



Universidad de **Nariño**
FUNDADA EN 1904



Universidad de **Nariño**

ACREDITADA EN ALTA CALIDAD
RESOLUCIÓN MEN 000022 - ENERO 11 DE 2023



FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE FÍSICA
TRABAJO DE GRADO



Clasificación de la sismicidad del volcán Galeras mediante redes neuronales artificiales tipo Transformer

Daniel Esteban Chapid Vela
Director: Dr. Oscar Ernesto Cadena Ibarra



Índice

1. Introducción
2. Problema y objetivo principal
3. Objetivos específicos
4. Marco teórico
5. Desarrollo metodológico
6. Resultados
7. Conclusiones

1. Introducción

1

¿Por qué es importante clasificar la sismicidad volcánica?

- Los volcanes activos producen distintos tipos de señales sísmicas.
- La dinámica interna de un volcán se manifiesta a través de señales detectables.
- Clasificar eventos sísmicos es clave para el monitoreo volcánico.
- Comprender la actividad volcánica depende de identificar correctamente sus señales.

2. Problema y objetivo principal

El proyecto pretende responder la siguiente pregunta:

¿Cómo se pueden clasificar de manera automática y precisa los distintos tipos de sismos volcánicos generados en el volcán Galeras, empleando redes neuronales artificiales tipo Transformer?

El objetivo general es:

Clasificar los diferentes tipos de sismos generados en el volcán Galeras (Nariño, Colombia) utilizando redes neuronales artificiales tipo Transformer.

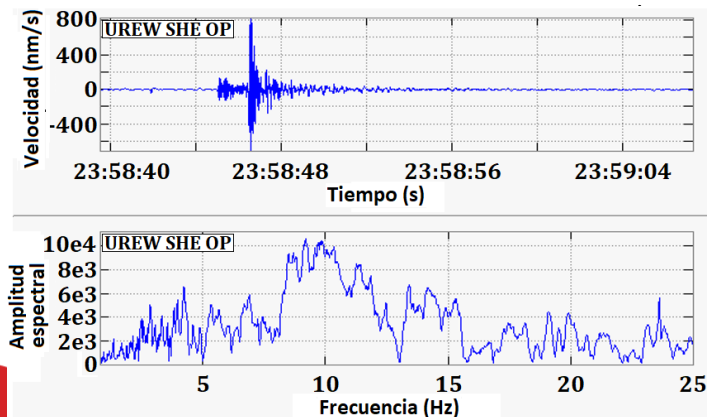
3. Objetivos específicos

- Seleccionar y preprocesar información sísmica del volcán Galeras para conformar los conjuntos de contexto y prueba requeridos por el modelo Transformer.
- Diseñar e implementar un flujo de trabajo automatizado que integre el preprocesamiento, aprendizaje en contexto y la evaluación del modelo Transformer para la clasificación de sismos del volcán Galeras.
- Emplear una red neuronal artificial tipo Transformer para el análisis de las características extraídas y la posterior clasificación de los sismos volcánicos del volcán Galeras.
- Evaluar el desempeño del modelo Transformer mediante datos de prueba, con el fin de validar su efectividad en la clasificación de sismos volcánicos.

4. Marco teórico

Volcán Galeras y su clasificación sísmica

Sismos volcano-tectónicos (VT)



Sismos de Largo periodo (LP)

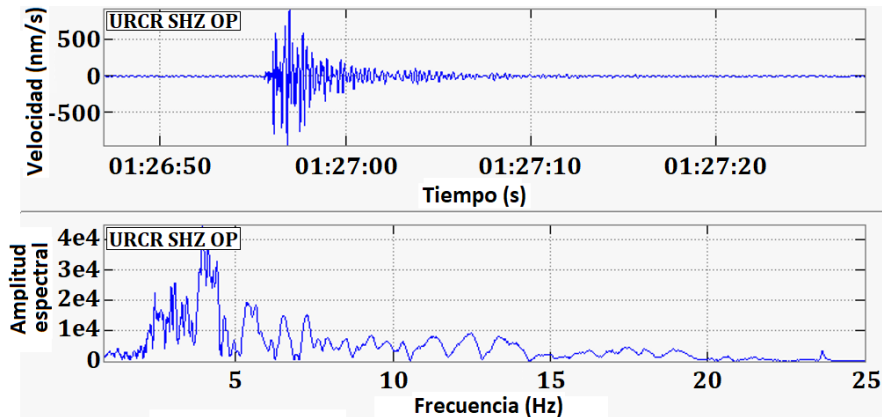
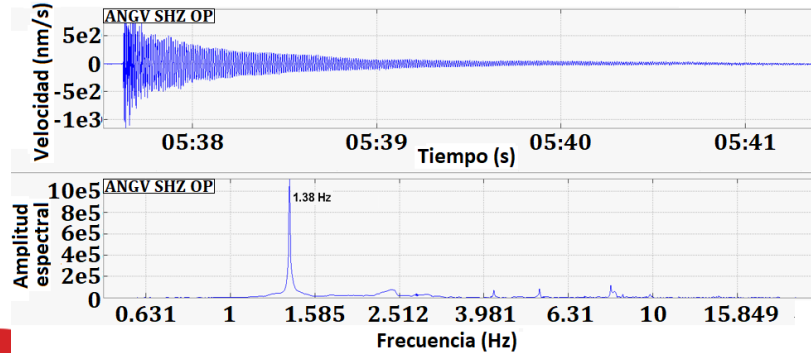


Figura 1. Sismo transitorios VT y LP registrado en la estación UREW. Tomado de [Cadena, 2021].

4. Marco teórico

Volcán Galeras y su clasificación sísmica

Sismos tipo Tornillo (TOR)



Tremor espasmódico (TRE)

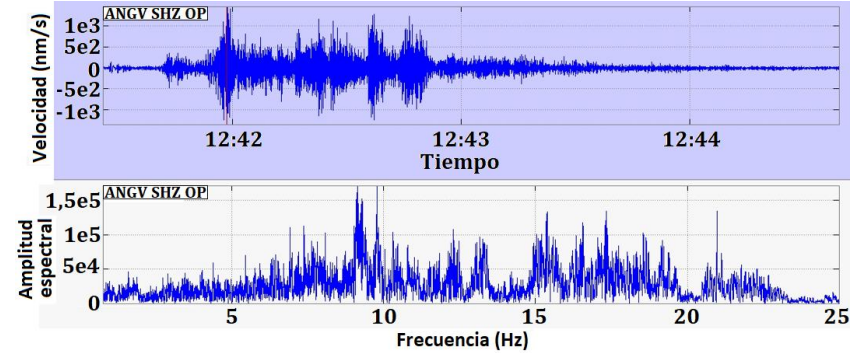


Figura 2. Sismo transitorio TOR y sismo continuo TRE registrado en la estación ANGV. Tomado de [Cadena, 2021].

4. Marco teórico

Arquitecturas de aprendizaje profundo

CNN

- Aprenden patrones locales en los datos.
- Eficientes para aprender características puntuales.

Arquitectura Transformer

- Procesan información de forma paralela.
- Analizan los datos en una sola toma.

RNN

- Procesan información de forma secuencia.
- Se limitan cuando contiene gran cantidad de datos.

