Universidad de Nariño
Facultad de Ciencias Exactas y Naturales
Departamento de Física

Universidad de Nariño
FUNDADA EN 1904

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

TRABAJO DE GRADO

Para optar por el título profesional de:

Físico

Daniel Esteban Chapid Vela

San Juan de Pasto, Colombia
14 mayo de 2026

Universidad de Nariño
Facultad de Ciencias Exactas y Naturales
Departamento de Física

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

Daniel Esteban Chapid Vela

TRABAJO DE GRADO

Director:

Oscar Ernesto Cadena Ibarra. PhD

San Juan de Pasto, Colombia
14 mayo de 2026

≪Las ideas y conclusiones aportadas en la tesis de grado son responsabilidad exclusiva de los autores≫

Artículo 1. del acuerdo No. 324 del 11 de Octubre de 1966, emanado por el Honorable Consejo
Directivo de la Universidad de Nariño

Nota de aceptación

_____

_____

_____

_____

_____

_____

Oscar Ernesto Cadena Ibarra

_____

Director

Jorge López Melo

_____

Jurado

Jaime Mauricio Bacca

_____

Jurado

San Juan de Pasto, mayo de 2026

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas que contribuyeron de alguna manera al desarrollo de este trabajo.

En primer lugar, agradezco a mi asesor, el Dr. Oscar Ernesto Cadena, por su paciencia, orientación y acompañamiento a lo largo de este proceso. Su apoyo y disposición fueron fundamentales para la realización de este trabajo.

De igual manera, agradezco profundamente a mi madre por el apoyo incondicional que me ha brindado en cada etapa de mi formación. Su esfuerzo, comprensión y confianza han sido una fuente constante de motivación para alcanzar este logro.

A todas las personas que, de una u otra forma, hicieron parte de este proceso, les expreso mi gratitud.

Dedicatoria

A mi madre, por su amor, su apoyo constante y por estar siempre a mi lado en cada paso de este camino.

Y a quienes, con su presencia y aliento, me impulsaron a seguir adelante en este viaje de aprendizaje y esfuerzo.

Resumen

La clasificación automática de sismicidad volcánica es un problema relevante en el monitoreo de sistemas activos, especialmente cuando se dispone de catálogos limitados y con un fuerte desbalance entre clases. En este trabajo se evalúa la capacidad de una arquitectura basada en *Transformer*, específicamente TabPFN, para clasificar diferentes tipos de sismos registrados en el volcán Galeras (Nariño, Colombia). Para ello, siguiendo las etapas del ciclo de vida de *Machine Learning*, se desarrolló un flujo de trabajo que integró la extracción de características, la construcción de conjuntos experimentales y la evaluación del modelo bajo cuatro escenarios: un entorno nativo desbalanceado, un contexto balanceado sin la clase minoritaria tornillo (TOR), un enfoque de aumento de datos con ruido instrumental y una estrategia de compensación mediante ponderación probabilística.

Los resultados muestran que el modelo alcanza un desempeño global cercano al 89,64 % de exactitud en el escenario nativo, con un F1-Score Macro de 89,83 %. En el análisis por categorías, se registraron valores de F1-Score promedio de 95,9 % para tremor (TRE), indicando el mejor comportamiento; 84,5 % para largo periodo (LP), evidenciando la mayor dificultad de discriminación; 88,4 % para volcano-tectónico (VT), mostrando un desempeño intermedio en la clasificación de este tipo de sismo; y 90,5 % para tornillo (TOR), a pesar de representar una fracción muy pequeña del catálogo. Para mitigar el marcado desbalance de esta última clase, se evaluaron dos estrategias de compensación. Por un lado, el aumento de datos mediante inyección de ruido instrumental introdujo un sobreajuste perjudicial, ya que alteró la morfología original de las señales y degradó la clasificación general. Por otro lado, el ajuste probabilístico del operador de decisión demostró ser una alternativa metodológicamente más estable, permitiendo calibrar la sensibilidad del modelo hacia este evento premonitorio sin contaminar el espacio de representación.

En conjunto, los resultados indican que TabPFN constituye una alternativa robusta y viable para la clasificación automática de sismicidad volcánica en contextos con pocos datos y clases severamente desbalanceadas.

Palabras clave: sismicidad volcánica, clasificación automática, TabPFN, ciclo de vida de machine learning, aprendizaje en contexto, volcán Galeras.

Abstract

The automatic classification of volcanic seismicity is a crucial challenge in monitoring active systems, particularly when dealing with limited catalogs and severe class imbalance. This work evaluates the performance of a Transformer-based architecture, specifically TabPFN, for classifying different types of seismic events recorded at the Galeras volcano (Nariño, Colombia). Following the stages of the Machine Learning life cycle, a workflow was developed that integrated feature extraction, experimental set construction, and model evaluation across four scenarios: a native imbalanced environment, a balanced context without the minority class Tornillo (TOR), a noise-based data augmentation approach, and a compensation strategy through probabilistic weighting.

The results show that the model achieves a global performance near 89.64 % accuracy in the native scenario, with an F1-Score Macro of 89.83 %. At the category level, average F1-Score values reached 95.9 % for tremor (TRE), indicating the highest performance; 84.5 % for long period (LP), evidencing the greatest discrimination difficulty; 88.4 % for volcano-tectonic (VT), showing intermediate classification performance; and 90.5 % for tornillo (TOR), despite its very small representation in the catalog. To mitigate the marked imbalance of this last class, two compensation strategies were evaluated. On one hand, data augmentation via instrumental noise injection introduced detrimental overfitting, as it altered the original morphology of the signals and degraded overall classification. On the other hand, the probabilistic adjustment of the decision operator proved to be a more methodologically stable alternative, allowing for the calibration of the model's sensitivity toward this premonitory event without contaminating the representation space.

Overall, the results indicate that TabPFN constitutes a robust and viable alternative for the automatic classification of volcanic seismicity in contexts with limited data and severely imbalanced classes.

Keywords: volcanic seismicity, automatic classification, TabPFN, machine learning life cycle, in-context learning, Galeras volcano.

Contenido

Índice de tablas

Índice de figuras

Glosario

OVSP: Observatorio Vulcanológico y Sismológico de Pasto, entidad encargada del monitoreo técnico del complejo volcánico Galeras.

Tremor (TRE): Señal sísmica continua y de larga duración, asociada al movimiento de fluidos o gases a través de los conductos volcánicos.

Largo Periodo (LP): Clase sísmica vinculada a cambios de presión en el sistema de fluidos del volcán, caracterizada por un contenido de baja frecuencia.

Volcano-Tectónico (VT): Sismo originado por el fracturamiento de roca sólida debido a esfuerzos internos, caracterizado por un inicio impulsivo y altas frecuencias.

Tornillo (TOR): Señal sísmica de alta importancia premonitoria, caracterizada por una forma de onda monocromática y un decaimiento exponencial muy marcado.

Envolvente analítica: Representación de la amplitud de una señal sísmica que permite analizar el crecimiento y decaimiento de la energía sin las oscilaciones de alta frecuencia.

MFCC: Coeficientes Cepstrales en las Frecuencias de Mel. Descriptores que capturan las características del espectro de una señal basándose en la percepción auditiva, útiles para la clasificación.

Aprendizaje automático: Subcampo de la inteligencia artificial, también denominado *Machine Learning*, centrado en el desarrollo de algoritmos que permiten a las computadoras identificar patrones y realizar tareas sin instrucciones explícitas.

Ciclo de vida de ML: Marco metodológico que abarca todas las etapas de un proyecto de aprendizaje automático, desde la preparación de los datos hasta la evaluación de los resultados.

Desbalance de clases: Situación estadística donde las categorías en un conjunto de datos no están representadas de manera equitativa, afectando el sesgo del clasificador hacia las clases mayoritarias.

Aumento de datos: Proceso de incrementar el tamaño de un conjunto de datos mediante la creación de variantes modificadas de los ejemplos existentes, como la adición de ruido instrumental.

Sobreajuste: Error de modelado, conocido en inglés como *Overfitting*, que ocurre cuando un algoritmo memoriza los detalles y el ruido de los datos de referencia proporcionados, perdiendo capacidad de generalización frente a datos nuevos.

Transformer: Arquitectura de red neuronal basada en mecanismos de atención que permite procesar datos de manera paralela, capturando relaciones complejas entre las variables.

Aprendizaje en contexto: Capacidad de un modelo para realizar inferencias inmediatas sobre nuevos datos utilizando ejemplos previos proporcionados como referencia, sin requerir una fase de actualización de pesos.

TabPFN: Arquitectura basada en *Transformer* preentrenada sintéticamente, diseñada específicamente para conjuntos de datos tabulares, que realiza predicciones rápidas mediante aprendizaje en contexto.

Exactitud (Accuracy): Proporción de predicciones correctas (tanto positivas como negativas) respecto al total de eventos analizados por el modelo.

Precisión: Proporción de eventos identificados correctamente por el modelo entre todos los eventos que el algoritmo asignó a esa categoría específica.

Exhaustividad (Recall): Capacidad del clasificador para identificar correctamente todos los eventos reales pertenecientes a una categoría específica dentro del catálogo.

F1-Score: Media armónica entre la precisión y la exhaustividad. Métrica principal en este estudio debido a su robustez para evaluar el rendimiento en escenarios con desbalance severo.

Capítulo 1

Introducción

El volcán Galeras, ubicado en las inmediaciones del municipio de Pasto (Nariño, Colombia), constituye uno de los sistemas volcánicos activos más complejos y de mayor riesgo del país debido a su proximidad a centros densamente poblados. Su historia eruptiva y su relevancia en el contexto volcánico colombiano han sido ampliamente documentadas en la literatura especializada [1]. Esta condición hace que su vigilancia constituya un componente fundamental para la gestión integral del riesgo en la región.

La clasificación de estas señales ha dependido históricamente de la revisión y validación por parte de expertos. Aunque este procedimiento ha sido esencial para construir los catálogos actuales, su aplicación sostenida presenta limitaciones prácticas. El crecimiento constante de la red de monitoreo y los incrementos repentinos en la sismicidad durante crisis volcánicas generan volúmenes de datos que dificultan su procesamiento oportuno. A esto se suma la subjetividad inherente a la interpretación de eventos con morfologías ambiguas. En este contexto, la automatización del análisis sismológico ha adquirido un papel cada vez más importante [5–7].

El catálogo de eventos empleado en esta investigación comprende registros clasificados por el Observatorio Vulcanológico y Sismológico de Pasto (OVSP) durante el período 2006–2010. La sismicidad analizada abarca cuatro categorías principales: largo periodo (LP), tremor espasmódico (TRE), volcano-tectónico (VT) y tornillo (TOR). El principal desafío de este conjunto de datos recae en su marcado desbalance de clases. Los eventos asociados al movimiento de fluidos, especialmente LP, predominan ampliamente, mientras que la clase TOR representa una fracción estadísticamente minúscula. Esta disparidad resulta crítica para el monitoreo, debido a que los eventos tipo tornillo presentan una fuerte relación con la presurización del conducto volcánico y se han descrito como señales asociadas a fases preeruptivas [2, 8].

En este contexto, el presente trabajo tiene como objetivo proponer y evaluar un sistema de clasificación automática enmarcado dentro de las etapas del ciclo de vida del *machine learning* [9, 10]. Esta aproximación metodológica asegura un desarrollo estructurado que abarca desde la selección y extracción de características en los dominios del tiempo y la frecuencia, hasta la evaluación sistemática de los modelos. Específicamente, la investigación busca determinar la viabilidad y el rendimiento de modelos tabulares basados en arquitecturas *Trans-*

former (como TabPFN) para clasificar sismicidad volcánica, enfrentando de manera directa el desafío analítico que impone la escasez de datos en las clases minoritarias [11, 12].
Para abordar el severo desbalance de la clase premonitoria sin comprometer la generalización del sistema, este estudio evalúa y contrasta dos estrategias independientes: las técnicas tradicionales de aumento de datos y la calibración probabilística del operador de decisión. Esta última estrategia permite ajustar la sensibilidad del modelo mediante un factor de ponderación empírico (γ), corrigiendo el sesgo mayoritario de forma puramente matemática sin contaminar ni alterar el espacio original de características de las señales [13, 14].
Para estructurar la presentación de este trabajo, el documento se organiza de la siguiente manera: inicialmente, se establece el marco referencial con los fundamentos sismológicos y computacionales que sustentan la investigación. Posteriormente, se da paso al desarrollo del trabajo, en el cual se sigue el ciclo de vida del *Machine Learning* para estructurar el procesamiento de la señal y el diseño experimental. Finalmente, se exponen los resultados obtenidos, seguidos de la interpretación de los datos en la discusión, para culminar con las conclusiones y recomendaciones del estudio.

Capítulo 2

Planteamiento del Problema

El volcán Galeras, ubicado en el departamento de Nariño, Colombia, es uno de los sistemas volcánicos más activos del país y, por tanto, requiere un monitoreo continuo debido a su potencial amenaza para las comunidades asentadas en su zona de influencia. Una de las principales herramientas para la vigilancia volcánica es el análisis de la sismicidad asociada al edificio volcánico, la cual permite inferir cambios en la dinámica interna del sistema. En particular, los sismos volcano-tectónicos (VT), de largo periodo (LP), tremor (TRE) y tornillo (TOR) constituyen señales de gran interés, ya que cada una presenta características físicas distintas y puede estar asociada con procesos internos específicos del volcán.

Sin embargo, la clasificación de este tipo de señales sigue siendo una tarea compleja. En la práctica, dicha labor suele depender de la interpretación experta de analistas, lo que implica una inversión considerable de tiempo y puede introducir variabilidad subjetiva en la identificación de los eventos. A esto se suma que los catálogos sísmicos volcánicos suelen estar desbalanceados, con clases minoritarias poco representadas, como ocurre en el caso de los eventos tornillo. Esta situación dificulta el rendimiento de modelos automáticos y afecta su capacidad para generalizar correctamente entre clases con firmas espectro-temporales parcialmente similares, especialmente en casos como LP y VT.

Ante esta problemática, se han empleado métodos automáticos basados en técnicas clásicas de procesamiento de señales y en redes neuronales artificiales como las convolucionales y recurrentes. Aunque estas arquitecturas han mostrado resultados favorables en distintas tareas de clasificación, su desempeño puede verse limitado cuando se requiere capturar relaciones complejas entre las características globales de una señal. En este contexto, las arquitecturas tipo *Transformer* constituyen una alternativa prometedora, ya que incorporan mecanismos de atención capaces de asignar relevancia a las diferentes partes de la información de entrada y modelar dependencias de largo alcance.

No obstante, pese al potencial de estas arquitecturas, su aplicación en la clasificación automática de sismicidad volcánica del volcán Galeras no ha sido suficientemente explorada en el ámbito de este trabajo. En particular, permanece abierta la pregunta de si un modelo basado en *Transformer* puede discriminar de manera efectiva las distintas clases sísmicas del Galeras bajo condiciones de desbalance y escasez de datos, manteniendo un desempeño consistente frente a la variabilidad natural del catálogo.

Por ello, se plantea la siguiente pregunta de investigación:

¿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?

Capítulo 3

Objetivos

3.1. General

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

3.2. 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.

Capítulo 4

Marco Referencial

4.1. El volcán Galeras y su actividad sísmica

El estudio y monitoreo de la sismicidad en entornos volcánicos difiere sustancialmente de la sismología tectónica clásica. Mientras que los sismos litosféricos son, en esencia, el resultado de la liberación de esfuerzos acumulados a lo largo de planos de falla, la sismicidad de un edificio volcánico como el Galeras es el reflejo directo de una compleja dinámica fluida y termodinámica. La geomorfología volcánica y sus procesos internos determinan que la interacción continua entre el ascenso magmático, la vaporización de acuíferos hidrotermales y el fracturamiento de roca, produzca un amplio catálogo de firmas sísmicas con características particulares [15, 16]. Para comprender esta variedad de señales, resulta fundamental establecer el marco físico e histórico en el cual se desarrollan, considerando que cada sistema volcánico posee una firma sísmica y estructural única.

4.1.1. Contexto geológico y localización

El volcán Galeras (conocido históricamente como Urcunina por las comunidades indígenas ancestrales) es un estratovolcán de composición predominantemente andesítica, emplazado en el departamento de Nariño, al suroccidente de Colombia ($1°13'$ N y $77°22'$ W). Constituye uno de los centros eruptivos más activos del segmento norte de la Cordillera de los Andes. Como se detalla en la Figura 4.1, su ubicación reviste un interés crítico en materia de gestión del riesgo, dado que se encuentra situado a menos de 9 km en línea recta de la ciudad de San Juan de Pasto, un centro urbano densamente poblado que supera el medio millón de habitantes [1].

Desde un punto de vista morfo-estructural, su edificación se fundamenta sobre el límite oriental de la depresión tectónica Cauca-Patía, alcanzando una cota máxima en su cono activo del orden de los 4.276 metros sobre el nivel del mar [1]. A nivel de geodinámica global, el complejo Volcánico Galeras (CVG) no es un sistema aislado; su génesis responde a la zona de subducción activa donde la placa oceánica de Nazca se sumerge por debajo de la placa continental Sudamericana [15]. Este marco tectónico le confiere una litología estratificada rica en flujos de lavas viscosas (andesitas y dacitas) y potentes depósitos de flujos piroclásticos.

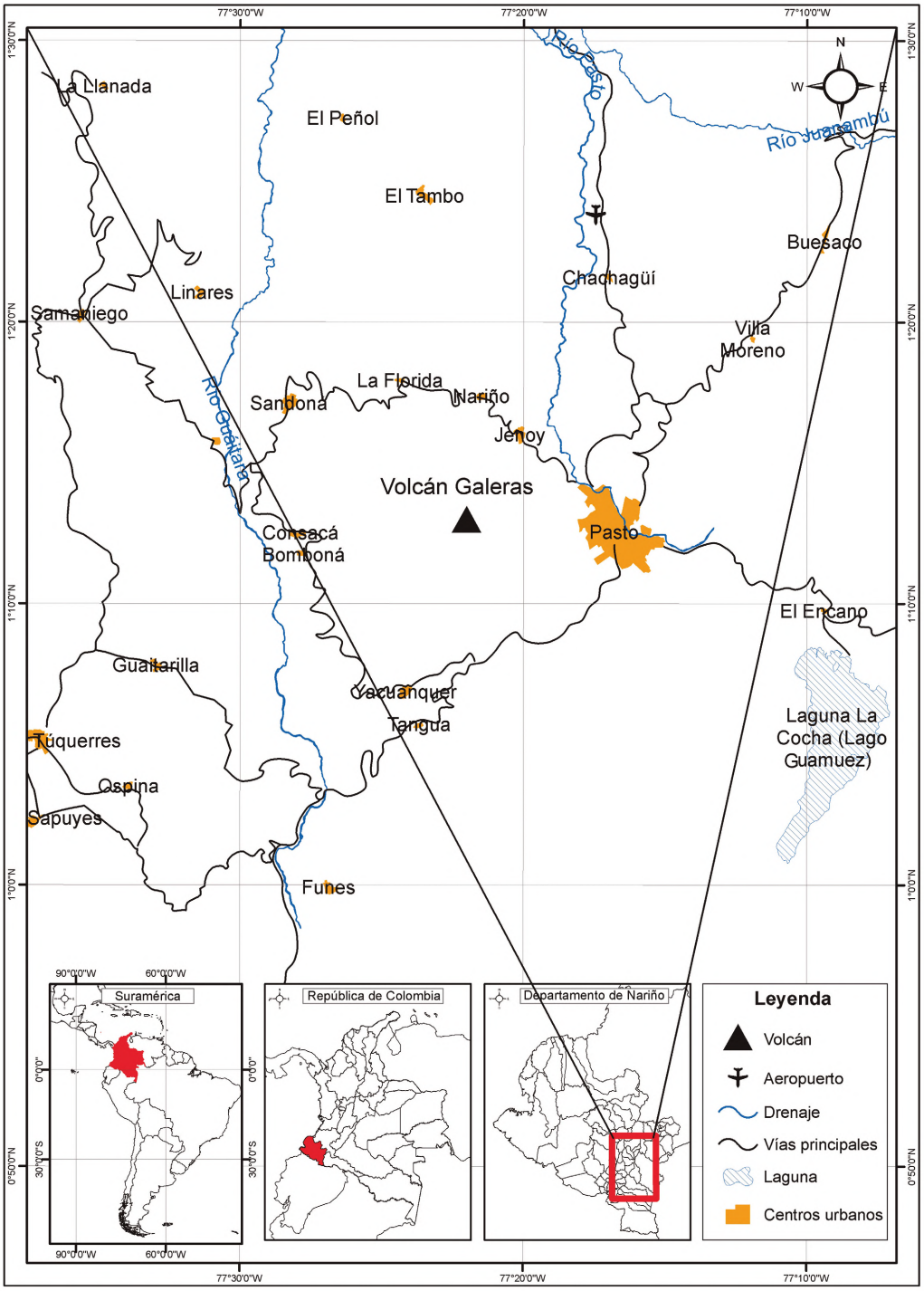


Figura 4.1: Localización geográfica del volcán Galeras en el departamento de Nariño, Colombia. El mapa principal destaca la proximidad del edificio volcánico a la ciudad de San Juan de Pasto y otros centros urbanos, así como la red hídrica principal. Los recuadros inferiores contextualizan la ubicación a nivel continental, nacional y departamental. Tomado de [1].

La estructura del edificio volcánico se encuentra profundamente disectada y debilitada por sistemas de fallas regionales mayores. Esta fracturación de la corteza adyacente no solo moldea la estructura morfológica del complejo, sino que modula significativamente las vías de ascenso de las intrusiones magmáticas e influye en la distribución tridimensional de los enjambres sísmicos volcano-tectónicos, incorporando un claro componente tectónico a la sismicidad de la zona.

4.1.2. Dinámica y fenomenología eruptiva

El volcán Galeras se consolida como uno de los sistemas más activos del continente americano, respaldado por un historial eruptivo que abarca más de quinientos años de documentación escrita. Esta alta recurrencia y su peligrosidad intrínseca motivaron su inclusión, durante la década de 1990, en el programa internacional Decade Volcanoes [1].

La dinámica interna contemporánea del edificio se define por una interacción constante e inestable entre un reservorio magmático superficial y un vigoroso sistema hidrotermal superior. Las investigaciones sobre su fenomenología reciente coinciden en clasificar su actividad predominantemente dentro de un estilo vulcaniano [16].
Las erupciones vulcanianas exhiben un comportamiento rítmico de acumulación y liberación súbita de energía. Este proceso ocurre debido a la cristalización del magma en el conducto central, lo que genera domos de lava que actúan como un sello parcial del sistema. Cuando la presión de los gases magmáticos profundos supera la resistencia mecánica de esta roca andesítica, el sistema experimenta una explosión violenta que proyecta material balístico y levanta densas columnas de ceniza.

Dada la naturaleza parcialmente sellada de este mecanismo, el volcán no siempre manifiesta cambios visuales evidentes antes de una erupción. Por lo tanto, la sismicidad constituye el principal indicador diagnóstico de su estado interno [1]. La caracterización rigurosa de precursores sísmicos, específicamente los enjambres de tornillos (TOR) y el incremento en la energía del tremor (TRE), constituye la herramienta analítica fundamental para anticipar y alertar sobre cambios críticos de presión en el cráter previos a erupciones explosivas.

4.1.3. Clasificación de eventos sísmicos en el Galeras

Como resultado de la dinámica magmática, hidrotermal y estructural descrita, el volcán Galeras genera una variedad de eventos sísmicos asociados a distintos mecanismos físicos que operan en su interior. La literatura especializada reconoce cinco tipos principales de eventos: sismos volcano-tectónicos (VT), de largo periodo (LP), tremor volcánico (TRE) y eventos tipo tornillo (TOR). Estos eventos se manifiestan a través de señales sísmicas volcánicas continuas, representadas principalmente por el TRE, y señales sísmicas volcánicas transitorias, categoría que comprende los eventos VT, LP y TOR.

La correcta identificación y clasificación de los eventos considerados es fundamental para la interpretación del estado de actividad del volcán y para el desarrollo de metodologías automáticas de clasificación sísmica. A continuación, se describen las principales características de cada tipo de evento abordado en este trabajo, de acuerdo con la literatura especializada y los criterios utilizados en la vigilancia volcánica.

Sismos Volcano-tectónicos (VT)

Los sismos volcano-tectónicos se originan por fractura frágil del material cortical en el edificio volcánico, inducida por cambios de esfuerzo asociados al movimiento de magma, enfriamiento o acumulación de esfuerzos tectónicos regionales. Estos eventos ocurren típicamente a profundidades de entre 2 y 10 km. Como se ilustra en la Figura 4.2, se distinguen por registros con arribos impulsivos donde las fases P y S son claramente identificables, seguidas por un decaimiento exponencial rápido de la amplitud en sus ondas de coda. Espectralmente, los VT están dominados por altas frecuencias (superiores a 5 Hz), con picos de mayor energía concentrados habitualmente entre los 6 y 8 Hz, aunque pueden alcanzar rangos de 15 a 20 Hz [2]. Su presencia suele asociarse a periodos de intrusión magmática y puede preceder o acompañar episodios eruptivos. El factor determinante para su clasificación es la localización del foco en las inmediaciones de la estructura volcánica [2, 16].

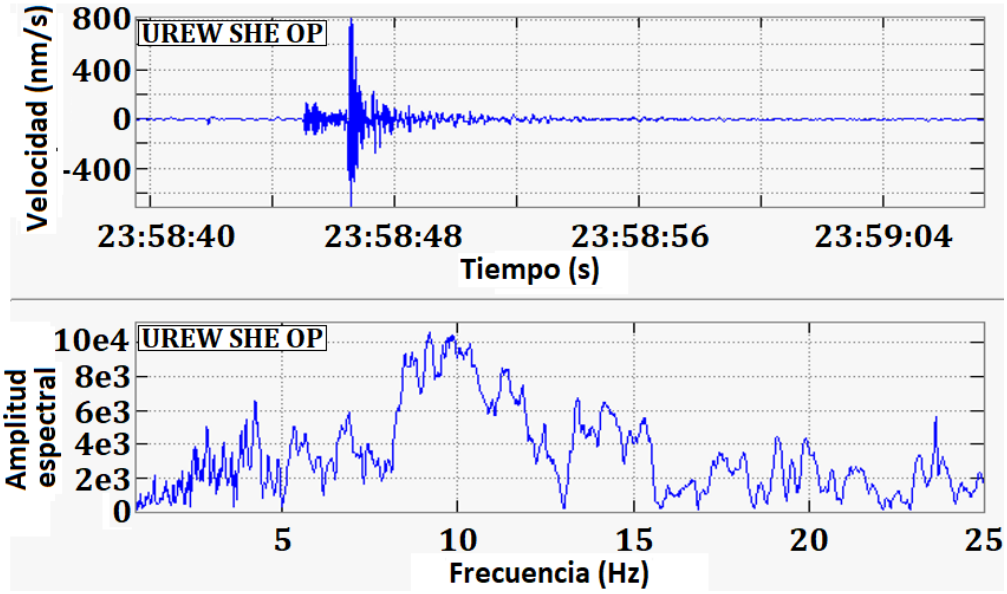


Figura 4.2: Sismo volcano-tectónico (VT) registrado en la estación UREW. Parte superior: sismograma con arribo impulsivo y rápido decaimiento. Parte inferior: espectro de amplitud dominado por altas frecuencias ($>$ 5 Hz). Tomado de [2].

Sismos de Largo periodo (LP)

Los sismos de largo periodo se caracterizan por señales dominadas por bajas frecuencias, generalmente menores a 5 Hz, con picos resonantes claros y componentes secundarias que pueden extenderse hasta aproximadamente 10–13 Hz. Tal como se observa en la Figura 4.3, su inicio suele ser emergente, especialmente en estaciones alejadas, y presentan duraciones prolongadas debido al lento decaimiento de la coda, lo que dificulta la identificación clara de la fase S. En algunos casos, se observa una señal precursora de baja energía seguida por un evento principal más energético, así como pulsos asociados a la superposición de modos armónicos o a oscilaciones pulsátiles de la fuente [2].
Estos eventos están vinculados a procesos controlados por la presencia y el movimiento de fluidos en el interior del volcán, donde variaciones transitorias de presión actúan como un mecanismo de excitación que induce la resonancia de fracturas, cavidades o conductos volcánicos saturados de fluido. La respuesta oscilatoria de estos sistemas confinados genera las señales características de largo periodo, tal como se ha documentado en diferentes contextos volcánicos [2, 5, 17]. En consecuencia, los sismos LP son interpretados como manifestaciones directas de la dinámica interna del sistema magmático-hidrotermal, constituyendo un indicador clave de cambios en el estado del volcán.

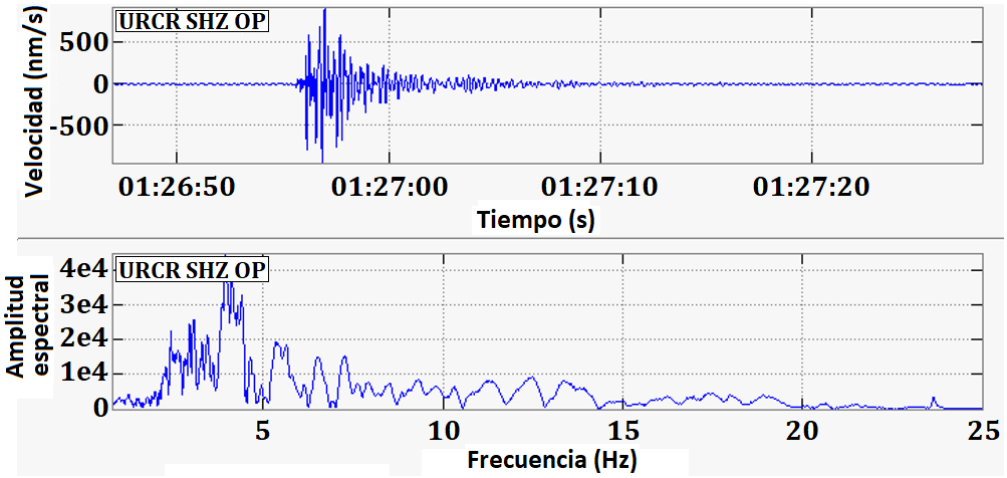


Figura 4.3: Sismo de largo periodo (LP) registrado en la estación URCR. Parte superior: sismograma con inicio emergente y duración prolongada. Parte inferior: espectro de amplitud con energía concentrada en bajas frecuencias (< 5 Hz). Tomado de [2].

Sismos tipo Tornillo (TOR)

Los sismos tipo tornillo son considerados un subtipo particular y altamente especializado de eventos de largo periodo, asociados a procesos de resonancia en grietas, cavidades o conductos volcánicos rellenos de fluido. Como se evidencia en la Figura 4.4, se caracterizan por un inicio impulsivo seguido de una coda muy prolongada, dominada por oscilaciones armónicas persistentes, lo que indica un decaimiento excepcionalmente lento de la amplitud y una baja

disipación de energía en el sistema resonante. En el dominio espectral, estos eventos presentan una distribución muy limitada de frecuencias, concentrando la energía en uno o pocos picos armónicos bien definidos (típicamente monocromáticos), rasgo distintivo frente a otros sismos LP [2, 5].

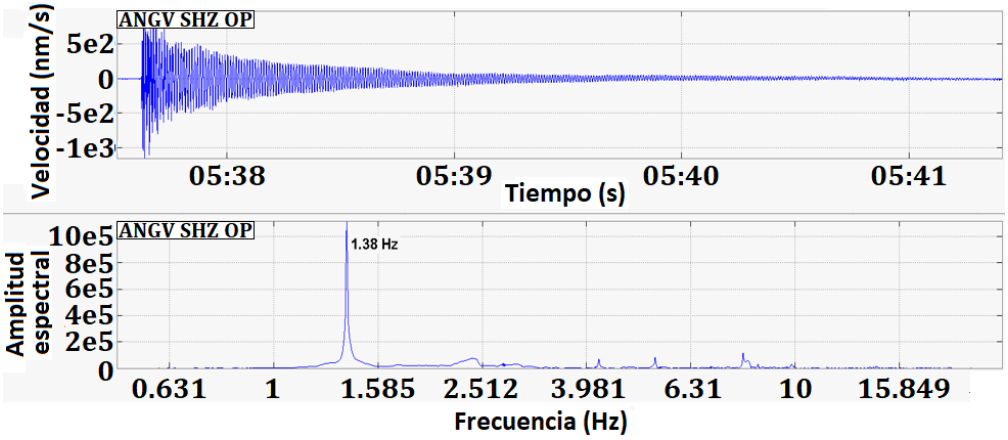


Figura 4.4: Sismo tipo tornillo (TOR) registrado en la estación ANGV. Parte superior: sismograma con decaimiento lento y oscilaciones prolongadas. Parte inferior: espectro de amplitud de carácter monocromático, con la energía concentrada en un pico bien definido de 1.38 Hz. Tomado de [2].

Debido a estas características, los TOR son interpretados como la respuesta de sistemas de fluidos altamente confinados y presurizados. En el volcán Galeras se han observado en estrecha asociación con etapas previas a erupciones explosivas, lo que los convierte en precursores sísmicos críticos, aunque su ocurrencia estadística suele ser extremadamente infrecuente [18].

Tremor volcánico (TRE)

El tremor volcánico corresponde a una forma de sismicidad cuasi-continua, asociada a una excitación persistente de la fuente. Se manifiesta como una señal sostenida en el tiempo con duraciones que pueden variar desde minutos hasta días, semanas o incluso meses. Su amplitud puede mantenerse relativamente constante o presentar variaciones graduales, reflejando cambios en la intensidad del proceso fluido subyacente. En algunos casos, el tremor puede originarse por la superposición continua de eventos discretos, como sismos LP disparados en rápida sucesión, que se encadenan hasta formar una señal aparentemente continua [2, 8].
Desde el punto de vista espectral, el tremor puede clasificarse en armónico cuando exhibe picos de frecuencia bien definidos y estables en el espectro. Este comportamiento, ilustrado en la Figura 4.5, evidencia una fuente de resonancia sostenida y regular en el interior del conducto volcánico.
Por otro lado, el tremor se clasifica como espasmódico cuando presenta una forma de onda irregular y una distribución espectral más difusa de banda ancha, tal como se expone en la Figura 4.6. Este tipo de sismicidad se asocia comúnmente a procesos de desgasificación,

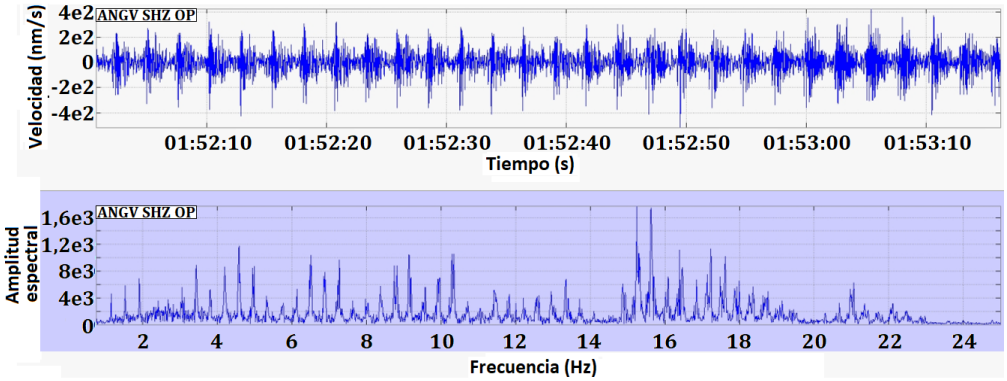


Figura 4.5: Tremor armónico registrado en la estación ANGV. Parte superior: sismograma con amplitud sostenida y oscilaciones regulares. Parte inferior: espectro caracterizado por múltiples picos de frecuencia estables y bien definidos. Tomado de [2].

ebullición hidrotermal y emisión de ceniza, constituyendo un indicador directo de la actividad volcánica en curso [5, 8].

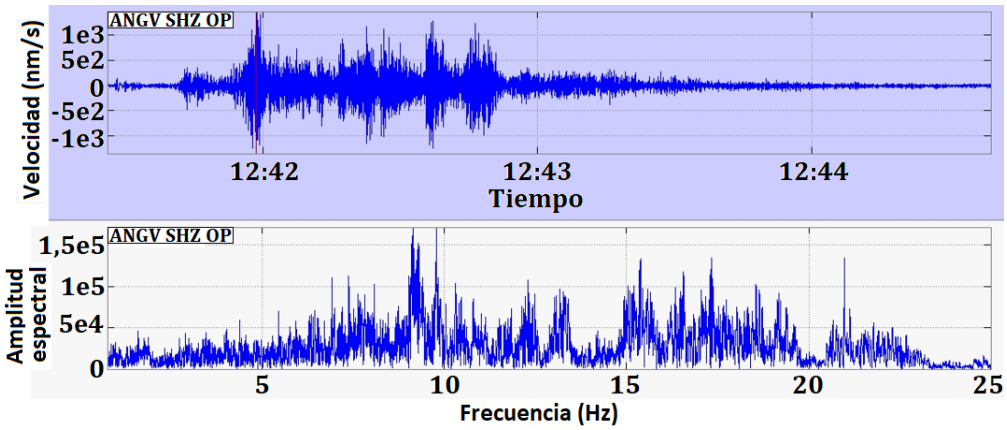


Figura 4.6: Tremor espasmódico registrado en la estación ANGV. Parte superior: sismograma de amplitud variable con forma de onda irregular. Parte inferior: espectro de banda ancha mostrando una distribución de energía difusa. Tomado de [2].

