UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

Clasificación de Eventos Sísmicos con Wav2Vec: Exploración y Evaluación de la tecnología de Reconocimiento de voz en el contexto Sísmico.

Paúl Andrés Quimbita Núñez

Ingeniería en Ciencias de la Computación

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero

Quito, 26 de noviembre de 2023

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA

Clasificación de Eventos Sísmicos con Wav2Vec: Exploración y Evaluación de la tecnología de Reconocimiento de voz en el contexto Sísmico.

Paúl Andrés Quimbita Núñez

Nombre del profesor, Título académico **Felipe Grijalva Arévalo, PhD**

Quito, 26 de noviembre de 2023

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Paúl Andrés Quimbita Núñez

Código: 00212513

Cédula de identidad: 1724492812

Lugar y fecha: Quito, 26 de noviembre de 2023

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en http://bit.ly/COPETheses.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on http://bit.ly/COPETheses.

RESUMEN

Este trabajo de fin de carrera aborda la clasificación de eventos sísmicos mediante la aplicación de la tecnología de reconocimiento de voz Wav2Vec. El enfoque se centra en la adaptación del modelo Wav2Vec 2.0, originalmente diseñado para procesamiento de lenguaje natural, para reconocer patrones sonoros característicos de eventos sísmicos. Se utiliza un conjunto de datos de eventos sísmicos registrados en las faldas del volcán Cotopaxi, y se detalla el preprocesamiento de los datos, la generación de audios a partir de las señales sísmicas y la configuración del modelo preentrenado. El modelo se ajusta fino para la clasificación de dos tipos de eventos sísmicos ("LP" y "VT") utilizando una arquitectura de clasificación. Los resultados muestran una sólida capacidad de generalización del modelo, con métricas como accuracy, F1 score y pérdida (loss) evaluadas en diferentes subconjuntos de prueba. Se discuten las implicaciones de estos resultados y se destaca el potencial de esta aproximación para mejorar la detección y clasificación de eventos sísmicos, contribuyendo así al avance en el campo de la sismología y la gestión de desastres naturales.

Palabras Clave: Clasificador, Wav2Vec 2.0, Evento Sísmico, Ajuste fino, Métricas, Audio, Preprocesamiento de datos, ESeismic.

ABSTRACT

This undergraduate thesis addresses the classification of seismic events through the application of Wav2Vec voice recognition technology. The focus is on adapting the Wav2Vec 2.0 model, originally designed for natural language processing, to recognize sound patterns characteristic of seismic events. A dataset of seismic events recorded at the foothills of Cotopaxi volcano is utilized, and details are provided on data preprocessing, audio generation from seismic signals, and the configuration of the pretrained model. The model is fine-tuned for the classification of two types of seismic events ("LP" and "VT") using a classification architecture. The results demonstrate a robust generalization capability of the model, with metrics such as accuracy, F1 score, and loss evaluated across different test subsets. The implications of these results are discussed, emphasizing the potential of this approach to enhance the detection and classification of seismic events, there by contributing to advancements in the field of seismology and natural disaster management.

Key Word: Classifier, Wav2Vec 2.0, Seismic Event, Fine Tuning, Metrics, Audio, Data Preprocessing, ESeismic.

Tabla de Contenido

Tabla de Ilustraciones

Índice de Tablas

1. Introducción

La clasificación de eventos sísmicos se ha convertido en una parte fundamental de la detección y comprensión de los fenómenos sísmicos. Actualmente, se emplean diversos enfoques basados en modelos de aprendizaje automático para llevar a cabo esta tarea de clasificación. Este estado del arte explora diferentes enfoques utilizados en esta tarea, incluyendo Convolutional Neural Networks (CNN) y modelos de Amplitud Ratio. Los modelos basados en CNN han demostrado un éxito que oscila entre el 91% y el 98% en la clasificación, mientras que los modelos de Amplitud Ratio alcanzan un rendimiento del 80% al 90% (Stangeland, 2021).

La información sísmica se obtiene a través de sensores que capturan ondas generadas durante los movimientos sísmicos en la propagación de ondas. Estas ondas pueden ser diferenciadas por su dirección de oscilación, ya sea longitudinal o transversal (Tibi, et al., 2019). Los modelos de clasificación analizan los datos recopilados de las formas de onda de eventos sísmicos y llevan a cabo la tarea de clasificación. Estos datos se almacenan en formatos como GeoJson, KML y NetCDF, que se utilizan para preservar información geoespacial (Peña et al., 2012). Sin embargo, existe una alternativa para el tratamiento de los datos de eventos sísmicos. Según Peña et al. (2012), mediante técnicas de procesamiento de señales, es posible transformar los datos sísmicos en señales de audio. Esta conversión facilita una comprensión más generalizada de la actividad sísmica.

Por otro lado, Wav2Vec 2.0 es un modelo de red neuronal preentrenado desarrollado por Facebook AI Research. Wav2Vec 2.0 se caracteriza por su entrenamiento inicial no supervisado en un conjunto de datos extenso de audio no etiquetado, seguido de un proceso de ajuste fino utilizando un conjunto más pequeño de datos etiquetados específicos para la tarea en cuestión (Baevski, et al., 2020). Proyectos recientes han adaptado este modelo preentrenado para diversas

tareas, aparte de la transcripción y el reconocimiento del habla, incluyendo "Emotion Recognition in Greek Speech Using Wav2Vec 2.0" (Anónimo, s.f.), "Music Genre Classification with Wav2Vec2" (Ajmain, 2021) y "Sound classification and localization using transformers" (Rosero, s.f.). Estos ejemplos destacan la versatilidad del modelo Wav2Vec 2.0 para diferentes aplicaciones.