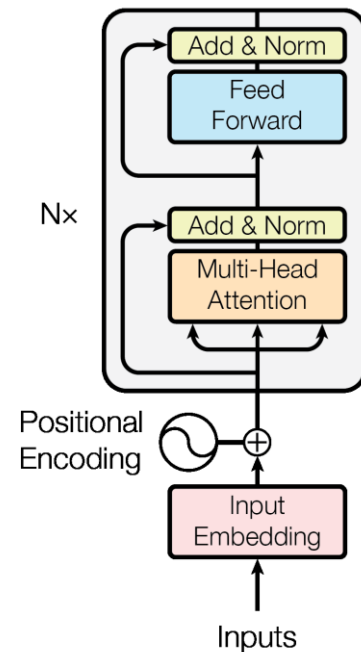


Figura 3. Arquitectura del modelo Transformer (codificador). Adaptado de [Vaswani, 2017].

Arquitectura Transformer

■ Mecanismo de atención

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

- Matriz de consultas.

$$\mathbf{Q} = \mathbf{X} \cdot \mathbf{W}^Q \in \mathbb{R}^{N \times d_k} \quad (2)$$

- Matriz de claves.

$$\mathbf{K} = \mathbf{X} \cdot \mathbf{W}^K \in \mathbb{R}^{N \times d_k} \quad (3)$$

- Matriz de valores.

$$\mathbf{V} = \mathbf{X} \cdot \mathbf{W}^V \in \mathbb{R}^{N \times d_v} \quad (4)$$

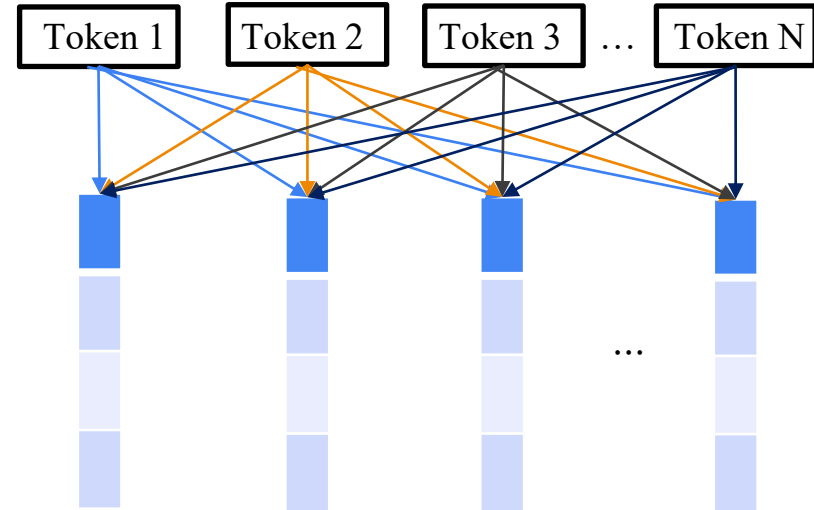
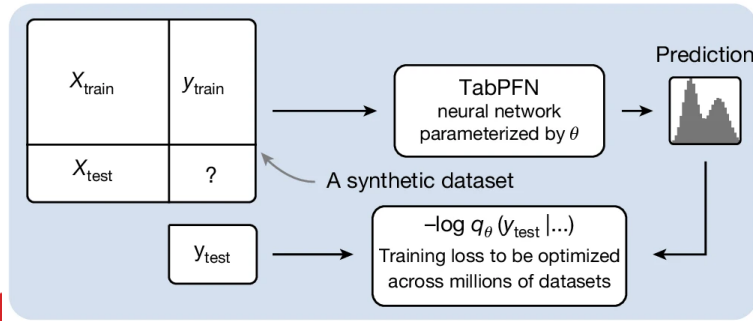


Figura 4. Auto-Atención: cada token actualiza su representación calculando su nivel de relevancia con todos los demás tokens de la secuencia.

Modelo TabPFN

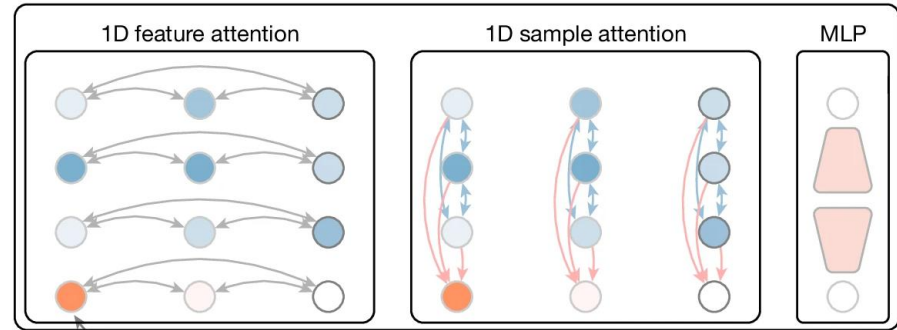
- Clasificador tabular
- Transformer pre-entrenado

TabPFN is trained on synthetic data to take entire datasets as inputs and predict in a forward pass



(a)

2D TabPFN layer (12x)

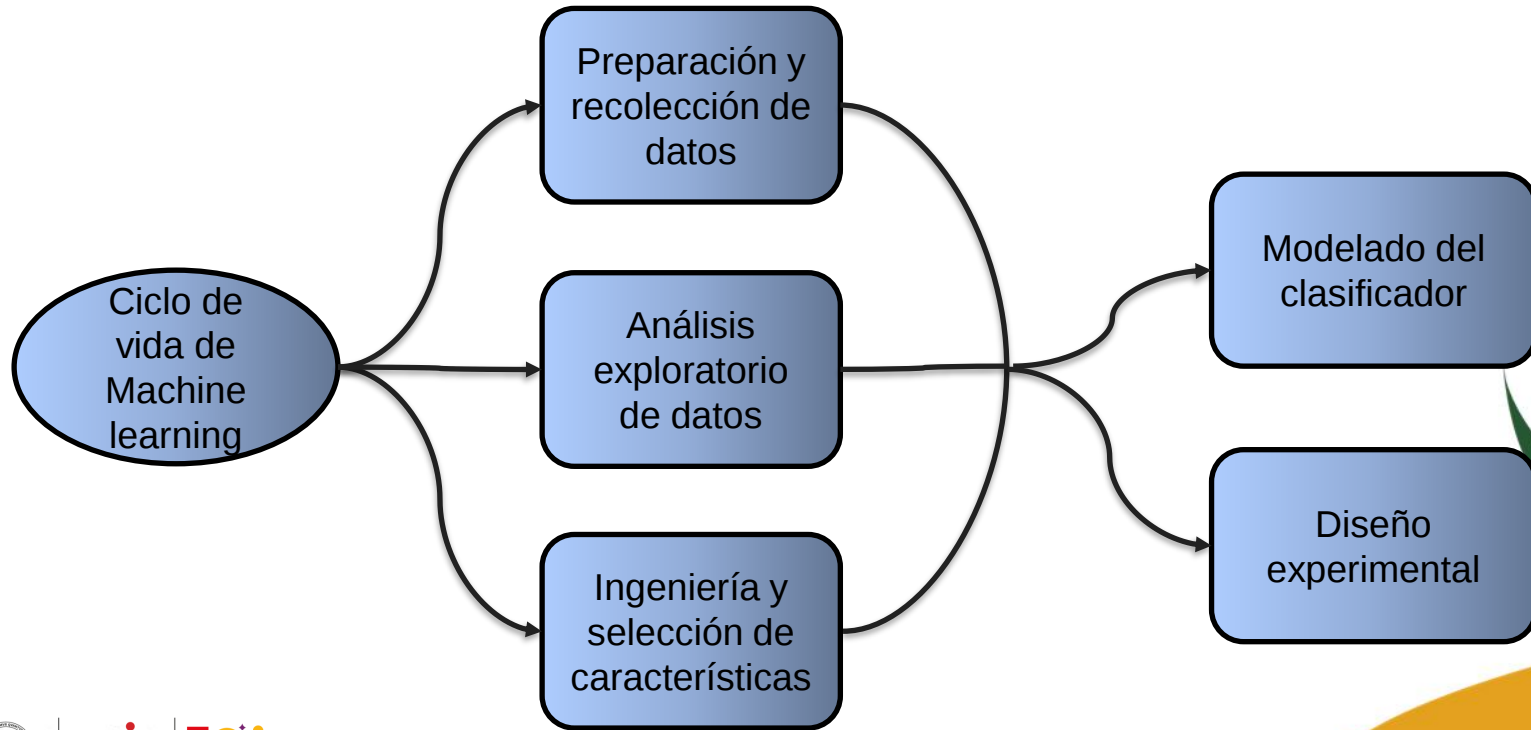


Each node represents one entry in the table

(b)

Figura 5. Panel (a): Flujo de preentrenamiento de TabPFN sobre datos sintéticos. Panel (b): Arquitectura interna de TabPFN con bloques 2D apilados que alternan atención sobre características y atención sobre muestras. Adaptado de [Hollmann et al., 2025].