Las marcadas diferencias temporales y frecuenciales expuestas entre los eventos VT, LP, TOR y TRE constituyen la base para su análisis computacional. Cuantificar matemáticamente propiedades como el decaimiento de la coda, la distribución de energía espectral o la concentración armónica permite extraer un conjunto de características numéricas representativas de cada tipo de evento. Este proceso de parametrización es el paso preliminar indispensable para que arquitecturas de aprendizaje profundo puedan clasificar automáticamente la actividad sísmica del volcán.

4.2. Procesamiento y parametrización de señales

El procesamiento digital de señales (DSP) constituye la base matemática para transformar datos sísmicos crudos en representaciones cuantitativas aptas para el análisis automático. La calidad de la extracción de características paramétricas, entendidas como descriptores escalares derivados de la señal, influye directamente en el rendimiento, la capacidad de generalización y la robustez de un clasificador automático aplicado a entornos de sismología volcánica [5, 19].
Esta sección presenta el fundamento matemático de cada etapa de acondicionamiento, así como los descriptores estadísticos que permiten caracterizar las propiedades temporales y espectrales representativas de los eventos sísmicos volcánicos.

4.2.1. Acondicionamiento de señales

Las series de tiempo registradas por las estaciones sismológicas contienen, además de la señal de interés, ruido microsísmico de fondo, deriva instrumental y desplazamientos basales. Para garantizar que la extracción de características opere sobre una base física homogénea, la señal discreta debe someterse a un proceso de acondicionamiento temporal, como se describe en [6, 7].

Detrending y eliminación de tendencias lineales

Los sensores sísmicos de banda ancha pueden presentar fluctuaciones que generan una componente de corriente continua (DC offset) y una tendencia lineal espuria a lo largo de la traza temporal [19]. Para aislar la vibración sísmica, el flujo de preprocesamiento aplica secuencialmente un ajuste de media nula y un ajuste polinomial de primer grado [5]. Sea $s[n]$ la señal observada, compuesta por la señal sísmica real $x[n]$ y una deriva lineal espuria:

$$s[n] = x[n] + p[n], \tag{4.1}$$

En este caso, la tendencia se modela como

$$p[n] = an + b. \tag{4.2}$$

El objetivo es estimar los coeficientes a y b que minimizan la suma de los errores cuadráticos mediante regresión por mínimos cuadrados [20]:

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \sum_{n=0}^{N-1} \left[s[n] - (an + b) \right]^2 . \tag{4.3}$$

Una vez estimada la tendencia, definida como

$$\hat{p}[n] = \hat{a}n + \hat{b}, \tag{4.4}$$

la señal corregida se obtiene como

$$x[n] = s[n] - \hat{p}[n]. \tag{4.5}$$

Esta corrección es necesaria para centrar la señal en una referencia nula, evitando que el desplazamiento basal afecte métricas de energía, asimetría o curtosis durante la extracción de características.

Filtrado digital: filtro Butterworth pasa-altas de fase cero

Con el fin de atenuar el contenido de muy baja frecuencia, asociado principalmente a acoplamientos atmosféricos, microseísmos oceánicos y otras perturbaciones remanentes, la señal se somete a un proceso de filtrado digital. De acuerdo con [5, 7], se emplea un filtro pasa-altas diseñado a partir de un prototipo IIR tipo Butterworth. Esta familia se selecciona por presentar una respuesta en magnitud máximamente plana (*maximally flat*) en la banda de paso, lo que evita la introducción de rizos artificiales (*ripples*) que puedan distorsionar el espectro de la señal [21].
La magnitud al cuadrado de la respuesta frecuencial del prototipo analógico Butterworth pasa-altas está dada por

$$|H(j\omega)|^2 = \frac{1}{1 + \left(\frac{\omega_c}{\omega}\right)^{2N_f}}, \tag{4.6}$$

donde ω_c representa la frecuencia angular de corte y N_f corresponde al orden del filtro. En esta configuración, la frecuencia de corte se fija en $0{,}7$ Hz, equivalente a $\omega_c = 2\pi(0{,}7)$ rad/s, con el fin de preservar la banda de interés sismológico mientras se atenúan las componentes espurias.
Sin embargo, los filtros IIR causales introducen de manera inherente una distorsión de fase no lineal, la cual desplaza temporalmente los arribos impulsivos (*onsets*) y altera la morfología de la onda. Para corregir este efecto y evitar distorsiones temporales en los arribos de las ondas P y S, se implementa una técnica de filtrado bidireccional en fase cero (*forward-backward filtering*). Al procesar la señal secuencialmente en dirección de avance y retroceso temporal, la respuesta en frecuencia efectiva resulta

$$Y(e^{j\omega}) = X(e^{j\omega})H(e^{j\omega})H(e^{-j\omega}) = X(e^{j\omega})|H(e^{j\omega})|^2. \tag{4.7}$$

Dado que $|H(e^{j\omega})|^2$ es real y no negativa, la fase efectiva es nula en las frecuencias donde la respuesta no se anula. Este esquema elimina la distorsión de fase y preserva la localización temporal de la señal, lo cual resulta esencial en la discriminación entre eventos volcano-tectónicos (VT) y eventos de largo periodo (LP), donde la forma del arranque constituye un rasgo diferenciador crítico [7, 21].

Remuestreo y teorema de Nyquist

Los repositorios sísmicos suelen reunir registros adquiridos a frecuencias de muestreo heterogéneas. Para que los vectores paramétricos resulten dimensionalmente consistentes, todas las trazas deben homologarse a una tasa de muestreo uniforme [6, 7].

El teorema de muestreo de Nyquist-Shannon establece que una señal limitada en banda a una frecuencia máxima $f_{\text{máx}}$ puede reconstruirse sin aliasing si la frecuencia de muestreo satisface $f_s \geq 2f_{\text{máx}}$. En la práctica, antes de reducir la tasa de muestreo se aplica un filtro antialiasing pasabajas y, posteriormente, se realiza el remuestreo.
Si el proceso se interpreta como una interpolación ideal sobre una rejilla M veces más densa, la señal remuestreada puede expresarse como

$$x_{\text{int}}[n] = \sum_{k=-\infty}^{\infty} x_{\text{filtrada}}[k] \operatorname{sinc}\left(\frac{n}{M} - k\right), \tag{4.8}$$

donde $\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$ es la función sinc normalizada. En cambio, si el procedimiento corresponde estrictamente a una decimación entera por un factor M, la operación discreta se escribe como

$$x_{\text{dec}}[n] = x_{\text{filtrada}}[nM]. \tag{4.9}$$

En los casos en que las frecuencias de muestreo original y objetivo no guardan una relación entera, el remuestreo se implementa de forma racional. Una tasa de muestreo objetivo que cumpla el criterio de Nyquist, con margen suficiente para la banda de transición del filtro antialiasing, garantiza que no se pierda información espectral relevante para la clasificación.

4.2.2. Características paramétricas en el dominio del tiempo

La dinámica física de la fuente volcánica se manifiesta de forma directa en la evolución de la amplitud y en la distribución temporal de la forma de onda. Mientras que los procesos de ruptura frágil, como los eventos volcano-tectónicos (VT), exhiben llegadas impulsivas y duraciones cortas, los eventos asociados a la dinámica de fluidos (LP y TRE) presentan inicios emergentes o sostenidos. Por su parte, señales de resonancia discreta como el tornillo (TOR) destacan por un prolongado y simétrico decaimiento de su envolvente. En este dominio, la representación paramétrica permite extraer descriptores estadísticos escalares que cuantifican estas marcadas diferencias geométricas y de distribución de energía [22]. Este enfoque opera directamente sobre la serie discreta, capturando la firma temporal intrínseca de cada evento sin requerir su transformación al dominio frecuencial, lo que proporciona una base cuantitativa esencial para discriminar entre los cuatro tipos de sismos analizados. A continuación, se definen las características temporales empleadas para interpretar y diferenciar la dinámica de los cuatro tipos de sismos analizados:

Amplitud cuadrática media (RMS)

La amplitud cuadrática media (*root mean square*, RMS) es una medida estándar de la magnitud de una señal sísmica discreta de longitud N_s y se define como:

$$x_{\text{RMS}} = \sqrt{\frac{1}{N_s} \sum_{n=1}^{N_s} x[n]^2} \tag{4.10}$$

Esta métrica constituye un indicador útil del nivel promedio de amplitud y energía de un evento, por lo que resulta apropiada para caracterizar la intensidad relativa de distintas clases de señales volcánicas [6, 21].

Factor de cresta (CF)

Para cuantificar la relación entre el valor pico y el nivel energético promedio, se emplea el factor de cresta (CF), útil para distinguir señales impulsivas de señales más continuas [21]:

$$\mathrm{CF} = \frac{\mathrm{máx}_n \, |x[n]|}{x_{\mathrm{RMS}}} \tag{4.11}$$

Señales impulsivas de corta duración, como los eventos VT, tienden a exhibir factores de cresta más elevados que señales como el tremor o los eventos de largo periodo, en las cuales la energía se distribuye de manera más sostenida en el tiempo.

Duración total

Además de la amplitud RMS y el factor de cresta, la caracterización básica de un sismo requiere extraer su duración total. Esta métrica se aproxima como:

$$T = \frac{N_s}{f_s} \tag{4.12}$$

donde f_s es la frecuencia de muestreo.

Amplitud pico absoluta

Por su parte, la amplitud pico absoluta ($A_{\mathrm{máx}}$) representa el valor máximo alcanzado por la señal discreta en todo el registro temporal. Se define como:

$$A_{\mathrm{máx}} = \max_n |x[n]| \tag{4.13}$$

Energía total y normalizada

A partir de las cantidades anteriores, la energía total de la traza se define como la suma de los cuadrados de las amplitudes:

$$E_{\mathrm{total}} = \sum_{n=1}^{N_s} x[n]^2 \tag{4.14}$$

Para obtener una magnitud adimensional independiente de la escala absoluta de la señal, la energía normalizada se define como [5]:

$$E_{\mathrm{norm}} = \frac{E_{\mathrm{total}}}{N_s \, A_{\mathrm{máx}}^2} \tag{4.15}$$

Con esta definición, el descriptor queda acotado en el intervalo $[0, 1]$.

Concentración temporal de energía

Finalmente, para cuantificar cuánta energía se concentra al inicio del evento, se emplea el ratio de energía en la primera mitad:

$$R_{\mathrm{E},1/2} = \frac{\sum_{n=1}^{\lfloor N_s/2 \rfloor} x[n]^2}{\sum_{n=1}^{N_s} x[n]^2} \tag{4.16}$$

Los eventos volcano-tectónicos (VT), cuyo mecanismo de fuente está asociado a fracturas repentinas, tienden a concentrar una fracción mayor de su energía en el primer segmento temporal, a diferencia de los eventos de largo periodo (LP) y del tremor volcánico, en los que la energía suele distribuirse de manera más sostenida [5, 19].

Tasa de cruces por cero (*zero-crossing rate*)

La tasa de cruces por cero (*zero-crossing rate*, ZCR) cuantifica el número de cambios de signo de la señal por unidad de tiempo [21]. En esta tesis se define como:

$$\mathrm{ZCR} = \frac{1}{T} \sum_{n=2}^{N_s} \mathbb{I}\{x[n]\, x[n-1] < 0\} \tag{4.17}$$

Asimetría de la señal (*skewness*)

Asumiendo la amplitud de la señal como una variable aleatoria, la asimetría cuantifica la desviación de la distribución de amplitudes respecto de la simetría alrededor de su media empírica μ. Se define como el tercer momento central estandarizado [23]:

$$\gamma_1 = \frac{\frac{1}{N_s} \sum_{n=1}^{N_s} (x[n] - \mu)^3}{\sigma^3} \tag{4.18}$$

donde σ es la desviación estándar de la señal. Una asimetría marcada en la señal cruda puede asociarse con un desarrollo temporal no balanceado de la fuente o con saturación instrumental.

Curtosis (*kurtosis*)

La curtosis mide el peso relativo de las colas de la distribución y la presencia de valores extremos. En este trabajo se adopta el *exceso de curtosis* (*Fisher kurtosis*), definido como el cuarto momento central estandarizado al cual se sustrae el valor de referencia de la distribución gaussiana, de modo que una distribución normal posee exceso de curtosis igual a cero [23]:

$$\gamma_2 = \frac{\frac{1}{N_s} \sum_{n=1}^{N_s} (x[n] - \mu)^4}{\sigma^4} - 3 \tag{4.19}$$

Bajo esta convención, los valores positivos indican una mayor concentración de amplitudes extremas o impulsos aislados respecto a la referencia gaussiana, comportamiento coherente con mecanismos de fractura frágil; valores negativos suelen asociarse con señales más distribuidas en el tiempo y con menor presencia de eventos extremos [6].

4.2.3. Características paramétricas en el dominio de la frecuencia

Mientras que el dominio temporal describe la evolución energética de la señal, el dominio de la frecuencia revela la distribución espectral asociada a los mecanismos de fuente. La transición de la señal desde el dominio temporal al espectral se fundamenta en el análisis de Fourier. Dada la señal sísmica discreta $x[n]$ de longitud N_s, la transformada discreta de Fourier (*discrete Fourier transform*, DFT) descompone la señal en sus componentes armónicas [21]:

$$X[k] = \sum_{n=0}^{N_s-1} x[n]\, e^{-j2\pi kn/N_s}, \quad k = 0, 1, \ldots, N_s - 1 \tag{4.20}$$

A partir de la DFT se estima la densidad espectral de potencia (*power spectral density*, PSD), denotada $P_{xx}(f_k)$, que cuantifica la distribución de energía en función de la frecuencia. Para reducir la varianza del estimador frente al ruido, se emplea el método de Welch [5,21], el cual divide la señal en segmentos traslapados, aplica una ventana a cada segmento, calcula un periodograma modificado y promedia los resultados. Esta estrategia reduce las discontinuidades en los bordes y mejora la estimación en señales sísmicas no estacionarias o de duración prolongada, como el tremor volcánico. Sobre $P_{xx}(f_k)$ se construyen los descriptores espectrales presentados a continuación.

Espectro normalizado de potencia

A partir de la estimación de la PSD ($P_{xx}(f_k)$), se define el espectro normalizado. Esta representación actúa como una distribución de probabilidad y es útil para derivar métricas estadísticas posteriores. Se calcula como:

$$P_{\text{norm}}(f_k) = \frac{P_{xx}(f_k)}{\sum_k P_{xx}(f_k)} \tag{4.21}$$

Frecuencia dominante

La frecuencia dominante (f_{dom}) corresponde al pico espectral de mayor energía de la señal:

$$k^* = \arg\max_k P_{xx}(f_k), \qquad f_{\text{dom}} = f_{k^*} \tag{4.22}$$

Centroide espectral

Por su parte, el centroide espectral representa el promedio ponderado de la distribución frecuencial, interpretado como el "centro de masa" espectral [5]:

$$f_{\text{centroide}} = \sum_{k} f_k \, P_{\text{norm}}(f_k) \tag{4.23}$$

Entropía espectral de Shannon

La entropía espectral de Shannon (H_s) cuantifica la dispersión o complejidad del espectro tratando $P_{\text{norm}}(f_k)$ como una distribución de probabilidad sobre el eje frecuencial [5, 23]:

$$H_s = -\sum_{k=1}^{K} P_{\text{norm}}(f_k) \log_2\big(P_{\text{norm}}(f_k) + \epsilon\big) \tag{4.24}$$

Planitud espectral de Wiener (*spectral flatness*)

La planitud espectral, o medida de Wiener, compara la media geométrica con la media aritmética de la PSD [21, 24]:

$$\text{Flatness} = \frac{\exp\Big(\frac{1}{K}\sum_{k=1}^{K} \ln\big(P_{xx}(f_k) + \epsilon\big)\Big)}{\frac{1}{K}\sum_{k=1}^{K} P_{xx}(f_k)} \tag{4.25}$$

Frecuencia de corte espectral (*spectral rolloff*)

Se define como la frecuencia f_r tal que una fracción prefijada ρ de la potencia total del espectro se acumula por debajo de ella [21]:

$$f_{\text{rolloff}} = \text{mín}\left\{ f_k : \sum_{f_{k'} \le f_k} P_{xx}(f_{k'}) \ge \rho \sum_{k} P_{xx}(f_k) \right\} \tag{4.26}$$

Conteo de picos espectrales (*peak count*)

Esta métrica complementa a la planitud al cuantificar el número de máximos locales prominentes en la PSD, permitiendo detectar la presencia de resonancias armónicas discretas características de los eventos tipo tornillo [8].

Razón espectral de baja y alta frecuencia

Cuantifica la proporción de energía contenida en dos bandas diferenciadas del espectro, lo cual permite contrastar algorítmicamente la dominancia de procesos de fluidos frente a fracturas tectónicas [5, 25]:

$$R_{\text{spec}} = \frac{\sum_{f_k \in [f_{L1}, f_{L2}]} P_{xx}(f_k)}{\sum_{f_k \in [f_{H1}, f_{H2}]} P_{xx}(f_k)} \tag{4.27}$$

Fracción de energía en baja frecuencia

Para resumir la distribución acumulada, esta medida calcula el porcentaje de la potencia total que se concentra por debajo de un umbral fijo f_c [5, 19]:

$$F_{\text{LF}} = \frac{\sum_{f_k \leq f_c} P_{xx}(f_k)}{\sum_k P_{xx}(f_k)} \tag{4.28}$$

Ancho de banda de energía concentrada (B_{70})

Se define como la extensión frecuencial mínima necesaria para acumular al menos el 70 % de la potencia total [5, 21]:

$$B_{70} = f_{\text{sup}} - f_{\text{inf}}. \tag{4.29}$$

donde $[f_{\text{inf}}, f_{\text{sup}}]$ es el intervalo contiguo de mínima anchura que contiene al menos el 70 % de la potencia total.

Estabilidad de la frecuencia dominante

Cuantifica la variación temporal del pico espectral mediante ventanas deslizantes con el método de Welch. Se define como la desviación estándar de las frecuencias dominantes obtenidas en cada ventana:

$$\sigma_{f_{\text{dom}}} = \text{std}\left\{ f_{\text{dom}}^{(l)} \right\}_{l=1}^{L} \tag{4.30}$$

4.2.4. Características de la envolvente analítica y cinemática

Múltiples características del proceso de extracción dependen de la envolvente analítica de la señal, obtenida mediante la transformada de Hilbert [21]. La señal analítica $z[n]$ se construye sumando a la señal original su transformada de Hilbert en cuadratura:

$$z[n] = x[n] + j\,\mathcal{H}\{x[n]\} \tag{4.31}$$

La envolvente instantánea $A[n]$, que describe la evolución suave de la amplitud del evento sísmico, se obtiene como el módulo de la señal analítica:

$$A[n] = |z[n]| = \sqrt{x[n]^2 + \mathcal{H}\{x[n]\}^2} \tag{4.32}$$

Esta envolvente es fundamental en sismología volcánica porque permite cuantificar la cinemática de la fase inicial e impulsiva del evento.

Relación señal a ruido (SNR)

Expresa la calidad del registro como la razón logarítmica entre el nivel de amplitud del sismo y el ruido de fondo previo al arribo energético [21]:

$$\mathrm{SNR} = 20\log_{10}\left(\frac{x_{\mathrm{RMS}}}{x_{\mathrm{RMS,\,ruido}}}\right) \quad \text{[dB]} \tag{4.33}$$

Tiempo de pico de la envolvente (*envelope peak time*)

Corresponde al intervalo temporal desde el inicio del registro hasta que se alcanza el máximo global de $A[n]$:

$$t_{\mathrm{peak}} = \frac{\arg\max_n A[n]}{f_s} \tag{4.34}$$

Tiempo de ataque (*attack time*)

Se define como el instante en que la envolvente supera por primera vez una fracción alta de su valor máximo [5, 7]:

$$n_{\mathrm{attack}} = \min\left\{n : A[n] \geq \alpha_{\mathrm{att}} A_{\max}\right\}, \qquad t_{\mathrm{attack}} = \frac{n_{\mathrm{attack}}}{f_s} \tag{4.35}$$

Pendiente de inicio (*onset slope*)

Cuantifica la tasa de acumulación de energía de la envolvente durante la fase temprana del evento. Se estima mediante regresión lineal de mínimos cuadrados sobre la secuencia de energía $A[n]^2$ en una ventana inicial T_{onset} [5]:

$$S_{\mathrm{onset}} = \hat{\beta}_1, \quad A[n]^2 \approx \hat{\beta}_0 + \hat{\beta}_1 t_n, \quad t_n = \frac{n}{f_s}, \quad n = 0, \ldots, \lfloor T_{\mathrm{onset}} f_s \rfloor \tag{4.36}$$

Duración de la coda (*coda duration*)

La coda sísmica es la parte decreciente de la señal que sigue a la llegada de las ondas directas y está asociada con dispersión multitrayecto y reverberaciones en el medio volcánico [6, 7]. La duración de la coda se estima como el tiempo transcurrido desde el índice de máxima envolvente $n_{\mathrm{peak}} = \arg\max A[n]$ hasta el primer instante en que la envolvente cae por debajo de un umbral relativo θ respecto a su valor máximo:

$$T_{\mathrm{coda}} = \frac{1}{f_s} \cdot \min\left\{n' : A[n_{\mathrm{peak}} + n'] < \theta \cdot A_{\max}, \quad n' > 0\right\} \tag{4.37}$$

Asimetría de la envolvente (*envelope skewness*)

Complementariamente a la asimetría de la señal cruda, se calcula la asimetría de la envolvente analítica $A[n]$. Aplicando la misma fórmula del tercer momento central estandarizado sobre la envolvente:

$$\gamma_{1,\text{env}} = \frac{\frac{1}{N_s}\sum_{n=1}^{N_s}(A[n]-\bar{A})^3}{\sigma_A^3} \tag{4.38}$$

donde $\bar{A}$ y σ_A son, respectivamente, la media y la desviación estándar de $A[n]$. Valores positivos indican que la distribución de la envolvente presenta una cola hacia la derecha, compatible con señales de arranque rápido y decaimiento prolongado [5].

Estabilidad de la frecuencia instantánea

La frecuencia instantánea se obtiene a partir de la derivada temporal de la fase desenvuelta de la señal analítica [26]. En este trabajo, $\phi_{\text{uw}}[n]$ denota la fase desenvuelta de la señal analítica. Su desviación estándar ($\sigma_{f_{\text{inst}}}$) actúa como un descriptor de la variabilidad de fase a lo largo del evento:

$$f_{\text{inst}}[n] = \frac{f_s}{2\pi}\left(\phi_{\text{uw}}[n] - \phi_{\text{uw}}[n-1]\right) \tag{4.39}$$

$$\sigma_{f_{\text{inst}}} = \sqrt{\frac{1}{N_s-1}\sum_{n=1}^{N_s}\left(f_{\text{inst}}[n] - \overline{f}_{\text{inst}}\right)^2} \tag{4.40}$$

4.2.5. Coeficientes cepstrales de Mel (MFCC)

Históricamente utilizados en el reconocimiento automático del habla, los coeficientes cepstrales en las frecuencias de Mel (MFCC) constituyen una representación compacta de la envolvente espectral. En este trabajo se emplean como descriptores para capturar patrones espectrales asociados a señales sísmicas volcánicas no estacionarias [21].

Principio cepstral

El *cepstrum* se define como la transformada de Fourier inversa del logaritmo del espectro de magnitud. Esta operación homomórfica transforma la relación multiplicativa entre la fuente y el medio de propagación en una representación aditiva, lo que permite separar la envolvente espectral general de las oscilaciones resonantes de mayor frecuencia [27].

Mapeo a la escala Mel

A diferencia del cepstrum estándar, que opera sobre una escala frecuencial lineal en hercios (Hz), el cálculo de los MFCC proyecta la frecuencia a una escala psicoacústica modificada,

conocida como escala Mel. El banco de filtros espaciados logarítmicamente incrementa la resolución en las frecuencias bajas y comprime el extremo superior del espectro. La traslación de una frecuencia lineal f a la escala de Mel m se define analíticamente como [27]:

$$m = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \tag{4.41}$$

Esta escala comprime la resolución en frecuencias altas y aumenta la resolución en la banda de baja frecuencia, donde suelen concentrarse rasgos discriminativos de las señales volcánicas.

Secuencia algebraica de extracción

La extracción secuencial de un vector paramétrico MFCC sobre una señal sísmica discreta se modela computacionalmente mediante las siguientes etapas estandarizadas en el procesamiento de señales [27]:

1. **Transformada de Fourier de tiempo corto (STFT):** generación del espectro de potencia $|X[k]|^2$, donde K representa el número de bins de frecuencia en la ventana de análisis.

2. **Filtrado en el banco Mel:** un banco de M filtros triangulares $H_m[k]$, distribuidos sobre la escala Mel, filtra el espectro:

$$S[m] = \sum_{k=0}^{K-1} |X[k]|^2 \, H_m[k] \tag{4.42}$$

3. **Logaritmo de energía:** se calcula $\log(S[m] + \epsilon)$ para comprimir el rango dinámico de la energía y evitar problemas numéricos cuando alguna banda presenta energía nula.

4. **Transformada discreta del coseno (DCT-II):** se decorrelacionan las bandas Mel y se obtienen los coeficientes MFCC mediante la proyección ortogonal:

$$c_n = \sum_{m=1}^{M} \log(S[m] + \epsilon) \, \cos\left[\frac{\pi n}{M}\,(m - 0{,}5)\right], \quad n = 1, 2, \ldots, N_{\text{MFCC}} \tag{4.43}$$

El resultado es un vector compacto y decorrelacionado que resume la firma espectral del sismo. En este trabajo, cada coeficiente MFCC se resume mediante dos estadísticos temporales: la media aritmética y la desviación estándar, con el fin de conformar una muestra de dimensión fija que capture tanto el valor central como la variabilidad del coeficiente a lo largo del evento sísmico.

4.2.6. Fundamentos de selección de características (*feature selection*)

Una vez extraídos los descriptores paramétricos en los dominios del tiempo, la frecuencia y el cepstrum, no todas las variables aportan la misma cantidad de información discriminativa. En muchos casos, varias características pueden estar fuertemente correlacionadas o contener información redundante, lo que incrementa la dimensionalidad del problema sin mejorar necesariamente la capacidad de clasificación. En este contexto, la selección de características permite identificar cuáles descriptores contribuyen de forma más relevante a la separación entre clases sismológicas, al tiempo que se reduce el riesgo de sobreajuste asociado a la maldición de la dimensionalidad [19, 28].
Entre los enfoques más utilizados para evaluar la relevancia de las variables se encuentran los métodos basados en árboles de decisión. En particular, los ensamblajes de tipo *random forest* ofrecen una medida interpretable de importancia de variables a partir de la reducción de impureza que producen durante la partición del espacio de características.

Árboles de decisión y ensamblajes (*random forest*)

Un algoritmo de bosques aleatorios (*random forest*) consiste en un ensamblaje de B árboles de decisión independientes [29, 30]. Cada árbol del ensamblaje particiona el espacio de características mediante divisiones binarias sucesivas con el objetivo de maximizar la pureza de clasificación y, en consecuencia, reducir la impureza local.
En un espacio sísmico multiclase generalizado $\mathcal{Y} = \{c_1, c_2, \ldots, c_C\}$, se define un nodo genérico m del árbol que contiene N_m muestras de entrenamiento.

Métrica de impureza de Gini e importancia paramétrica

La regla de partición óptima para dividir los nodos del árbol obedece a la métrica de impureza de Gini, $I_G(t)$. Esta función cuantifica el grado de mezcla de clases dentro de un nodo. La métrica de Gini se define como [29]:

$$I_G(t) = 1 - \sum_{k=1}^{C} (p_{tk})^2 \tag{4.44}$$

donde p_{tk} es la proporción de muestras del nodo t que pertenecen a la clase k.
La partición en cada nodo se ejecuta maximizando el decremento ponderado de impureza $\Delta I_G(t)$, definido como:

$$\Delta I_G(t) = I_G(t) - \frac{N_{t_L}}{N_t} I_G(t_L) - \frac{N_{t_R}}{N_t} I_G(t_R) \tag{4.45}$$

donde t_L y t_R son los nodos hijos izquierdo y derecho, y N_t, N_{t_L} y N_{t_R} representan sus respectivos tamaños. La puntuación de importancia de una característica (*mean decrease impurity*) de la variable θ_j sobre todo el bosque aleatorio se obtiene como el promedio de la disminución acumulada de impureza:

$$\mathrm{Imp}(\theta_j) = \frac{1}{B}\sum_{b=1}^{B}\sum_{t\in T_b:\,\theta_j(t)=\theta_j} \Delta I_G(t) \tag{4.46}$$

donde B es el número de árboles del ensamblaje, T_b denota el conjunto de nodos del árbol b y $\theta_j(t)$ indica que la variable θ_j fue utilizada como criterio de partición en el nodo t.

4.3. Arquitecturas de aprendizaje profundo

El procesamiento avanzado de señales converge con las ciencias de la computación mediante clasificadores capaces de interpretar características paramétricas extraídas previamente, como los parámetros en el dominio del tiempo, en el dominio de la frecuencia o los coeficientes MFCC. En esta sección se presenta la evolución conceptual desde los algoritmos supervisados multiclase tradicionales hacia las arquitecturas basadas en autoatención, culminando en TabPFN como modelo fundacional para datos tabulares. En este marco, el conjunto etiquetado suministrado durante la inferencia se interpreta como un conjunto de contexto, mientras que el modelo mismo conserva sus pesos preentrenados y no requiere reoptimización local para cada nuevo catálogo.

4.3.1. Clasificación supervisada multiclase

El reconocimiento automático de sismos volcánicos se enmarca en un problema canónico de aprendizaje supervisado multiclase. Matemáticamente, se busca aproximar una función f_θ parametrizada mediante un conjunto de parámetros θ, la cual mapea un vector de características sísmicas $\mathbf{x} \in \mathbb{R}^d$ hacia un conjunto discreto de clases $y \in \{c_1, c_2, \ldots, c_C\}$, donde C representa el número total de clases geológicas a discriminar [7].

En un esquema computacional de clasificación, el modelo ajusta sus límites de decisión optimizando una función de pérdida (*loss function*). En problemas de clasificación multiclase excluyente, el estándar predominante es la pérdida de entropía cruzada categórica (*categorical cross-entropy*) [31]. Para el cómputo de esta pérdida en una muestra individual i, la etiqueta de clase y_i se codifica algebraicamente como un vector *one-hot* $\mathbf{y}_i \in \{0,1\}^C$ y se compara con el vector de probabilidades predichas $\hat{\mathbf{y}}_i$, asumido con componentes positivas y normalizadas:

$$\ell_{\mathrm{CE}}(\mathbf{y}_i, \hat{\mathbf{y}}_i) = -\sum_{c=1}^{C} y_{i,c}\ln(\hat{y}_{i,c}) \tag{4.47}$$

La pérdida promedio sobre el conjunto de N muestras se define como:

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\ln(\hat{y}_{i,c}), \tag{4.48}$$

donde $y_{i,c}$ es la componente c-ésima del vector de referencia y $\hat{y}_{i,c}$ es la probabilidad predicha por el modelo para dicha clase [7]. Los modelos tradicionales, como las *support vector machines* o los perceptrones multicapa, optimizan funciones objetivo que permiten separar distribuciones complejas; sin embargo, en registros con alta tasa de ruido estocástico inherente a los sismómetros, estas arquitecturas suelen requerir grandes volúmenes de datos limpios y balanceados. Esta condición rara vez se cumple en catálogos observacionales reales, lo que hace indispensable la adopción de esquemas de aprendizaje más robustos.

4.3.2. De CNN y RNN al Transformer

Antes de la adopción de mecanismos de atención global, la extracción de características profundas en geofísica se abordaba principalmente mediante dos enfoques. Por un lado, las redes neuronales convolucionales (*convolutional neural networks*, CNN) demostraron alta eficacia al procesar representaciones bidimensionales de la señal, como espectrogramas. Sin embargo, su campo receptivo fijo limita la captura de dependencias de largo alcance entre distintas fases del sismo sin incrementar de forma considerable la complejidad paramétrica de la red [32, 33]. Por otro lado, los datos sísmicos también se modelaron como series de tiempo no estacionarias, lo que orientó su procesamiento hacia las redes neuronales recurrentes (*recurrent neural networks*, RNN).

Una arquitectura RNN clásica procesa la secuencia de manera iterativa a partir del estado oculto previo $\mathbf{h}_{t-1}$ [32, 34]. Su actualización se expresa como:

$$\mathbf{h}_t = \tanh(\mathbf{W}_{hx}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h)\,, \tag{4.49}$$

donde $\mathbf{x}_t$ es la entrada en el instante t, $\mathbf{W}_{hx}$ y $\mathbf{W}_{hh}$ son matrices de pesos, y $\mathbf{b}_h$ es el sesgo del estado oculto.

El principal desafío de este esquema radica en que las colas de dispersión profunda, como las codas de sismos de largo periodo, pueden abarcar cientos o miles de puntos temporales. Durante la propagación hacia atrás a través del tiempo (*backpropagation through time*, BPTT), el gradiente asociado a dependencias lejanas se obtiene como el producto de múltiples jacobianos consecutivos. En consecuencia, cuando dichas matrices tienen norma menor que uno, el gradiente tiende a disminuir exponencialmente con la separación temporal [32]:

$$\frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_k} = \prod_{\tau=k+1}^{t} \frac{\partial \mathbf{h}_\tau}{\partial \mathbf{h}_{\tau-1}}, \qquad k < t. \tag{4.50}$$

Para mitigar esta limitación, Hochreiter y Schmidhuber introdujeron las arquitecturas con memoria a corto y largo plazo (*long short-term memory*, LSTM) [35]. Una red LSTM regula el flujo de información mediante compuertas multiplicativas. La puerta de olvido $\mathbf{f}_t$ controla qué fracción del estado previo se conserva:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f)\,, \tag{4.51}$$

donde $[\mathbf{h}_{t-1}, \mathbf{x}_t]$ denota la concatenación de ambos vectores. La puerta de entrada $\mathbf{i}_t$ regula la incorporación de nueva información, mientras que la celda candidata $\tilde{\mathbf{c}}_t$ propone el contenido potencialmente actualizado:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i)\,, \tag{4.52}$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c)\,. \tag{4.53}$$

El mecanismo de memoria a largo plazo se preserva en el estado de celda $\mathbf{c}_t$, que combina la retención del pasado con la información nueva:

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t. \tag{4.54}$$

Finalmente, la puerta de salida $\mathbf{o}_t$ filtra el estado de celda para producir el estado oculto actualizado:

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o)\,, \tag{4.55}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \tag{4.56}$$

Aunque el camino aditivo del estado de celda permite que el gradiente fluya con mayor estabilidad, la naturaleza recurrente obliga a la arquitectura a procesar la secuencia estrictamente de forma secuencial. Esta restricción limita la capacidad del modelo para relacionar el arribo inicial (*onset*) del evento con oscilaciones que ocurren decenas de segundos después en la coda [36]. Esta limitación, compartida en parte con el alcance local de las CNN, motivó un avance fundamental en el aprendizaje profundo: la paralelización del contexto computacional mediante mecanismos de atención.

4.3.3. Transformer y autoatención

El cambio de paradigma en el modelado de eventos no estacionarios se consolidó con la arquitectura Transformer [3] (véase la Figura 4.7). Su pilar algorítmico abandona las iteraciones temporales y las convoluciones espaciales en favor de un mecanismo de cómputo de dependencias globales denominado autoatención de producto punto escalado (*scaled dot-product self-attention*).

Embeddings y codificación posicional

Antes de que la red pueda procesar una secuencia de entrada, la matriz debe proyectarse hacia un espacio latente continuo. Dada una secuencia con N elementos y dimensionalidad original d_{input}, una transformación lineal genera la matriz de embedding $\mathbf{X}_{\text{emb}} \in \mathbb{R}^{N \times d_{\text{model}}}$, donde d_{model} es la dimensión del modelo.
A diferencia de las redes recurrentes, el mecanismo de atención por sí solo no incorpora información explícita sobre el orden secuencial; en consecuencia, es equivariante a permutaciones.

