Segment Size(5s) Crossover Segments(3)	0.02	0.01	0.00	0.01	0.01	0.01	0.45	-2.49	-0.40	-2.06	2.54	-0.39	0.35	0.40	0.37	0.30	0.36	0.36
xiv	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	1.65	0.83	0.76	1.03		1.07	1.51	4.59	2.90	1.29		2.57	0.35	0.41	0.35	0.30	0.35	0.35
xv	Genetic Algorithm Segment Size(5s) Crossover Segments(5)	0.02	0.01	0.01	0.02	0.01	0.01	1.06	-0.48	-0.32	-2.37	-1.77	-0.77	0.35	0.40	0.37	0.33	0.40	0.37
xvi	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	0.90	0.63	0.55	0.41		0.63	3.99	1.91	0.08	4.25		2.56	0.36	0.41	0.35	0.30	0.35	0.35
xvii	Genetic Algorithm Signal Percent(40%) Crossover Segments(10)	0.02	0.02	0.04	0.01	0.01	0.02	-0.28	-3.37	-0.14	1.87	-0.70	-0.52	0.34	0.41	0.32	0.29	0.34	0.34
xviii	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	0.97	0.67	0.73	0.36		0.68	4.43	-0.28	3.54	0.97		2.16	0.35	0.41	0.33	0.30	0.34	0.34
xix	Genetic Algorithm Signal Percent(45%) Crossover Segments(10)	0.01	0.00	0.00	0.01	0.01	0.01	0.91	0.75	-0.73	2.75	0.86	0.91	0.31	0.41	0.35	0.30	0.34	0.34
xx	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)	1.53	1.48	1.85	1.21		1.52	6.23	4.41	6.33	2.06		4.76	0.37	0.41	0.33	0.30	0.34	0.34
xxi	SpecAugment Frequency Percent(0%-10%) Frequency Masks(2) Time Percent(0%-10%) Time Masks(2)	0.00	0.01	0.02	0.01	0.01	0.01	-1.21	2.46	-0.33	-1.04	1.03	0.18	0.34	0.40	0.35	0.31	0.34	0.35
xxii	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Masks(2)	3.84	3.45	3.08	3.22		3.40	16.73	11.34	7.82	6.15		10.51	0.35	0.41	0.33	0.31	0.35	0.35
xxiii	SpecAugment Frequency Percent(0%-15%) Frequency Masks(2) Time Percent(0%-15%) Time Masks(2)	0.01	0.01	0.00	0.01	0.01	0.01	0.31	1.29	0.47	0.72	-0.61	0.44	0.35	0.40	0.36	0.32	0.39	0.37
xxiv	Interpolation AE Signal Percent(40%-60%)	2.31	2.46	2.18	1.88		2.21	9.48	3.04	4.03	7.12		5.92	0.38	0.44	0.36	0.33	0.38	0.38
xxv	Interpolation AE Signal Percent(40%-60%)	0.01	0.01	0.01	0.01	0.01	0.01	-0.22	0.62	-0.89	-0.44	0.03	-0.18	0.43	0.48	0.41	0.37	0.41	0.42

- [2] 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).
- [3] Malfante, M. *et al.* Automatic classification of volcano seismic signatures. *Journal of Geophysical Research: Solid Earth* **123**, 10–645 (2018).
- [4] Curilem, M., Canário, J. P., Franco, L. & Rios, R. A. Using cnn to classify spectrograms of seismic events from llaima volcano (chile). In *2018 International Joint Conference on Neural Networks (IJCNN)*, 1–8 (2018).
- [5] Titos, M., Bueno, A., García, L., Benítez, M. C. & Ibañez, J. Detection and classification of continuous volcano-seismic signals with recurrent neural networks. *IEEE Transactions on Geoscience and Remote Sensing* **57**, 1936–1948 (2018).
- [6] Venegas, P., Pérez, N., Benítez, D., Lara-Cueva, R. & Ruiz, M. Combining filter-based feature selection methods and gaussian mixture model for the classification of seismic events from cotopaxi volcano. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **12**, 1991–2003 (2019).
- [7] 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).
- [8] Lara, P. E. E. *et al.* Automatic multichannel volcano-seismic classification using machine learning and emd. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **13**, 1322–1331 (2020).
- [9] Salazar, A., Arroyo, R., Pérez, N. & Benítez, D. Deep-learning for volcanic seismic events classification. In *2020 IEEE Colombian Conference on Applications of Computational Intelligence (IEEE ColCACI 2020)*, 1–6 (IEEE, 2020).
- [10] Titos, M., Bueno, A., García, L., Benítez, C. & Segura, J. C. Classification of isolated volcano-seismic events based on inductive transfer learning. *IEEE Geoscience and Remote Sensing Letters* **17**, 869–873 (2019).
- [11] 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).
- [12] 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>.