5. Desarrollo metodológico



1. Preparación y recolección de datos

- 44 913 Señales sísmicas
- Años de clasificación 2006-2010

Clase	2006	2007	2008	2009	2010	Total	%
LP	1 854	3 334	11 866	8 112	2 998	28 164	62,7
TRE	--	7 058	4 454	--	--	11 512	25,6
VT	440	690	558	711	2 643	5 042	11,2
TOR	64	85	11	27	8	195	0,5
Total	2 358	11 167	16 889	8 850	5 649	44 913	100

Tabla 1. Distribución de eventos sísmicos por año y clase.

1. Preparación y recolección de datos

■ Preprocesamiento

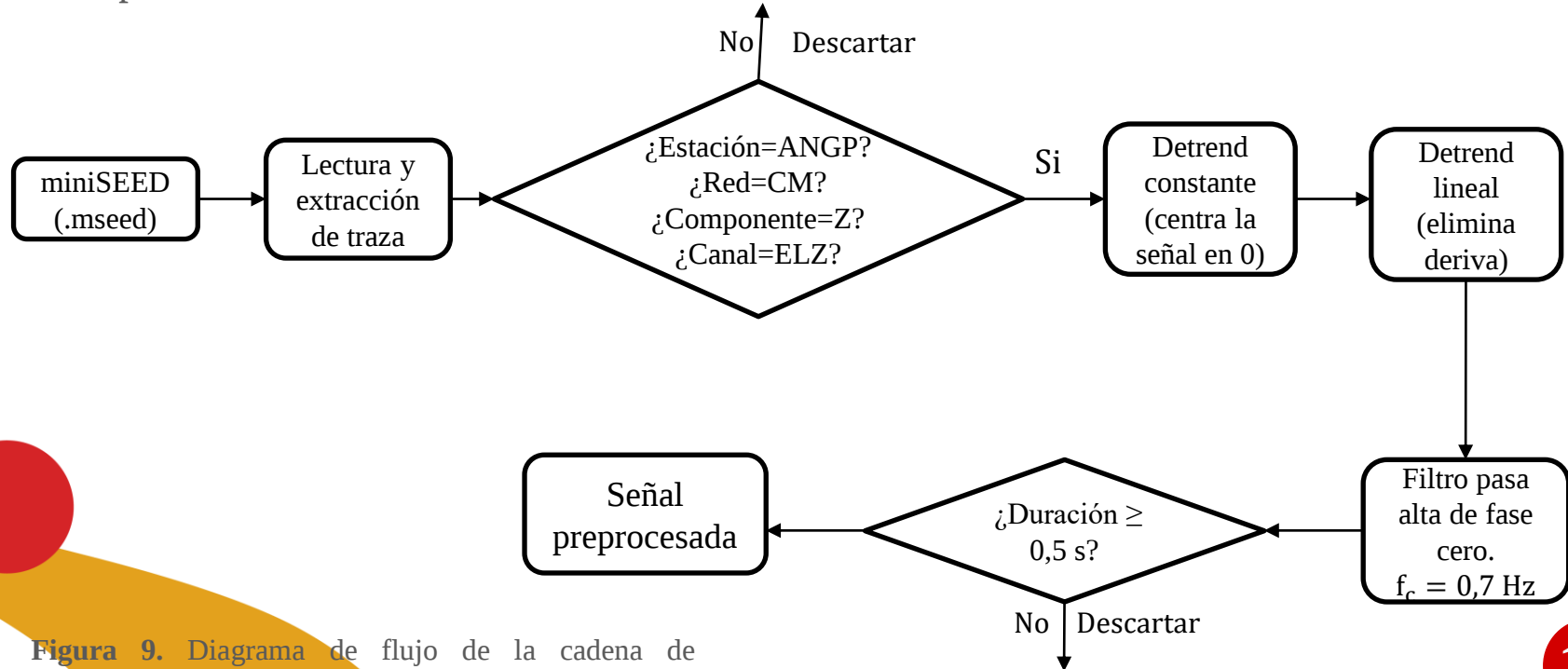
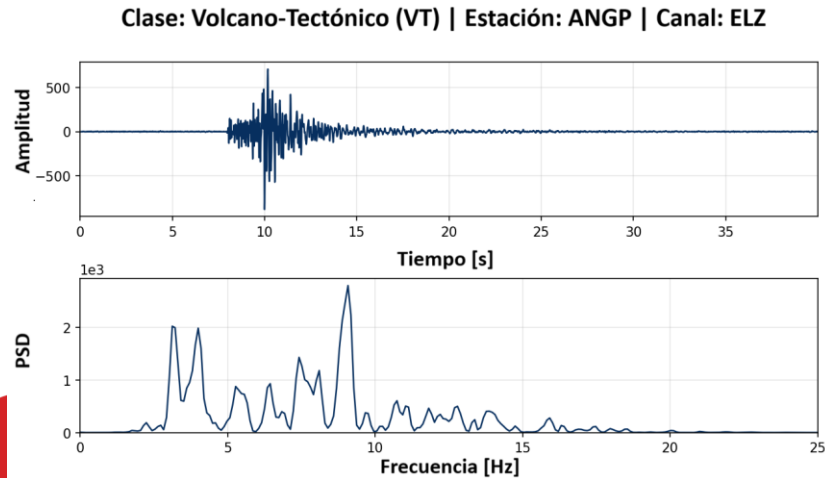


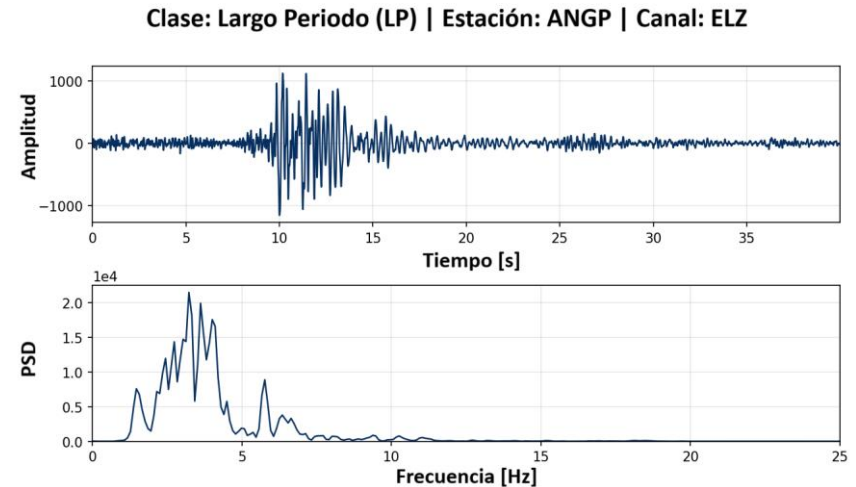
Figura 9. Diagrama de flujo de la cadena de preprocesamiento aplicada a cada traza miniSEED.