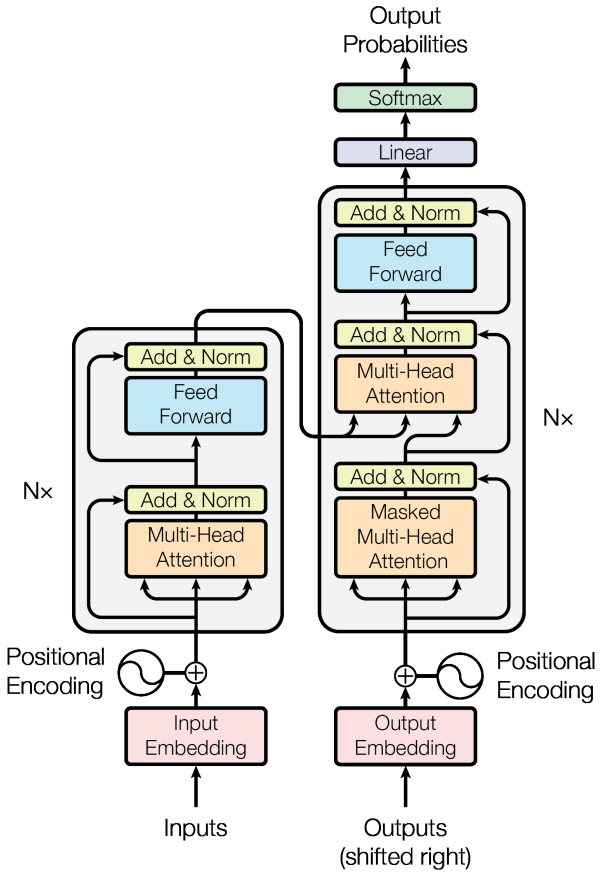


Figura 4.7: Arquitectura del modelo Transformer. Se ilustra la estructura conformada por el codificador (izquierda) y el decodificador (derecha), la cual prescinde de recurrencias y se basa íntegramente en mecanismos de autoatención. Tomado de [3]

Para secuencias donde el orden temporal es relevante, como texto o formas de onda crudas, se inyecta una codificación posicional $\mathrm{PE} \in \mathbb{R}^{N \times d_{\text{model}}}$. La formulación original [3] define esta codificación mediante funciones sinusoidales entrelazadas, donde $i = 0, 1, \ldots, \frac{d_{\text{model}}}{2} - 1$:

$$\mathrm{PE}_{(\text{pos},2i)} = \sin\left(\frac{\text{pos}}{10000^{2i/d_{\text{model}}}}\right) \tag{4.57}$$

$$\mathrm{PE}_{(\text{pos},2i+1)} = \cos\left(\frac{\text{pos}}{10000^{2i/d_{\text{model}}}}\right) \tag{4.58}$$

La matriz resultante codifica tanto el contenido como el orden secuencial:

$$\mathbf{X} = \mathbf{X}_{\text{emb}} + \mathrm{PE} \tag{4.59}$$

Matrices Q, K y V

La matriz de entrada $\mathbf{X}$ se somete a tres proyecciones lineales que generan las consultas ($\mathbf{Q}$), las claves ($\mathbf{K}$) y los valores ($\mathbf{V}$) [3]:

$$\mathbf{Q} = \mathbf{X}\mathbf{W}^Q, \quad \mathbf{K} = \mathbf{X}\mathbf{W}^K, \quad \mathbf{V} = \mathbf{X}\mathbf{W}^V \tag{4.60}$$

La atención escalada se define como:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V} \tag{4.61}$$

La operación de atención computa el grado de afinidad entre cada elemento de la secuencia. En particular, el producto $\mathbf{Q}\mathbf{K}^\top \in \mathbb{R}^{N\times N}$ genera una matriz cuadrada que evalúa la correlación de cada instante con todos los demás en un único paso computacional paralelo, con menor dependencia secuencial que las RNN [3]. Además, su complejidad temporal crece cuadráticamente con la longitud de la secuencia: $\mathcal{O}(N^2 d_k)$.
El término de escala $\sqrt{d_k}$ estabiliza los gradientes durante la retropropagación, mientras que la función *softmax* normaliza los pesos de atención a lo largo de cada fila. Al multiplicar esta matriz de probabilidades por los valores $\mathbf{V}$, la red conserva las características más relevantes para la tarea de clasificación [3, 36].

Atención multicabezal

La atención multicabezal permite proyectar la información de entrada en varios subespacios representacionales, de modo que el modelo pueda capturar dependencias complementarias en paralelo [3]. Para ello, la arquitectura se divide en h cabezales de atención independientes. Cada i-ésimo cabezal aplica una proyección lineal distinta sobre la matriz de entrada $\mathbf{X}$, mediante matrices de pesos $\mathbf{W}_i^Q \in \mathbb{R}^{d_{\text{model}}\times d_k}$, $\mathbf{W}_i^K \in \mathbb{R}^{d_{\text{model}}\times d_k}$ y $\mathbf{W}_i^V \in \mathbb{R}^{d_{\text{model}}\times d_v}$:

$$\text{head}_i = \text{Attention}(\mathbf{X}\mathbf{W}_i^Q, \mathbf{X}\mathbf{W}_i^K, \mathbf{X}\mathbf{W}_i^V) \tag{4.62}$$

Las salidas de los h cabezales se concatenan y se proyectan nuevamente al espacio de dimensión del modelo mediante una matriz de salida $\mathbf{W}^O \in \mathbb{R}^{h\,d_v\times d_{\text{model}}}$:

$$\text{MultiHead}(\mathbf{X}) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)\mathbf{W}^O \tag{4.63}$$

Posteriormente, la salida de la atención multicabezal se procesa mediante una red neuronal prealimentada (*feed-forward network*, FFN), aplicada de forma independiente a cada posición. La FFN se define como:

$$\text{FFN}(\mathbf{x}) = \text{ReLU}(\mathbf{x}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \tag{4.64}$$

Cada subcapa, ya sea de atención o de la FFN, se integra mediante una conexión residual seguida de normalización por capas (*Layer Normalization*) [37, 38]:

$$\mathbf{x}' = \text{LayerNorm}(\mathbf{x} + \text{Sublayer}(\mathbf{x})) \tag{4.65}$$

Esta combinación de atención en múltiples subespacios, procesamiento posición a posición y conexiones residuales constituye el núcleo del desempeño de las arquitecturas Transformer, permitiendo su exitosa adaptación a tareas de clasificación tabular y aprendizaje en contexto.

4.3.4. TabPFN y aprendizaje en contexto

Transferir el rendimiento predictivo de las arquitecturas Transformer [3] al dominio de las clasificaciones geofísicas ha presentado desafíos, debido a la naturaleza de las bases de datos tabulares. En este contexto, diversos estudios han observado que los modelos neuronales profundos pueden presentar un desempeño menor que los algoritmos potenciados por gradiente (*Gradient Boosting*), especialmente en tablas con características numéricas heterogéneas y sin sesgos inductivos espaciales o temporales [39]. Para abordar esta problemática, arquitecturas recientes como TabPFN estructuran la inferencia mediante aprendizaje en contexto sobre distribuciones sintéticas previas [4,40]. La Figura 4.8 resume este enfoque, en el cual se ilustra el esquema general de preentrenamiento e inferencia sobre conjuntos tabulares completos y la arquitectura interna del modelo basada en atención alternante entre características y muestras. A lo largo de este documento se reserva el término *entrenamiento* exclusivamente para la fase de preentrenamiento masivo realizada por los autores del modelo sobre distribuciones sintéticas previas, mientras que las muestras etiquetadas proporcionadas al modelo durante la clasificación del catálogo sismológico se denominan sistemáticamente *conjunto de contexto* ($\mathcal{D}_{\text{ctx}}$), siguiendo la nomenclatura original [4].

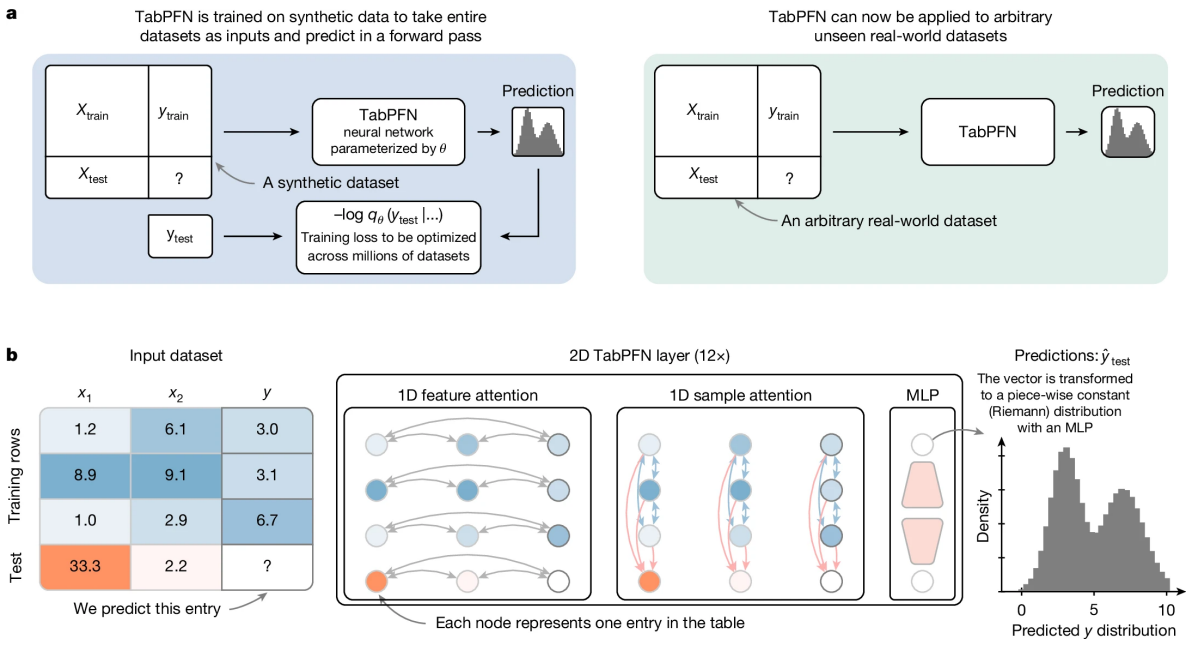


Figura 4.8: Esquema general de TabPFN y su arquitectura interna. (a) Flujo de preentrenamiento sobre datos sintéticos e inferencia sobre una tabla tabular con muestras etiquetadas de contexto y muestra de prueba en un solo *forward pass*. (b) Arquitectura basada en atención alternante entre atributos y muestras para procesar la tabla de entrada y producir la distribución predictiva. Tomado de [4]

Heterogeneidad tabular y representación por atributo

Mientras que una señal temporal pura posee una dependencia secuencial intrínseca, los parámetros sísmicos extraídos conforman una matriz de datos $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$. Cada registro sísmico $\mathbf{x}_i \in \mathbb{R}^d$ contiene d características heterogéneas que pueden incluir descriptores temporales, espectrales y cepstrales. Estos atributos no comparten una escala ni una distribución común [4, 39].

Para procesar esta estructura, el modelo proyecta la información empírica hacia un espacio latente continuo. De forma esquemática, cada valor escalar $x_{i,j}$ puede representarse mediante una transformación lineal compartida y un vector asociado al atributo:

$$\mathbf{e}_{i,j} = x_{i,j}\mathbf{w} + \mathbf{p}_j, \tag{4.66}$$

donde $\mathbf{w} \in \mathbb{R}^k$ es un vector de proyección compartido, $\mathbf{p}_j \in \mathbb{R}^k$ identifica el atributo j, y k denota la dimensión del espacio latente. Esta expresión debe interpretarse como una simplificación conceptual del codificado interno del modelo, que en la práctica emplea una transformación no lineal por celda en lugar de una proyección estrictamente lineal [4].

Para las muestras etiquetadas proporcionadas como contexto, esta representación se complementa con la información de su etiqueta de clase, la cual se integra como un token adicional al mismo tensor de entrada. El embedding de identidad por columna $\mathbf{p}_j$ se aleatoriza sistemáticamente durante el preentrenamiento, lo que induce en el modelo una tolerancia aprendida a permutaciones en el eje de atributos sin eliminar la necesidad de dicha señal de identidad [4].

Preentrenamiento con datos sintéticos y aprendizaje en contexto

El concepto subyacente de las redes ajustadas a partir de distribuciones previas (*Prior-Fitted Networks*, PFN) radica en trasladar la carga computacional desde el ajuste convencional del modelo hacia una fase de preentrenamiento masivo. Esta idea se fundamenta en la demostración teórica de que las arquitecturas atencionales pueden aproximar la inferencia bayesiana posterior [41]. Formalmente, el modelo busca aproximar la distribución predictiva posterior del proceso generativo latente:

$$p(y \mid \mathbf{x}, \mathcal{D}) = \int p(y \mid \mathbf{x}, \phi)\, p(\phi \mid \mathcal{D})\, d\phi, \tag{4.67}$$

donde ϕ denota los parámetros o el mecanismo generativo subyacente, cuya distribución a posteriori se rige por la regla de Bayes:

$$p(\phi \mid \mathcal{D}) \propto p(\mathcal{D} \mid \phi)\, p(\phi). \tag{4.68}$$

En lugar de ajustar iterativamente sus parámetros minimizando una pérdida empírica sobre el catálogo sismológico de interés, TabPFN preentrena su capacidad de inferencia mediante conjuntos de datos generados sintéticamente. Adaptando la notación original al marco de este trabajo, el objetivo de preentrenamiento (*prior-fitting*) puede expresarse como [40]:

$$\mathcal{L}_{\mathrm{PFN}} = \mathbb{E}_{(\mathcal{D}_{\mathrm{ctx}}, \mathbf{x}_{\mathrm{test}}, y_{\mathrm{test}}) \sim p(\mathcal{D})} \left[-\log q_\theta(y_{\mathrm{test}} \mid \mathbf{x}_{\mathrm{test}}, \mathcal{D}_{\mathrm{ctx}}) \right], \tag{4.69}$$

donde $p(\mathcal{D})$ es una distribución previa sobre conjuntos de datos sintéticos y q_θ representa la distribución predictiva aproximada por el modelo. En la práctica, TabPFN v2 y TabPFN-2.5 generan estas tablas sintéticas a partir de modelos causales estructurales (*structural causal models*, SCM), de modo que el preentrenamiento expone al modelo a una amplia variedad de mecanismos generativos posibles antes de su despliegue [4, 12].
A través de este procedimiento, el modelo aprende un algoritmo general de inferencia. Posteriormente, ante la consulta de un evento geofísico real, TabPFN realiza aprendizaje en contexto (*in-context learning*), aproximando la distribución posterior sin requerir actualizaciones de gradiente sobre los datos analizados [4, 41].
Como consecuencia directa de este paradigma, los parámetros θ del Transformer permanecen congelados durante toda la aplicación del modelo al catálogo sismológico de interés. La fase de ajuste al conjunto de contexto no involucra retropropagación alguna, sino únicamente el almacenamiento de $\mathcal{D}_{\text{ctx}}$ como evidencia empírica; la fase de inferencia, por su parte, se resuelve mediante un único *forward pass* sobre la tabla combinada contexto-prueba, sin actualizaciones de gradiente ni modificación de los pesos [4].

Atención alternante e invariancia permutacional

Como se observa en la Figura 4.8(b), la principal innovación arquitectónica de la familia TabPFN durante la inferencia es su mecanismo de atención alternante [4, 40]. En términos generales, las N_{ctx} muestras etiquetadas del conjunto de contexto y la muestra de prueba se integran en un tensor tridimensional de entrada $\mathbf{X} \in \mathbb{R}^{(N_{\text{ctx}}+1)\times(d+1)\times k}$, donde $N_{\text{ctx}}+1$ indexa las muestras, $d+1$ contempla los d atributos sísmicos más una columna adicional que codifica la etiqueta de clase, y k es la dimensión del espacio latente. En las filas de contexto esta columna adicional contiene el embedding de la etiqueta observada y_i, mientras que en la fila de prueba se utiliza un vector especial $\mathbf{m}_{\text{mask}} \in \mathbb{R}^k$, aprendido durante el preentrenamiento, que codifica la ausencia de etiqueta. La representación latente asociada a este vector tras atravesar las capas del Transformer constituye el estado $\mathbf{h}_{\text{test}}$ que alimenta la cabeza de clasificación [4].
Para procesar este tensor, las capas del Transformer alternan atención sobre características y sobre muestras [4]. Cada bloque atencional reutiliza la operación estándar de atención escalada, definida previamente en la ecuación (4.61), pero aplicada de forma especializada sobre cada eje del tensor de entrada. La particularidad arquitectónica de TabPFN consiste en alternar dos tipos de autoatención:

1. **Atención sobre características:** el modelo aplica autoatención entre las diferentes columnas de un mismo sismo. Para la capa l, esta operación puede escribirse como:

$$\mathbf{X}^{\left(l+\frac{1}{2}\right)}_{i,:,:} = \text{Attn}_{\text{feat}}\left(\mathbf{X}^{(l)}_{i,:,:}\right), \tag{4.70}$$

   donde el superíndice $\left(l+\frac{1}{2}\right)$ denota el estado intermedio de la representación tras la atención sobre atributos. Esta operación permite modelar dependencias locales entre características espectrales, temporales y cepstrales de un mismo evento sísmico. Dicha

atención se implementa mediante la misma formulación general de la ecuación (4.61), aplicada sobre la dimensión de atributos.

2. **Atención sobre muestras:** posteriormente, el modelo aplica autoatención a lo largo del eje de las muestras para una característica dada, comparando la muestra de prueba con las instancias etiquetadas del conjunto de contexto:

$$\mathbf{X}^{(l+1)}_{:,j,:} = \mathrm{Attn}_{\mathrm{samp}}\left(\mathbf{X}^{\left(l+\frac{1}{2}\right)}_{:,j,:}\right), \tag{4.71}$$

la cual también se basa en la operación general definida en la ecuación (4.61), pero aplicada sobre la dimensión de las muestras.

El tratamiento de los datos como conjuntos no ordenados tiene sus raíces en arquitecturas diseñadas para la invariancia sobre conjuntos, como el *Set Transformer* [42]. A través de esta atención alternante en dos direcciones, TabPFN trata las filas como un conjunto no ordenado y presenta robustez frente a permutaciones en el eje de las columnas, inducida por la aleatorización sistemática del orden de los atributos durante el preentrenamiento [4]. Esto permite tratar el catálogo sísmico empírico como un conjunto matemático y no como una secuencia temporal, independizando la predicción de cualquier ordenamiento artificial en la matriz de datos.

A diferencia de los Transformers aplicados a secuencias de lenguaje natural, TabPFN no incorpora codificaciones posicionales sinusoidales ni rotatorias. La única señal de identidad estructural proviene del embedding por columna $\mathbf{p}_j$ descrito en la ecuación (4.66), el cual actúa como un identificador de atributo más que como un índice ordinal; el eje de las muestras, por su parte, se trata explícitamente como un conjunto no ordenado, en coherencia con la naturaleza no secuencial de un catálogo sismológico [4, 42].

Internamente, cada capa de TabPFN encapsula los componentes estándar del Transformer descritos en la Sección 4.3.3, atención multicabezal (ec. (4.63)), red prealimentada por posición (ec. (4.64)) y conexiones residuales con normalización por capas (ec. (4.65)), pero los aplica de manera duplicada: una vez sobre el eje de atributos (ec. (4.70)) y otra sobre el eje de muestras (ec. (4.71)). En particular, TabPFN-2.5 apila entre 18 y 24 de estos bloques alternantes, manteniendo la topología base de TabPFN v2 [4, 12]. Adicionalmente, la transformación que mapea cada celda escalar al espacio latente (ec. (4.66)) se implementa como una pequeña red prealimentada por celda, y no como una proyección estrictamente lineal [4].

Los hiperparámetros arquitectónicos, número de capas, dimensión latente k y número de cabezales, forman parte del *checkpoint* preentrenado y permanecen fijos durante toda la aplicación al catálogo sísmico; el usuario únicamente controla la composición del conjunto de contexto $\mathcal{D}_{\mathrm{ctx}}$ y un conjunto reducido de banderas de inferencia, descritas posteriormente en el capítulo metodológico.

En esta investigación se emplea la versión TabPFN-2.5 [12][1]. Esta versión conserva la formulación base de aprendizaje en contexto y atención alternante, pero introduce mejoras orientadas a ampliar la capacidad de procesamiento de datos.

[1] A nivel de implementación, TabPFN-2.5 se usa por defecto en la librería oficial `tabpfn` a partir de la versión 6.0.0 [43].

En particular, TabPFN-2.5 fue diseñado para operar con conjuntos de datos de hasta $50\,000$ filas y $2\,000$ características [12]. Su costo de inferencia escala como

$$\mathcal{O}\big(r^2 \min(c, 500) + r \min(c, 500)^2\big)\,, \tag{4.72}$$

donde r es el número de filas y c el número de columnas [12]. Esta complejidad explica por qué la capacidad de contexto constituye un aspecto central de la escalabilidad del modelo.

Inferencia probabilística y representación de decisión

Mediante la evaluación contextual, la arquitectura procesa el tensor combinado y arroja una estimación condicionada a la evidencia empírica [41]:

$$q_\theta(y \mid \mathbf{x}_{\text{test}}, \mathcal{D}_{\text{ctx}}) = \mathcal{F}_\theta(\mathbf{x}_{\text{test}}, \mathcal{D}_{\text{ctx}}), \tag{4.73}$$

donde $\mathcal{F}_\theta$ representa el modelo preentrenado que estima la distribución posterior sobre las clases geofísicas.

La inferencia culmina cuando la red condensa la información extraída en el vector de representación final $\mathbf{h}_{\text{test}} \in \mathbb{R}^k$, correspondiente a la posición del token $\mathbf{m}_{\text{mask}}$ tras las capas de atención. Para traducir este vector a *logits* $(\mathbf{z})$, se aplica una proyección lineal final hacia un espacio de dimensión C (número total de clases):

$$\mathbf{z} = \mathbf{h}_{\text{test}} \mathbf{W}_{\text{out}} + \mathbf{b}_{\text{out}}, \tag{4.74}$$

donde $\mathbf{W}_{\text{out}} \in \mathbb{R}^{k \times C}$ y $\mathbf{b}_{\text{out}} \in \mathbb{R}^C$. La cabeza de clasificación de TabPFN admite de forma nativa hasta diez clases en una única *forward pass* [4], capacidad que cubre con holgura las cuatro categorías sismológicas consideradas en este trabajo (LP, TRE, VT y TOR).

Para calibrar la incertidumbre y obtener la probabilidad de pertenencia a cada clase c, la arquitectura aplica la función *softmax*, la cual puede estar ajustada por un parámetro de temperatura (T):

$$\hat{p}_c = P(y_{\text{test}} = c \mid \mathbf{x}_{\text{test}}, \mathcal{D}_{\text{ctx}}) = \frac{\exp(z_c/T)}{\sum_{j=1}^{C} \exp(z_j/T)}. \tag{4.75}$$

En la formulación empleada en este estudio, la temperatura adopta el valor $T = 0{,}9$ fijado por defecto en TabPFN-2.5 [12]. Este valor, calibrado por los autores durante el preentrenamiento sobre los conjuntos sintéticos, es inferior a la unidad y, por tanto, afila ligeramente la distribución predictiva respecto a un *softmax* no escalado; así, los *scores* reportados conservan las propiedades de calibración originales del modelo preentrenado y no requieren recalibración posterior.

Finalmente, para transicionar de esta distribución continua a una clasificación discreta, el modelo aplica la regla de máxima probabilidad obteniendo la etiqueta predicha $\hat{y}$:

$$\hat{y} = \arg\max_{c \in \mathcal{C}} \hat{p}_c, \tag{4.76}$$

donde $\mathcal{C}$ representa el conjunto de categorías sismológicas discretas.

4.4. Esquemas de validación cruzada

En problemas de clasificación supervisada aplicados a la geofísica, la confiabilidad de un modelo no depende únicamente de su capacidad de ajuste, sino también de la solidez del esquema empleado para estimar su generalización sobre datos no observados. En particular, arquitecturas de alta capacidad como los *Transformers* pueden capturar patrones complejos pero también sobreajustarse a regularidades espurias presentes en los datos empíricos, por lo que resulta indispensable adoptar estrategias de validación rigurosas que eviten la fuga de información entre los conjuntos de contexto y evaluación [29, 44]. En esta sección se presentan los fundamentos teóricos de los esquemas de validación empleados en este trabajo, desde el Hold-Out clásico hasta la Validación Cruzada de Monte Carlo (MCCV), esquema adoptado para la evaluación experimental.

4.4.1. Fundamentos de la validación cruzada

La estimación del error de generalización requiere aislar las muestras con las que se construye la frontera de decisión de aquellas con las que se evalúa. A continuación se describen los esquemas considerados, desde el más simple hasta el adoptado en este trabajo.

Hold-Out y K-Fold como antecedentes

La métrica convencional de retención (*Hold-Out*), que divide el conjunto de eventos una sola vez, produce estimaciones del error con alta varianza, fuertemente dependientes de la partición aleatoria inicial [29].
Para mitigar este problema, tradicionalmente se ha empleado la validación cruzada en k iteraciones (*k-fold Cross-Validation*), la cual particiona el conjunto empírico total $\mathcal{D}$ en k subconjuntos mutuamente excluyentes y de tamaño aproximadamente equitativo:

$$\mathcal{D} = \bigcup_{i=1}^{k} \mathcal{D}_i, \qquad \mathcal{D}_i \cap \mathcal{D}_j = \emptyset \quad \forall \quad i \neq j \tag{4.77}$$

El riesgo empírico cruzado se calcula promediando la función de pérdida $\mathcal{L}$ sobre las k iteraciones [29]:

$$\hat{R}_{\mathrm{CV}} = \frac{1}{k} \sum_{i=1}^{k} \mathcal{L}\Big(f_{\hat{\theta}(\mathcal{D} \setminus \mathcal{D}_i)}(\mathbf{X}_i), \mathbf{y}_i \Big), \tag{4.78}$$

donde $f_{\hat{\theta}(\mathcal{D} \setminus \mathcal{D}_i)}$ representa el modelo configurado con todos los datos disponibles excepto el i-ésimo pliegue ($\mathcal{D} \setminus \mathcal{D}_i$), el cual se evalúa posteriormente sobre las características $\mathbf{X}_i$ y las etiquetas $\mathbf{y}_i$ de dicho pliegue excluido.
En la variante estratificada (*Stratified k-fold*), cada subconjunto $\mathcal{D}_i$ respeta la distribución proporcional de clases del conjunto total, evitando la subrepresentación de categorías minoritarias. Sin embargo, este esquema impone particiones de tamaño fijo, lo que puede resultar

problemático cuando alguna clase es tan escasa que ciertos subconjuntos quedan sin representantes de esa categoría [44].

Validación Cruzada de Monte Carlo (MCCV)

El esquema adoptado en este trabajo es la Validación Cruzada de Monte Carlo (*Monte Carlo Cross-Validation*, MCCV), también conocida como *Repeated Random Subsampling* [29, 44]. A diferencia del k-fold, la MCCV no divide el conjunto en subconjuntos fijos de igual tamaño; en su lugar, realiza R divisiones aleatorias e independientes, cada una controlada por una semilla pseudoaleatoria distinta. En cada repetición r, se asigna aleatoriamente una proporción ρ del conjunto total como subconjunto de contexto, reservando el resto para evaluación:

$$\mathcal{D}_{\text{ctx}}^{(r)} \subset \mathcal{D}, \quad \mathcal{D}_{\text{test}}^{(r)} = \mathcal{D} \setminus \mathcal{D}_{\text{ctx}}^{(r)}, \quad \frac{|\mathcal{D}_{\text{ctx}}^{(r)}|}{|\mathcal{D}|} = \rho \tag{4.79}$$

En el marco de TabPFN, este subconjunto corresponde al conjunto de contexto suministrado al modelo durante la inferencia probabilística, mientras que el complemento se reserva para evaluación. A diferencia del k-fold, los subconjuntos de prueba de distintas repeticiones pueden solaparse:

$$\mathcal{D}_{\text{test}}^{(r)} \cap \mathcal{D}_{\text{test}}^{(r')} \neq \emptyset, \quad r \neq r' \tag{4.80}$$

Este comportamiento es esperado y no constituye un error metodológico; refleja el muestreo estocástico independiente de cada partición [44]. El estimador del error de generalización se obtiene como el promedio sobre las R repeticiones:

$$\hat{R}_{\text{MCCV}} = \frac{1}{R} \sum_{r=1}^{R} \mathcal{L}\left(f_\theta^{\mathcal{D}_{\text{ctx}}^{(r)}}(\mathbf{X}_{\text{test}}^{(r)}), \mathbf{y}_{\text{test}}^{(r)} \right) \tag{4.81}$$

La varianza de las métricas sobre las R repeticiones constituye un indicador de la estabilidad del clasificador ante cambios en la composición del contexto, cuantificando su robustez empírica [44].

Estratificación

Para que cada subconjunto de prueba sea representativo de la distribución real de clases en el catálogo sísmico, la MCCV se aplica en su variante estratificada: cada repetición garantiza que la proporción de cada clase en $\mathcal{D}_{\text{ctx}}^{(r)}$ y $\mathcal{D}_{\text{test}}^{(r)}$ sea aproximadamente igual a la del conjunto completo $\mathcal{D}$ [29]:

$$\frac{|\mathcal{D}_{c,\text{ctx}}^{(r)}|}{|\mathcal{D}_{\text{ctx}}^{(r)}|} \approx \frac{|\mathcal{D}_{c,\text{test}}^{(r)}|}{|\mathcal{D}_{\text{test}}^{(r)}|} \approx \frac{|\mathcal{D}_c|}{|\mathcal{D}|}, \qquad \forall\, c \in \mathcal{Y} \tag{4.82}$$

donde $\mathcal{Y} = \{c_1, c_2, \ldots, c_C\}$ denota el conjunto de clases geofísicas {LP, TOR, TRE, VT}. La estratificación es particularmente crítica en catálogos con la clase tornillo (TOR), cuya

frecuencia es muy inferior a la de las demás clases, ya que sin ella algunas repeticiones podrían generar subconjuntos de prueba sin ningún evento de esa categoría [19].

Peligro de fuga de datos

La fuga de datos (*data leakage*) ocurre cuando información del conjunto de evaluación, o del proceso físico que lo genera, se filtra inadvertidamente hacia el conjunto de contexto, sesgando de forma artificial y optimista las métricas de rendimiento predictivo [29].
En el análisis de series temporales y sismología volcánica, este riesgo se manifiesta cuando muestras que comparten una dependencia temporal o física estrecha se distribuyen simultáneamente entre el conjunto de contexto y el de prueba. Un ejemplo clásico ocurre durante el procesamiento de señales: la extracción paramétrica a partir de ventanas temporales solapadas sobre una misma traza continua genera vectores de características altamente correlacionados. Si estos vectores dependientes se separan mediante un muestreo puramente aleatorio, el modelo es capaz de memorizar la firma transitoria específica del segmento temporal, en lugar de aprender características físicas generalizables al comportamiento global de la fuente sísmica [45].
Para prevenir este fenómeno dentro del esquema MCCV, el diseño experimental debe garantizar una partición estructurada que preserve la estricta independencia estadística entre los subconjuntos en cada iteración. Esta precaución es especialmente crítica cuando se emplean técnicas de aumento de datos, donde múltiples versiones sintéticas se derivan de un mismo evento original: si la señal base quedara en el contexto y sus variantes aumentadas en la prueba, el modelo reconocería patrones previamente memorizados, invalidando las métricas de rendimiento [46]. La restricción de no-fuga exige que, durante cualquier repetición r, un grupo completo, compuesto por el evento original y todas sus derivaciones sintéticas, se asigne íntegramente al contexto o íntegramente a la evaluación, sin dividirse:

$$\forall\, G_m \in \mathcal{G}, \quad G_m \subseteq \mathcal{D}_{\text{ctx}}^{(r)} \quad \vee \quad G_m \subseteq \mathcal{D}_{\text{test}}^{(r)} \tag{4.83}$$

Esta restricción garantiza la total independencia estadística entre la matriz de contexto y las muestras de validación en cada iteración de la MCCV, asegurando que el modelo sea evaluado exclusivamente sobre sismos cuya dinámica de fuente es completamente desconocida para la red [29, 46].

4.4.2. Métricas de evaluación

En este apartado se describen las métricas estadísticas empleadas para medir la eficacia del modelo durante la fase de inferencia sobre el conjunto de prueba. Dado el contexto multiclase de la sismología volcánica y la importancia de evaluar correctamente cada tipo de evento, el análisis se fundamenta en los conteos derivados de la matriz de confusión, lo que permite una caracterización robusta del rendimiento algorítmico más allá de la proporción general de aciertos.

Matriz de confusión

La matriz de confusión $\mathbf{M} \in \mathbb{Z}^{C\times C}$ es una tabla cuadrada donde el elemento M_{ij} representa el número de eventos de la clase real i que el clasificador asignó a la clase predicha j. La diagonal principal M_{ii} corresponde a las predicciones correctas de cada clase, mientras que los elementos fuera de la diagonal representan los errores de clasificación [47]. Para una clase determinada c, se derivan los siguientes elementos:

- **Verdaderos Positivos (**TP_c**)**: Eventos reales de clase c correctamente clasificados como clase c.
- **Falsos Negativos (**FN_c**)**: Eventos reales de clase c clasificados incorrectamente como otra clase.
- **Falsos Positivos (**FP_c**)**: Eventos de otras clases clasificados erróneamente como clase c.
- **Verdaderos Negativos (**TN_c**)**: Eventos que no pertenecen a la clase c y fueron correctamente clasificados como no pertenecientes a c.

Exactitud global (*Accuracy*)

La exactitud global mide la proporción de predicciones correctas sobre el total de muestras evaluadas [47]:

$$\mathrm{Accuracy} = \frac{\sum_{c=1}^{C} \mathrm{TP}_c}{N} \tag{4.84}$$

donde TP_c es el número de verdaderos positivos de la clase c y N es el número total de muestras. Sin embargo, en problemas de clasificación con distribuciones asimétricas, un algoritmo que asigne la clase mayoritaria a todas las observaciones puede alcanzar una exactitud aparentemente alta sin detectar ninguna clase minoritaria. En consecuencia, esta métrica resulta insuficiente como indicador fiable de rendimiento en conjuntos desbalanceados.

Métrica de Precisión (*Precision*)

Cuantifica qué proporción de las predicciones asignadas a una clase específica corresponde verdaderamente a dicha categoría, penalizando las falsas alarmas [47]:

$$\mathrm{Precision}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c} \tag{4.85}$$

Métrica de Exhaustividad (*Recall*)

Mide la capacidad del modelo para detectar la totalidad de los eventos reales de una clase, penalizando las omisiones. Representa la proporción de eventos verdaderos que el algoritmo logró capturar respecto al total real de esa categoría [47]:

$$\text{Recall}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c} \tag{4.86}$$

Puntuación F1 (*F1-Score*)

El *F1-Score* combina Precisión y Exhaustividad mediante su media armónica, evitando que la maximización unilateral de una métrica compense la deficiencia de la otra. Se calcula por clase [47]:

$$\text{F1}_c = 2 \cdot \frac{\text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \tag{4.87}$$

F1-Score Macro

Para obtener una evaluación global independiente del volumen de cada clase, se emplea el *Macro F1-Score*, que estima el rendimiento general del clasificador promediando los *F1-Score* individuales de cada clase con igual peso [47]:

$$\text{F1}_{\text{macro}} = \frac{1}{C} \sum_{c=1}^{C} \text{F1}_c \tag{4.88}$$

Al tratar todas las clases con igual peso, el *Macro F1-Score* obliga al modelo a desempeñarse satisfactoriamente en cada categoría por separado; un clasificador que ignore las clases minoritarias no podrá obtener un valor alto en esta métrica, aunque su exactitud global sea elevada.

F1-Score Ponderado (*Weighted F1*)

El *Weighted F1-Score* complementa al *Macro F1* al ponderar cada *F1-Score* individual por el soporte (número de muestras reales) de la clase correspondiente [48]:

$$\text{F1}_{\text{weighted}} = \sum_{c=1}^{C} \frac{n_c}{N} \text{F1}_c \tag{4.89}$$

donde n_c es el número de muestras reales de la clase c y $N = \sum_{c=1}^{C} n_c$ es el total de muestras evaluadas. Esta métrica resulta útil cuando se desea que el rendimiento global refleje la contribución proporcional de cada clase al conjunto de datos, ofreciendo una perspectiva complementaria al *Macro F1* en contextos de desbalance moderado.

4.5. Ponderación probabilística de desbalance

En el análisis de catálogos sismológicos observacionales, el desbalance intrínseco de las clases geofísicas puede introducir un sesgo estadístico en la distribución predictiva, desplazando

los límites de decisión hacia las categorías mayoritarias. Para arquitecturas de inferencia directa, una alternativa analítica rigurosa consiste en implementar una ponderación probabilística a posteriori [49]. Este enfoque compensa la asimetría del catálogo operando exclusivamente sobre la regla de decisión final: una vez que el modelo genera el vector de probabilidades $\hat{\mathbf{p}}_i$ para cada muestra, sus componentes se modulan mediante un factor de escala w_c específico de cada clase. De esta manera, se amplifica la señal predictiva de las categorías minoritarias sin alterar la matriz de contexto empírico ni la representación latente del modelo [13].