- [13] 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). URL <https://www.sciencedirect.com/science/article/pii/S0377027319306171>.
- [14] 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).
- [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).
- [16] Titos, M., Garcia, L., Kowsari, M. & Benitez, C. Toward knowledge extraction in classification of volcano-seismic events: Visualizing hidden states in recurrent neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **15**, 2311–2325 (2022).
- [17] 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.
- [18] Salazar, P., Yupanqui, F., Meneses, C., Layana, S. & Yáñez, G. Multi-station automatic classification of seismic signatures from the lascar volcano database. *Natural Hazards and Earth System Sciences Discussions* **2022**, 1–27 (2022).
- [19] 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).
- [20] Espinosa-Curilem, C., Curilem, M. & Basualto, D. A framework for real-time volcano-seismic event recognition based on multi-station seismograms and semantic segmentation models. *arXiv preprint arXiv:2410.20595* (2024).
- [21] Malfante, M. *et al.* Machine learning for volcano-seismic signals: Challenges and perspectives. *IEEE Signal Processing Magazine* **35**, 20–30 (2018).
- [22] Wassermann, J. Volcano seismology In *New manual of seismological observatory practice 2 (NMSOP-2)* 1–77 (Deutsches GeoForschungsZentrum GFZ, 2012).
- [23] Iglesias, G., Talavera, E., González-Prieto, Á., Mozo, A. & Gómez-Canaval, S. Data augmentation techniques in time series domain: a survey and taxonomy. *Neural Computing and Applications* **35**, 10123–10145 (2023). URL <http://dx.doi.org/10.1007/s00521-023-08459-3>.

- [24] Güven, S. A., Şahin, E. & Talu, M. F. Image-to-image translation with cnn based perceptual similarity metrics. *Computer Science* **9**, 84–98 (2024).
- [25] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* **30** (2017). URL <https://arxiv.org/abs/1706.08500>.
- [26] Kynkäänniemi, T., Karras, T., Aittala, M., Aila, T. & Lehtinen, J. The role of imagenet classes in fréchet inception distance (2023). URL <https://arxiv.org/abs/2203.06026>. arXiv:2203.06026.
- [27] Wang, W., Sun, Y. & Halgamuge, S. Improving mmd-gan training with repulsive loss function (2019). URL <https://arxiv.org/abs/1812.09916>. arXiv:1812.09916.
- [28] Kim, J., Kim, M., Kang, H. & Lee, K. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. *arXiv preprint arXiv:1907.10830* (2020). URL <https://arxiv.org/abs/1907.10830>.
- [29] Bińkowski, M., Sutherland, D. J., Arbel, M. & Gretton, A. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401* (2021). URL <https://arxiv.org/abs/1801.01401>.
- [30] Wang, Z., Simoncelli, E. P. & Bovik, A. C. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, Vol. 2, 1398–1402 (Ieee, 2003).
- [31] Ma, K. *et al.* Group mad competition-a new methodology to compare objective image quality models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1664–1673 (2016).
- [32] Arsénio, M., Vigário, R. & Mota, A. M. Recovering image quality in low-dose pediatric renal scintigraphy using deep learning. *Journal of Imaging* **11**, 88 (2025).
- [33] Odena, A., Olah, C. & Shlens, J. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, 2642–2651 (PMLR, 2017). URL <https://arxiv.org/abs/1610.09585>. arXiv:1610.09585.
- [34] Guo, Y. *et al.* Auto-embedding generative adversarial networks for high resolution image synthesis. *IEEE Transactions on Multimedia* **21**, 2726–2737 (2019).
- [35] Fahim, M. A.-N. I. & Jung, H. Y. A lightweight gan network for large scale fingerprint generation. *IEEE Access* **8**, 92918–92928 (2020).

967 [36] Torbunov, D. *et al.* Rethinking cyclegan: Improving quality of gans for unpaired
968 image-to-image translation. *arXiv preprint arXiv:2303.16280* (2023).
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012