2. Análisis exploratorio de datos (EDA)

Validación visual



(a)

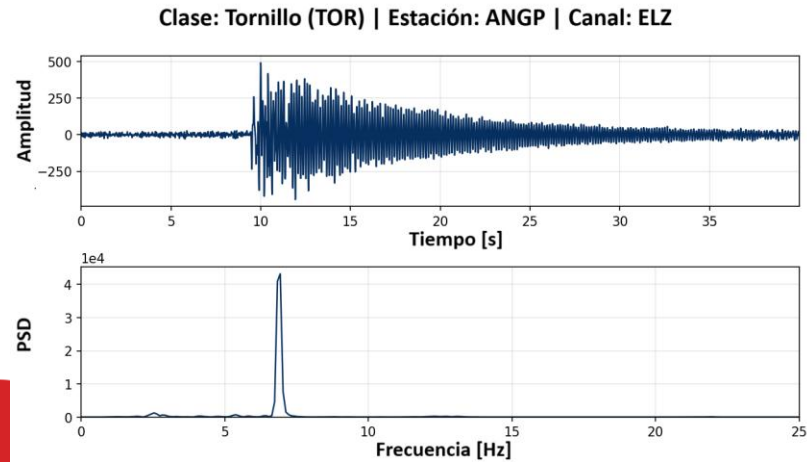


(b)

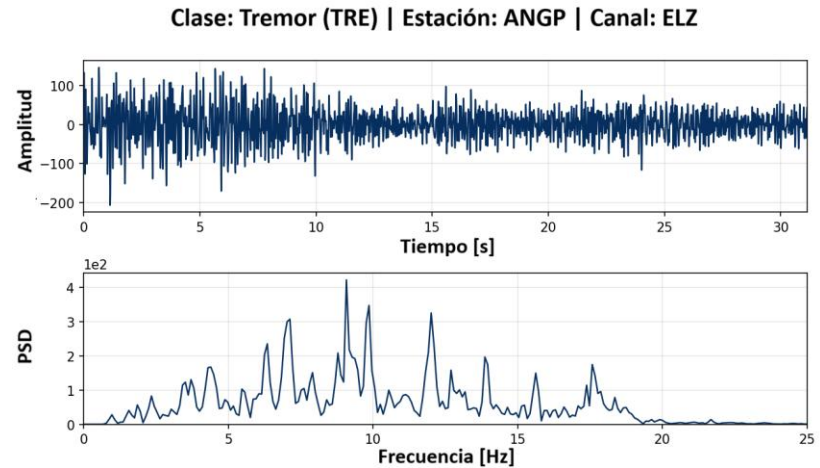
Figura 6. Sismograma y espectro de eventos a) VT y b) LP obtenidos de el catalogo proporcionado por OVSP.

2. Análisis exploratorio de datos (EDA)

Validación visual



(c)



(d)

Figura 7. Sismograma y espectro de eventos c) TOR y d) TRE obtenidos de el catalogo proporcionado por OVSP.

3. Ingeniería y selección de características

- Partición y validación cruzada de Monte Carlo (MCCV)
 - Proporción: 80 % contexto / 20 % prueba
 - 4 repeticiones con semillas independientes : {42, 123, 456, 789}.
- Estrategia de muestreo

$MAX_{TRZ} = 12\ 500$

Caso 1 — clase minoritaria por debajo del cupo ($n_{min} < x$):

$$N_{per-clas} = \frac{MAX_{TRZ} - n_{min}}{C - 1} \quad (5)$$

Caso 2 — clases balanceadas:

$$N_{per-clas} = \frac{MAX_{TRZ}}{C} \quad (6)$$

3. Ingeniería y selección de características

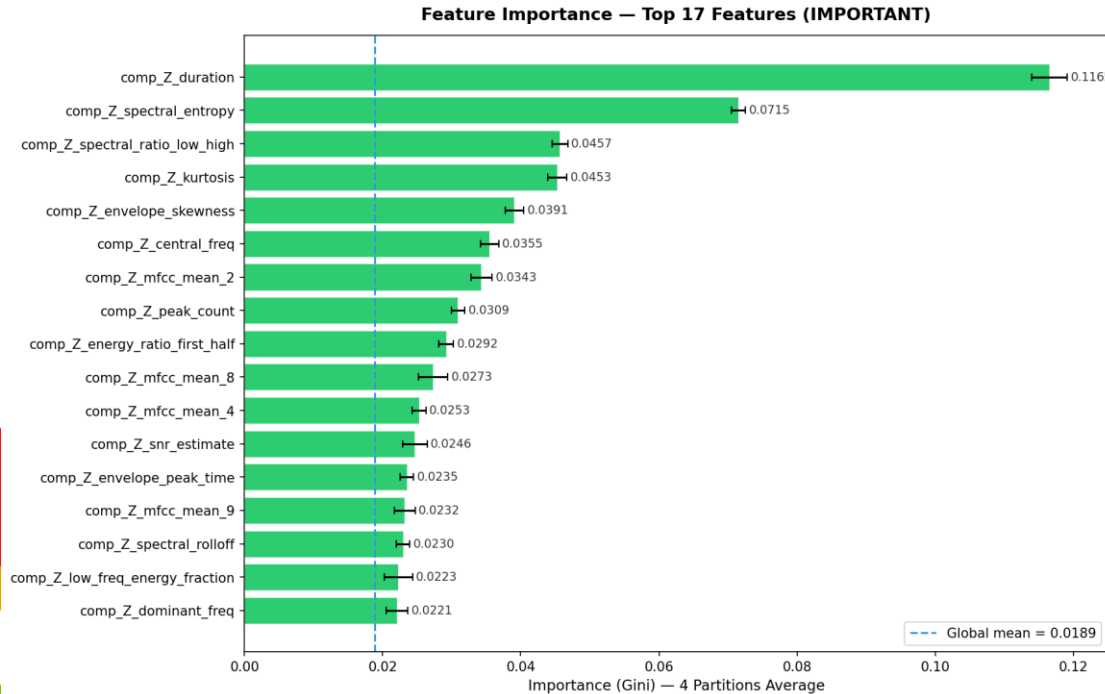
Conjunto de características

Grupo de características	Cantidad
Dominio temporal	10
Dominio frecuencial	10
Coeficientes Cepstrales en Frecuencias de Mel. (MFCC)	26
Envolvente analítica y cinemática	7
Total	53

Tabla 3. Grupos de características del vector θ_i

3. Ingeniería y selección de características

■ Selección de características mediante random forest



Criterio de selección:

$$\overline{\text{Imp}}(\theta_j) \geq \frac{1}{d} \sum_{j=1}^d \overline{\text{Imp}}(\theta_j) \quad (7)$$

Umbral global: $\text{Imp}_{\text{global}} \approx 0,0189$

- Reducción: $53 \rightarrow 17$ (-67,9 %)
- Vector final: $\theta_i \in \mathbb{R}^{17}$

Figura 8. Ranking de importancia promedio de las 17 características seleccionadas.