4.5.1. El factor de suavizado γ y la regla de decisión ponderada

Para ajustar la intensidad de esta compensación, se introduce un factor de ponderación w_c para cada clase. El núcleo de este cálculo se fundamenta en la heurística de frecuencias inversas implementada por defecto en la librería scikit-learn [48], la cual se encuentra modulada por un hiperparámetro de suavizado $\gamma \in [0, 1]$ para este diseño experimental. El peso para una clase c se define formalmente como:

$$w_c = \left(\frac{N}{C \cdot n_c} \right)^{\gamma} \tag{4.90}$$

donde N es el total de muestras en el contexto, C el número total de clases y n_c la cantidad de ejemplares de la clase específica c presentes en el conjunto de datos. El exponente γ controla la agresividad del balanceo según los siguientes límites operativos [13]:

- **Sin balanceo ($\gamma = 0$):** Todos los pesos se reducen a la unidad ($w_c = 1$). El modelo opera bajo su configuración base, realizando un $\arg\max$ directo sobre las probabilidades originales sin intervención externa.
- **Compensación total ($\gamma = 1$):** Se aplica la ponderación inversamente proporcional a la frecuencia de clase, equivalente al balanceo estricto de scikit-learn [48]. Es el régimen de mayor intensidad, diseñado para corregir el desbalance más severo al amplificar la puntuación de las clases minoritarias, como los eventos tipo tornillo (TOR), aproximándose a un aprendizaje sensible a los costos puro.

Los valores intermedios de γ permiten un ajuste fino para mitigar el riesgo de que el ruido en categorías con muy pocos ejemplos domine la decisión.

Desde una perspectiva bayesiana, el modelo incorpora las frecuencias del contexto como una estimación de la probabilidad *a priori* $P(c)$. En catálogos donde una clase crítica representa una fracción mínima del total, su probabilidad posterior suele quedar suprimida por la dominancia estadística de las clases mayoritarias [49].
La ponderación mediante el factor w_c contrarresta matemáticamente este sesgo empírico. Al modular la sensibilidad de la regla de decisión, el sistema previene que las señales minoritarias sean sistemáticamente enmascaradas por la prevalencia de las clases mayoritarias, alineándose con los principios de calibración probabilística multiclase [50].

Regla de decisión ponderada (Weighted Argmax)

Una vez computado el vector de probabilidades $\hat{\mathbf{p}}_i \in [0,1]^C$ generado por el modelo fundacional TabPFN, se construye el vector de puntuaciones ponderadas $\mathbf{s}_i \in \mathbb{R}^C$ mediante la ponderación componente a componente con los factores de clase correspondientes:

$$s_{i,c} = w_c \cdot \hat{p}_{i,c}, \quad \forall\, c \in \{1, \ldots, C\} \tag{4.91}$$

La clase predicha final se obtiene seleccionando el índice que maximiza dicha puntuación ponderada dentro del conjunto de etiquetas $\mathcal{Y} = \{\text{LP, TOR, TRE, VT}\}$:

$$\hat{y}_i = \underset{c \in \mathcal{Y}}{\arg\max}\, s_{i,c} = \underset{c \in \mathcal{Y}}{\arg\max}\, [w_c \cdot \hat{p}_{i,c}] \tag{4.92}$$

La Ecuación 4.92 ajusta los valores de las probabilidades de cada clase para dar una mayor importancia relativa a las clases minoritarias. De este modo, cada factor w_c actúa como un amplificador que regula qué tan sensible es el modelo ante eventos que, debido a su baja frecuencia, suelen quedar enmascarados por las clases dominantes

4.5.2. Impacto sobre las métricas de evaluación

La ponderación posterior altera el equilibrio entre la precisión y la exhaustividad (*Recall*) de cada clase. Para la clase minoritaria tornillo, el incremento del hiperparámetro de suavizado γ implica una mayor propensión a la asignación de dicha etiqueta incluso cuando su probabilidad posterior no ponderada es baja, lo que conlleva un aumento de la exhaustividad a expensas de una reducción en la precisión [49]. En términos cualitativos, esta relación se define por un comportamiento inverso donde:

$$\gamma \uparrow \quad \Longrightarrow \quad \text{Recall}_{\text{TOR}} \uparrow, \quad \text{Precision}_{\text{TOR}} \downarrow \tag{4.93}$$

Este desplazamiento refleja un cambio en la sensibilidad del sistema diseñado para favorecer la detección de eventos críticos. En el caso de las clases mayoritarias, el efecto es opuesto, pudiendo generarse omisiones adicionales que se manifiestan como un deterioro marginal en sus respectivos valores de Exhaustividad. El impacto neto sobre el F1 Macro depende, por tanto, del equilibrio entre la ganancia de detección en la clase minoritaria y la degradación de las métricas globales del sistema, especialmente en escenarios donde la clase crítica cuenta con escasa representación en el catálogo [13, 49].

Es fundamental reconocer que este factor de ponderación introduce un compromiso metodológico: si bien permite equilibrar la importancia de las clases minoritarias en la regla de decisión, este ajuste condiciona el desempeño de las categorías dominantes. Al desplazar la frontera de decisión para maximizar la sensibilidad hacia señales críticas, se induce un incremento de falsos positivos en las clases mayoritarias, evidenciando que el reequilibrio del catálogo no constituye una operación neutra, sino una reconfiguración de la sensibilidad del clasificador.

Capítulo 5

Desarrollo del trabajo

El desarrollo del presente trabajo se estructuró siguiendo las fases fundamentales del ciclo de vida de un proyecto de Machine Learning (ML). La adopción de esta metodología garantiza un enfoque ordenado, empírico y reproducible, estrictamente alineado con el objetivo de clasificar la sismicidad del volcán Galeras mediante una arquitectura basada en Transformers. En este capítulo se detallan los procedimientos experimentales ejecutados, comenzando con la recolección, acondicionamiento y preprocesamiento de los registros sísmicos oficiales. Posteriormente, se presenta un análisis exploratorio de los datos para comprender la naturaleza de las señales, seguido del proceso de ingeniería y selección de características. Finalmente, se describe el modelado TabPFN, detallando el diseño experimental para evaluar su capacidad predictiva bajo distintos escenarios.

5.1. Preparación y recolección de datos

En esta sección se describe detalladamente el conjunto de datos y su estructura computacional. Para abordar el problema de clasificación automática definido en este trabajo, resulta indispensable disponer de un catálogo de eventos sísmicos debidamente etiquetados que represente de manera fiel la diversidad de mecanismos de fuente que operan en el volcán Galeras. Por consiguiente, a continuación se detalla el origen de la información, la distribución cuantitativa de las categorías sismológicas analizadas y los criterios de selección aplicados sobre los registros oficiales empleados para conformar la matriz de experimentación.

Origen y procedencia de los registros

Los registros sísmicos utilizados en este trabajo fueron proporcionados por el Observatorio Vulcanológico y Sismológico de Pasto (OVSP) del Servicio Geológico Colombiano (SGC), entidad encargada del monitoreo instrumental continuo del volcán Galeras. El catálogo comprende eventos registrados durante el período 2006–2010, etiquetados manualmente por analistas expertos del observatorio, quienes clasificaron cada evento según su firma sismológica conforme a las categorías definidas en la sección 4.1.3 del Capítulo 4

Los datos originales, almacenados en los sistemas de adquisición del OVSP, fueron convertidos al formato estándar miniSEED por personal autorizado del observatorio, garantizando así la integridad temporal y la preservación de los metadatos instrumentales (estación, red, componente, canal y frecuencia de muestreo) necesarios para el procesamiento posterior. Esta conversión constituye una etapa previa e independiente al presente trabajo.

Formato de almacenamiento: miniSEED

Las señales sísmicas se encuentran almacenadas en formato miniSEED (`.mseed`), un estándar internacional definido por la International Federation of Digital Seismograph Networks (FDSN) para el intercambio de series temporales digitales en sismología [51]. El formato miniSEED es una versión reducida del estándar SEED (*Standard for the Exchange of Earthquake Data*) que contiene exclusivamente los datos de la forma de onda junto con metadatos mínimos necesarios para su interpretación.

Cada archivo miniSEED encapsula una traza sísmica: una serie temporal discreta de amplitudes registradas por un sensor específico durante un evento determinado. Los metadatos integrados en cada archivo incluyen:

- **Código de estación:** identificador del sitio de registro (e.g., `ANGP`, `CUFP`).
- **Red sísmica:** código de la red instrumental a la que pertenece la estación (e.g., `CM` para la red del SGC).
- **Localización:** código de localización del sensor dentro de la estación.
- **Canal:** codifica el tipo de sensor y la orientación del componente (e.g., `HHZ` indica un sismómetro de banda ancha, componente vertical).
- **Frecuencia de muestreo:** tasa de adquisición del sensor en Hz.

La lectura y manipulación de estos archivos se realiza mediante la biblioteca ObsPy, una herramienta de código abierto ampliamente adoptada en la comunidad sismológica para el procesamiento de datos sísmicos en Python.

Estructura del dataset

El dataset presenta una estructura de cuatro niveles que organiza los registros de acuerdo con el tipo de evento sísmico, el año de ocurrencia, el evento individual y las trazas por estación. La estructura general se ilustra a continuación:
El nombre de cada carpeta de evento codifica la marca temporal del registro en formato `AAMMDDHHMM`, seguido de un sufijo que identifica la clase del evento:
Por ejemplo, la carpeta `0603072112GVA` corresponde a un evento Volcano-Tectónico tipo A registrado el 7 de marzo de 2006 a las 21:12 horas.

```
Dataset/
|-- LP/                                          <- Largo Periodo
|   |-- 06/                                      <- Año 2006
|   |   |-- 0605290429GLP/                       <- Evento individual
|   |   |   |-- ANGP.CM.20.ELZ.05290429_estaciones.mseed
|   |   |   |-- CUFP.CM.00.HHZ.05290429_estaciones.mseed
|   |   |   `-- ...
|   |   `-- ...
|   |-- 07/, 08/, 09/, 10/
|-- TOR/                                         <- Tornillo
|-- TRE/                                         <- Tremor
`-- VT/                                          <- Volcano-Tectónico
```

Figura 5.1: Estructura de directorios del conjunto de datos sísmicos.

Sufijo	**Tipo de evento**
`GLP`	Long Period
`GTO`	Tornillo
`GTR`	Tremor
`GVA`	Volcano-Tectónico A

Tabla 5.1: Sufijos de clasificación en los nombres de las carpetas de eventos del catálogo del OVSP.

Los nombres de los archivos miniSEED dentro de cada evento siguen la convención sismológica estándar del FDSN:

`ESTACIÓN.RED.LOCALIZACIÓN.CANAL`

Por ejemplo, el archivo `CUFP.CM.00.HHZ.05290429_estaciones.mseed` corresponde a la estación `CUFP`, perteneciente a la red `CM` (SGC), localización `00`, canal `HHZ` (sismómetro de banda ancha, componente vertical).

El componente del canal codifica la orientación del sensor según la convención:

Canal	**Componente**
`HHZ`, `ELZ`, `BHZ`	Vertical
`HHN`, `ELN`	Norte–Sur (horizontal)
`HHE`, `ELE`	Este–Oeste (horizontal)

Tabla 5.2: Orientación de las componentes sísmicas según el código de canal FDSN.

El número de archivos miniSEED dentro de cada carpeta de evento no es constante, dado que depende de las estaciones que registraron el fenómeno, los canales disponibles en cada

estación y la calidad del registro. Algunos eventos contienen más de 20 archivos (multi-estación, tri-componente), mientras que otros poseen un único archivo cuando solamente una estación registró el evento o fue la única señal considerada válida durante la clasificación original.

Distribución cuantitativa del catálogo

El catálogo crudo proporcionado por el OVSP se centra en cuatro categorías principales de eventos sísmicos compilados a lo largo de los años. La Tabla 5.3 presenta la distribución histórica reportada por el observatorio.

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 5.3: Distribución de eventos sísmicos por año y clase.

Cabe señalar que la clase Tremor (TRE) solo incluye registros de los años 2007 y 2008 en el catálogo proporcionado por el OVSP. Esto no implica que no se haya presentado actividad de tremor durante los años 2006, 2009 y 2010, sino que el volumen de registros disponibles para esos dos años ya era suficiente para conformar una representación adecuada de esta clase en el conjunto de datos.

La tabla 5.3 evidencia un desbalance estructural marcado en la distribución de clases. La categoría LP es la más representada, con el 62,7 % del total, mientras que la clase TOR es la menos frecuente, con apenas el 0,5 %. Esta asimetría implica que cualquier algoritmo de clasificación deberá enfrentar una distribución fuertemente desbalanceada.
Es importante reconocer que este desbalance no obedece a un error de muestreo ni a una omisión en la recolección temporal, sino que refleja de manera fiel la dinámica observada en el sistema volcánico del Galeras. Los eventos LP, asociados a procesos de largo período, se registran con mayor frecuencia, mientras que la sismicidad tipo TOR, considerada una variante especial de la sismicidad LP y caracterizada por señales armónicas cuasi-monocromáticas de larga duración, aparece con mucha menor frecuencia en el catálogo.

Distribución instrumental del catálogo

Antes de definir los criterios de filtrado, es necesario analizar la distribución de los registros sísmicos en función de las variables instrumentales. Este análisis permite identificar qué combinación de parámetros maximiza la cobertura de eventos para las cuatro clases seleccionadas, prestando especial atención a la clase minoritaria tornillo (TOR).

Dado que cada traza sísmica está identificada de manera única por una configuración específica de `Estación.Red.Localización.Canal`, resulta fundamental analizar cuáles configuraciones ofrecen la mayor disponibilidad y completitud de registros. La Tabla 5.4 presenta las 15 combinaciones instrumentales con mayor cantidad de registros en el directorio general, detalladas por cada clase de evento.

Combinación (Est.Red.Loc.Canal)	LP	TOR	TR	VA	Total
URCP.CM.20.ELZ	27 697	192	11 197	4 888	43 974
ANGP.CM.20.ELZ	27 697	193	11 197	4 884	43 971
URCP.CM.20.ELN	27 694	192	11 197	4 888	43 971
URCP.CM.20.ELE	27 694	192	11 196	4 888	43 970
NARP.CM.20.ELZ	27 688	191	11 196	4 889	43 964
OBVR.00.00.000	27 692	190	11 197	4 885	43 964
OBNR.00.00.000	27 691	190	11 197	4 885	43 963
OBER.00.00.000	27 692	190	11 194	4 886	43 962
CONP.CM.20.ELZ	27 523	191	11 196	4 877	43 787
IRIG.00.00.000	27 037	189	11 510	4 741	43 477
COBP.CM.20.ELZ	27 190	192	11 197	4 885	43 464
CAVZ.00.00.000	26 310	185	11 134	4 573	42 202
CAVE.00.00.000	26 306	185	11 131	4 574	42 196
CAVN.00.00.000	26 299	186	11 134	4 576	42 195
CUFP.CM.00.HHZ	23 201	168	11 155	4 286	38 810

Tabla 5.4: Las 15 combinaciones instrumentales con mayor número de trazas en el catálogo del OVSP.

Las estaciones de corto período `URCP` y `ANGP`, específicamente sus componentes verticales `ELZ`, son las que logran la mayor cobertura global de eventos. Adicionalmente, resulta crítico evaluar el comportamiento frente a la clase minoritaria TOR, la cual representa un desafío de desbalance estructural: la combinación `ANGP.CM.20.ELZ` es la única de todo el listado que logra registrar el mayor número absoluto de eventos tipo tornillo.

Selección de estación, red y componente

Un aspecto fundamental para garantizar la homogeneidad del conjunto de datos es la selección coherente de los parámetros de adquisición. Alimentar al clasificador con trazas provenientes de instrumentos heterogéneos introduciría sesgos instrumentales, de modo que el modelo aprendería a distinguir diferencias técnicas entre sensores en lugar de diferencias físicas asociadas a los mecanismos de fuente sísmica.

Con base en el análisis de la Tabla 5.4, se seleccionó la configuración específica que maximiza simultáneamente la representación de la clase minoritaria y la homogeneidad instrumental:

Parámetro	Valor	Justificación
Estación	`ANGP`	Estación de referencia del OVSP ubicada en la zona proximal del volcán. Registra aproximadamente el 99.0 % de los eventos tornillo (193/195), la mayor cobertura para esta clase entre todas las estaciones.
Red	`CM`	Red instrumental del Servicio Geológico Colombiano. Los sensores de la red siguen procedimientos de calibración y mantenimiento estandarizados.
Componente	`Z`	La componente vertical suele maximizar la relación señal-ruido en eventos volcánicos. El canal correspondiente para la estación `ANGP` es `ELZ` (localización `20`).

Tabla 5.5: Criterios de selección instrumental aplicados al catálogo del OVSP.

Resumen del conjunto de datos

La Tabla 5.6 sintetiza las especificaciones del conjunto de datos que alimenta las etapas subsecuentes del pipeline.

Especificación	Valor
Fuente	OVSP / Servicio Geológico Colombiano
Formato	miniSEED (`.mseed`)
Clases	LP, TOR, TRE, VT
Estación	ANGP
Red	CM (SGC)
Componente	Z (vertical)
Frecuencias de muestreo	100 Hz
Etiquetado	Personal autorizado del OVSP

Tabla 5.6: Resumen de las especificaciones del conjunto de datos utilizado.

5.1.1. Acondicionamiento y preprocesamiento de señales

Las señales sísmicas registradas en formato miniSEED contienen, además de la vibración volcánica de interés, artefactos instrumentales y ruido ambiental que deben eliminarse antes de la extracción de características paramétricas. La calidad del acondicionamiento determina directamente la fidelidad de los descriptores estadísticos y, en consecuencia, el rendimiento del clasificador.

En este apartado se describe la cadena de preprocesamiento implementada, especificando los valores concretos de los hiperparámetros seleccionados y su justificación sismológica. Cada traza se somete a las siguientes tres etapas:

1. Eliminación del offset de corriente continua (detrend constante).
2. Eliminación de la tendencia lineal espuria (detrend lineal).
3. Filtrado pasa-alto de fase cero.

La Tabla 5.7 resume los hiperparámetros configurados para esta etapa.

Hiperparámetro	**Valor**	**Referencia config**
Frecuencia de corte (pasa-altas)	0,7 Hz	`HIGHPASS_FREQ`
Frecuencia de muestreo objetivo	100,0 Hz	`TARGET_SAMPLERATE`
Ventana de ruido (SNR)	10 % inicial	`NOISE_WINDOW_FRACTION`
Ventana de onset	1,5 s	`ONSET_WINDOW_SEC`
Muestras mínimas de onset	5	`ONSET_MIN_SAMPLES`

Tabla 5.7: Hiperparámetros de la etapa de preprocesamiento de señales.

Eliminación de tendencias (Detrending)

Conforme al modelo matemático presentado en la Ecuación 4.1 del marco teórico, la señal observada $s[n]$ contiene una componente de vibración sísmica real $x[n]$ superpuesta a una deriva instrumental $p[n] = an + b$. La implementación aplica dos operaciones de `detrend` sucesivas mediante la biblioteca ObsPy:

1. **Detrend constante**: sustrae la media aritmética global de la traza, eliminando el *offset* de corriente continua (DC) generado por la polarización del sensor, centrando la señal en el cero.
2. **Detrend lineal**: estima los coeficientes $\hat{a}$ y $\hat{b}$ por mínimos cuadrados (Ecuación 4.3) y sustrae la recta ajustada, eliminando derivas instrumentales progresivas asociadas a fluctuaciones térmicas, cambios de presión barométrica o envejecimiento electrónico del sensor.

La aplicación en dos pasos (primero constante, luego lineal) mejora la estabilidad numérica de la regresión, dado que al operar sobre una señal previamente centrada en cero, la estimación de la pendiente $\hat{a}$ es menos sensible a valores extremos de offset.

Filtrado pasa-altas de fase cero

La señal centrada se procesa mediante un filtro IIR tipo Butterworth pasa-altas, con frecuencia de corte $f_c = 0{,}7$ Hz y filtrado en fase cero (`zerophase=True`).
La selección de f_c responde a un compromiso entre la reducción de ruido de muy baja frecuencia y la preservación del contenido espectral de interés:

- **Ruido atenuado** ($f < 0{,}7$ Hz): el microsismo oceánico y otras contribuciones de muy baja frecuencia quedan fuertemente atenuados.
- **Contenido preservado** ($f > 0{,}7$ Hz): los eventos LP concentran buena parte de su energía entre 1 y 5 Hz. En el caso de los VT, el contenido es predominantemente superior a 5 Hz. En consecuencia, el corte en 0,7 Hz busca reducir el ruido sin afectar de forma sustancial el contenido de interés.

Como se describe en el Capítulo 4 (sección 4.2.1), el filtrado en fase cero se logra mediante dos pasadas sobre la señal, una en sentido directo y otra en sentido inverso, con lo cual se cancela el desfase introducido por el filtro. Esta propiedad es fundamental para preservar la morfología temporal de los arribos impulsivos de las ondas P y S en los eventos VT, cuya detección depende del instante en que ocurre el cambio abrupto de amplitud.

Criterios de exclusión de trazas

No todas las trazas del catálogo son aptas para la extracción de características. Se establecen los siguientes criterios de exclusión automática:

- **Duración mínima:** se descartan trazas inferiores a 0,5 s (< 50 muestras a 100 Hz). Si bien la ventana de Welch se ajusta dinámicamente a la longitud de la señal ($N_{seg} = \text{mín}(256, N_s)$), por debajo de este umbral temporal se pierde la resolución frecuencial mínima requerida para un cálculo válido de la Densidad Espectral de Potencia (PSD).
- **Archivos corruptos:** se excluyen archivos miniSEED que no pueden ser leídos por la biblioteca ObsPy, sea por corrupción de datos, formato inválido o registros incompletos.

La Figura 5.2 sintetiza la cadena completa de acondicionamiento aplicada a cada traza individual antes de la extracción de características.

5.2. Análisis exploratorio de datos (EDA)

Previo a la fase de extracción de características, se realizó un análisis exploratorio de datos (EDA) con el fin de evaluar la integridad, coherencia física y calidad del conjunto de datos recopilado. Esta etapa permitió verificar que los eventos etiquetados en el catálogo original correspondieran efectivamente con la dinámica geofísica esperada para cada tipo de sismo volcánico.

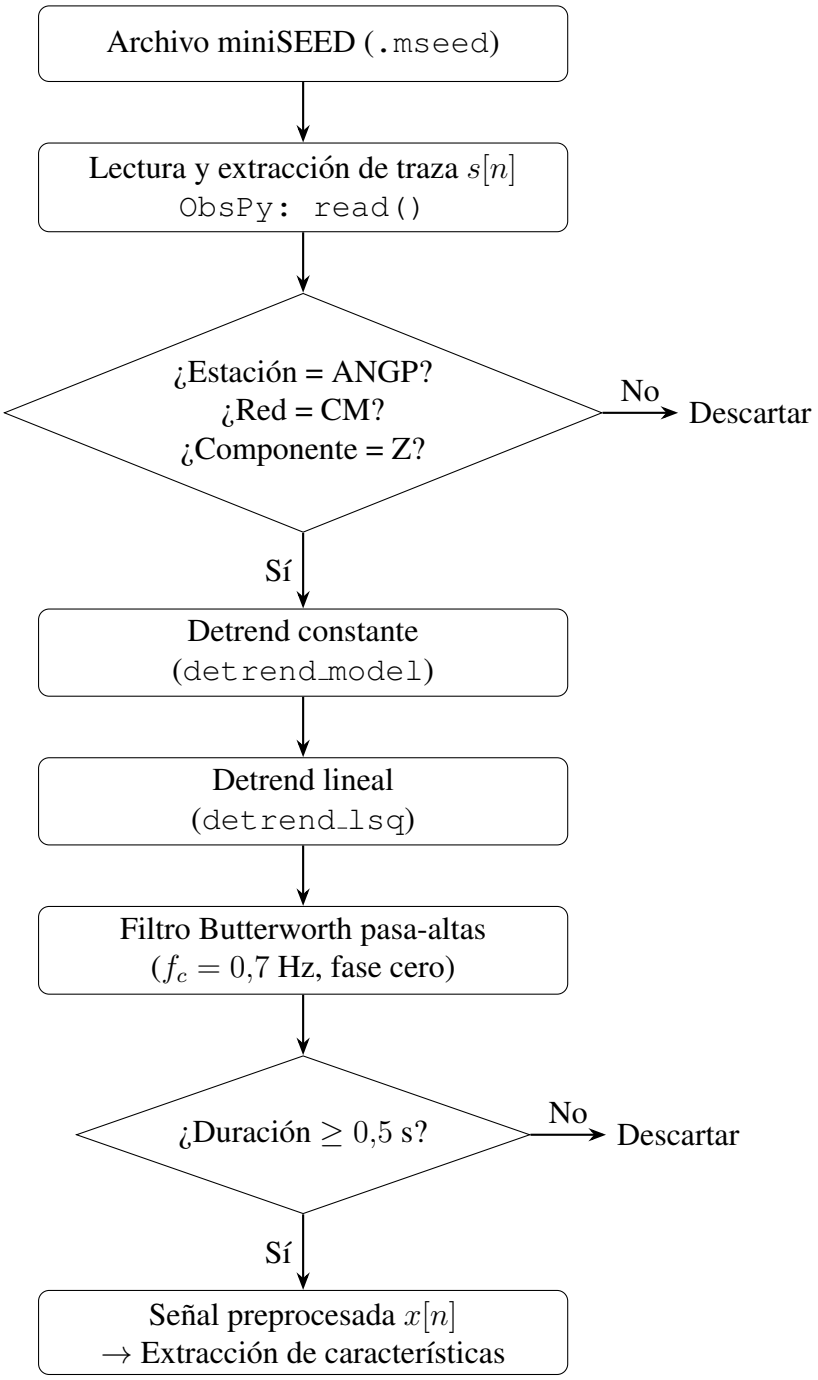


Figura 5.2: Diagrama de flujo de la cadena de preprocesamiento aplicada a cada traza mini-SEED.

Para validar el *dataset*, se seleccionaron ejemplos representativos de cada clase (LP, TRE, VT y TOR) extraídos directamente del catálogo de trabajo. Para cada evento se generaron tres representaciones complementarias: el sismograma de la forma de onda, el espectro de amplitud y el espectrograma, con el propósito de inspeccionar su morfología temporal, sus frecuencias dominantes y la evolución temporal-frecuencial de la energía, respectivamente.
Las validaciones visuales detalladas de cada clase se presentan en el Apéndice A. En conjunto, el contraste observado entre las representaciones temporales y espectrales sugiere que el preprocesamiento preservó las firmas físicas características de cada tipo de sismo, manteniendo la variabilidad estructural necesaria para la posterior fase de modelado y clasificación.

5.3. Ingeniería y selección de características

Una vez validada empíricamente la integridad y diversidad morfológica de las señales sísmicas durante el análisis exploratorio, el siguiente paso en el ciclo de vida del proyecto corresponde a la ingeniería de características (*Feature Engineering*). Dado que la arquitectura del modelo TabPFN requiere entradas tabulares, no es posible introducir la forma de onda cruda directamente al clasificador. Por lo tanto, es imperativo proyectar las trazas sísmicas hacia un espacio de características matemáticas que capturen la esencia física del fenómeno subyacente.
Esta fase se divide en dos etapas secuenciales. En la primera etapa, se ejecuta una extracción exhaustiva de descriptores paramétricos abarcando el dominio del tiempo, el dominio de la frecuencia, la envolvente analítica y los coeficientes cepstrales en la escala Mel (MFCC). En la segunda etapa, para mitigar la maldición de la dimensionalidad, reducir el ruido y optimizar el costo computacional, se aplica un algoritmo de selección de características basado en la importancia relativa de un ensamble de *Random Forest* [30]. Este procedimiento garantiza que el vector final que alimentará al modelo clasificador contenga únicamente la información más discriminativa y redundante mínima entre las distintas clases (LP, TRE, VT y TOR).

5.3.1. Extracción de características

En esta sección se describe la extracción de características a partir de la traza sísmica preprocesada. Cada traza se transforma en un vector numérico de dimensión finita $\boldsymbol{\theta}_i \in \mathbb{R}^d$, donde d es el número de descriptores que la representan cuantitativamente. Este vector conforma la representación que será procesada por el clasificador TabPFN. A lo largo de esta sección, n denota el índice de muestra discreto y $t = n/f_s$ el tiempo continuo equivalente, con f_s la frecuencia de muestreo de la traza.
La extracción produce un total de 53 descriptores por traza, organizados en cuatro familias funcionales: dominio temporal, dominio frecuencial, coeficientes cepstrales Mel (MFCC) y envolvente analítica con descriptores adicionales. Cada familia se describe en las subsecciones siguientes. Al final de la sección se tratan dos aspectos operativos transversales a todas las familias: el manejo de valores faltantes y la serialización del vector resultante en formato Parquet. Los 53 descriptores son sometidos posteriormente a un proceso de reducción de

dimensionalidad.

Hiperparámetros de la extracción

La Tabla 5.8 resume los parámetros configurables que gobiernan los algoritmos de extracción. Cada parámetro se encuentra centralizado en el archivo `config_extract.py`, lo que permite la reproducibilidad exacta de los resultados y su justificación física para el entorno volcánico.

Parámetro	Valor	Uso
N_MFCC	13	Número de coeficientes cepstrales
WELCH_NPERSEG	256	Longitud de ventana para el estimador de Welch
SPECTRAL_ROLLOFF	0.85	Percentil de energía acumulada (ρ)
ENERGY_BAND_PCT	0.70	Fracción de energía para ancho de banda
CODA_THRESHOLD	0.10	Umbral relativo de decaimiento de la coda
SNR_MIN_FOR_CODA	3.0 dB	SNR mínimo para estimar la coda
DOMFREQ_WINDOW_SEC	2.0 s	Ventana deslizante para estabilidad de f_{dom}
DOMFREQ_OVERLAP	0.50	Traslape entre ventanas deslizantes
SPECRATIO_LOW	1.0–3.0 Hz	Banda de baja frecuencia del ratio espectral
SPECRATIO_HIGH	8.0–20.0 Hz	Banda de alta frecuencia del ratio espectral
ONSET_WINDOW_SEC	1.5 s	Ventana fija para la pendiente de onset
NOISE_WINDOW_FRACTION	0.10	Fracción para estimar ruido de fondo

Tabla 5.8: Hiperparámetros base utilizados para los algoritmos de extracción.

Características del dominio temporal

El análisis en el dominio temporal permite describir la evolución de la señal sísmica a partir de 10 descriptores paramétricos, fundamentados teóricamente en el apartado 4.2.2. Estos cuantifican la estadística directa de la forma de onda y la distribución macroscópica de su energía (Tabla 5.9).

Características del dominio frecuencial

El análisis en el dominio frecuencial comprende 10 descriptores paramétricos, fundamentados teóricamente en la Sección 4.2.3, los cuales caracterizan la distribución de la energía espectral y la estabilidad de las frecuencias principales (Tabla 5.10).
La implementación computacional del espectro se fundamentó en el método de Welch, configurado con una ventana de 256 muestras (`WELCH_NPERSEG`). A 100 Hz, esto equivale a ventanas de 2.56 segundos, un hiperparámetro que ofrece la resolución espectral necesaria para discriminar eventos de baja frecuencia (LP), permitiendo a la vez promediar segmentos para mitigar la varianza del ruido.

#	Característica	Descripción
1	`rms`	Amplitud cuadrática media
2	`crest_factor`	Factor de cresta (razón pico a RMS)
3	`duration`	Duración total en segundos
4	`peak_amplitude`	Amplitud máxima absoluta
5	`total_energy`	Energía total de la señal
6	`normalized_total_energy`	Energía normalizada (adimensional)
7	`energy_ratio_first_half`	Ratio de energía concentrada en la primera mitad
8	`zero_crossing_rate`	Tasa de cruces por cero
9	`skewness`	Asimetría (tercer momento estadístico)
10	`kurtosis`	Curtosis (cuarto momento estadístico, exceso)

Tabla 5.9: Características paramétricas extraídas en el dominio del tiempo.

#	Característica	Descripción
1	`dominant_freq`	Frecuencia dominante (máxima energía espectral)
2	`central_freq`	Centroide frecuencial (promedio ponderado)
3	`spectral_entropy`	Entropía espectral de Shannon
4	`spectral_flatness`	Planitud espectral de Wiener
5	`spectral_rolloff`	Frecuencia de caída o corte (rolloff) al 85 %
6	`peak_count`	Conteo de picos espectrales prominentes
7	`spectral_ratio_low_high`	Razón de energías de baja y alta frecuencia
8	`low_freq_energy_fraction`	Fracción de energía espectral bajo los 5 Hz
9	`energy_band_70`	Ancho de banda que concentra el 70 % de la energía
10	`dominant_freq_stability`	Desviación estándar de f_{dom} sobre ventanas deslizantes

Tabla 5.10: Características extraídas en el dominio frecuencial.

A partir de este espectro, la métrica de caída (`spectral_rolloff`) se acotó definiendo un percentil de energía acumulada del 85 % (`SPECTRAL_ROLLOFF`). Utilizar un umbral más restrictivo que el estándar comercial evita que el descriptor sea arrastrado por el ruido de alta frecuencia instrumental. Asimismo, el ancho de banda principal se delimitó al 70 % de la energía total (`ENERGY_BAND_PCT`). La razón espectral requirió fronteras absolutas: 1.0–3.0 Hz para baja frecuencia (`SPECRATIO_LOW`) y 8.0–20.0 Hz para alta frecuencia (`SPECRATIO_HIGH`), forzando al algoritmo a contrastar directamente la banda de resonancia de fluidos frente a la banda de ruptura frágil.
Para evaluar la variabilidad espectral (`dominant_freq_stability`), el algoritmo segmentó la señal en ventanas de análisis de 2.0 s (`DOMFREQ_WINDOW_SEC`) con un traslape del 50 % (`DOMFREQ_OVERLAP`), obteniendo una estimación local estable de la frecuencia dominante capaz de representar la estructura de eventos tipo tornillo.

Características de la envolvente analítica y cinemática

A partir de la señal analítica calculada mediante la transformada de Hilbert, fundamentada teóricamente en la Sección 4.2.4, se extrajeron 7 descriptores orientados a caracterizar la mor-

fología dinámica, el arribo, la disipación del evento y la variabilidad de su fase (Tabla 5.11).

#	Característica	Descripción
1	`snr_estimate`	Razón Señal a Ruido (SNR estimada en decibelios)
2	`envelope_peak_time`	Tiempo, en segundos, hasta alcanzar la máxima amplitud
3	`attack_time`	Tiempo en alcanzar el 90 % del máximo de la envolvente
4	`onset_slope`	Pendiente inicial de acumulación de energía
5	`coda_duration`	Duración de la disipación o coda sísmica
6	`envelope_skewness`	Asimetría calculada sobre la envolvente analítica
7	`inst_freq_std`	Variabilidad de la frecuencia instantánea de la señal

Tabla 5.11: Descriptores extraídos a partir de la transformada de Hilbert.

La estimación automatizada de la duración de la coda (`coda_duration`) se reguló mediante dos hiperparámetros restrictivos: un umbral de decaimiento (`CODA_THRESHOLD`) fijado en el 10 % de la amplitud máxima, y una relación señal-ruido mínima (`SNR_MIN_FOR_CODA`) de 3.0 dB. El umbral garantiza que la medición cruce la barrera de las fluctuaciones transitorias antes de considerar finalizado el evento, capturando la reverberación real de los sismos LP y TOR, mientras que el umbral de SNR detiene el cálculo si el ruido microsísmico es muy alto, asignando un valor nulo para no contaminar el vector de características.
Para la cinemática de arribo, la métrica `attack_time` se definió mediante un umbral relativo del 90 % respecto al máximo de la envolvente, reduciendo la sensibilidad a espigas transitorias de ruido instrumental. Finalmente, la pendiente de inicio (`onset_slope`) se estimó mediante regresión lineal restringida a una ventana temporal fija de 1.5 s (`ONSET_WINDOW_SEC`) desde el inicio de la traza, evitando que en eventos cortos se incorporen porciones de decaimiento que sesguen la medición.

Coeficientes cepstrales en la escala Mel (MFCC)

La extracción de la firma cepstral, cuyo fundamento teórico se desarrolla en la Sección 4.2.5, se implementó restringiendo el número de coeficientes a $N_{\text{MFCC}} = 13$. Este valor se adoptó como una elección metodológica habitual, debido a que permite modelar la envolvente macroscópica del espectro con una representación compacta, mientras que los coeficientes de orden superior tienden a ser más sensibles a variaciones finas asociadas al ruido de alta frecuencia.
Dado que la longitud del conjunto de datos es heterogénea, con eventos cortos tipo VT frente a episodios prolongados de tremor, los hiperparámetros de extracción temporal se configuraron de manera adaptativa según la longitud N_s de cada traza:

- **Tamaño de la FFT:** $N_{\text{FFT}} = \max(32, \min(512, N_s))$. Esta cota evita definir ventanas de análisis mayores que la propia señal en eventos extremadamente cortos y mantiene un rango operativo estable para el cálculo espectral.
- **Tamaño de salto (*hop length*):** $\max(8, \lfloor N_{\text{FFT}}/4 \rfloor)$. Esta relación establece un traslape aproximado del 75 %, preservando la continuidad temporal de la representación sin depender de una longitud fija de la señal.