Modelado del clasificador

- Preparación de datos por partición

1. Carga del conjunto de contexto

$$x_{\text{ctx}} \in \mathbb{R}^{(10\,000 \times 17)}, y_{\text{ctx}} \in \mathbb{R}^{(10\,000)} \quad (8)$$

2. Imputación de NaN

$$\begin{aligned} \mu &= \text{mean}(x_{\text{ctx}}, \text{axis} = 0) \in \mathbb{R}^{17} \\ x_{\text{ctx}}[\text{NaN}] &\leftarrow \mu \end{aligned} \quad (9)$$

3. Almacenamiento del contexto.

4. Imputación de NaN en prueba

$$x_{\text{test}} \in \mathbb{R}^{(2\,500 \times 17)}, \quad x_{\text{test}}[\text{NaN}] \leftarrow \mu \quad (10)$$

Diseño experimental

- Estrategia experimental progresiva

4 Escenarios experimentales

#	Configuración	Objetivo
1	4 clases, desbalance nativo	Línea base bajo desbalance
2	3 clases balanceadas	Techo de rendimiento sin TOR
3	4 clases + aumento de datos para clase TOR	Compensación vía datos sintéticos
4	4 clases + ponderación por argmax	Compensación vía posprocesamiento

Tabla 4. Configuración y objetivo de los cuatro escenarios experimentales.

Diseño experimental

- Experimento 1: 4 clases con desbalance nativo

$$N_{\text{per_class}} = \frac{12\,500 - 195}{3} \rightarrow 4\,101, \quad \text{Residuo} = 2$$

Clase	Eventos disponibles	Muestreo	Muestras validas	Contexto 80%	Prueba 20%
LP	28 164	4 102	4 038	3 230	808
TOR	195	195	193	154	39
TRE	11 512	4 102	3986	3 189	797
VT	5 042	4 101	3976	3 181	795
Total	44 913	12500	12 193	9 754	2 439

Tabla 5. Distribución de muestras en el Experimento 1 (desbalance nativo del catálogo OVSP).

Diseño experimental

- Experimento 2: 3 clases balanceadas (sin TOR)

$$N_{\text{per_class}} = \frac{12\ 500}{3} \rightarrow 4\ 166, \quad \text{Residuo} = 2$$

Clase	Eventos disponibles	Muestreo	Muestras validas	Contexto 80%	Prueba 20%
LP	28 164	4 167	4 100	3 280	820
TRE	11 512	4 167	4 058	3 246	812
VT	5 042	4 166	4 043	3 234	809
Total	44 718	12500	12 201	9 760	2 441

Tabla 6. Distribución de muestras en el Experimento 2 (3 clases balanceadas, sin TOR).

Diseño experimental

- Procedimiento de aumento de datos para la clase TOR

1. Selección y normalización de la señal base

$$A_{\text{ref}} = \max_{0 \leq n \leq N_s} |x_{\text{base}}[n]|, \quad \tilde{x}_{\text{base}}[n] = \frac{x_{\text{base}}[n]}{A_{\text{ref}}} \quad (11)$$

2. de selección del preámbulo

$$L \geq 300, \quad \frac{\text{Peak}}{\text{RMS}} < 3,5 \quad (12)$$

3. Extracción de ruido

$$A_{\text{ruido}} = \max_{0 \leq n \leq N_s} |x_{\text{ruido}}[n]|, \quad \tilde{x}_{\text{ruido}}[n] = \frac{x_{\text{ruido}}[n]}{A_{\text{ruido}}} \quad (13)$$

Diseño experimental

- Procedimiento de aumento de datos para la clase TOR

4. Definición del factor de mezcla

$$\alpha \in \text{baja, media, alta} \quad (14)$$

5. Restricción de la perturbación

$$\rho(\alpha) = \frac{\text{RMS}(\alpha \tilde{x}_{\text{ruido}})}{\text{RMS}(\tilde{x}_{\text{base}})} \quad (15)$$

6. Construcción de la señal aumentada

$$\tilde{x}_{\text{aug}}[n] = \tilde{x}_{\text{base}}[n] + \alpha \cdot \tilde{x}_{\text{ruido}}[n] \quad (16)$$

7. Reescalado final

$$x_{\text{aug}}[n] = A_{\text{ref}} \cdot \tilde{x}_{\text{aug}}[n] \quad (17)$$

Diseño experimental

■ Visualización de datos sintéticos

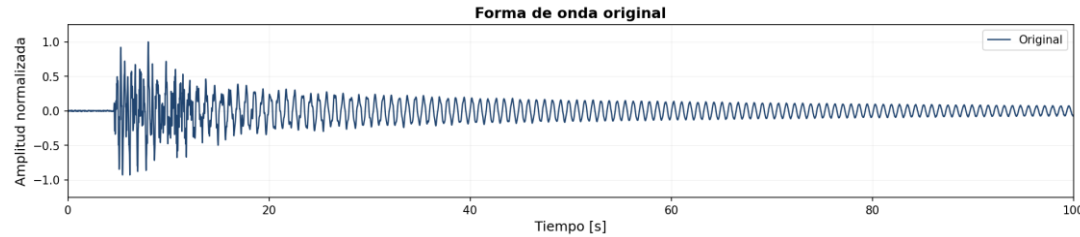
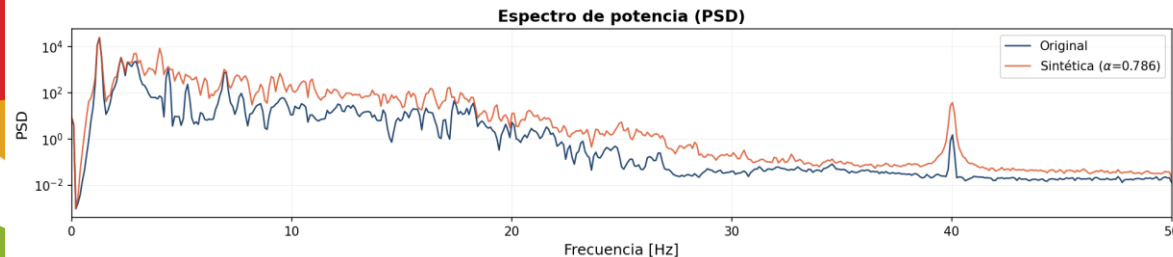
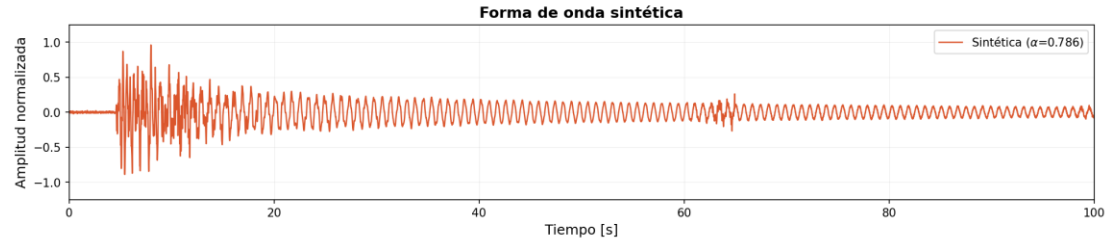


Figura 9 . Forma de onda original, forma de onda sintética ($\alpha = 0.786$) y espectro de potencia comparativo (PSD) de un evento sísmico registrado en la estación ANGP.



2 985 sismos sintéticos
generados de TOR

Diseño experimental

- Agrupamiento por evento (prevención de fuga de datos)

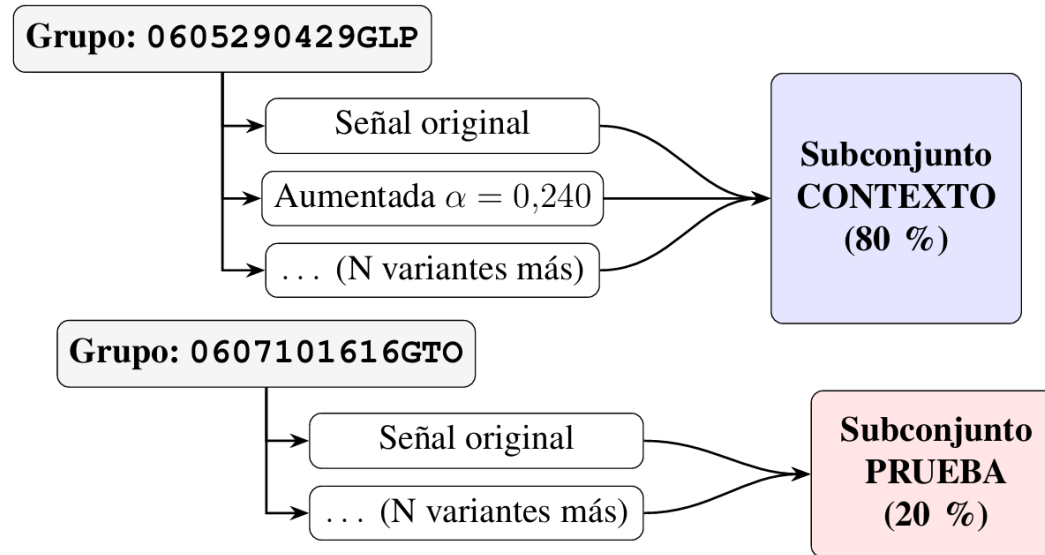


Figura 10 . Esquema de asignación agrupada por sismo base.

Diseño experimental

- Experimento 3: muestreo con TOR sintético

$$N_{\text{per-clas}} = \frac{12\ 500}{4} = 3\ 125$$

Clase	Eventos disponibles	Muestreo	Muestras validas
LP	28 164	3 125	3 079
TOR	3 180	3 125	3 123
TRE	11 512	3 125	3 046
VT	5 042	3 125	3 011
Total	47 898	12 500	12 250

Tabla 7. Distribución de eventos disponibles, muestreados y grupos únicos en el Experimento 3.

Diseño experimental

- Experimento 3: distribución por partición (1 y 2)

Clase	Contexto 80%		Prueba 20%	
	Muestras	Grupos	Muestras	Grupos
LP	2 463	2 463	616	616
TOR	2 454	154	669	39
TRE	2 437	2 437	609	609
VT	2 409	2 409	602	602
Total	9 763	7463	2 496	1 866

(a)

Clase	Contexto 80%		Prueba 20%	
	Muestras	Grupos	Muestras	Grupos
LP	2 463	2 463	616	616
TOR	2 484	154	639	39
TRE	2 437	2 437	609	609
VT	2 409	2 409	602	602
Total	9 793	7463	2 466	1 866

(b)

Tabla 8. Distribución de muestras por clase. a) Partición 1, b) Partición 2.

Diseño experimental

- Experimento 3: distribución por partición (3 y 4)

Clase	Contexto 80%		Prueba 20%	
	Muestras	Grupos	Muestras	Grupos
LP	2 463	2 463	616	616
TOR	2 498	154	625	39
TRE	2 437	2 437	609	609
VT	2 409	2 409	602	602
Total	9 807	7 463	2 452	1 866

(c)

Clase	Contexto 80%		Prueba 20%	
	Muestras	Grupos	Muestras	Grupos
LP	2 463	2 463	616	616
TOR	2 478	154	645	39
TRE	2 437	2 437	609	609
VT	2 409	2 409	602	602
Total	9 787	7 463	2 472	1 866

(d)

Tabla 9. Distribución de muestras por clase. c) Partición 3, d) Partición 4.

Diseño experimental

■ Experimento 4: argmax ponderado

Operador de decisión modificado

$$\hat{y}_i = \underset{c \in \{LP, TRE, TOR, VT\}}{\operatorname{argmax}} (\hat{p}_{i,c} \cdot w_c) \quad (18)$$

Cálculo del peso por clase:

$$w_c = \left(\frac{N}{C \cdot n_c} \right)^\gamma \quad (29)$$

- N : número total de muestras del contexto
- C : número de clases (4)
- n_c : número de muestras de la clase c en el contexto
- γ : exponente que regula la intensidad del efecto de la ponderación ($\in \{0; 0,5; 1\}$)

Diseño experimental

■ Experimento 4: niveles de compensación evaluados

Clase	n_c	$w_c(\gamma = 0, 0)$	$w_c(\gamma = 0, 5)$	$w_c(\gamma = 1, 0)$
LP	3 230	1,000	0,869	0,755
TOR	154	1,000	3,979	15,834
TRE	3 189	1,000	0,874	0,765
VT	3 181	1,000	0,8755	0,767

Tabla 10. Factores de ponderación w_c calculados para el contexto de $N = 9\,754$ muestras con $C = 4$ clases.

5. Resultados

Experimento 1: línea base con desbalance nativo

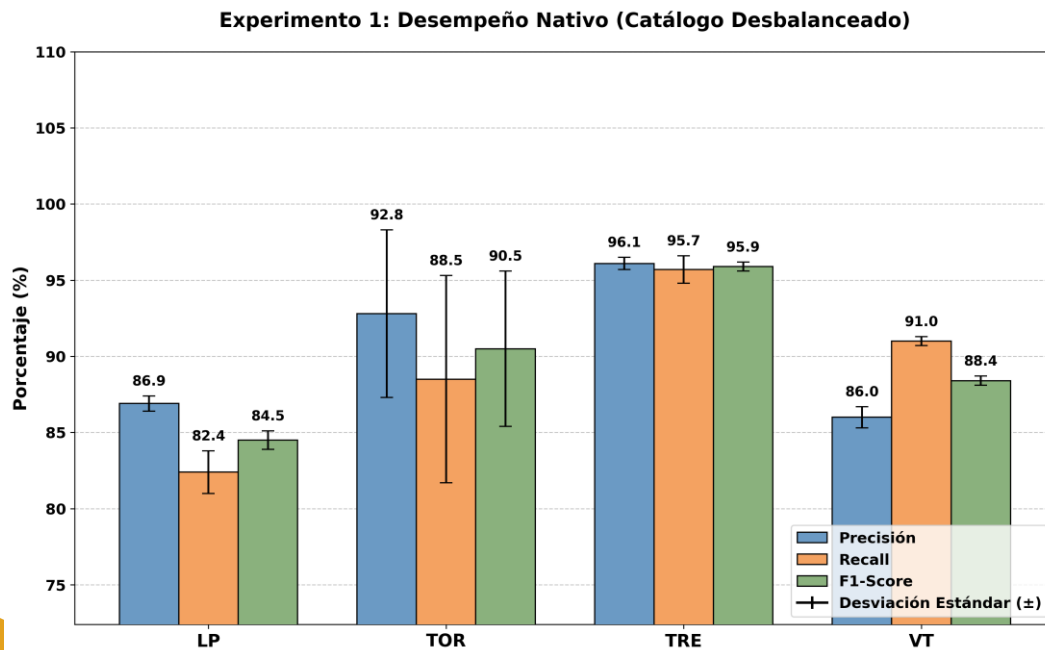


Figura 11. Desempeño por clase del Experimento 1 sobre el catálogo OVSP con desbalance nativo

Experimento 2: 3 clases balanceadas (sin TOR)