Para integrar esta matriz bidimensional a la estructura tabular del modelo, se calcularon la media y la desviación estándar de cada coeficiente sobre el eje temporal. De esta manera, la representación final quedó compuesta por 26 descriptores, capaces de capturar tanto la firma global como la variabilidad interna de cada evento.

Resumen del vector de características

La Tabla 5.12 sintetiza los 53 descriptores finales extraídos por cada traza, agrupados según su dominio físico y matemático de procedencia.

Grupo de características	**Cantidad**
Dominio temporal	10
Dominio frecuencial	10
Coeficientes MFCC (media y desviación)	26
Envolvente analítica y cinemática	7
Total	**53**

Tabla 5.12: Resumen de la dimensionalidad del vector de características paramétricas ($d = 53$).

Tratamiento de valores faltantes (NaN)

Ciertas características pueden generar valores `NaN` bajo condiciones específicas de la señal:

- `coda_duration`: indefinida cuando $SNR < 3{,}0$ dB.
- `onset_slope`: indefinida cuando la regresión lineal resulta numéricamente inestable, por ejemplo, en señales excepcionalmente cortas o con amplitud aproximadamente constante.
- `inst_freq_std`: indefinida cuando la señal contiene menos de 10 muestras válidas de frecuencia instantánea, es decir, valores dentro del rango $[0,\ f_s/2]$.
- Características frecuenciales: pueden resultar indefinidas en señales con energía espectral nula, condición que en la práctica no se observa dado el preprocesamiento previo aplicado a cada traza.

Los valores faltantes no se imputan durante la etapa de extracción: se almacenan en los archivos Parquet como `NaN` para preservar la trazabilidad de los datos. La imputación se realiza en una etapa posterior del flujo de clasificación, donde las medias del conjunto de contexto se calculan y aplican de forma independiente para cada partición, evitando así la fuga de información entre subconjuntos (*data leakage*).

Serialización y almacenamiento

Las características extraídas se almacenan en formato Apache Parquet, un formato columnar orientado al almacenamiento y análisis de datos, que ofrece compresión mediante Snappy y conservación de tipos numéricos, como `float64`. Cada fila del archivo corresponde a un vector de características individual y contiene:

- Las d características paramétricas, con prefijo de componente (`comp_Z_`).
- La etiqueta de clase: `label` $\in \{\text{LP}, \text{TOR}, \text{TRE}, \text{VT}\}$.
- Metadatos de trazabilidad: `event_group` (identificador del evento base), `event_id` (carpeta del evento), `station_code` (estación) y `file_name` (archivo miniSEED original).

La inclusión de metadatos en el mismo archivo permite verificar *a posteriori* la procedencia de cada muestra de contexto o prueba, facilitando la auditoría del flujo de procesamiento y la detección de anomalías en vectores asociados a eventos específicos. Una vez consolidados estos archivos, el paso siguiente consiste en definir el esquema de partición que estructura los datos para la evaluación experimental.

Partición y validación cruzada

Para la evaluación del modelo, se adopta una proporción 80 % contexto / 20 % prueba (`PARTITION_CTX_FRAC = 0.80`). Esta partición se realiza a nivel de evento (*event-grouping*) para evitar que vectores derivados de un mismo sismo se mezclen entre ambos subconjuntos. Además, se estratifica por clase, de modo que cada clase mantenga su peso relativo:

$$\frac{|\mathcal{D}_{c,\text{ctx}}|}{|\mathcal{D}_{\text{ctx}}|} \approx \frac{|\mathcal{D}_{c,\text{test}}|}{|\mathcal{D}_{\text{test}}|} \approx \frac{|\mathcal{D}_c|}{|\mathcal{D}|}, \qquad \forall\, c \in \mathcal{Y}. \tag{5.1}$$

Con el fin de estimar la variabilidad del rendimiento y no depender de una única división, el esquema anterior se repite cuatro veces bajo el principio de la validación cruzada de Monte Carlo (MCCV), empleando semillas pseudoaleatorias distintas en cada iteración (Tabla 5.13). Dado que cada partición se obtiene por sorteo independiente, los conjuntos de prueba de diferentes iteraciones pueden solaparse, comportamiento esperado y no constitutivo de error metodológico.

Partición	Semilla	Propósito
1	42	Variación pseudoaleatoria 1
2	123	Variación pseudoaleatoria 2
3	456	Variación pseudoaleatoria 3
4	789	Variación pseudoaleatoria 4

Tabla 5.13: Semillas del generador pseudoaleatorio para la división 80/20 de cada partición.

La estructura de salida física generada para cada una de las cuatro particiones refleja directamente este esquema de separación, organizando los archivos Parquet en carpetas diferenciadas por subconjunto y clase:

```
data_processed/partitions/
|-- Partition_1/
|   |-- context_80/
|   |   `-- features_context.parquet   <- 80 % (contexto)
|   `-- test_20/
|       |-- LP.parquet                 <- 20 % prueba (LP)
|       |-- TOR.parquet                <- 20 % prueba (TOR)
|       |-- TRE.parquet                <- 20 % prueba (TRE)
|       `-- VT.parquet                 <- 20 % prueba (VT)
|-- Partition_2/ ...
|-- Partition_3/ ...
`-- Partition_4/ ...
```

Figura 5.3: Estructura de directorios de las particiones de datos procesadas.

El archivo `features_context.parquet` contiene las instancias correspondientes al 80 % de los eventos asignados al conjunto de contexto, mientras que cada archivo de prueba contiene exclusivamente instancias de una clase específica.

5.3.2. Selección de características mediante Random Forest

El vector inicial de 53 descriptores de la sección 5.3.1 contiene información que puede llegar a ser redundante y características con bajo poder discriminativo. Se empleó un ensamblaje Random Forest para obtener un ranking de importancia y seleccionar el subconjunto que alimenta al clasificador TabPFN.

Configuración del ensamblaje

La Tabla 5.14 detalla los hiperparámetros empleados. El bosque se entrenó de forma independiente sobre el conjunto de contexto de cada una de la 4 particiones.

Hiperparámetro	**Valor**	**Justificación**
Número de árboles (B)	500	Estabiliza el ranking de importancia
Balanceo de clases	`balanced`	Pesos inversamente proporcionales a n_c
Semilla aleatoria	42	Reproducibilidad
Paralelización	`n_jobs = -1`	Todos los núcleos disponibles
Criterio de partición	Impureza de Gini	Métrica estándar (Ec. 4.44)

Tabla 5.14: Hiperparámetros de Random Forest utilizado para el ranking de características.

La opción `class_weight = "balanced"` asigna a cada clase un peso inversamente proporcional a su frecuencia en el conjunto de contexto, $w_c \propto 1/n_c$, con el fin de compensar el desbalance del catálogo y reducir el sesgo hacia la clase mayoritaria (LP).

Procedimiento y criterio de selección

Para cada partición $k \in \{1, \ldots, N_{\text{part}}\}$ se carga `features_context.parquet`, se imputan los valores faltantes (`NaN`) con la media de cada columna calculada sobre el mismo conjunto, y se entrena el modelo Random Forest. La importancia de cada característica θ_j se extrae como la disminución promedio ponderada de la impureza de Gini, siguiendo la formulación presentada en la Ec. (4.46) del marco teórico. Las importancias obtenidas en las N_{part} particiones se organizan en una matriz $\mathbf{I} \in \mathbb{R}^{N_{\text{part}} \times 53}$ y se promedian por columna:

$$\overline{\text{Imp}}(\theta_j) = \frac{1}{N_{\text{part}}} \sum_{k=1}^{N_{\text{part}}} \text{Imp}_k(\theta_j). \tag{5.2}$$

La desviación estándar inter-partición $\sigma_{\text{Imp}}(\theta_j)$ actúa como indicador de estabilidad del ranking. Con base en las importancias obtenidas, se seleccionaron las características cuya importancia promedio supera la media global del conjunto:

$$\theta_j \in \mathcal{S} \iff \overline{\text{Imp}}(\theta_j) \geq \frac{1}{d} \sum_{j=1}^{d} \overline{\text{Imp}}(\theta_j), \qquad d = 53 \tag{5.3}$$

La media global se estableció en $\overline{\text{Imp}}_{\text{global}} \approx 0{,}0189$.

Características seleccionadas

La Figura 5.4 muestra el segmento superior del ranking de importancia, correspondiente a las 17 características seleccionadas. El ranking faltante se presenta en el Apéndice B (Figuras B.1 y B.2). La Tabla 5.15 resume el subconjunto final retenido.

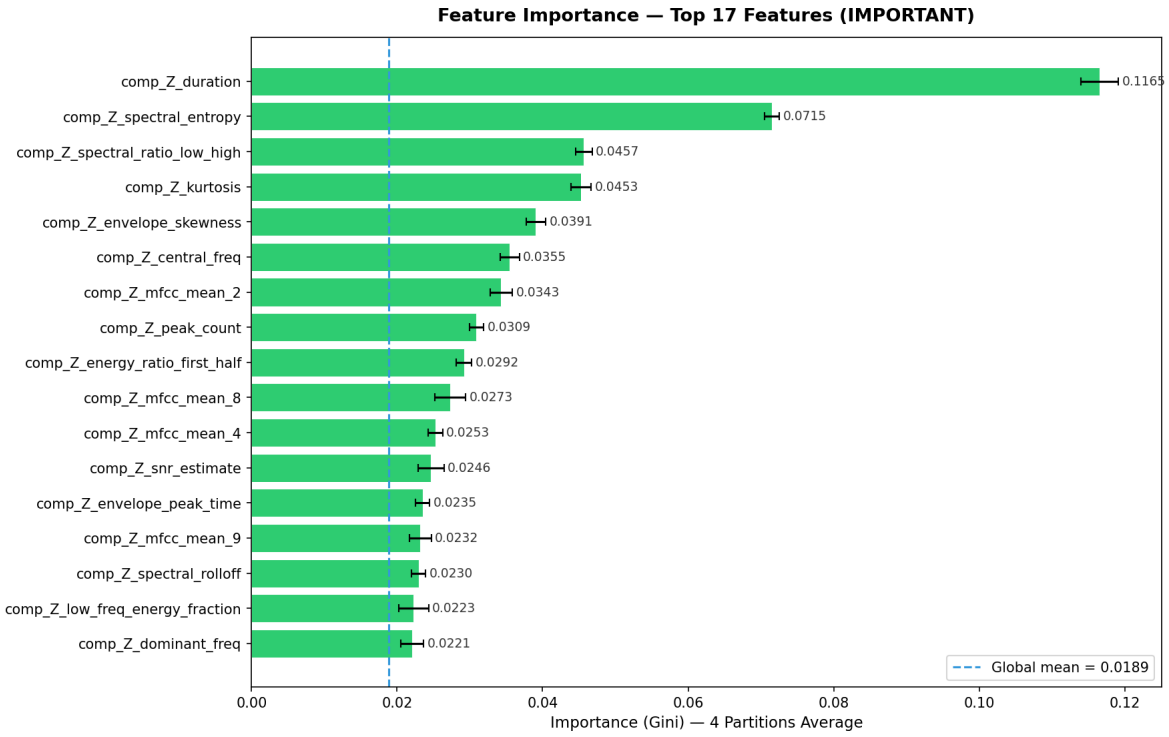


Figura 5.4: Ranking de importancia de las características 1 a 17. Las barras que superan la línea punteada azul, correspondiente a la media global, conforman el subconjunto seleccionado.

El ranking es consistente con la naturaleza física de los eventos analizados. La duración ocupa el primer lugar, lo que sugiere que la escala temporal del evento aporta una capacidad de discriminación relevante. Las variables frecuenciales también presentan una alta contribución, en particular la entropía espectral, que refleja la dispersión de la energía en el espectro. En cuanto a los coeficientes MFCC, solo cuatro medias de orden bajo fueron retenidas, mientras que no aparecen desviaciones estándar entre las primeras 17 variables, lo que sugiere que, para este conjunto de datos y bajo el criterio de selección empleado, la envolvente espectral promedio aporta más información discriminante que su variabilidad interna.

Vector final de características

La selección reduce el espacio de características de $d = 53$ a $d' = 17$ descriptores, lo que equivale a una reducción del 67,9 % en la dimensionalidad, y conserva el 63,9 % de la importancia acumulada del ranking basado en la métrica de Gini.

$$\boldsymbol{\theta}_i = [\theta_{i,1},\, \theta_{i,2},\, \ldots,\, \theta_{i,17}]^\top \in \mathbb{R}^{17} \tag{5.4}$$

donde $\theta_{i,j}$ denota la j-ésima característica seleccionada de la Tabla 5.15, evaluada sobre la i-ésima traza del conjunto de datos. Este vector $\boldsymbol{\theta}_i$ corresponde al registro $\mathbf{x}_i \in \mathbb{R}^d$ con $d = 17$ que ingresa como muestra al clasificador TabPFN descrito en la siguiente sección,

Rank	Característica	$\overline{\text{Imp}}$ (Gini)	Dominio
1	comp_Z_duration	0.1165	Temporal
2	comp_Z_spectral_entropy	0.0715	Frecuencial
3	comp_Z_spectral_ratio_low_high	0.0457	Frecuencial
4	comp_Z_kurtosis	0.0453	Temporal
5	comp_Z_envelope_skewness	0.0391	Temporal
6	comp_Z_central_freq	0.0355	Frecuencial
7	comp_Z_mfcc_mean_2	0.0343	Cepstral
8	comp_Z_peak_count	0.0309	Frecuencial
9	comp_Z_energy_ratio_first_half	0.0292	Temporal
10	comp_Z_mfcc_mean_8	0.0273	Cepstral
11	comp_Z_mfcc_mean_4	0.0253	Cepstral
12	comp_Z_snr_estimate	0.0246	Adicional
13	comp_Z_envelope_peak_time	0.0235	Adicional
14	comp_Z_mfcc_mean_9	0.0232	Cepstral
15	comp_Z_spectral_rolloff	0.0230	Frecuencial
16	comp_Z_low_freq_energy_fraction	0.0223	Frecuencial
17	comp_Z_dominant_freq	0.0221	Frecuencial
	Importancia acumulada	**0.6393**	

Tabla 5.15: Las 17 características seleccionadas y su importancia promedio sobre 4 particiones.

manteniendo así la coherencia entre la representación paramétrica y la entrada esperada por el modelo de aprendizaje en contexto.

5.4. Modelado del clasificador

Siguiendo el ciclo de vida del proyecto, una vez consolidada la matriz de características, la siguiente fase corresponde al diseño algorítmico y empírico del clasificador. La presente sección describe detalladamente la estructura de las pruebas computacionales implementadas. En primer lugar, se define la estrategia de partición y muestreo de los datos, indispensable para garantizar una validación representativa y evitar el filtrado de información. Posteriormente, se establece la configuración operativa de los modelos, destacando la particularidad de la arquitectura TabPFN, la cual opera bajo el paradigma de aprendizaje en contexto (*in-context learning*).

5.4.1. Estrategia de muestreo

El procedimiento de muestreo parte de un cupo teórico por clase calculado como

$$x = \left\lfloor \frac{\texttt{MAX\_TRZ}}{C} \right\rfloor, \tag{5.5}$$

donde `MAX_TRZ` representa el número máximo de trazas a conservar en el experimento, cuyo valor se justifica en la configuración de TabPFN (apartado 5.4.2), y C es el número de clases incluidas. A continuación, se evalúa la cardinalidad mínima del conjunto de clases, $n_{\text{mín}} = \text{mín}_c(n_c)$. Si se cumple que $n_{\text{mín}} \geq x$, el conjunto se considera balanceado y se seleccionan x trazas por clase. En caso contrario, la clase minoritaria se conserva completa y el muestreo de las clases restantes se ajusta repartiendo el cupo disponible entre las demás clases.

En este caso, con $C = 4$ y `MAX_TRZ` = 12 500, según la Tabla 5.3 la clase minoritaria corresponde a TOR, con $n_{\text{mín}} = 195$. Por tanto, el cupo disponible para las clases mayoritarias se calcula como

$$N_{\text{per\_class}} = \left\lfloor \frac{\texttt{MAX\_TRZ} - n_{\text{mín}}}{C - 1} \right\rfloor = \left\lfloor \frac{12\,500 - 195}{3} \right\rfloor \approx 4\,101. \tag{5.6}$$

El residuo de la división se asigna a las clases mayoritarias para alcanzar el límite exacto de muestras. El procedimiento de muestreo queda de la forma:

$$\mathcal{E}_c = \begin{cases} \text{Sample}(\mathcal{E}_c^{\text{total}},\, 4\,101,\, \text{seed} = 42) & \text{si } |\mathcal{E}_c^{\text{total}}| > 4\,101 \\ \mathcal{E}_c^{\text{total}} & \text{si } |\mathcal{E}_c^{\text{total}}| \leq 4\,101 \end{cases} \tag{5.7}$$

donde $\mathcal{E}_c^{\text{total}}$ es el conjunto completo de muestras de la clase c y $\mathcal{E}_c$ el subconjunto resultante.

5.4.2. Clasificador TabPFN

Con los subconjuntos de datos formalmente particionados mediante las estrategias de muestreo descritas, las matrices resultantes, compuestas por los 17 descriptores paramétricos y consolidadas en formato binario (*Parquet*) para cada pliegue de validación, constituyen la entrada directa para la fase de modelado. Operativamente, el flujo de inferencia computacional se orquesta mediante el módulo `Transformer.py`, el cual se encarga de cargar e inyectar secuencialmente los conjuntos de contexto y prueba en el algoritmo clasificador.
A diferencia de los algoritmos de clasificación tradicionales, TabPFN opera estrictamente bajo el paradigma de aprendizaje en contexto (*in-context learning*). Esto implica que el modelo no ejecuta actualizaciones de parámetros internos durante esta etapa; en su lugar, utiliza el conjunto de soporte exclusivamente como base empírica de referencia. A partir de esta matriz de contexto, condiciona sus predicciones y evalúa su rendimiento sobre los datos reservados para prueba mediante un proceso de inferencia directa.

Este apartado describe la configuración interna del clasificador, el esquema de preprocesamiento aplicado a las características de entrada, la generación de los vectores de probabilidad y el criterio matemático empleado para la decisión final.

Configuración del clasificador

El clasificador se inicializa con los hiperparámetros detallados en la Tabla 5.16. Se utiliza la versión `tabpfn` 6.4.1, cuyo modelo por defecto corresponde a TabPFN-2.5 [12]. De acuerdo con la documentación oficial de los desarrolladores, esta arquitectura está diseñada para operar óptimamente con conjuntos de hasta 50 000 muestras y 2 000 características.

Hiperparámetro	Valor	Descripción
`n_estimators`	4	Número de estimadores utilizados en el ensamblaje interno del modelo.
`random_state`	42	Semilla para garantizar reproducibilidad.
`n_preprocessing_jobs`	`-1`	Paralelización del preprocesamiento interno.
`ignore_pretraining_limits`	`True`	Permite operar con conjuntos de datos fuera de los límites de preentrenamiento recomendados.

Tabla 5.16: Hiperparámetros de inicialización del clasificador TabPFN-2.5.

Cabe precisar que el límite de `MAX_TRZ`=12 500 trazas por experimento, de las cuales 10 000 se destinan al contexto y 2 500 a la prueba, no corresponde a una restricción intrínseca del modelo (TabPFN-2.5 admite hasta 50 000 muestras y 2 000 atributos), sino a una decisión metodológica derivada del hardware disponible: en pruebas preliminares, superar aproximadamente 10 000 instancias de contexto generaba inestabilidad por consumo excesivo de memoria, comprometiendo la reproducibilidad de las ejecuciones.
La activación de la bandera `ignore_pretraining_limits=True` responde, por su parte, a una causa independiente del tamaño teórico del modelo. Dado que las inferencias se ejecutan en CPU, la librería `tabpfn` aplica por defecto una validación preventiva que bloquea ejecuciones con más de 1 000 muestras de contexto, con el fin de proteger al usuario frente a tiempos de cómputo prohibitivos. Puesto que el conjunto de contexto de este estudio supera ese umbral operativo aun siendo muy inferior a los límites del preentrenamiento, se habilita explícitamente esta bandera, lo cual permite que el clasificador proceda con el conjunto de 10 000 instancias sin que la validación interna interrumpa la ejecución.

Preprocesamiento e imputación de valores faltantes

Antes de alimentar el modelo, la matriz de características se somete a un procedimiento de unificación e imputación conforme a los siguientes pasos:

1. Se identifican los valores infinitos (`Inf`), producto de eventuales anomalías numéricas durante el cálculo de descriptores, como divisiones por cero, y se reemplazan por `NaN` para tratarlos uniformemente como valores faltantes.

2. Se calcula el vector de medias por columna $\boldsymbol{\mu} \in \mathbb{R}^d$ exclusivamente sobre el conjunto de contexto de la partición, ignorando todos los valores `NaN`.

3. Cada valor `NaN` se reemplaza por la media correspondiente de su columna: $\theta_{ij} \leftarrow \mu_j$.

4. El mismo vector de medias se reutiliza durante la imputación del conjunto de prueba, sin recalcularlo con información de la evaluación.

Esta separación estricta garantiza que ninguna estadística del conjunto de prueba participe en el cálculo de las medias de imputación, preservando la independencia estadística entre el contexto y la evaluación.

Mecanismo de inferencia y atención en contexto

En términos prácticos, el proceso de clasificación no se basa en un ajuste iterativo de los parámetros del modelo, sino en una comparación dinámica que ocurre durante la ejecución del script `Transformer.py`. Al ingresar un vector de características $\boldsymbol{\theta}_i$, el cual contiene los 17 descriptores paramétricos de una señal desconocida, el algoritmo lo sitúa frente a la matriz de contexto. Esta matriz actúa como una memoria de trabajo compuesta por el subconjunto de contexto de la partición.
Internamente, aplicando el mecanismo de atención, TabPFN evalúa la similitud matemática entre la muestra de prueba y cada instancia del contexto. En este flujo, el vector de prueba $\boldsymbol{\theta}_i$ asume el rol de una consulta (*query*) que interroga a las claves (*keys*) de la matriz de contexto. Esta interacción permite que el modelo identifique la clase de la señal mediante la analogía estructural con los sismos conocidos en el contexto.

Generación de probabilidades y criterio de decisión

La salida de este proceso de atención se traduce en un vector de probabilidad estimada $\hat{\mathbf{p}}_i$, cuya estructura matemática corresponde a la distribución de confianza definida en la Ecuación 4.75. Este vector representa la compatibilidad de la muestra de prueba con cada una de las cuatro categorías sismológicas según el conocimiento contenido en el contexto:

$$\hat{\mathbf{p}}_i = \big[P(\text{LP} \mid \boldsymbol{\theta}_i),\ P(\text{TOR} \mid \boldsymbol{\theta}_i),\ P(\text{TRE} \mid \boldsymbol{\theta}_i),\ P(\text{VT} \mid \boldsymbol{\theta}_i)\big] \tag{5.8}$$

Finalmente, la etiqueta definitiva $\hat{y}_i$ se obtiene mediante el operador *argmax*, el cual selecciona la clase correspondiente a la componente con el valor máximo dentro del vector de probabilidades $\hat{\mathbf{p}}_i$:

$$\hat{y}_i = \arg\max(\hat{\mathbf{p}}_i) \tag{5.9}$$

Este flujo asegura que cada clasificación sea una decisión informada por la totalidad de los datos de contexto disponibles en esa partición específica.

Flujo completo por partición

El proceso de evaluación de una partición k se ejecuta mediante el script `Transformer.py`, indicando la partición requerida como argumento, y comprende las siguientes etapas:

1. Lectura del contexto: se carga `features_context.parquet` de la partición k.

2. Imputación: se calculan las medias $\boldsymbol{\mu}$ del contexto y se imputan los `NaN`.

3. Preparación del modelo con el contexto: `model.fit(X_context, y_context)`.

4. Evaluación sobre el conjunto de prueba: para cada archivo `{clase}.parquet` del conjunto de prueba:

   a) Se imputan los `NaN` con las mismas medias $\boldsymbol{\mu}$ del contexto.

   b) Se generan los vectores de probabilidad en lotes de 500.

   c) Se clasifica por argmax directo.

   d) Se registran los aciertos y errores en la matriz de confusión.

5. Cómputo de métricas: Accuracy, Precision, Recall, F1 por clase, F1 Macro y F1 Ponderado.

6. Almacenamiento: se exportan la matriz de confusión como imagen (`.png`), el resumen de métricas en formato tabular (`.csv`) y los vectores de probabilidad (embeddings) en formato (`.txt`).

5.5. Diseño experimental y estrategias de evaluación

Establecida la arquitectura algorítmica del clasificador, esta sección detalla la configuración empírica de las pruebas computacionales. Dado que el principal desafío en el análisis del volcán Galeras radica en el desbalance extremo de los eventos sismológicos, se estructuró una metodología de evaluación progresiva basada en cuatro escenarios experimentales: (1) clasificación bajo el desbalance nativo del catálogo, (2) reducción del espacio predictivo a las tres clases dominantes, (3) balanceo mediante técnicas de aumento de datos, y (4) reajuste analítico mediante ponderación probabilística. Esta estructura permite aislar el impacto de cada estrategia y medir de forma rigurosa el rendimiento del modelo frente a distribuciones asimétricas de información.

5.5.1. Experimento 1: línea base

Como se documenta en la sección 5.1, aunque las clases LP, TRE y VT cuentan con un volumen suficiente de datos para mantener un equilibrio estadístico entre sí, el catálogo del OVSP presenta una limitación natural importante: la clase tornillo (TOR) representa apenas el 0,5 % de los registros totales. Este primer experimento se evalúa bajo las condiciones de desbalance nativo del catálogo. En su estado original, la distribución de los datos presenta una asimetría particular: tres de las clases sismológicas se presentan en proporciones similares, mientras que la categoría minoritaria TOR se incorpora conservando estrictamente su baja proporción original.

El objetivo principal de esta configuración es caracterizar el comportamiento y la capacidad predictiva del modelo TabPFN ante la inclusión de una única clase severamente subrepresentada. Al evaluar el clasificador bajo estas condiciones específicas, se establece una línea base de rendimiento que permite aislar el impacto directo que tiene la escasez de registros de la clase TOR en la capacidad de generalización del algoritmo.

Muestreo con clase minoritaria

Para la ejecución de este experimento, la conformación del conjunto de datos se rigió estrictamente por el esquema de partición estándar descrito en el apartado 5.4.1. Dado que la clase TOR cuenta con un número de instancias ($n_{\text{mín}} = 195$) significativamente inferior al cupo teórico ideal, se aplicó la regla de ajuste definida en la Ecuación 5.6. Como resultado, se incorporó la totalidad de los eventos de la clase minoritaria, mientras que las clases mayoritarias fueron submuestreadas a aproximadamente $4\,101$–$4\,102$ instancias cada una, garantizando el límite global de $12\,500$ trazas sin alterar la representatividad física del catálogo original.
La Tabla 5.17 resume los eventos disponibles en el catálogo y los seleccionados tras aplicar el muestreo.

Clase	Eventos disponibles	Muestreo
LP	28 164	4 102
TOR	195	195
TRE	11 512	4 102
VT	5 042	4 101
Total	**44 913**	**12 500**

Tabla 5.17: Distribución de los registros en el catálogo original y subconjunto seleccionado bajo el escenario de desbalance nativo.

Extracción de características y filtrado

Sobre los eventos muestreados se ejecuta el proceso de extracción y filtrado de calidad miniSEED. No todos los eventos originales producen instancias válidas; aquellos cuyas señales no superan los umbrales de calidad técnica se descartan. La Tabla 5.18 muestra, por clase, la relación entre los eventos muestreados inicialmente y las muestras válidas (vectores de características) finalmente extraídas.

Partición contexto-prueba

La partición de los datos y la validación cruzada se ejecutaron siguiendo el protocolo de muestreo aleatorio estratificado detallado en el apartado 5.3.1. Se realizaron cuatro repeticiones con una proporción 80/20 a nivel de evento, empleando las semillas independientes 42, 123, 456 y 789 para asegurar la reproducibilidad del experimento. Gracias a que la estratificación preserva la consistencia de las proporciones de clase, la distribución muestral

Clase	Eventos muestreados	Muestras válidas
LP	4 102	4 038
TOR	195	193
TRE	4 102	3 986
VT	4 101	3 976
Total	**12 500**	**12 193**

Tabla 5.18: Muestras por clase tras superar los filtros de calidad miniSEED.

permanece constante en las cuatro iteraciones, limitando la variabilidad estadística exclusivamente al orden de selección controlado por la semilla pseudoaleatoria. El conteo final de muestras por subconjunto, el cual se mantiene invariable para todas las repeticiones, se encuentra consolidado en la Tabla 5.19.

Clase	Muestras válidas	Contexto (80 %)	Prueba (20 %)
LP	4 038	3 230	808
TOR	193	154	39
TRE	3 986	3 189	797
VT	3 976	3 181	795
Total	**12 193**	**9 754**	**2 439**

Tabla 5.19: Distribución de muestras válidas en los subconjuntos de contexto y prueba.

Tras evaluar el conjunto de prueba, TabPFN genera un vector de probabilidad estimada $\hat{\mathbf{p}}_i$ para cada muestra, cuantificando matemáticamente la confianza asignada a cada clase a partir del contexto disponible (Ec. 5.8). En este experimento, el vector de salida se interpreta como la distribución posterior aproximada sobre las cuatro categorías sismológicas consideradas. La etiqueta final se obtiene aplicando el operador $\arg\max$, que selecciona la clase asociada al mayor valor de probabilidad.

En términos operativos, esto permite asignar la categoría más probable a cada evento del conjunto de prueba sin requerir ajuste iterativo de los parámetros del modelo. La evaluación estadística detallada de los resultados obtenidos en este escenario se presenta en el Capítulo 6.

5.5.2. Experimento 2: entorno balanceado de tres clases

Para aislar el efecto del desbalance severo analizado en el escenario anterior, este segundo experimento plantea un entorno de evaluación restringido exclusivamente a las tres clases mayoritarias (LP, TRE y VT). Al excluir temporalmente la clase tornillo (TOR), es posible aprovechar el amplio volumen de registros de estas tres categorías para configurar un conjunto de datos balanceado que alcanza el límite operativo definido por `MAX_TRZ`.

La ejecución de TabPFN bajo este entorno controlado tiene como propósito establecer una línea base de rendimiento del clasificador frente a los eventos sísmicos más frecuentes, aislando las distorsiones causadas por la escasez de datos de la clase TOR. De este modo, se define

un referente empírico sólido que permitirá, en las fases posteriores del estudio, contrastar y medir el impacto real de reintroducir la clase minoritaria mediante técnicas de aumento de datos y estrategias de ponderación probabilística.

Muestreo balanceado

Para la conformación de este escenario, el procedimiento se rigió estrictamente por la estrategia de partición descrita en el apartado 5.4.1. Dado que las clases LP, TRE y VT superan ampliamente el volumen de registros requerido, ninguna de ellas activa la condición de ajuste para clases minoritarias. En consecuencia, el cupo muestral se determinó aplicando la ecuación base 5.5 para el caso de $C = 3$ categorías:

$$N_{\text{per\_class}} = \left\lfloor \frac{\texttt{MAX\_TRZ}}{3} \right\rfloor = \left\lfloor \frac{12\,500}{3} \right\rfloor = 4\,166. \tag{5.10}$$

La aplicación de este criterio resulta en un submuestreo equitativo donde el residuo de la división se distribuyó de forma unitaria entre dos de las clases para alcanzar con exactitud el límite global de `MAX_TRZ = 12 500`. La Tabla 5.20 consolida el volumen de eventos disponibles en el catálogo original y las instancias finales seleccionadas tras aplicar este procedimiento de balanceo.

Clase	Eventos disponibles	Muestreo
LP	28 164	4 167
TRE	11 512	4 167
VT	5 042	4 166
Total	**44 718**	**12 500**

Tabla 5.20: Eventos disponibles en el catálogo y eventos seleccionados tras el muestreo balanceado con tres clases.

Extracción de características y filtrado

Sobre los eventos muestreados se ejecuta el proceso de extracción y filtrado de calidad miniSEED. La Tabla 5.21 muestra, por clase, la relación entre los eventos muestreados inicialmente y los vectores de características (muestras) finalmente extraídos.

Clase	Eventos muestreados	Muestras válidas
LP	4 167	4 100
TRE	4 167	4 058
VT	4 166	4 043
Total	**12 500**	**12 201**

Tabla 5.21: Muestras extraídas por clase tras superar los filtros de calidad miniSEED.

Partición contexto-prueba

El proceso de división y validación cruzada para este segundo escenario replica rigurosamente la configuración metodológica aplicada en el primer experimento (apartado 5.5.1). Se conserva el esquema de partición estratificada 80/20 a nivel de evento y la evaluación sobre las cuatro particiones de Monte Carlo, una por cada semilla independiente.
Al mantener esta directriz, se asegura que la variabilidad estadística se limite exclusivamente a la asignación pseudoaleatoria, preservando las proporciones equitativas en cada iteración. La Tabla 5.22 detalla la distribución volumétrica resultante para los subconjuntos de contexto y prueba correspondientes a este entorno de tres clases.

Clase	Muestras	Contexto (80 %)	Prueba (20 %)
LP	4 100	3 280	820
TRE	4 058	3 246	812
VT	4 043	3 234	809
Total	**12 201**	**9 760**	**2 441**

Tabla 5.22: Distribución de muestras por clase y partición para el subconjunto de tres clases (LP, TRE y VT).

Inferencia probabilística y cierre experimental

De manera análoga al procedimiento estructurado para el primer escenario, el modelo evalúa las instancias frente a su respectiva matriz de contexto. No obstante, al suprimir la clase minoritaria, el vector de probabilidad estimada $\hat{\mathbf{p}}_i$ reduce su dimensionalidad a tres componentes ($C = 3$):

$$\hat{\mathbf{p}}_i = \left[P(\text{LP} \mid \boldsymbol{\theta}_i),\ P(\text{TRE} \mid \boldsymbol{\theta}_i),\ P(\text{VT} \mid \boldsymbol{\theta}_i)\right] \tag{5.11}$$

El procedimiento de análisis y selección para estos vectores sigue el mismo criterio descrito en el primer experimento. En este sentido, el operador matemático $\arg\max$ continúa ejerciendo la función de decisión final, asignando a cada muestra la etiqueta correspondiente a la clase con mayor masa probabilística.
Con la consolidación de estas predicciones concluye el flujo de procesamiento para este segundo escenario experimental. Los resultados derivados de estas asignaciones, así como el impacto predictivo de operar en un entorno estadísticamente balanceado, se presentarán en detalle en el Capítulo 6.

5.5.3. Procedimiento para el aumento de datos

El aumento de datos (*data augmentation*) en series temporales constituye una estrategia para incrementar la diversidad del conjunto de muestras sin necesidad de adquirir nuevas observaciones [46]. En este proyecto, el procedimiento metodológico tuvo como objetivo generar

señales sintéticas tipo tornillo (TOR) a partir de registros reales para ampliar su representación, conservando la coherencia temporal, espectral y estructural de las trazas originales.
A diferencia del ruido artificial puramente estocástico, que podría alterar de forma no deseada descriptores espectrales relevantes, esta propuesta utilizó ruido real extraído de preámbulos de otras señales sísmicas registradas por el mismo instrumento. Para ello, se implementó una estrategia de superposición aditiva, controlando explícitamente la intensidad de la perturbación mediante el factor de mezcla α y una restricción energética basada en la raíz cuadrática media (*RMS*).

Selección y lectura de señales base

La primera etapa consistió en localizar las señales tipo tornillo (TOR) disponibles en la estructura de directorios. Se identificaron los archivos compatibles con la estación `ANGP`, la red `CM` y el canal `ELZ`, en formato miniSEED. Cada señal base se representó como un vector unidimensional discreto $x_{\text{base}}[n]$ de longitud N_s, a partir del cual se conservó la información de cabecera original (estación, red, canal, ubicación, tasa de muestreo y tiempo de inicio).

Normalización por amplitud máxima

Con el fin de hacer comparable la mezcla entre la señal base y el ruido, cada traza debió ser reescalada dimensionalmente. El algoritmo calculó la amplitud de referencia de la señal base, definida como el máximo valor absoluto de la traza:

$$A_{\text{ref}} = \max_{0 \le n < N_s} |x_{\text{base}}[n]| \,, \tag{5.12}$$

con la condición $A_{\text{ref}} > 0$. A partir de esta magnitud, la señal base se normalizó para obtener una traza adimensional con rango unitario:

$$\tilde{x}_{\text{base}}[n] = \frac{x_{\text{base}}[n]}{A_{\text{ref}}}. \tag{5.13}$$

Este procedimiento permitió proyectar la señal al intervalo aproximado $[-1, 1]$, conservando intacta la morfología de la onda original y manteniendo, en la memoria de procesamiento, el factor de escala A_{ref} necesario para el reescalado posterior de la muestra final.