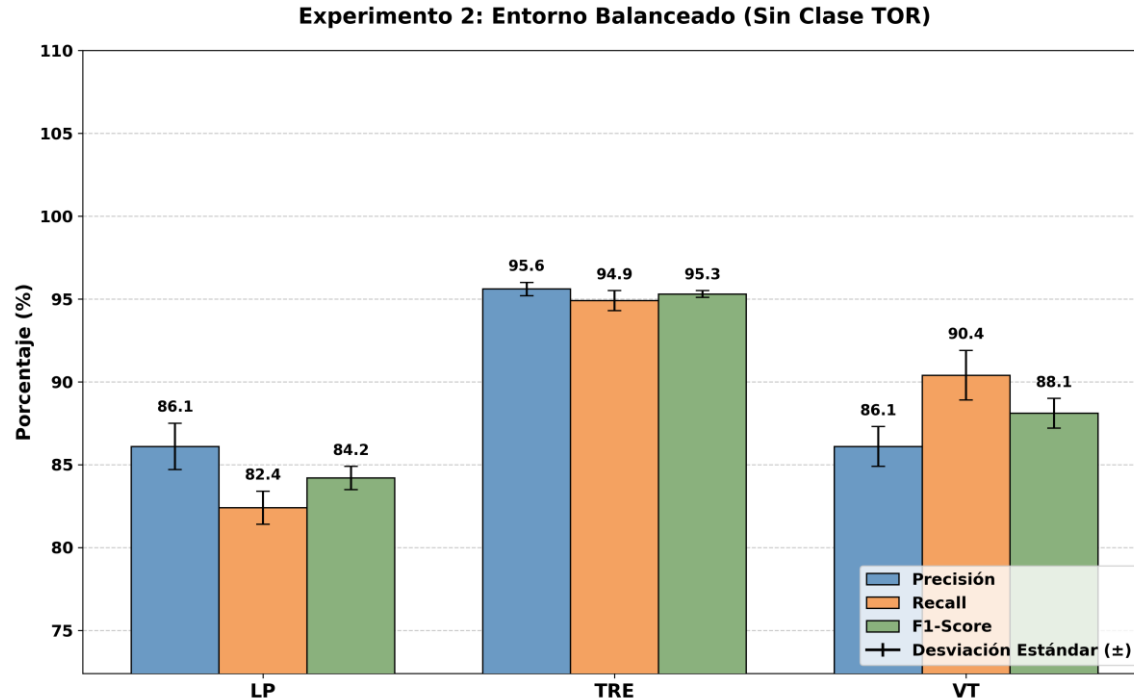


Figura 12. Desempeño del clasificador sobre tres clases balanceadas; cuantifica el costo predictivo del desbalance en el Exp. 1.

Experimento 3: aumento de datos (TOR sintético)

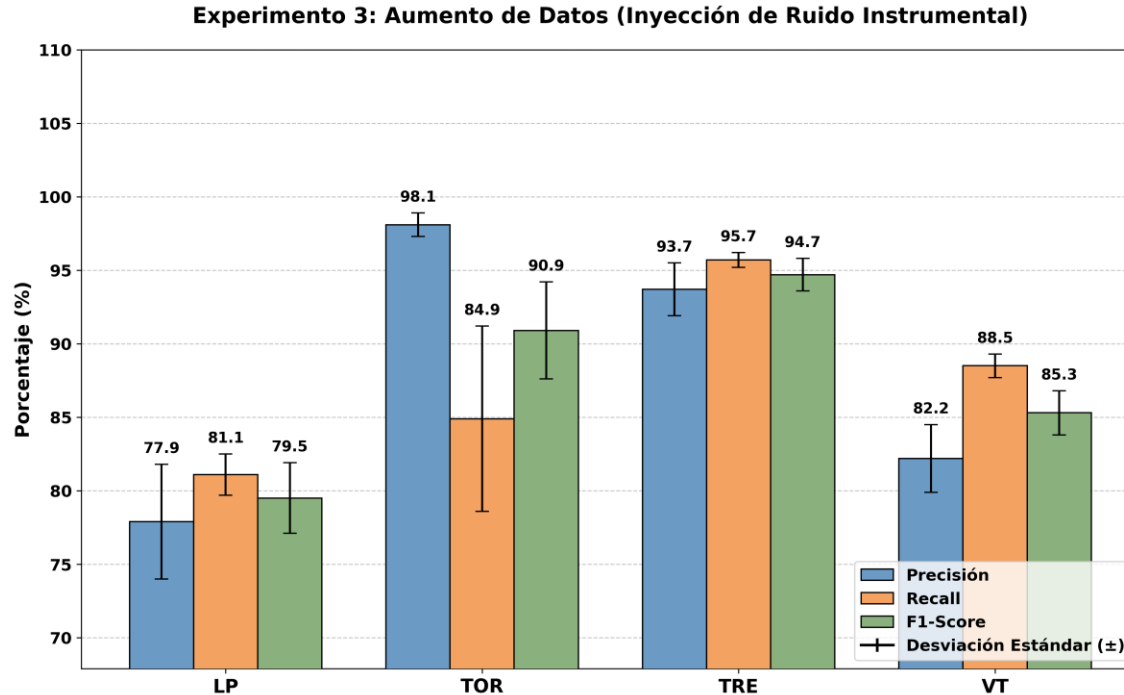


Figura 13. Desempeño por clase del Experimento 3 con muestras sintéticas TOR generadas por superposición aditiva.

Experimento 4A: argmax ponderado moderado ($\gamma = 0,5$)

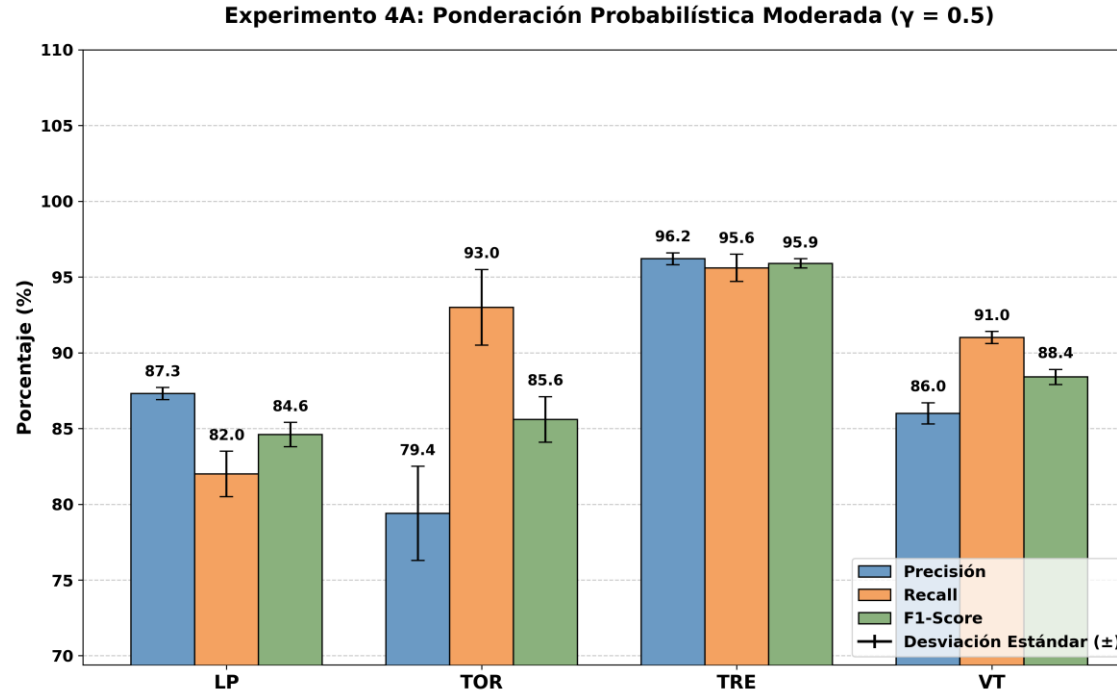


Figura 14. Desempeño del Experimento 4A con compensación moderada ($\gamma = 0,5$).