Procesamiento del ruido instrumental

Debido a que un solo registro sísmico rara vez contiene un segmento de ruido de fondo lo suficientemente extenso para igualar la longitud N_s del evento base, el algoritmo implementó una fase de extracción selectiva de preámbulos. El ruido se extrajo de señales pertenecientes a las clases LP, TRE y VT de la misma estación. Para garantizar que estos segmentos representaran fielmente el ruido de fondo, se aplicó el flujo de detección y validación ilustrado en la Figura 5.5.

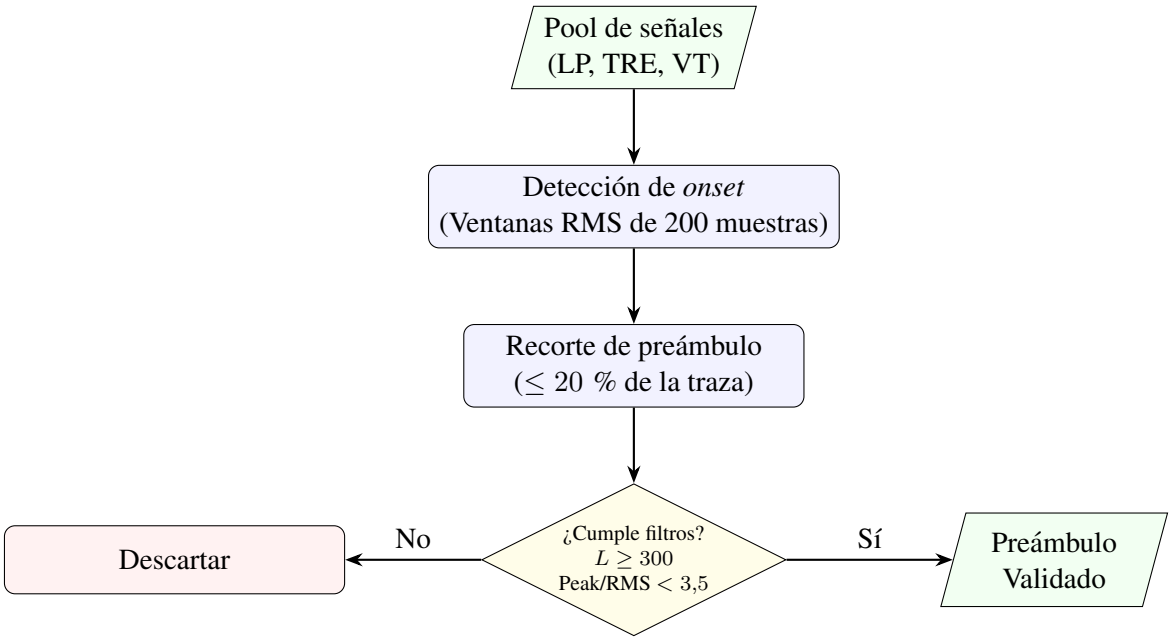


Figura 5.5: Flujo de extracción y validación de preámbulos de ruido instrumental.

El algoritmo estimó el instante de arribo (*onset*) mediante el análisis de energía en ventanas deslizantes. La calidad de este preámbulo se aseguró mediante filtros de estacionariedad (relación pico/RMS $< 3{,}5$ y longitud ≥ 300 muestras), reduciendo la probabilidad de incluir transitorios del evento principal.
Una vez validados, los preámbulos se articularon secuencialmente. Para mitigar variaciones estacionales, se seleccionaron preámbulos registrados en una ventana máxima de 7 días respecto al sismo base (ver Figura 5.6).
Tras alcanzar la longitud requerida N_s, el vector de ruido se normalizó de forma análoga a la señal base, dividiéndolo por su propia amplitud máxima para garantizar que ambas señales fueran comparables antes de la mezcla:

$$A_{\text{ruido}} = \max_{0 \leq n < N_s} |x_{\text{ruido}}[n]|\,, \qquad \tilde{x}_{\text{ruido}}[n] = \frac{x_{\text{ruido}}[n]}{A_{\text{ruido}}}. \tag{5.14}$$

Control de intensidad y mezcla final

La etapa final consistió en la integración matemática de los componentes. Para evitar que la adición de ruido ocultara la resonancia característica del evento base, el factor de escala α se restringió mediante la razón de energía relativa $\rho(\alpha)$, la cual cuantifica el porcentaje de la potencia final que representa la perturbación frente a la señal original utilizando la raíz cuadrática media (***RMS***):

$$\rho(\alpha) = \frac{\text{RMS}(\alpha \tilde{x}_{\text{ruido}})}{\text{RMS}(\tilde{x}_{\text{base}})} = \alpha \frac{\text{RMS}(\tilde{x}_{\text{ruido}})}{\text{RMS}(\tilde{x}_{\text{base}})}. \tag{5.15}$$

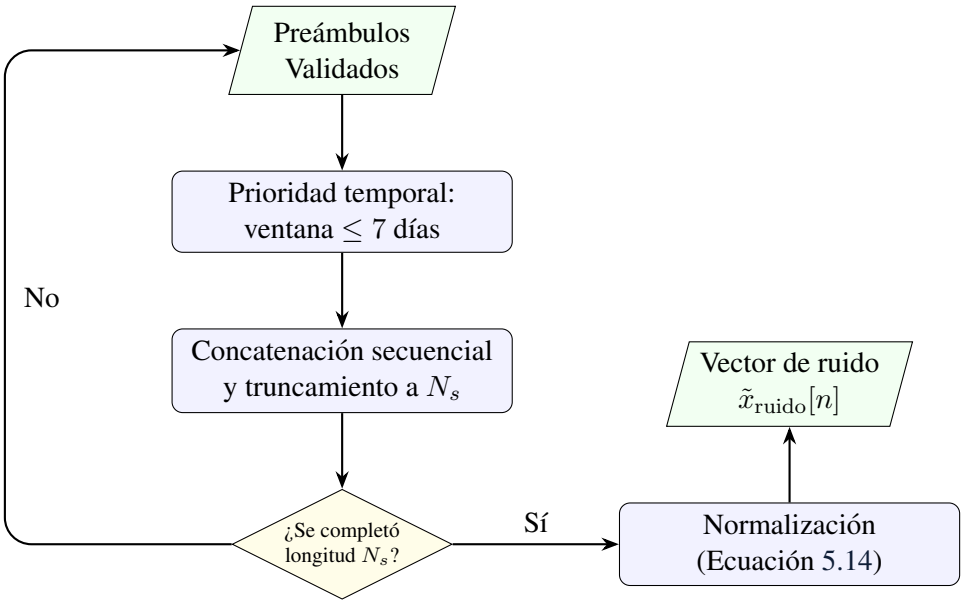


Figura 5.6: Flujo de ensamblaje, priorización y normalización del vector de ruido final.

Al imponer un rango de contaminación aceptable entre un límite mínimo y máximo ($\rho_{\text{mín}}$ y $\rho_{\text{máx}}$), se obtienen los límites operativos para calcular dinámicamente el factor de escala α:

$$\alpha \in \left[\rho_{\text{mín}} \frac{\text{RMS}(\tilde{x}_{\text{base}})}{\text{RMS}(\tilde{x}_{\text{ruido}})},\ \rho_{\text{máx}} \frac{\text{RMS}(\tilde{x}_{\text{base}})}{\text{RMS}(\tilde{x}_{\text{ruido}})} \right]. \tag{5.16}$$

Para el presente diseño experimental, se definieron $\rho_{\text{mín}} = 0{,}05$ y $\rho_{\text{máx}} = 0{,}80$. El límite inferior garantiza variabilidad estadística, mientras que el superior actúa como un techo de seguridad físico. Para capturar diversos escenarios, el valor definitivo de α se seleccionó aleatoriamente dentro de tres bandas de intensidad: baja $[0{,}35, 0{,}50)$, media $[0{,}50, 0{,}65)$ y alta $[0{,}65, 0{,}80)$.
Una vez validado el factor α, la señal aumentada en el dominio normalizado se obtuvo mediante la superposición lineal:

$$\tilde{x}_{\text{aug}}[n] = \tilde{x}_{\text{base}}[n] + \alpha\, \tilde{x}_{\text{ruido}}[n]. \tag{5.17}$$

Finalmente, para recuperar la dimensión física del evento, la señal sintética en el dominio normalizado ($\tilde{x}_{\text{aug}}[n]$) se reescaló utilizando la amplitud de referencia original A_{ref} calculada en la Ecuación 5.12:

$$x_{\text{aug}}[n] = A_{\text{ref}}\, \tilde{x}_{\text{aug}}[n]. \tag{5.18}$$

Al sustituir las definiciones previas en esta operación, se obtiene la expresión expandida definitiva del evento aumentado:

$$x_{\text{aug}}[n] = x_{\text{base}}[n] + \alpha\, A_{\text{ref}}\, \tilde{x}_{\text{ruido}}[n]. \tag{5.19}$$

Esta formulación asegura que la morfología fundamental se mantenga anclada a la dinámica del sismo base.
Las 2 985 muestras sintéticas resultantes se exportaron en formato miniSEED. Este volumen fue deliberadamente superior al cupo objetivo de la clase TOR en el catálogo balanceado, anticipando que una fracción de las señales generadas no superaría los filtros técnicos de calidad miniSEED. De este modo, con los 2 985 datos sinteticos y 195 datos originales se garantizó que las muestras válidas tras el filtrado fueran suficientes para equiparar la clase minoritaria con las clases mayoritarias.

Almacenamiento, trazabilidad y control de calidad

Las señales aumentadas se almacenaron conservando íntegramente los metadatos de la estación ANGP. Para garantizar una trazabilidad rigurosa, se implementó una jerarquía de directorios donde el nombre de cada archivo incluye un sufijo que codifica el valor del escalar α empleado (ver Figura 5.7).

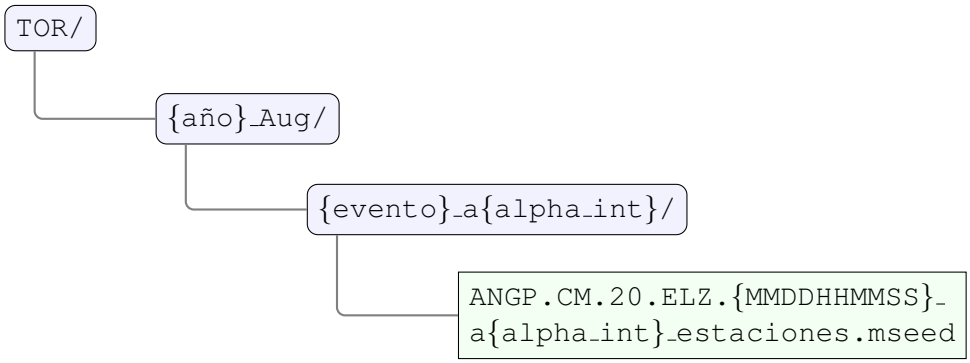


Figura 5.7: Jerarquía de almacenamiento y nomenclatura de los archivos sintéticos.

Como validación cualitativa final, se evaluó la preservación de la morfología característica mediante inspección visual. La Figura 5.8 ilustra un evento base tipo tornillo (TOR) crudo frente a su variante aumentada con nivel de perturbación bajo ($\alpha = 0{,}358$). Las comparaciones para los escalares medio y alto se documentan en el Apéndice C.

5.5.4. Experimento 3: entorno con datos sinteticos

Una vez corroborada la integridad física de las señales sintéticas, este tercer escenario evalúa el desempeño de TabPFN al incorporar dicho volumen de datos para compensar la clase minoritaria TOR. El objetivo central es determinar si el incremento en el soporte estadístico logra redefinir las fronteras de decisión del clasificador a favor de los eventos premonitorios, mejorando efectivamente la capacidad predictiva respecto a la línea base establecida en el experimento 1. La tabla 5.23 resume los eventos disponibles del catalogo original, incluyendo los datos sinteticos para la clase TOR, en el cual fueron válidos 3 127/3 180.

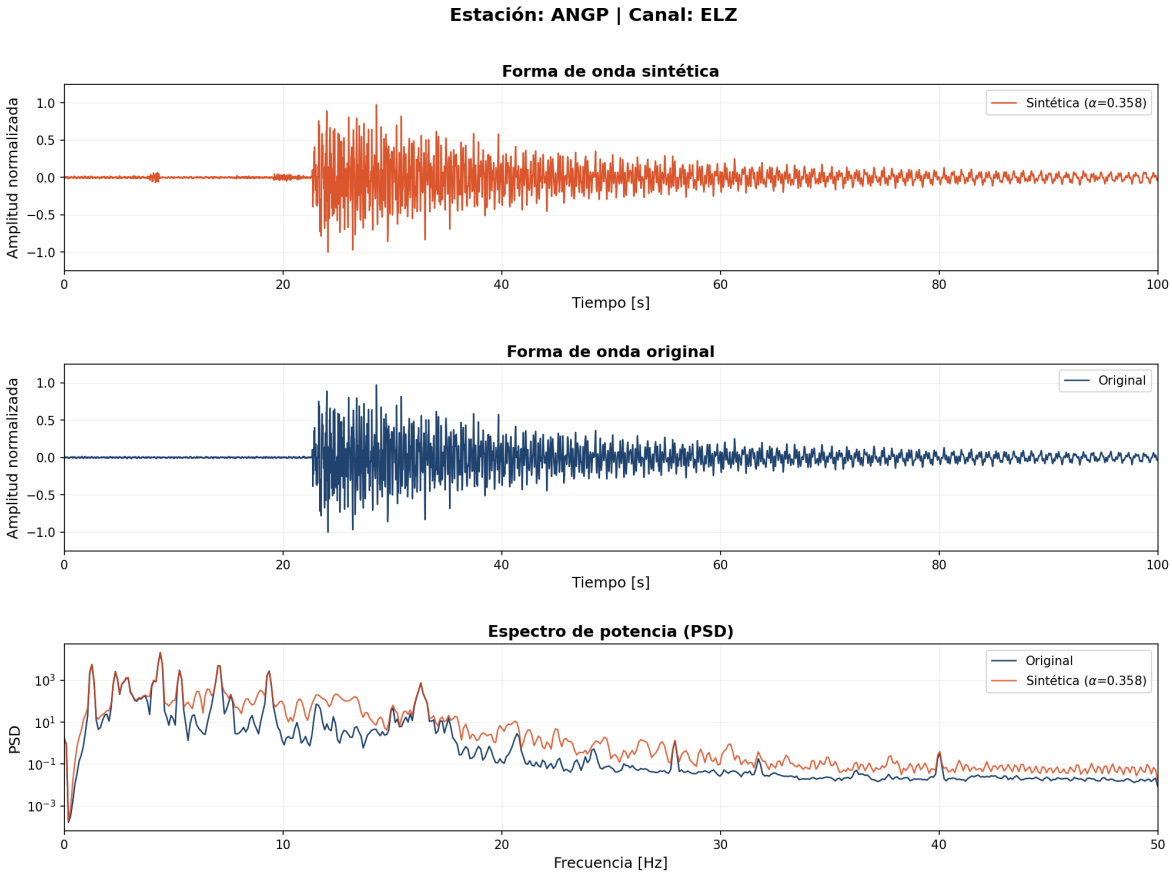


Figura 5.8: Comparación temporal y espectral de un evento tipo tornillo (TOR) registrado en la estación ANGP (componente ELZ) frente a su variante aumentada con nivel de perturbación bajo ($\alpha = 0{,}358$). Los paneles superior y central muestran la amplitud normalizada en el dominio del tiempo para la señal sintética y la original, respectivamente. El panel inferior presenta la densidad espectral de potencia (PSD) comparativa de ambas trazas. Evento registrado el 09 de enero de 2007.

Clase	Eventos disponibles
LP	28 164
TOR	3 127
TRE	11 512
VT	5 042
Total	**47 845**

Tabla 5.23: Distribución de los registros en el catálogo con datos sinteticos.

Partición agrupada y distribución del conjunto aumentado

Al contar con múltiples versiones sintéticas derivadas de un mismo evento original, una partición estándar a nivel de muestras individuales introduciría una fuga de información entre los subconjuntos de contexto y prueba. Como se establece formalmente en el marco teórico (sección 4.4.1, ecuación 4.83), la restricción de no fuga exige que el evento original y todas sus derivaciones sintéticas se asignen íntegramente a un único subconjunto en cada iteración. Para cumplir esta restricción, se implementa una estrategia de partición agrupada tomando como unidad base el evento sísmico original. A cada registro se le extrae el sufijo de aumento para asignarle una clave identificadora común:

$$\texttt{event\_group} = \texttt{base\_event\_id}(\texttt{event\_id}) \tag{5.20}$$

Con esta definición, la señal original y todas sus copias sintéticas comparten el mismo identificador de grupo. La partición estratificada se ejecuta estrictamente sobre estos grupos, garantizando que las variantes generadas nunca se separen de su señal de origen, como se ilustra en la Figura 5.9.

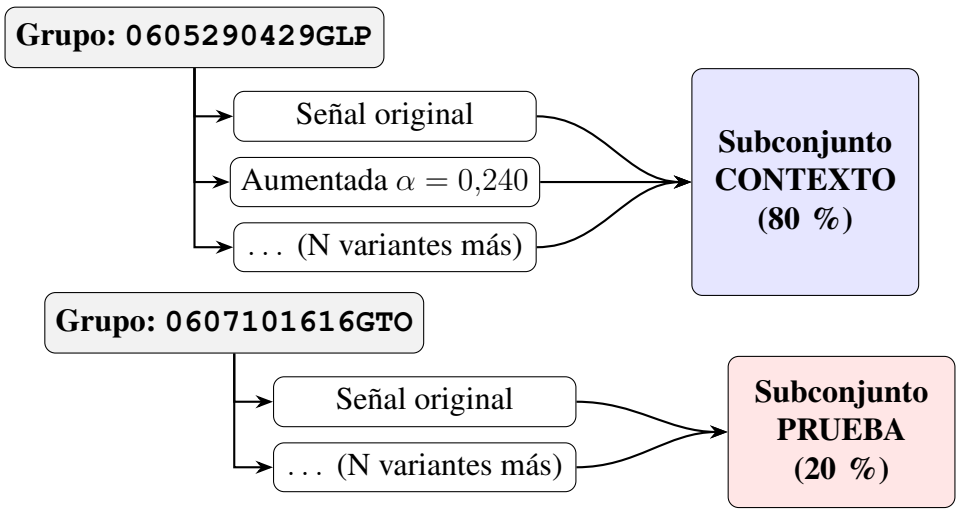


Figura 5.9: Esquema de asignación agrupada por sismo base: el evento original y sus variantes sintéticas se asignan íntegramente al mismo subconjunto, previniendo la fuga de información entre contexto y prueba.

Dado que en esta configuración todas las clases se encuentran balanceadas, el procedimiento de selección sigue la lógica de muestreo descrita en la sección 5.4.1, equivalente a la empleada en el experimento 2. Tras el proceso de extracción de características y filtrado de calidad, la Tabla 5.24 resume la conformación del conjunto de datos final. El catálogo consolidado contiene 12 259 muestras provenientes de 9 329 grupos únicos. En particular, la clase TOR presenta 3 123 muestras asociadas a únicamente 193 grupos reales, lo que evidencia el volumen de copias sintéticas generadas.

La proporción media entre instancias y grupos para TOR se define como:

$$\frac{3\,123}{193} \approx 16{,}18 \tag{5.21}$$

Este valor indica que, en promedio, cada evento base de la clase TOR aporta aproximadamente 16 muestras al catálogo, la señal original más 15 variantes aumentadas válidas. Este promedio no es un valor fijo: el número exacto de muestras por grupo varía según cuántas combinaciones sintéticas fueron rechazadas durante la augmentación o descartadas en la fase de extracción.

Esta variabilidad en el tamaño interno de los grupos repercute directamente en el volumen final de cada partición. La división estratificada 80/20 se calcula únicamente sobre el volumen de grupos, utilizando semillas independientes (42, 123, 456 y 789) para cada iteración. Por ejemplo, el 80 % de los 193 grupos base de TOR equivale aproximadamente a 154 grupos para contexto y 39 para prueba. Una vez sorteados los grupos, el conteo de instancias es la suma del contenido de los eventos seleccionados; dado que el tamaño interno fluctúa, se generan variaciones matemáticas entre iteraciones. Las Tablas 5.25 a 5.28 exponen las distribuciones exactas para cada repetición de la validación cruzada.

Clase	Eventos muestreados	Muestras válidas	Grupos únicos
LP	3 125	3 079	3 079
TOR	3 125	3 123	193
TRE	3 125	3 046	3 046
VT	3 125	3 011	3 011
Total	**12 500**	**12 259**	**9 329**

Tabla 5.24: Distribución global del catálogo aumentado: relación entre eventos muestreados, muestras válidas tras la extracción de características y grupos únicos asignados.

	Contexto (80 %)		Prueba (20 %)	
Clase	**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**	**7 463**	**2 496**	**1 866**

Tabla 5.25: Distribución de grupos y muestras para la Partición 1 (Semilla 42).

	Contexto (80 %)		Prueba (20 %)	
Clase	**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**	**7 463**	**2 466**	**1 866**

Tabla 5.26: Distribución de grupos y muestras para la Partición 2 (Semilla 123).

	Contexto (80 %)		Prueba (20 %)	
Clase	**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**

Tabla 5.27: Distribución de grupos y muestras para la Partición 3 (Semilla 456).

	Contexto (80 %)		Prueba (20 %)	
Clase	**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**

Tabla 5.28: Distribución de grupos y muestras para la Partición 4 (Semilla 789).

Con la definición del esquema de partición agrupada y la distribución de los conjuntos de contexto y prueba concluye el diseño metodológico de este tercer escenario. A diferencia de

los experimentos anteriores, la matriz de contexto presentada al clasificador incorpora ahora una representación estadísticamente equiparable de las cuatro categorías sismológicas, con TOR respaldada por señales sintéticas que preservan su morfología física pero amplían su diversidad paramétrica. Esta condición permite evaluar si TabPFN, operando bajo un contexto balanceado, es capaz de establecer fronteras de decisión más robustas para la clase premonitoria sin degradar el rendimiento sobre las clases mayoritarias. El desempeño predictivo del modelo frente a este escenario particular se detallará en el Capítulo 6.

5.5.5. Experimento 4: clasificación por *argmax* ponderado

Este cuarto escenario experimental se propone como una alternativa algorítmica al aumento de datos. Su objetivo es equilibrar la sensibilidad del clasificador TabPFN frente a la clase minoritaria sin recurrir a la generación de muestras sintéticas. Para ello, las etapas de muestreo y extracción de características son idénticas a las del Experimento 1 (Tabla 5.18), así como la distribución de muestras en los subconjuntos de contexto y prueba (Tabla 5.19).

La variación metodológica radica exclusivamente en la etapa de posprocesamiento probabilístico. En lugar de aplicar la regla de decisión estándar, se modifica el vector de probabilidades de salida $\hat{\mathbf{p}}_i$ multiplicando cada componente por un factor de clase w_c antes de seleccionar la etiqueta final: la clase asignada es aquella cuya probabilidad ponderada resulta máxima. Esta intervención no altera el modelo ni su contexto; opera únicamente sobre la decisión, amplificando la puntuación de TOR para compensar su subrepresentación. Los fundamentos teóricos, la formulación matemática del factor w_c y su hiperparámetro de suavizado γ se detallan en la sección 4.5.

Para evaluar el impacto de esta intervención, se ejecutaron dos configuraciones independientes del hiperparámetro de suavizado: un punto intermedio ($\gamma = 0{,}5$, Experimento A) y la compensación total ($\gamma = 1{,}0$, Experimento B). Cada configuración produce un conjunto de predicciones distinto, permitiendo comparar el efecto progresivo de la ponderación sobre la frontera de decisión. El límite de neutralidad ($\gamma = 0{,}0$) no se ejecutó como escenario independiente, dado que equivale exactamente a la línea base del Experimento 1.

A partir de la distribución de muestras de la Tabla 5.19, los factores de ponderación para cada clase se calculan mediante la expresión introducida en la Ecuación 4.90:

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

donde $N = 9\,754$ es el total de muestras en el contexto, $C = 4$ el número de clases y n_c la cantidad de muestras de la clase c. La Tabla 5.29 presenta los valores resultantes para los niveles de γ evaluados, conservando $\gamma = 0{,}0$ a modo ilustrativo.

Clase	n_c	w_c ($\gamma = 0{,}0$)	w_c ($\gamma = 0{,}5$)	w_c ($\gamma = 1{,}0$)
LP	3 230	1,0000	0,8689	0,7550
TOR	154	1,0000	3,9792	15,8344
TRE	3 189	1,0000	0,8744	0,7647
VT	3 181	1,0000	0,8755	0,7666

Tabla 5.29: Factores de ponderación w_c calculados para el contexto de $N = 9\,754$ muestras con $C = 4$ clases, en función del hiperparámetro γ.

Una vez definidos estos factores, el ajuste se aplica de forma sistemática sobre las probabilidades de salida de cada muestra en el conjunto de prueba, generando dos conjuntos de predicciones independientes para $\gamma = 0{,}5$ y $\gamma = 1{,}0$. El impacto de esta estrategia sobre la frontera de decisión se analiza en el Capítulo 6.

Disponibilidad de código

El código fuente desarrollado en esta investigación para el preprocesamiento de señales, extracción de características y la implementación del flujo de trabajo con TabPFN, así como los *scripts* para la generación de las figuras, se encuentran disponibles de forma pública para garantizar la reproducibilidad de los resultados.
El repositorio puede ser consultado en la plataforma GitHub a través del siguiente enlace: `https://github.com/danielchapid/Clasificacion-Sismica-Galeras`.

Capítulo 6

Resultados

Este capítulo describe los resultados obtenidos en los cuatro escenarios experimentales diseñados para evaluar la capacidad de TabPFN en la categorización de sismos volcánicos del Galeras. La toma de decisiones del algoritmo se fundamenta en la generación de vectores de probabilidad (*embeddings* probabilísticos) para cada instancia evaluada frente a la matriz de contexto.
En los apartados siguientes se expone la efectividad de este mecanismo midiendo el porcentaje de acierto que resulta de aplicar el operador matemático $\arg\max$ sobre dichos vectores. Debido al alto volumen de resultados generados por cada partición de la validación cruzada, el texto principal consolida las métricas predictivas expresando su media aritmética y desviación estándar ($\mu \pm \sigma$). Esta aproximación permite resumir la información y cuantificar la estabilidad estadística del algoritmo, a la par que se detalla el comportamiento de los errores cruzados por iteración.
Las matrices de confusión para cada experimento se encuentran documentadas exhaustivamente en el Apéndice D. Cabe destacar que todos los porcentajes de acierto discutidos a lo largo de este capítulo corresponden a la normalización por filas de dichas matrices. Es decir, los porcentajes presentados se calculan dividiendo el conteo absoluto de una predicción entre el soporte total real de la clase evaluada.

6.1. Experimento 1: evaluación base

Este escenario evalúa el rendimiento del modelo frente al catálogo en su distribución nativa. El objetivo es caracterizar el comportamiento de la arquitectura ante la escasez extrema de la clase tornillo (TOR). La cantidad exacta de muestras asignadas para contexto y prueba en este experimento corresponde a la distribución formalizada previamente en la Tabla 5.19 de la sección 5.5.1 del Capítulo 5.

Porcentaje de acierto por clase (*Recall*)

La Tabla 6.1 refleja, para cada una de las cuatro particiones independientes, el porcentaje de eventos predichos correctamente tras aplicar el operador $\arg\max$ sobre los vectores de

probabilidad de cada archivo de prueba. Esta métrica evidencia la capacidad del modelo para reconocer y recuperar la firma de cada sismo dentro de la distribución nativa del catálogo.

Clase	Partición 1	Partición 2	Partición 3	Partición 4
LP	81,06 %	83,29 %	83,91 %	81,19 %
TOR	92,31 %	87,18 %	94,87 %	79,49 %
TRE	96,11 %	95,23 %	94,60 %	96,74 %
VT	90,82 %	91,07 %	90,82 %	91,45 %

Tabla 6.1: Porcentaje de acierto (*Recall*) en la predicción final desglosado por partición para cada clase evaluada.

Rendimiento predictivo por clase

Para consolidar el análisis del rendimiento, la Tabla 6.2 detalla las métricas predictivas para cada categoría sismológica. Con el objetivo de cuantificar la estabilidad estadística del algoritmo frente a las distintas variaciones del conjunto de datos, los resultados se reportan expresando la media aritmética y la desviación estándar ($\mu \pm \sigma$) calculadas a partir de las cuatro particiones independientes.

Clase	Precisión ($\mu \pm \sigma$)	Recall ($\mu \pm \sigma$)	F1-Score ($\mu \pm \sigma$)
LP	$86{,}9\,\% \pm 0{,}5\,\%$	$82{,}4\,\% \pm 1{,}4\,\%$	$84{,}5\,\% \pm 0{,}6\,\%$
TOR	$92{,}8\,\% \pm 5{,}5\,\%$	$88{,}5\,\% \pm 6{,}8\,\%$	$90{,}5\,\% \pm 5{,}1\,\%$
TRE	$96{,}1\,\% \pm 0{,}4\,\%$	$95{,}7\,\% \pm 0{,}9\,\%$	$95{,}9\,\% \pm 0{,}3\,\%$
VT	$86{,}0\,\% \pm 0{,}7\,\%$	$91{,}0\,\% \pm 0{,}3\,\%$	$88{,}4\,\% \pm 0{,}3\,\%$

Tabla 6.2: Métricas de evaluación por clase en el Experimento 1, presentadas como media $\pm$ desviación estándar sobre las cuatro particiones.

Métricas globales de evaluación

Para complementar el análisis por categoría, la Tabla 6.3 consolida el rendimiento general del sistema. La inclusión simultánea de estas métricas permite contrastar la exactitud global (*Accuracy*) frente al F1-Score Macro, garantizando una evaluación transparente frente al desbalance de las clases del Galeras.

En síntesis, el modelo operando sobre el catálogo en su distribución nativa demuestra un desempeño predictivo global estable. Estadísticamente, la clase TRE registra las métricas más altas y consistentes, mientras que LP presenta los valores inferiores de F1-Score dentro del grupo mayoritario. Por su parte, la clase minoritaria TOR logra ser identificada con alta sensibilidad, aunque su puntaje de *Recall* exhibe una marcada varianza a través de las distintas particiones evaluadas.

Métrica Global	**P1**	**P2**	**P3**	**P4**	**Promedio ($\mu \pm \sigma$)**
Accuracy global	89,34 %	89,79 %	89,83 %	89,59 %	89,64 % ± 0,22 %
F1 Macro	89,74 %	89,18 %	91,65 %	88,73 %	89,83 % ± 1,28 %
F1 Ponderado	89,29 %	89,78 %	89,84 %	89,52 %	89,61 % ± 0,26 %

Tabla 6.3: Consolidado de métricas de desempeño global por partición (P1 a P4) y su respectivo promedio estadístico.

6.2. Experimento 2: evaluación con tres clases en entorno balanceado

Este escenario evalúa el rendimiento del modelo tras excluir la clase minoritaria TOR y operar sobre un conjunto de contexto balanceado. El objetivo es caracterizar el comportamiento predictivo de la arquitectura en un entorno equitativo de tres clases (LP, TRE y VT), eliminando el sesgo estadístico inducido por la escasez de datos. La cantidad exacta de muestras asignadas para contexto y prueba en este experimento corresponde a la distribución formalizada previamente en la Tabla 5.22 de la sección 5.5.2 del Capítulo 5.

Porcentaje de acierto por clase (*Recall*)

La Tabla 6.4 refleja, para cada una de las cuatro particiones independientes, el porcentaje de eventos predichos correctamente tras aplicar el operador $\arg\max$ sobre los vectores de probabilidad de cada archivo de prueba. Esta métrica evidencia la capacidad del modelo para reconocer y recuperar la firma de cada sismo en ausencia de desbalance.

Clase	**Partición 1**	**Partición 2**	**Partición 3**	**Partición 4**
LP	81,59 %	81,95 %	82,20 %	83,78 %
TRE	95,07 %	94,83 %	94,09 %	95,57 %
VT	91,59 %	91,59 %	88,63 %	89,62 %

Tabla 6.4: Porcentaje de acierto en la predicción final desglosado por partición para cada clase en el entorno balanceado.

Rendimiento predictivo por clase

Para consolidar el análisis del rendimiento, la Tabla 6.5 detalla las métricas predictivas para cada categoría sismológica. Con el objetivo de cuantificar la estabilidad estadística del algoritmo frente a las distintas variaciones del conjunto de datos, los resultados se reportan expresando la media aritmética y la desviación estándar ($\mu \pm \sigma$) calculadas a partir de las cuatro particiones independientes.

Clase	**Precisión** ($\mu \pm \sigma$)	**Recall** ($\mu \pm \sigma$)	**F1-Score** ($\mu \pm \sigma$)
LP	$86{,}1\,\% \pm 1{,}4\,\%$	$82{,}4\,\% \pm 1{,}0\,\%$	$84{,}2\,\% \pm 0{,}7\,\%$
TRE	$95{,}6\,\% \pm 0{,}4\,\%$	$94{,}9\,\% \pm 0{,}6\,\%$	$95{,}3\,\% \pm 0{,}2\,\%$
VT	$86{,}1\,\% \pm 1{,}2\,\%$	$90{,}4\,\% \pm 1{,}5\,\%$	$88{,}1\,\% \pm 0{,}9\,\%$

Tabla 6.5: Métricas de evaluación por clase en el Experimento 2, presentadas como media $\pm$ desviación estándar sobre las cuatro particiones.

Métricas globales de evaluación

Para complementar el análisis por categoría, la Tabla 6.6 consolida el rendimiento general del sistema. La inclusión simultánea de estas métricas permite contrastar la exactitud global (*Accuracy*) frente al F1-Score Macro, garantizando una evaluación transparente de la capacidad del clasificador.

Métrica Global	**P1**	**P2**	**P3**	**P4**	**Promedio** ($\mu \pm \sigma$)
Accuracy global	$89{,}39\,\%$	$89{,}43\,\%$	$88{,}28\,\%$	$89{,}64\,\%$	$89{,}19\,\% \pm 0{,}62\,\%$
F1 Macro	$89{,}37\,\%$	$89{,}41\,\%$	$88{,}31\,\%$	$89{,}63\,\%$	$89{,}18\,\% \pm 0{,}60\,\%$
F1 Ponderado	$89{,}36\,\%$	$89{,}39\,\%$	$88{,}30\,\%$	$89{,}62\,\%$	$89{,}17\,\% \pm 0{,}59\,\%$

Tabla 6.6: Consolidado de métricas de desempeño global por partición (P1 a P4) y su respectivo promedio estadístico para el Experimento 2.

La evaluación sobre un entorno equitativo de tres clases arroja un comportamiento predictivo altamente consistente. Las métricas globales convergen sin dispersiones significativas entre particiones, consolidando a TRE como la categoría con mayor facilidad de clasificación y a LP como la de menor rendimiento relativo dentro de este subconjunto. Este escenario confirma empíricamente la estabilidad de la arquitectura en ausencia de desbalance. Precisamente, al demostrar que representaciones de contexto equilibradas logran mantener un balance casi constante en las métricas de evaluación, estos resultados sientan la base metodológica y dan paso a los Experimentos 3 y 4, cuyo propósito será intentar replicar esta estabilidad para la clase minoritaria mediante técnicas de compensación sintética y probabilística.

6.3. Experimento 3: evaluación con aumento de datos

Este tercer escenario evalúa el rendimiento del modelo tras aplicar técnicas de aumento de datos (inyección de ruido instrumental) sobre la clase minoritaria tornillo. El objetivo es caracterizar la capacidad predictiva de la arquitectura en un entorno donde las cuatro clases (LP, TRE, VT y TOR) se encuentran balanceadas artificialmente, induciendo al modelo a extraer patrones de firmas espectrales sintéticamente ruidosas. La cantidad de muestras asignadas corresponde a la distribución formalizada en el apartado 5.5.4 del Capítulo 5.

Porcentaje de acierto por clase (*Recall*)

La Tabla 6.7 refleja, para cada una de las cuatro particiones independientes, el porcentaje de eventos predichos correctamente tras aplicar el operador $\arg\max$ sobre los vectores de probabilidad. Esta métrica cuantifica la capacidad del clasificador para recuperar la firma de cada sismo tras la alteración del espacio de características original.

Clase	Partición 1	Partición 2	Partición 3	Partición 4
LP	82,14 %	80,19 %	79,71 %	82,47 %
TOR	82,96 %	77,00 %	90,88 %	88,84 %
TRE	95,89 %	95,24 %	95,40 %	96,22 %
VT	88,21 %	89,70 %	87,87 %	88,21 %

Tabla 6.7: Porcentaje de acierto en la predicción final desglosado por partición para el escenario con aumento de datos.

Rendimiento predictivo por clase

Para consolidar el análisis, la Tabla 6.8 detalla las métricas predictivas para cada categoría sismológica. Los resultados se reportan expresando la media aritmética y la desviación estándar ($\mu \pm \sigma$) obtenidas a lo largo de las cuatro particiones, permitiendo evaluar el impacto del ruido sintético en la estabilidad del algoritmo.