Experimento 4B: argmax ponderado máximo ($\gamma = 1,0$)

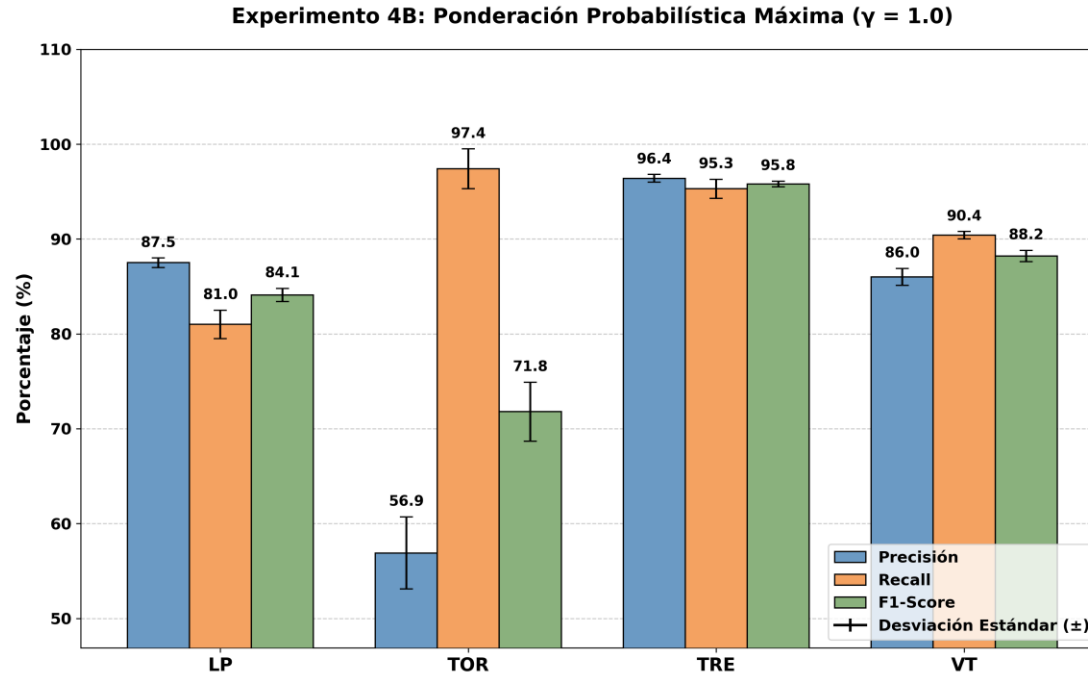


Figura 15. Desempeño del Experimento 4B con compensación máxima ($\gamma = 1,0$), evidenciando el límite del posprocesamiento agresivo.

Comparación global entre experimentos

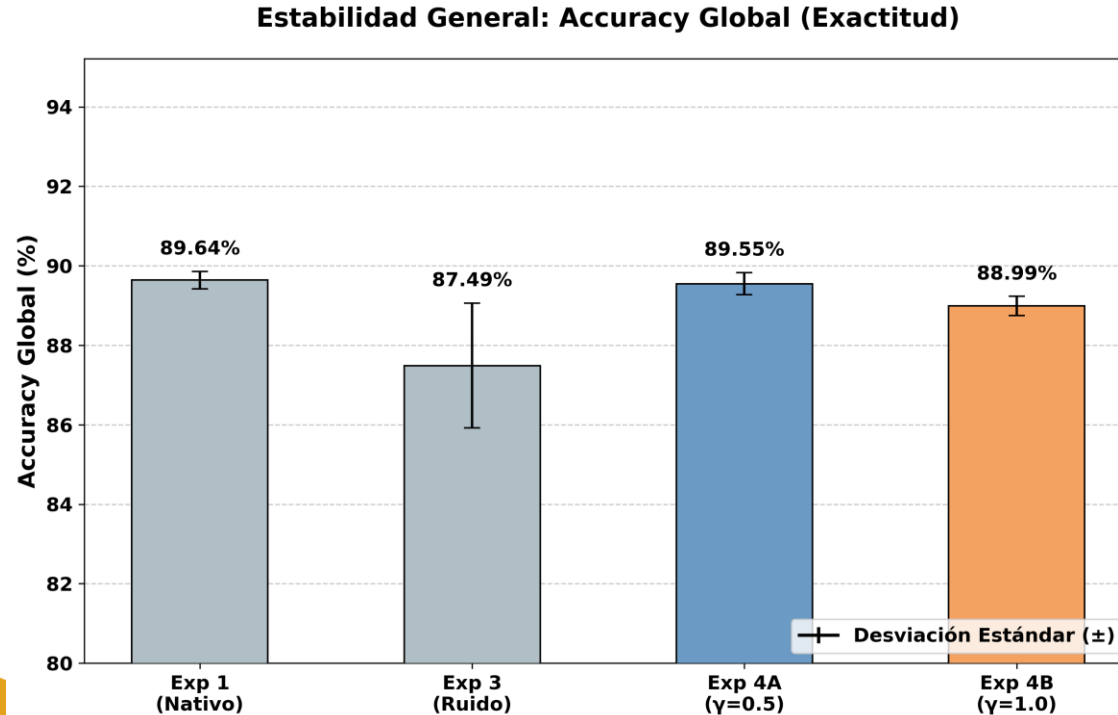


Figura 16. Accuracy global ($\pm\sigma$) por experimento sobre 4 particiones MCCV.

5. Conclusiones

- El modelo Transformer pre-entrenado TabPFN demostró ser una alternativa robusta y viable para la clasificación de la sismicidad del volcán Galeras, alcanzando una exactitud global cercana al 89 % sobre las cuatro clases (LP, TOR, TRE y VT) sin requerir reentrenamiento.
- La integración del preprocesamiento, la reducción del espacio a 17 características mediante Random Forest y la validación por Monte Carlo Cross-Validation, conformaron un flujo automatizado, reproducible y libre de fuga de información.
- El rendimiento del modelo está ligado a la separabilidad física de cada clase: el tremor (TRE) obtuvo el mejor desempeño por su firma espectral inconfundible, mientras que LP, VT y TOR al compartir características similares presentaron confusión.
- El desbalance de la clase tornillo (TOR) impone un límite estructural que las estrategias algorítmicas atenúan pero no eliminan: el aumento por ruido contaminó el espacio de características, y la ponderación probabilística con $\gamma = 0.5$ y $\gamma = 1.0$ elevó el recall a costa de reducir progresivamente la precisión.

Bibliografía

Cepeda, H. (2020). Quinientos años de documentación histórica de actividad del volcán Galeras: escenarios eruptivos. Servicio Geológico Colombiano. <https://doi.org/10.32685/9789585279421>

Cadena, O. E. (2021). Modelos de fuente de sismicidad LP para la actividad del volcán Galeras 2004-2010 (Colombia). Recuperado de: <https://repositorio.unal.edu.co/handle/unal/80915>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. <https://arxiv.org/abs/1706.03762>

N. Hollmann, S. Müller, L. Purucker, A. Krishnakumar, M. Körfer, S. B. Hoo, R. T. Schirrmeister, and F. Hutter, “Accurate predictions on small data with a tabular foundation model,” Nature, vol. 637, pp. 319–326, 1 2025.

¡Gracias!



Universidad de Nariño
FUNDADA EN 1904



Universidad de Nariño
ACREDITADA EN ALTA CALIDAD
RESOLUCIÓN MEN 000527 - ENERO 11 DE 2023