Clase	Precisión ($\mu \pm \sigma$)	Recall ($\mu \pm \sigma$)	F1-Score ($\mu \pm \sigma$)
LP	$77{,}9\,\% \pm 3{,}9\,\%$	$81{,}1\,\% \pm 1{,}4\,\%$	$79{,}5\,\% \pm 2{,}4\,\%$
TOR	$98{,}1\,\% \pm 0{,}8\,\%$	$84{,}9\,\% \pm 6{,}3\,\%$	$90{,}9\,\% \pm 3{,}3\,\%$
TRE	$93{,}7\,\% \pm 1{,}8\,\%$	$95{,}7\,\% \pm 0{,}5\,\%$	$94{,}7\,\% \pm 1{,}1\,\%$
VT	$82{,}2\,\% \pm 2{,}3\,\%$	$88{,}5\,\% \pm 0{,}8\,\%$	$85{,}3\,\% \pm 1{,}5\,\%$

Tabla 6.8: Métricas de evaluación por clase en el Experimento 3, presentadas como media $\pm$ desviación estándar sobre las cuatro particiones.

Métricas globales de evaluación

La Tabla 6.9 consolida el rendimiento general del sistema bajo el paradigma de datos aumentados.

La evaluación sobre el conjunto con aumento de datos refleja una mayor dispersión estadística general. Si bien la clase TRE sostiene un desempeño invariante, las clases LP y VT experimentan un descenso en sus puntuaciones de Precisión y F1-Score. Para el caso particular de la clase TOR, la inyección de ruido instrumental fija una Precisión marcadamente elevada, pero induce una variabilidad considerable en su *Recall* a través de las particiones.

Métrica Global	**P1**	**P2**	**P3**	**P4**	**Promedio ($\mu \pm \sigma$)**
Accuracy global	87,18 %	85,40 %	88,46 %	88,92 %	87,49 % ± 1,57 %
F1 Macro	87,24 %	85,53 %	88,49 %	88,96 %	87,56 % ± 1,53 %
F1 Ponderado	87,31 %	85,51 %	88,51 %	89,01 %	87,59 % ± 1,55 %

Tabla 6.9: Consolidado de métricas de desempeño global por partición (P1 a P4) y su respectivo promedio estadístico tras el aumento de datos.

6.4. Experimento 4: evaluación por *argmax* ponderado

Este cuarto escenario aborda el problema del desbalance de datos sin alterar la matriz de contexto, interviniendo exclusivamente en la etapa de posprocesamiento probabilístico. A continuación, se presentan los resultados obtenidos al aplicar dos niveles de compensación algorítmica sobre el operador de decisión final.

6.4.1. Experimento A: evaluación con $\gamma = 0{,}5$ en la regla de decisión

Este escenario evalúa el rendimiento del modelo frente al catálogo empleando un factor de balance $\gamma = 0{,}5$. El objetivo es caracterizar el comportamiento de la arquitectura y el impacto sobre la regla de decisión (*argmax*) al inyectar una compensación moderada.

Recall por clase y por partición

La Tabla 6.10 refleja, para cada una de las cuatro particiones independientes, el porcentaje de eventos predichos correctamente tras aplicar el operador $\arg\max$ ponderado (Ec. 4.92) sobre los vectores de probabilidad crudos de cada archivo de prueba. Esta métrica evidencia la capacidad del modelo para reconocer y recuperar la firma de cada sismo bajo la influencia del factor de suavizado.

Clase	**Partición 1**	**Partición 2**	**Partición 3**	**Partición 4**
LP	80,32 %	82,92 %	83,66 %	81,19 %
TOR	94,87 %	92,31 %	94,87 %	89,74 %
TRE	95,98 %	95,11 %	94,60 %	96,49 %
VT	90,94 %	91,32 %	90,44 %	91,32 %

Tabla 6.10: Recall por clase y por partición en la predicción final bajo $\gamma = 0{,}5$.

Métricas por clase agregadas entre particiones

Para consolidar el análisis del rendimiento, la Tabla 6.11 detalla las métricas predictivas para cada categoría sismológica. Con el objetivo de cuantificar la estabilidad estadística del algoritmo frente a la alteración de la frontera de decisión, los resultados se reportan expresando la

media aritmética y la desviación estándar ($\mu \pm \sigma$) calculadas a partir de las cuatro particiones independientes.

Clase	**Precisión** ($\mu \pm \sigma$)	**Recall** ($\mu \pm \sigma$)	**F1-Score** ($\mu \pm \sigma$)
LP	$87{,}3\,\% \pm 0{,}4\,\%$	$82{,}0\,\% \pm 1{,}5\,\%$	$84{,}6\,\% \pm 0{,}8\,\%$
TOR	$79{,}4\,\% \pm 3{,}1\,\%$	$93{,}0\,\% \pm 2{,}5\,\%$	$85{,}6\,\% \pm 1{,}5\,\%$
TRE	$96{,}2\,\% \pm 0{,}4\,\%$	$95{,}6\,\% \pm 0{,}9\,\%$	$95{,}9\,\% \pm 0{,}3\,\%$
VT	$86{,}0\,\% \pm 0{,}7\,\%$	$91{,}0\,\% \pm 0{,}4\,\%$	$88{,}4\,\% \pm 0{,}5\,\%$

Tabla 6.11: Precisión, recall y F1-score por clase, reportados como media $\pm$ desviación estándar sobre las cuatro particiones, bajo $\gamma = 0{,}5$.

Métricas globales de evaluación

Para complementar el análisis por categoría, la Tabla 6.12 consolida el rendimiento general del sistema. La inclusión simultánea de estas métricas permite observar el efecto directo del parámetro γ sobre la exactitud global (*Accuracy*), el F1-Score Macro y ponderado.

Métrica Global	**P1**	**P2**	**P3**	**P4**	**Promedio** ($\mu \pm \sigma$)
Accuracy global	$89{,}13\,\%$	$89{,}79\,\%$	$89{,}63\,\%$	$89{,}63\,\%$	$89{,}55\,\% \pm 0{,}28\,\%$
F1 Macro	$88{,}14\,\%$	$88{,}36\,\%$	$89{,}04\,\%$	$88{,}84\,\%$	$88{,}60\,\% \pm 0{,}42\,\%$
F1 Ponderado	$89{,}08\,\%$	$89{,}78\,\%$	$89{,}64\,\%$	$89{,}57\,\%$	$89{,}52\,\% \pm 0{,}30\,\%$

Tabla 6.12: Consolidado de métricas de desempeño global por partición (P1 a P4) y su respectivo promedio estadístico bajo $\gamma = 0{,}5$.

La aplicación de una ponderación moderada ($\gamma = 0{,}5$) sobre la regla de decisión conserva la estabilidad del desempeño global del sistema. A nivel de categorías, las métricas de las clases dominantes (LP, TRE y VT) permanecen estadísticamente estables respecto a su estado natural, mientras que la clase TOR refleja un intercambio empírico evidente: un incremento directo en su *Recall* a expensas de una disminución en su precisión.

6.4.2. Experimento B: evaluación con $\gamma = 1{,}0$ en la regla de decisión

Este escenario evalúa el rendimiento del modelo aplicando un factor de balance total ($\gamma = 1{,}0$). El objetivo es analizar el comportamiento del operador $\arg\max$ ponderado cuando se compensa matemáticamente el desbalance en su máxima expresión, inyectando la distribución natural completa en el cálculo probabilístico. Se busca observar los límites metodológicos de la ponderación y su efecto sobre la estabilidad predictiva de las clases.

Recall por clase y por partición

La Tabla 6.13 refleja, para cada una de las cuatro particiones independientes, el porcentaje de eventos predichos correctamente tras aplicar el operador de salida bajo compensación

total. Esta métrica evidencia el impacto radical que tiene la máxima amplificación sobre la capacidad de retención de las clases minoritarias frente a las dominantes.

Clase	Partición 1	Partición 2	Partición 3	Partición 4
LP	79,70 %	82,18 %	82,30 %	79,70 %
TOR	97,44 %	97,44 %	94,87 %	100,00 %
TRE	95,73 %	94,86 %	94,23 %	96,36 %
VT	90,69 %	90,69 %	89,81 %	90,44 %

Tabla 6.13: Recall por clase y por partición en la predicción final bajo $\gamma = 1{,}0$.

Métricas por clase agregadas entre particiones

Para consolidar el análisis del rendimiento bajo este sesgo forzado, la Tabla 6.14 detalla las métricas predictivas para cada categoría sismológica. Los resultados se reportan expresando la media aritmética y la desviación estándar ($\mu \pm \sigma$) calculadas a partir de las cuatro particiones independientes, evidenciando el compromiso estadístico generado.

Clase	Precisión ($\mu \pm \sigma$)	Recall ($\mu \pm \sigma$)	F1-Score ($\mu \pm \sigma$)
LP	$87{,}5\,\% \pm 0{,}5\,\%$	$81{,}0\,\% \pm 1{,}5\,\%$	$84{,}1\,\% \pm 0{,}7\,\%$
TOR	$56{,}9\,\% \pm 3{,}8\,\%$	$97{,}4\,\% \pm 2{,}1\,\%$	$71{,}8\,\% \pm 3{,}1\,\%$
TRE	$96{,}4\,\% \pm 0{,}4\,\%$	$95{,}3\,\% \pm 1{,}0\,\%$	$95{,}8\,\% \pm 0{,}3\,\%$
VT	$86{,}0\,\% \pm 0{,}9\,\%$	$90{,}4\,\% \pm 0{,}4\,\%$	$88{,}2\,\% \pm 0{,}6\,\%$

Tabla 6.14: Precisión, recall y F1-score por clase, reportados como media $\pm$ desviación estándar sobre las cuatro particiones, bajo $\gamma = 1{,}0$.

Métricas globales de evaluación

Para complementar el análisis por categoría, la Tabla 6.15 consolida el rendimiento general del sistema. Esta tabla permite cuantificar cómo la alteración drástica del equilibrio probabilístico en la etapa de inferencia repercute sobre métricas altamente sensibles al desbalance, como el F1-Score Macro.

Métrica Global	P1	P2	P3	P4	Promedio ($\mu \pm \sigma$)
Accuracy global	$88{,}81\,\%$	$89{,}34\,\%$	$88{,}85\,\%$	$88{,}97\,\%$	$88{,}99\,\% \pm 0{,}24\,\%$
F1 Macro	$85{,}77\,\%$	$84{,}77\,\%$	$84{,}27\,\%$	$85{,}03\,\%$	$84{,}96\,\% \pm 0{,}63\,\%$
F1 Ponderado	$88{,}80\,\%$	$89{,}44\,\%$	$88{,}96\,\%$	$89{,}00\,\%$	$89{,}05\,\% \pm 0{,}27\,\%$

Tabla 6.15: Consolidado de métricas de desempeño global por partición (P1 a P4) y su respectivo promedio estadístico bajo $\gamma = 1{,}0$.

La evaluación bajo una ponderación máxima ($\gamma = 1{,}0$) expone una alteración radical en la distribución probabilística de las predicciones. Los datos confirman que esta configuración empuja al clasificador a maximizar la detección de la clase TOR, logrando un *Recall* cercano a la totalidad de las muestras, pero provocando simultáneamente una caída severa en su precisión. En contraste, las clases mayoritarias logran mantener sus indicadores de reconocimiento en rangos funcionalmente intactos.

Capítulo 7

Discusión

Los resultados expuestos en el capítulo 6 permiten articular tres ejes interpretativos fundamentales: la correspondencia entre la arquitectura de aprendizaje en contexto y la naturaleza física de las señales sísmicas del Galeras, el análisis crítico de las estrategias de compensación empleadas, y el reconocimiento de las limitaciones estructurales que delimitan el techo de rendimiento alcanzable en este problema.

7.1. Correspondencia entre la arquitectura y la física de la señal

El desempeño diferencial entre clases no depende exclusivamente del diseño algorítmico, sino que refleja en gran medida la separabilidad intrínseca que exhiben las representaciones espectro-temporales de cada tipo de sismo en el espacio de características construido. La clase tremor (TRE) emerge de forma sistemática y estable como la más fácilmente reconocible en todos los experimentos. En el contexto de este catálogo, TRE corresponde a tremor espasmódico, cuya dinámica tiene una caracterización geofísica precisa: se manifiesta como ráfagas de energía que pueden durar desde decenas de segundos hasta más de 30 minutos, frecuentemente acompañadas por eventos de alta frecuencia superpuestos y codas extensas [52]. Ante un mecanismo de aprendizaje en contexto como TabPFN (cuya decisión descansa en la similitud métrica entre el vector de prueba y la distribución conjunta de los ejemplos de contexto), esta morfología compleja pero altamente distintiva se traduce en vectores de probabilidad con una moda dominante, lo que la hace robusta frente a cambios en la composición del contexto.

La clase LP, en contraste, concentra sistemáticamente los valores inferiores de *Recall* y *F1-Score* entre las clases mayoritarias a lo largo de todos los experimentos, una tendencia que persiste incluso cuando el problema se reduce a tres clases en el Experimento 2. Físicamente, los eventos de periodo largo se asocian a la resonancia de grietas o conductos llenos de fluidos, presentando un inicio emergente, poca o nula claridad en el arribo de la fase S, y un contenido frecuencial dominante entre 1 y 5 Hz, con escasa energía por encima de los 13 Hz.

La resistencia del algoritmo a separar esta clase con alta precisión sugiere que las señales LP del catálogo exhiben una heterogeneidad interna considerable. Distintos mecanismos de resonancia y variaciones en la trayectoria de propagación pueden coexistir bajo esta misma categoría, difuminando las fronteras de decisión matemáticas. La red neuronal no puede aprender a separar de manera tajante aquello que el espacio de características no aísla por sí solo.

La clase VT presenta un comportamiento complementario: su *Recall* es consistentemente superior a su precisión, lo que indica que el modelo tiende a sobreestimar la asignación de esta etiqueta. A diferencia de los eventos impulsados por fluidos, los eventos volcano-tectónicos responden a la fractura frágil de la roca, evidenciando arribos P y S mucho más claros y un contenido frecuencial dominante más alto (típicamente entre 5 y 15 Hz, o con energía significativa por encima de los 10 Hz) [53]. Este cruce de métricas apunta a una superposición parcial en el espacio de representación entre VT y LP, especialmente en regímenes de baja energía. La frontera de decisión entre ambas clases es difusa, y TabPFN, al operar por similitud, hereda esa ambigüedad física de la señal original.

El comportamiento fluctuante de la clase TOR en el Experimento 1 (alta precisión media pero elevada varianza en el *Recall* entre particiones) es la expresión cuantitativa de un problema de soporte estadístico insuficiente. Físicamente, los eventos tipo tornillo presentan formas de onda cuasi-monocromáticas o con uno a tres picos de frecuencia estrechos, una coda larga y un decaimiento lento [2, 54]. Esta forma de onda tan característica y su morfología espectral le proporcionan al algoritmo un patrón identificable, lo que explica su elevada precisión. Sin embargo, al contar con una cantidad sumamente reducida de eventos reales distribuidos en el contexto, la composición de cada iteración altera de manera significativa la representación matemática que el modelo tiene de esta clase. Esto desplaza la frontera de decisión de forma impredecible ante cualquier ligera variación natural en la señal, causando la inestabilidad observada en el *Recall*.

7.2. Debate crítico de las estrategias de compensación

Los Experimentos 3 y 4 ponen a prueba dos mecanismos cualitativamente distintos para abordar el desbalance: la intervención en el espacio de características mediante aumento de datos (ruido) y la intervención en la etapa de decisión mediante la ponderación probabilística (γ). Sus resultados permiten argumentar las consecuencias metodológicas de cada aproximación.

Aumento de datos por inyección de ruido

La inyección de ruido instrumental para sintetizar nuevos eventos TOR resuelve el problema aritmético del desbalance, pero introduce una tensión estructural en el espacio de características. El ruido añadido, aunque modelado teóricamente sobre el fondo sísmico real, altera las formas de onda cuasi-monocromáticas y las codas que caracterizan a los tornillos reales. Al

ampliar la clase TOR con señales cuya energía ha sido parcialmente aleatorizada, se satura una región del espacio de características que no corresponde fielmente a la distribución natural del evento, sino a una distribución expandida y difusa en torno a ella.

Las consecuencias son observables y asimétricas entre clases. La precisión de LP se deteriora significativamente bajo el Experimento 3, mientras que su *Recall* se mantiene. Esto indica que la distribución expandida de TOR comienza a solaparse con la nube de puntos LP en el espacio de representación, provocando que el modelo clasifique algunos eventos de resonancia de fluidos como tornillos. Mientras que la adición de ruido afecta severamente a las clases con mayor ambigüedad física, la categoría TRE permanece al margen de este efecto. Su naturaleza de banda ancha y su alta intensidad le otorgan una cohesión matemática que la vuelve inmune a la perturbación sintética. Adicionalmente, la precisión de TOR bajo ruido escala a valores cercanos a 98,1 %, mientras que su *Recall* se vuelve inestable: el modelo aprendió a reconocer el patrón del ruido sintético como la huella exclusiva de la clase, pero esa representación conservadora no captura con fidelidad todos los eventos reales de prueba, especialmente aquellos cuyos picos espectrales originales son más sutiles o están enmascarados por la propia actividad del volcán.
Esta asimetría revela un fenómeno de sobreajuste a la perturbación sintética: la red ajusta su métrica de comparación a las muestras artificiales, que son numéricamente abundantes, y reduce su capacidad de generalización hacia los tornillos reales.

Ponderación probabilística del operador de decisión

La compensación vía factor γ opera en la fase final del proceso de inferencia, interviniendo únicamente en la transformación del vector de probabilidad antes de aplicar la función $\arg\max$. Esta característica metodológica hace que su análisis sea diametralmente distinto al del aumento de datos.

Con $\gamma = 0{,}5$, la perturbación introducida en la frontera de decisión es suficiente para elevar el *Recall* de TOR por encima de su estado base, con un coste de precisión acotado y un impacto prácticamente nulo sobre las clases mayoritarias. La moderación de este intercambio sugiere que en este rango, la probabilidad nativa asignada por el modelo a TOR ya era funcionalmente informativa, y el γ actúa como un mecanismo de calibración empírica que corrige la penalización probabilística impuesta por el desbalance de clases.
Con $\gamma = 1{,}0$, la compensación máxima revela el límite operativo del mecanismo: la precisión de TOR colapsa a 56,9 %, mientras que su *Recall* roza la unidad. Este fenómeno no es un fallo computacional, sino la consecuencia algebraica de amplificar la probabilidad asignada a TOR hasta un punto en que el algoritmo declara un evento premonitorio con tal agresividad que acepta una tasa de falsos positivos operativamente inviable, catalogando como TOR a una porción no trivial de eventos LP y VT transitorios. El F1-Score Macro cae de forma apreciable en este escenario, demostrando que la ganancia en la detección de la clase minoritaria se adquiere a un costo desproporcionado para la estabilidad global del sistema.
La comparación entre ambos niveles de γ demuestra que la ponderación del vector de proba-

bilidades es una herramienta robusta únicamente cuando la probabilidad base generada por el modelo contiene información discriminativa genuina. Si esa información no existe en el origen físico de la señal, forzar el equilibrio matemático solo produce una sobreestimación sistemática.

7.3. Limitaciones estructurales del problema

El conjunto de resultados exhibe una notable convergencia de las métricas globales en torno a un rango de exactitud (*Accuracy*) del 87 % al 90 %, independientemente del experimento evaluado. Esta convergencia refleja la dominancia numérica de las clases mayoritarias: LP, TRE y VT constituyen la inmensa mayoría del catálogo, y cualquier estrategia que las clasifique correctamente ancla la métrica global en ese rango, aislando el impacto real que sufra la clase TOR. En este sentido, el techo de rendimiento observado no está dictado por la capacidad expresiva de TabPFN, sino por la estructura estadística del catálogo sísmico del Galeras y por los límites geofísicos inherentes a la superposición de las señales.

La clase LP representa la limitación más difícil de superar sin una reconfiguración de los datos de entrada. Su *F1-Score* no logra elevarse significativamente en ningún escenario, lo que señala que el vector de características seleccionado, si bien es informativo, no captura de forma absoluta toda la variabilidad que diferencia al período largo de los eventos de ruptura frágil en rangos de baja frecuencia. Esta barrera indica que la complejidad del fenómeno físico volcánico supera la capacidad de dimensionalidad del espacio actual para representarlo de forma unívoca y libre de traslapes.

El desbalance extremo de la clase TOR constituye asimismo un límite estructural que las estrategias algorítmicas no logran neutralizar por completo. Las técnicas de compensación matemática pueden reconfigurar el balance entre precisión y recall, pero no pueden fabricar la información discriminativa de los fenómenos físicos ausentes. La rareza de este evento impone un límite epistémico sobre la capacidad de cualquier clasificador estadístico para generalizarlo de forma absoluta sin incrementar el margen de error probabilístico.

Finalmente, el comportamiento del Experimento 2 (donde la exclusión de TOR y el equilibrio del contexto produjeron métricas de las clases mayoritarias virtualmente idénticas a las del Experimento 1) arroja una análisis fundamental: la presencia de la clase tornillo (TOR), reconocida por su carácter premonitorio en la actividad del Galeras, no degrada la identificación de las dinámicas dominantes, pero el balance perfecto tampoco incrementa el rendimiento algorítmico sobre ellas. Esto ratifica que el desafío central en el modelamiento del volcán Galeras no radica meramente en optimizar arquitecturas computacionales, sino en gestionar la morfología propia de una sismicidad cuyas fronteras físicas entre clases son difusas por naturaleza.

Capítulo 8

Conclusiones y recomendaciones

El objetivo general de este trabajo fue clasificar los diferentes tipos de sismos generados en el volcán Galeras mediante una arquitectura basada en *Transformer*. Para ello, se desarrolló un flujo de trabajo que integró la selección de características, la construcción de conjuntos experimentales y la evaluación del modelo TabPFN. Los resultados demuestran que esta arquitectura, por sí sola y en su entorno nativo, posee un funcionamiento altamente competitivo. Alcanzando una exactitud (*Accuracy*) global cercana al 89 % en su configuración nativa, TabPFN demostró ser una alternativa robusta y viable para la clasificación automática de sismicidad volcánica, destacándose por su capacidad de extraer patrones útiles y operar de manera eficiente en contextos con disponibilidad limitada de datos y distribuciones severamente desbalanceadas.

El análisis por categorías reveló que el rendimiento del modelo está profundamente ligado a la separabilidad física de las señales. La clase tremor (TRE) presentó sistemáticamente los mejores valores de *F1-Score*, con métricas de Precisión y *Recall* consistentemente elevadas en todos los escenarios, confirmando que su dinámica de banda ancha e intensidad genera una representación inconfundible para el algoritmo. En contraste, las clases de periodo largo (LP) y volcano-tectónicas (VT) evidenciaron niveles de confusión persistentes. Esta dificultad de discriminación confirma que ambas dinámicas comparten regímenes de frecuencia y características parcialmente superpuestas en el espacio de representación actual, lo que restringe la separación matemática absoluta.

Adicionalmente, la evaluación del Experimento 2 sobre un escenario equitativo de tres clases cumplió un doble propósito metodológico: estableció una referencia del rendimiento alcanzable en ausencia de desbalance y, simultáneamente, demostró que las clases dominantes (LP, TRE y VT) son robustas al desbalance, manteniendo un desempeño prácticamente invariante respecto al entorno nativo. Esta evidencia permitió delimitar con precisión el origen del problema: la dificultad del catálogo no radica en un desbalance global que afecte a todas las clases, sino en la escasez estructural específica de la clase tornillo (TOR). Por esta razón, los Experimentos 3 y 4 se diseñaron como intervenciones focalizadas sobre TOR, mediante aumento sintético de datos y ponderación probabilística de la decisión final, respectivamente,

en lugar de continuar manipulando el equilibrio global del contexto.

Respecto al manejo de la clase minoritaria tornillo (TOR), la estrategia de aumento de datos mediante inyección de ruido instrumental produjo efectos adversos en la generalización. Aunque este método permitió equilibrar numéricamente el conjunto de contexto y prueba, introdujo una tensión en el espacio de características que diluyó la morfología original de la señal. Esto indujo un sobreajuste a la perturbación sintética y degradó la clasificación de los eventos circundantes, demostrando empíricamente que la alteración artificial de la onda no es el camino óptimo para compensar eventos con frecuencias discretas tan particulares.
La estrategia de ponderación probabilística sobre la regla de decisión final, evaluada bajo dos niveles del factor γ, exhibió un comportamiento monotónico predecible. Con un factor moderado ($\gamma = 0{,}5$) se logró elevar la sensibilidad hacia la clase tornillo sin alterar de forma significativa las métricas de las clases dominantes, configurando una operación funcionalmente útil para escenarios en los que prima la detección de eventos premonitorios. Por el contrario, la compensación máxima ($\gamma = 1{,}0$) expuso el límite operativo del mecanismo: la precisión de TOR se redujo hasta el 56,9 %, evidenciando que forzar el equilibrio matemático en ausencia de información discriminativa adicional sólo reconfigura el compromiso entre precisión y exhaustividad sin generar capacidad de separación nueva en el espacio de características.
Al contrastar la exactitud (*Accuracy*) nativa de TabPFN con estudios previos basados en arquitecturas profundas tradicionales, como redes convolucionales y recurrentes (LSTM, GRU) que reportan desempeños entre el 90 % y el 94 % en la clasificación de señales sismo-volcánicas [33,34], se evidencia el valor metodológico de la presente propuesta. Aunque las métricas no son estrictamente comparables debido a las diferencias en los catálogos y al número de categorías evaluadas, la capacidad de TabPFN para lograr un desempeño sólido sobre un problema multiclase severamente desbalanceado respalda su eficacia y utilidad práctica en la caracterización automática de la actividad volcánica.

A partir de las conclusiones expuestas, y con el fin de superar las limitaciones estructurales identificadas a lo largo de la investigación, se formulan tres líneas de trabajo futuro orientadas, respectivamente, a la representación de la señal, a la generación de datos sintéticos y al ajuste probabilístico de la decisión:

1. **Enriquecimiento del espacio de características:** se recomienda construir un espacio paramétrico ampliado que mejore la discriminación entre los cuatro tipos de sismos. La incorporación de un mayor número de descriptores temporales y frecuenciales resulta fundamental para resolver la superposición matemática observada entre los eventos de resonancia de fluidos y los de ruptura frágil, logrando capturar con mayor precisión las sutilezas morfológicas de cada categoría.

2. **Reformulación de las técnicas de aumento de datos:** frente al desafío persistente del desbalance de clases, resulta pertinente reevaluar el enfoque de aumento utilizado. En lugar de la inyección de ruido de fondo, se sugiere explorar métodos que preserven íntegramente la dinámica natural de las señales y, sobre dicha base, evaluar nuevamente

el rendimiento de TabPFN, con el fin de determinar si tales alternativas logran equilibrar el catálogo sin inducir la contaminación del espacio de características observada en el presente estudio.

3. **Calibración del operador de decisión ponderado:** como aproximación complementaria a la manipulación directa de la señal, se sugiere continuar explorando el ajuste probabilístico del operador de decisión. En particular, se recomienda buscar un factor γ que se acople a la distribución específica de cada conjunto de datos analizado, permitiendo así calibrar la detección de eventos minoritarios según las necesidades operativas de cada catálogo volcánico.

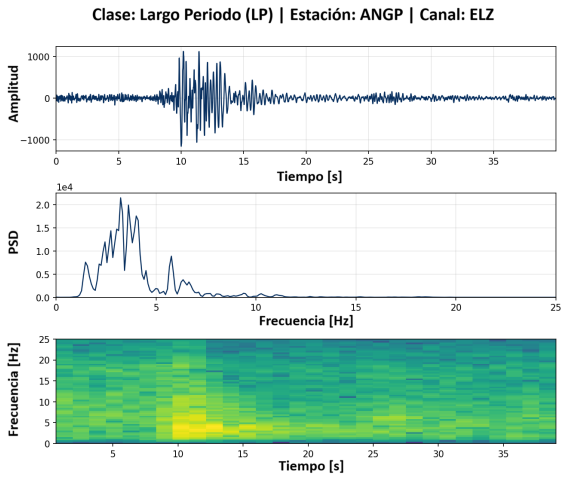
En conjunto, estas tres líneas conforman una hoja de ruta complementaria: la primera atiende la representación de los datos en su origen, la segunda la generación controlada de instancias adicionales, y la tercera la regulación final de la regla de decisión. Su exploración articulada permitiría avanzar de manera sistemática en la caracterización automática de la sismicidad del volcán Galeras.

APÉNDICES

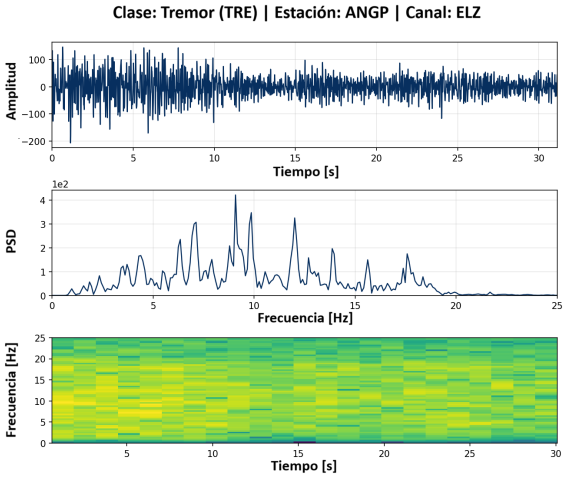
Apéndice A

Caracterización visual de las clases sísmicas

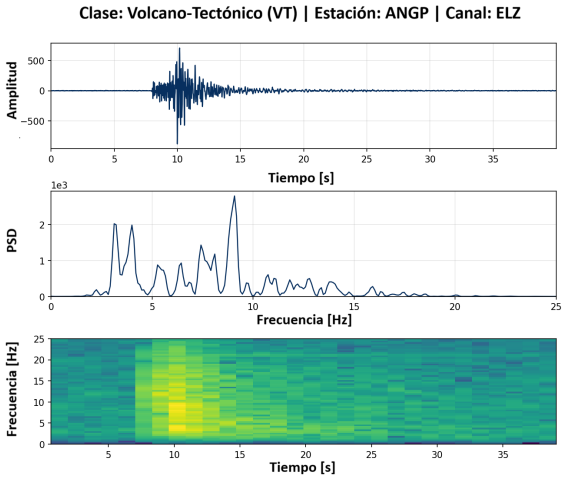
En este apéndice se exponen las representaciones visuales de los diferentes tipos de sismos registrados en el volcán Galeras.



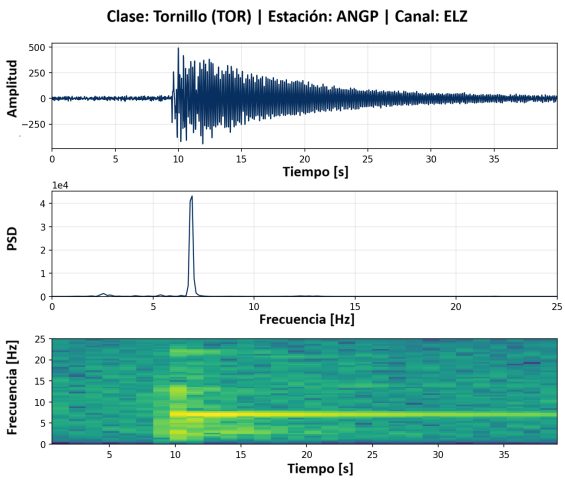
(a) Largo Periodo (LP) - 16/11/2008



(b) Tremor volcánico (TRE) - 11/07/2007



(c) Volcano-Tectónico (VT) - 30/06/2010



(d) Tornillo (TOR) - 25/09/2009

Figura A.1: Caracterización espectro-temporal de las cuatro clases sísmicas del volcán Galeras en la componente ELZ de la estación ANGP. Se contrasta la resonancia de los eventos LP y TOR, la emisión sostenida del TRE y la naturaleza impulsiva de alta frecuencia del VT.

Apéndice B

Ranking complementario de importancia de características

En este apéndice se presenta la continuación del ranking de importancia de las 53 características, correspondiente a los bloques no mostrados en la Sección 5.3.2.

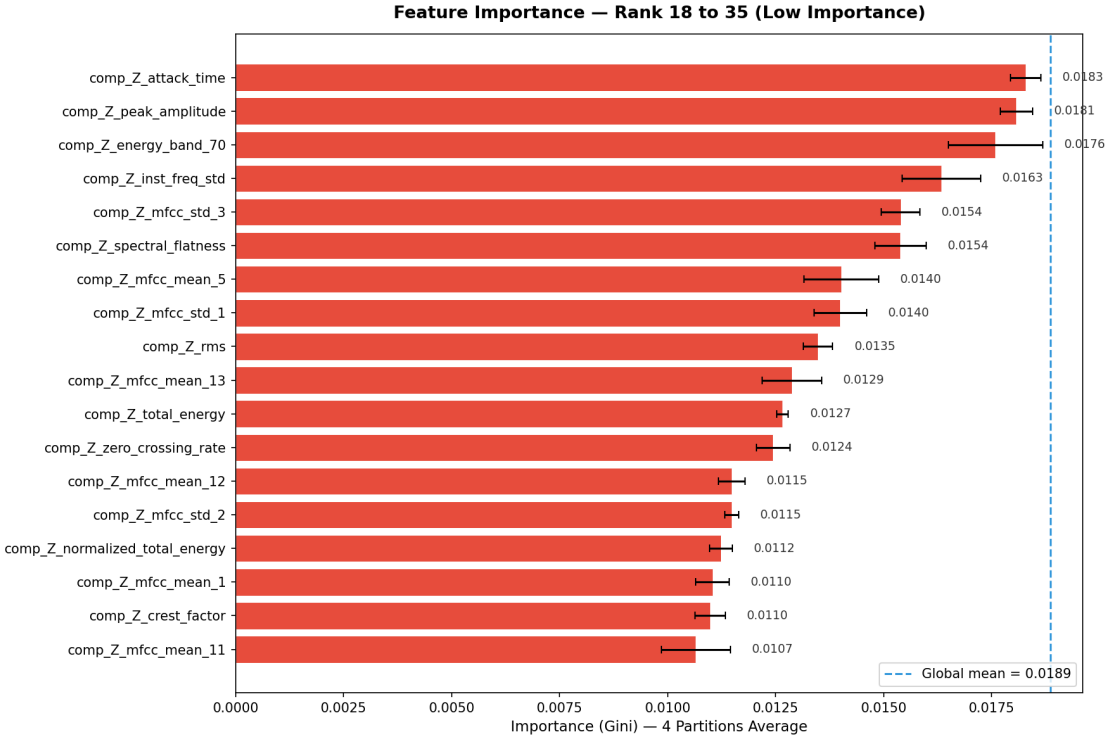


Figura B.1: Ranking de importancia, características 18 a 35.

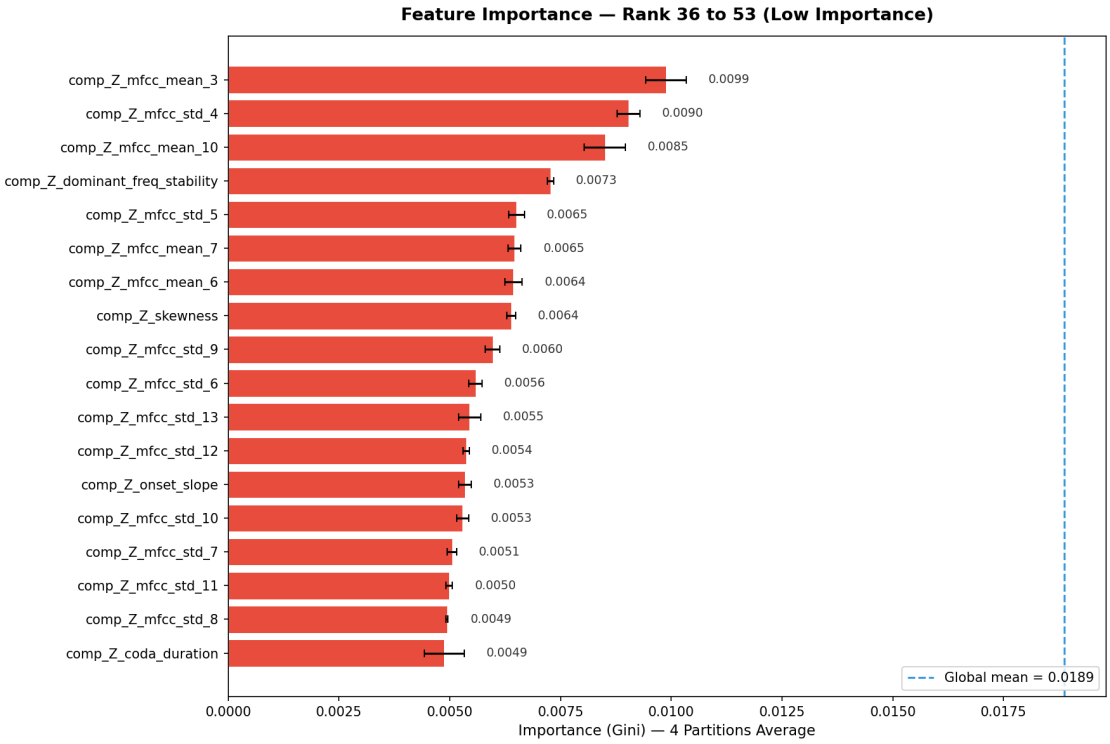


Figura B.2: Ranking de importancia, características 36 a 53.

Apéndice C

Validación visual del proceso de generación de datos sintéticos

En esta sección se presentan las validaciones visuales correspondientes a los escalares de perturbación medio y alto, aplicados sobre eventos base distintos de la clase tornillo (TOR). Estas comparaciones complementan el ejemplo mostrado en la Sección 5.5.4.

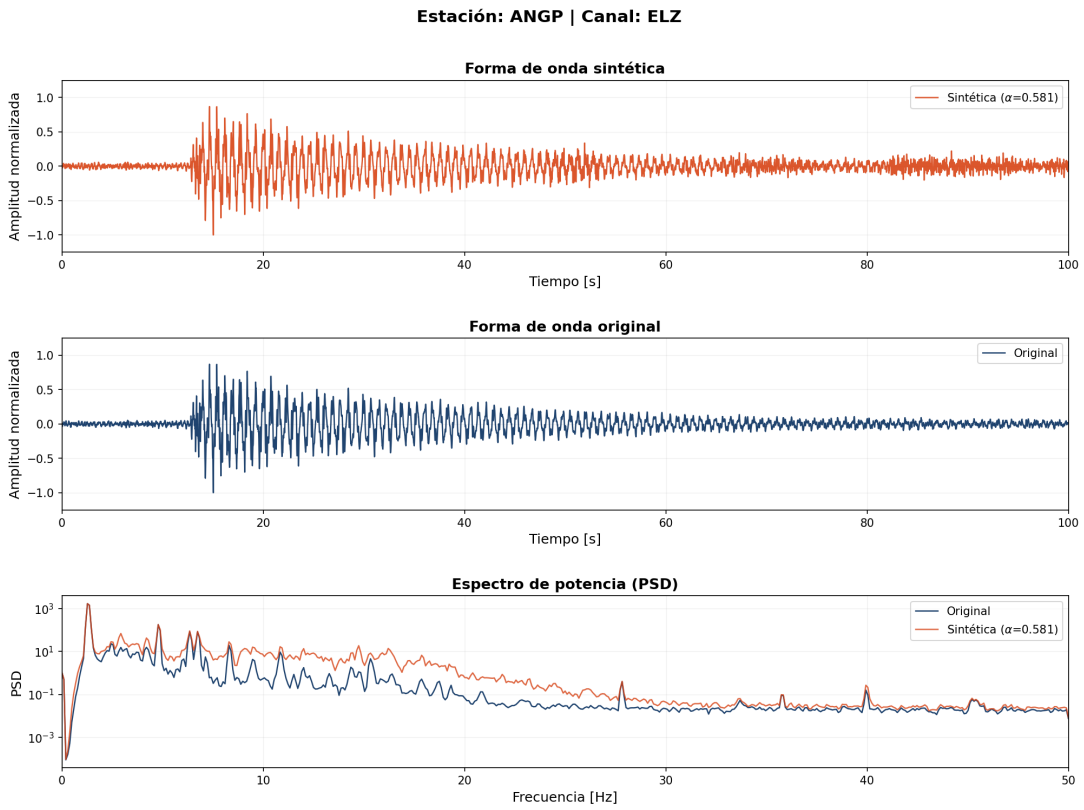


Figura C.1: Comparación temporal y espectral de un evento tipo tornillo (TOR) registrado en la estación ANGP (componente ELZ) frente a su variante aumentada con nivel de perturbación medio ($\alpha = 0{,}581$). Los paneles superior y central muestran la amplitud normalizada en el dominio del tiempo para la señal sintética y la original, respectivamente. El panel inferior presenta la densidad espectral de potencia (PSD) comparativa de ambas trazas. Evento registrado el 15 de febrero de 2007.

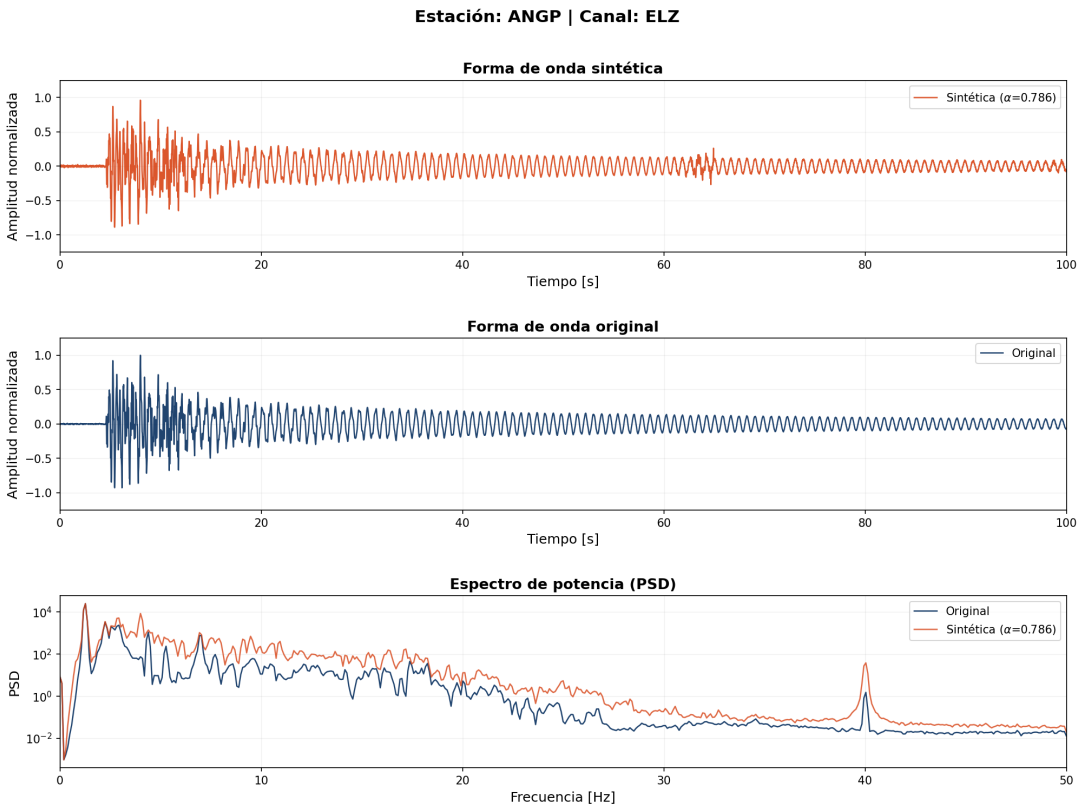
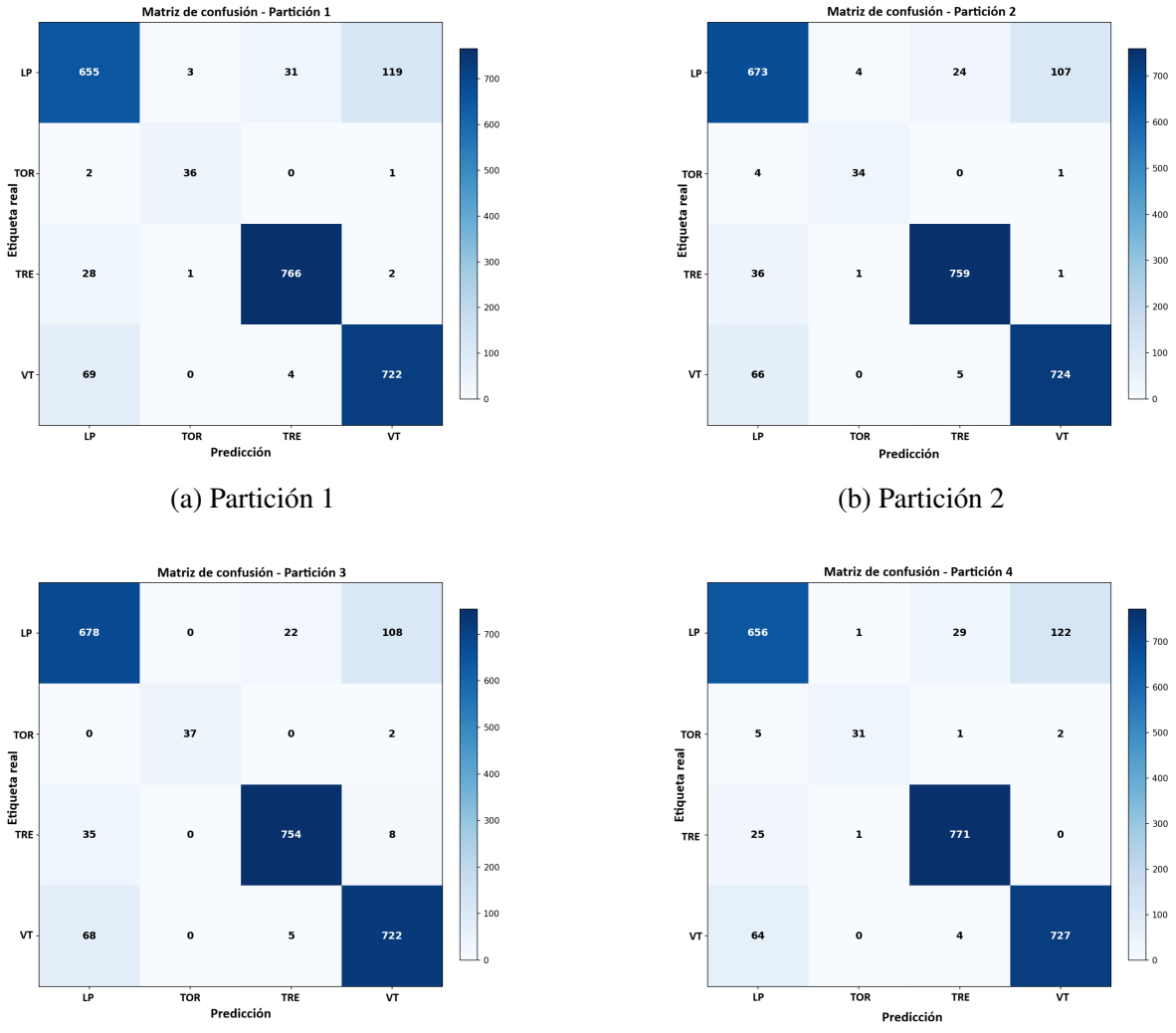


Figura C.2: Comparación temporal y espectral de un evento tipo tornillo (TOR) registrado en la estación ANGP (componente ELZ) frente a su variante aumentada con nivel de perturbación alto ($\alpha = 0{,}786$). Los paneles superior y central muestran la amplitud normalizada en el dominio del tiempo para la señal sintética y la original, respectivamente. El panel inferior presenta la densidad espectral de potencia (PSD) comparativa de ambas trazas. Evento registrado el 31 de diciembre de 2006.

Apéndice D
Matrices de confusión

En este apéndice se presentan en detalle las matrices de confusión obtenidas en cada una de las cuatro particiones (MCCV) para los cuatro experimentos diseñados en esta investigación.

D.1. Experimento 1: línea base



(a) Partición 1

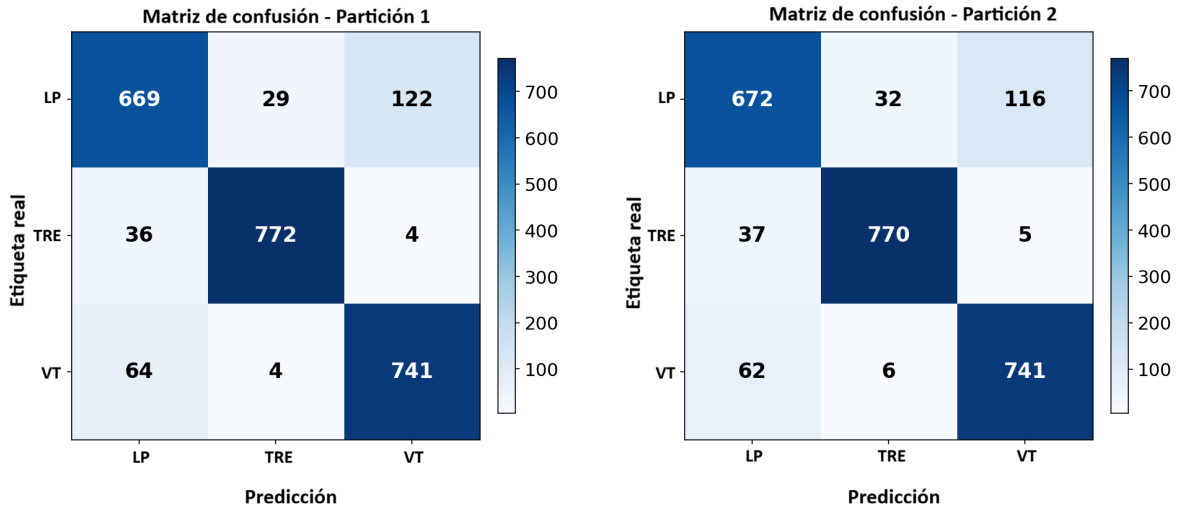
(b) Partición 2

(c) Partición 3

(d) Partición 4

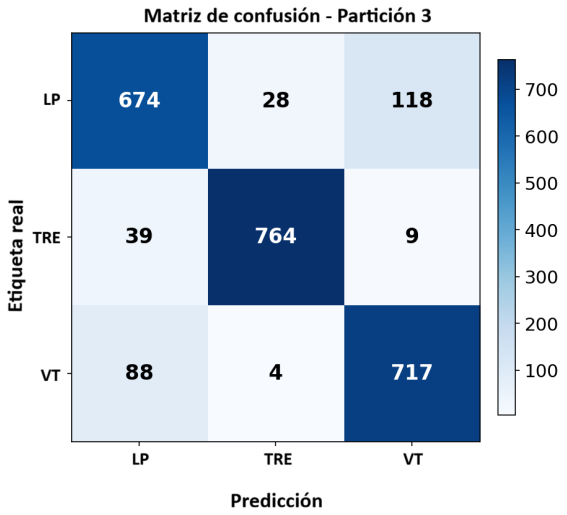
Figura D.1: Matrices de confusión resultantes de las cuatro particiones del Experimento 1. Las celdas diagonales representan las predicciones correctas, mientras que los valores fuera de la diagonal ilustran la confusión cruzada entre clases.

D.2. Experimento 2: entorno balanceado sin la clase tornillo

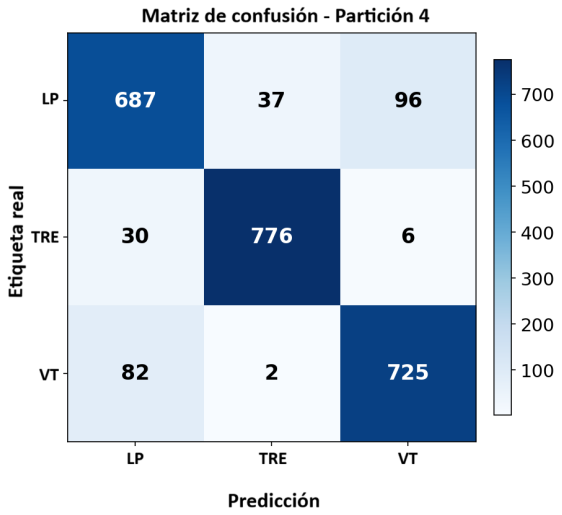


(a) Partición 1

(b) Partición 2



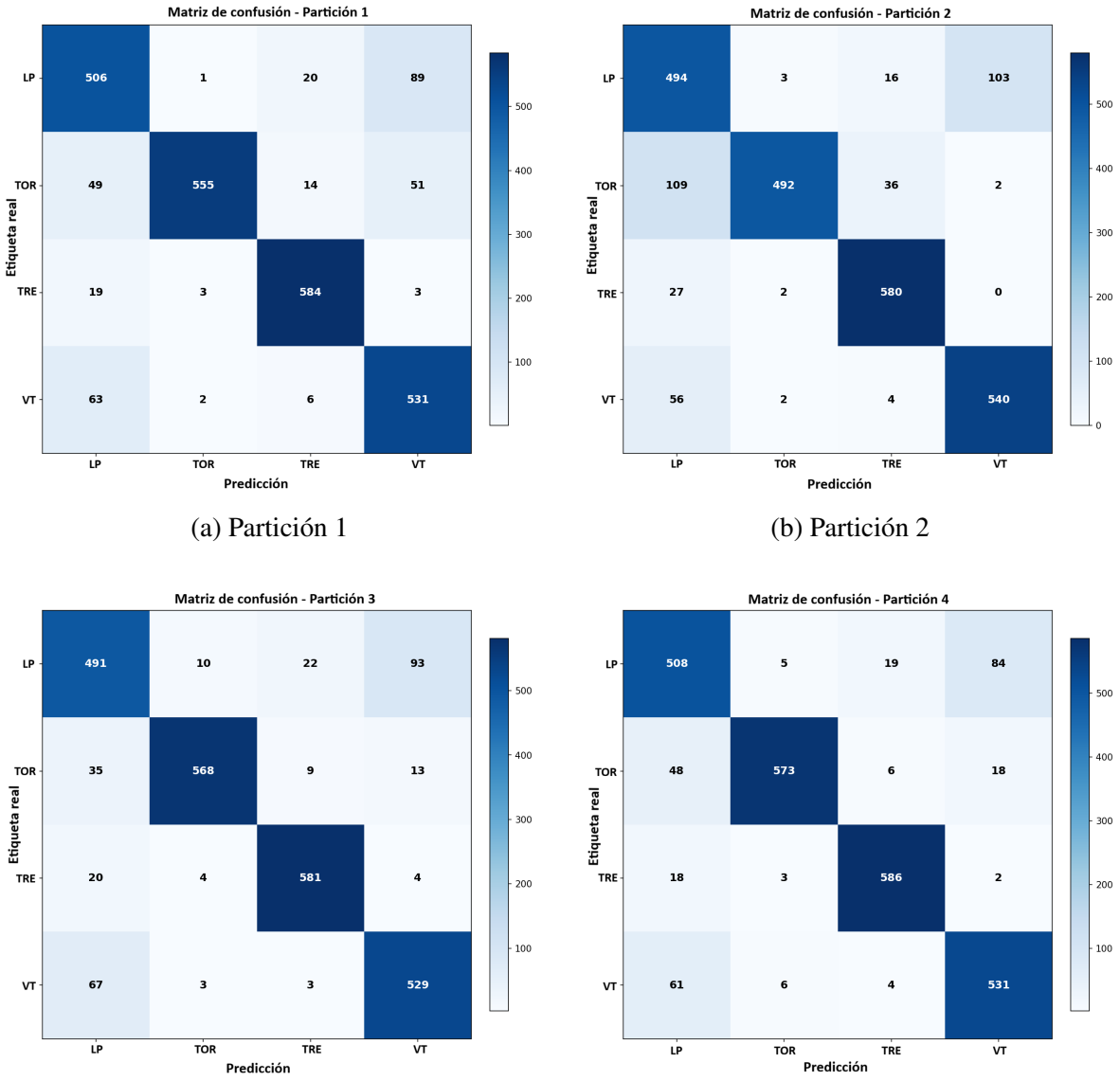
(c) Partición 3



(d) Partición 4

Figura D.2: Matrices de confusión resultantes de las cuatro particiones del Experimento 2, evidenciando la superposición física entre las clases LP y VT.

D.3. Experimento 3: aumento de datos mediante inyección de ruido



(a) Partición 1

(b) Partición 2

(c) Partición 3

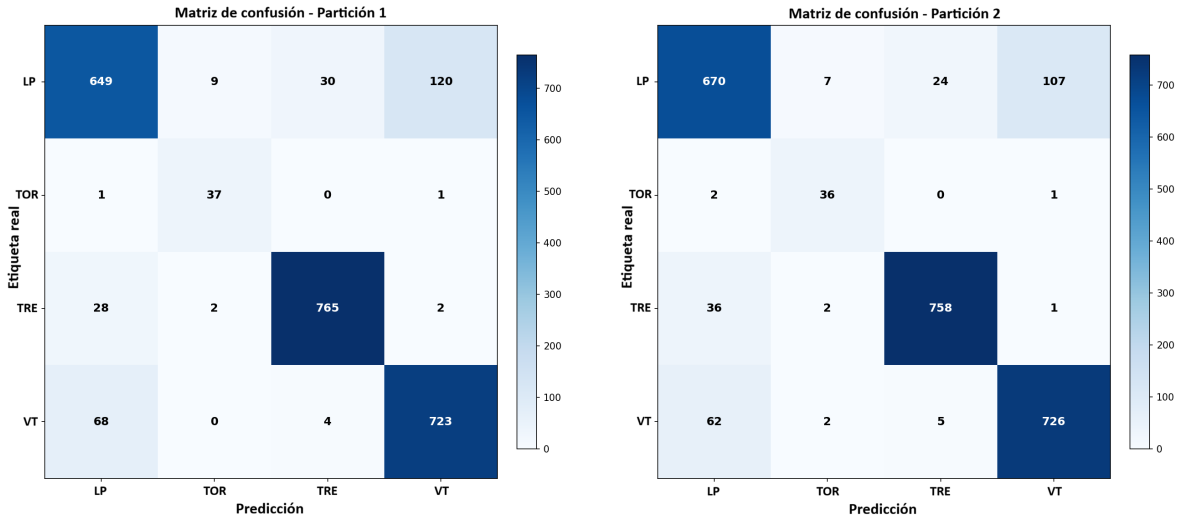
(d) Partición 4

Figura D.3: Matrices de confusión resultantes de las cuatro particiones del Experimento 3. Se observa el impacto del ruido sintético en la dispersión de los aciertos, especialmente en la clase tornillo.

D.4. Experimento 4: ponderación probabilística

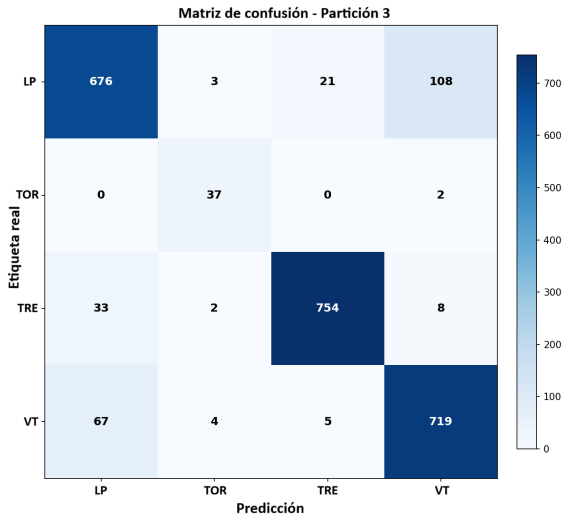
Este apartado detalla el comportamiento predictivo del modelo al aplicar diferentes configuraciones en la ponderación de clases. Se presentan los resultados obtenidos para dos escenarios de ajuste del hiperparámetro γ.

D.4.1. Matrices de confusión con $\gamma = 0{,}5$

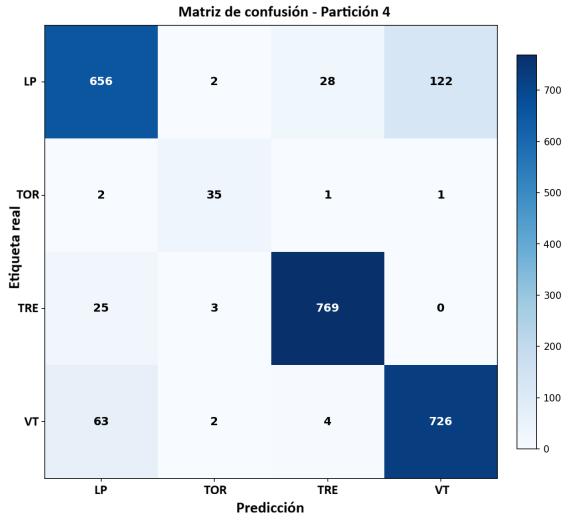


(a) Partición 1

(b) Partición 2



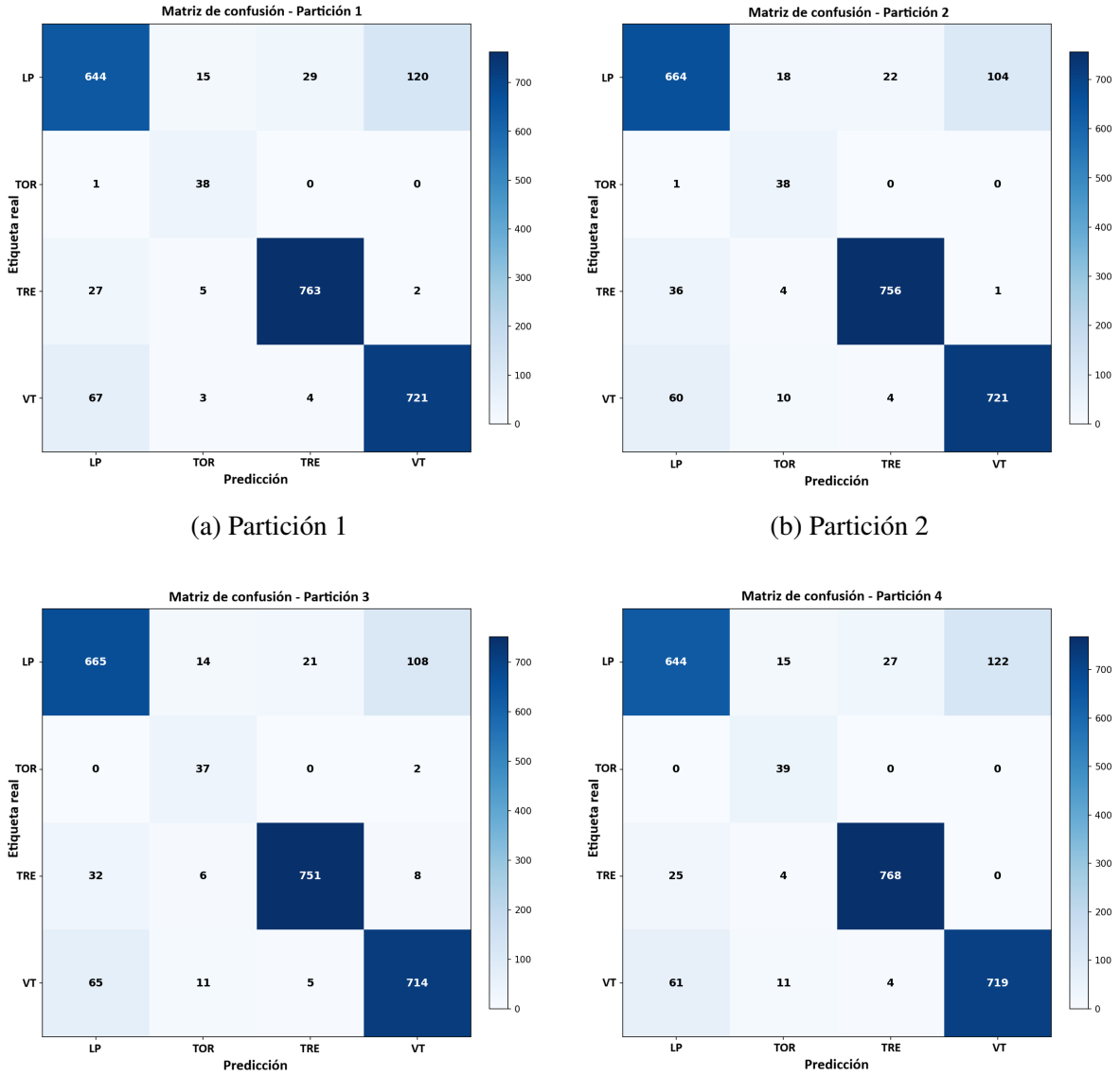
(c) Partición 3



(d) Partición 4

Figura D.4: Matrices de confusión de las cuatro particiones del Experimento 4 aplicando ponderación probabilística con un factor $\gamma = 0{,}5$.

D.4.2. Matrices de confusión con $\gamma = 1{,}0$



(a) Partición 1

(b) Partición 2

(c) Partición 3

(d) Partición 4

Figura D.5: Matrices de confusión de las cuatro particiones del Experimento 4 aplicando ponderación probabilística con un factor $\gamma = 1{,}0$.

Bibliografía

[1] H. Cepeda, *Quinientos años de documentación histórica de actividad del volcán Galeras: escenarios eruptivos*. Servicio Geológico Colombiano, 2020.

[2] O. E. C. Ibarra, "Modelos de fuente de sismicidad lp para la actividad del volcán galeras 2004-2010 (colombia)," Ph.D. dissertation, Universidad Nacional de Colombia, 2021.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Łukasz Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 12 2017, pp. 5998–6008.

[4] 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.

[5] M. Malfante, M. D. Mura, J. I. Mars, J. Métaxian, O. Macedo, and A. Inza, "Automatic classification of volcano seismic signatures," *Journal of Geophysical Research: Solid Earth*, vol. 123, pp. 10 645–10 658, 12 2018.

[6] D. Tan, D. Fee, A. J. Hotovec-Ellis, J. D. Pesicek, M. M. Haney, J. A. Power, and T. Girona, "Volcanic earthquake catalog enhancement using integrated detection, matched-filtering, and relocation tools," *Frontiers in Earth Science*, vol. 11, 5 2023.

[7] D. Fee, D. Tan, J. Lyons, M. Sciotto, A. Cannata, A. Hotovec-Ellis, T. Girona, A. Wech, D. Roman, M. Haney, and S. D. Angelis, "A generalized deep learning model to detect and classify volcano seismicity," *Volcanica*, vol. 8, pp. 305–323, 6 2025.

[8] Y. Yukutake, R. Honda, M. Harada, R. Doke, T. Saito, T. Ueno, S. Sakai, and Y. Morita, "Analyzing the continuous volcanic tremors detected during the 2015 phreatic eruption of the hakone volcano," *Earth, Planets and Space*, vol. 69, p. 164, 12 2017.

[9] K. J. Bergen, P. A. Johnson, M. V. de Hoop, and G. C. Beroza, "Machine learning for data-driven discovery in solid earth geoscience," *Science*, vol. 363, no. 6433, p. eaau0323, 2019.

[10] R. Ashmore, R. Calinescu, and C. Paterson, "Assuring the machine learning lifecycle: Desiderata, methods, and challenges," *arXiv preprint arXiv:1905.04223*, 2019.

[11] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, pp. 1263–1284, 9 2009.

[12] L. Grinsztajn, K. Flöge, O. Key, F. Birkel, B. Roof, P. Jund, B. Jäger, A. Hayler, D. Safaric, S. Alessi, F. Jablonski, M. Manium, R. Yu, A. Garg, J. Robertson, S. B. Hoo, V. Moroshan, M. Bühler, L. Purucker, C. Cornu, L. C. Wehrhahn, A. Bonetto, S. Gambhir,

N. Hollmann, and F. Hutter, "Tabpfn-2.5: Advancing the state of the art in tabular foundation models," *arXiv preprint arXiv:2511.08667*, 2025.

[13] C. Elkan, "The foundations of cost-sensitive learning," in *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 17, no. 1, 2001, pp. 973–978.

[14] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.

[15] P. Carrión-Mero, N. Montalván-Burbano, N. Paz-Salas, and F. Morante-Carballo, "Volcanic geomorphology: A review of worldwide research," *Geosciences*, vol. 10, p. 347, 9 2020.

[16] H. Sigurdsson, B. Houghton, S. McNutt, H. Rymer, and J. Stix, Eds., *The Encyclopedia of Volcanoes*. San Diego, CA: Academic Press, 2015.

[17] S. Petrosino and P. Cusano, "Low frequency seismic source investigation in volcanic environment: the mt. vesuvius atypical case," *Advances in Geosciences*, vol. 52, pp. 29–39, 6 2020.

[18] D. Martinez, R. Torres, D. Seidl, M. Hellweg, and H. Rademacher, "Tornillo seismic events at galeras volcano, colombia: a summary and new information from broadband three-component measurements," *Annals of Geophysics*, vol. 42, 11 1999.

[19] P. Salazar, F. Yupanqui, C. Meneses, S. Layana, and G. Yáñez, "Multi-station automatic classification of seismic signatures from the lascar volcano database," *Natural Hazards and Earth System Sciences*, vol. 23, pp. 991–1006, 3 2023.

[20] D. A. R. Palomino and Y. A. Vera, "Análisis de las señales sísmicas registradas en el volcán ubinas periodo 22 de abril al 16 de mayo del 2008." Instituto Geológico Minero y Metalúrgico. Dirección de Geología Ambiental y Riesgo Geológico., Tech. Rep., 2008.

[21] A. Oppenheim, R. W. O. Alan V/Schafer, and R. W. Schafer, *Discrete-Time Signal Processing*, 2011.

[22] G. Cortés, M. C. Benítez, L. García, I. Álvarez, and J. M. Ibanez, "A comparative study of dimensionality reduction algorithms applied to volcano-seismic signals," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 1, pp. 253–263, 2016.

[23] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes : the art of scientific computing*. Cambridge University Press, 2007.

[24] S. Dubnov, "Generalization of spectral flatness measure for non-gaussian linear processes," *IEEE Signal Processing Letters*, vol. 11, pp. 698–701, 8 2004.

[25] M. G. Tsipouras, "Spectral information of eeg signals with respect to epilepsy classification," *EURASIP Journal on Advances in Signal Processing*, p. 10, 2019.

[26] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. i. fundamentals," *Proceedings of the IEEE*, vol. 80, pp. 520–538, 4 1992.

[27] L. Rabiner and R. Schafer., *Theory and Applications of Digital Speech Processing*, 2011.

[28] Y. Yukutake, A. Kim, and T. Ohminato, "Reappraisal of volcanic seismicity at the kirishima volcano using machine learning," *Earth, Planets and Space*, vol. 75, p. 183, 12 2023.

[29] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2008.

[30] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 10 2001.

[31] C. M. Bishop, *Pattern recognition and machine learning*. Springer Science + Business Media, 2009.

[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, 2016.

[33] D. A. Salas, M. Zumbado, J. Pacheco, M. Mora, L. van der Laat, and E. Meneses, "Automatic classification of seismo-volcanic signals with deep learning: The case of turrialba volcano," in *2023 IEEE 5th International Conference on BioInspired Processing (BIP)*. IEEE, 11 2023, pp. 1–6.

[34] M. Titos, A. Bueno, L. Garcia, M. C. Benitez, and J. Ibanez, "Detection and classification of continuous volcano-seismic signals with recurrent neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 1936–1948, 4 2019.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.

[36] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, "Transformers in time series: A survey," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI 2023)*. International Joint Conferences on Artificial Intelligence (IJCAI) Organization, 8 2023, pp. 6778–6786.

[37] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[39] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?" in *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2022.

[40] N. Hollmann, S. Müller, K. Eggensperger, and F. Hutter, "A transformer that solves small tabular classification problems in a second," 2023.

[41] S. Müller, N. Hollmann, S. Pineda, J. Grabocka, and F. Hutter, "Transformers can do bayesian inference," in *International Conference on Learning Representations*, 2022.

[42] J. Lee, Y. Lee, J. Kim, A. Kosiorek, S. Choi, and Y. W. Teh, "Set transformer: A framework for attention-based permutation-invariant neural networks," in *International Conference on Machine Learning*, 2019.

[43] Prior Labs, "Tabpfn changelog," https://github.com/PriorLabs/TabPFN/blob/main/CHANGELOG.md, gitHub repository, accessed 29 Apr. 2026.

[44] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statistics Surveys*, vol. 4, 2010.

[45] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, "Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking," *Nature Communications*, vol. 11, p. 3952, 8 2020.

[46] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, and H. Xu, "Time series data augmentation for deep learning: A survey," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, 8 2021, pp. 4653–4660.

[47] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.

[48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[49] M. Saerens, P. Latinne, and C. Decaestecker, "Adjusting the outputs of a neural network to new a priori probabilities: a simple procedure," *Neural computation*, vol. 14, no. 1, pp. 21–41, 2002.

[50] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *International conference on machine learning*. PMLR, 2017, pp. 1321–1330.

[51] International Federation of Digital Seismograph Networks (FDSN), "Fdsn miniseed 3," https://docs.fdsn.org/projects/miniseed3, consultado el 23 de abril de 2026.

[52] S. R. McNutt, "Volcanic seismology," *Annual Review of Earth and Planetary Sciences*, vol. 33, pp. 461–491, 2005.

[53] B. A. Chouet, "Long-period volcano seismicity: its source and use in eruption forecasting," *Nature*, vol. 380, no. 6572, pp. 309–316, 1996.

[54] H. Kumagai and B. A. Chouet, "Acoustic properties of a crack containing magmatic or hydrothermal fluids," *Journal of Geophysical Research: Solid Earth*, vol. 105, no. B11, pp. 25 493–25 512, 2000. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JB900273