Basándonos en estos modelos y con la creciente comprensión del procesamiento de señales sísmicas, se ha propuesto la posibilidad de utilizar el modelo Wav2Vec preentrenado y realizar un fine tuning para llevar a cabo tareas de clasificación basadas en el reconocimiento de patrones sonoros en eventos sísmicos. Esta perspectiva abre nuevas oportunidades para la aplicación de tecnologías de procesamiento de voz en el campo de la sismología.

La detección temprana y la clasificación precisa de eventos sísmicos han adquirido una importancia vital en el campo de la sismología, contribuyendo significativamente a mejorar la gestión de desastres naturales. Este proyecto se propone explorar la adaptación del modelo de aprendizaje automático auto supervisado Wav2Vec 2.0 al ámbito de la sismología. La premisa subyacente es que este modelo tiene el potencial de reconocer patrones sonoros característicos de eventos sísmicos, aprovechando su capacidad previamente probada en el reconocimiento de voz. Se considera que analizar y ajustar este modelo en el contexto de la clasificación de eventos sísmicos podría abrir nuevas perspectivas para mejorar significativamente esta tarea.

Este proyecto no solo responde a la imperiosa necesidad de encontrar mejores modelos para el reconocimiento de eventos sísmicos, crucial en la gestión de desastres naturales de este tipo, sino que también busca explorar el potencial de las nuevas tecnologías como un enfoque revolucionario en la realización de tareas de esta índole. En última instancia, su objetivo es

contribuir al avance en la sismología y a la capacidad de anticipar y responder eficazmente a eventos sísmicos, salvaguardando vidas y propiedades.

2. Materiales y Metodología

Se propuso un modelo de aprendizaje automático que permite la clasificación de señales sísmicas recolectadas por un sismógrafo. Esto se logra mediante un preprocesamiento de los datos que los convierte en representaciones de audio, seguido por la generación del nuevo modelo mediante el fine tuning del modelo preentrenado Wav2Vec2.0 (Meta, 2020). Este último fue desarrollado inicialmente para el Procesamiento del Lenguaje Natural (NLP por sus siglas en inglés), pero se adaptó para realizar la tarea de clasificación sobre los audios generados durante el preprocesamiento de los datos recolectados por los sismógrafos. Durante el desarrollo del proyecto, tanto para el preprocesamiento de datos como para la generación del modelo, se utilizó el lenguaje de programación Python. Las siguientes secciones proporcionarán más detalles acerca del dataset utilizado, el preprocesamiento de los datos y el modelo propuesto para llevar a cabo la tarea especificada en los objetivos de este proyecto.

2.1. Dataset

Para el desarrollo del proyecto, se utilizaron dos conjuntos de datos pertenecientes al repositorio ESeismic, que contiene eventos sísmicos registrados en las faldas del volcán Cotapaxi (Pérez et al., 2020a). El primer conjunto de datos, *SeisBenchV1*, se empleó tanto para entrenar el modelo de clasificación de eventos sísmicos, así como conjunto de prueba durante el entrenamiento. Para ambos conjuntos se realizó el procesamiento de las señales, el remuestreo de los datos y generación de los audios que serán usados en el training y test sobre el modelo de aprendizaje automático. A continuación, se explica de manera detalla el preprocesamiento de los datos sobre cada uno de los dataset usados en este proyecto.

Para el primer conjunto de datos de *SeisBenchV1*, se realizó inicialmente un cambio de formato para la lectura de los datos. Cada uno de los datos recolectados de los sismógrafos se guardó en archivos de Matlab con la extensión *.mat*. La estructura de datos, que contenía la información de los eventos sísmicos, consistía en una matriz de diccionarios y arreglos anidados que se registraban por la estación que recolectó los datos, el canal en el cual se registraron los datos y los propios datos del evento registrado, tal como se muestra en el ejemplo de la Figura 1.

Figura 1. Ejemplo archivo .mat de los datos recolectados.

5 BNAS Min NaN Max NaN Mean NaN Std Dev NaN Missing 235 Class table	6 BREF Min NaN Max NaN Mean NaN Std Dev NaN Missing 238 Class table	7 BTAM Min NaN Max NaN Mean NaN Std Dev NaN Missing 235 Class table	8 BVC2 Min NaN Max NaN Mean NaN Std Dev NaN Missing 244 Class table	9 Voting_Detection Min 28 Max 82 Mean 55 Unique 2 Missing 0 Class categorical	10 Event_Identifier Min 1 Max 1 Mean 1 Unique 110 Missing 0 Class cell
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'f5111a3507f03815692478db0cf8e51e'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'ac4871bb62f79aed4c66f6086e63504d'
1x7 table	1x7 table	1x7 table	1x7 table	Confirmed	'7e7ea43c735bbfe9e9c1fe73ed97a7e3'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'dfb7b835bca1f713193f426ed4858f7a'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'cc6c980a6bf65f8340d07ec2a3e59f14'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'e4068f31b41b9be40e60332355bdaa4c'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'97c779b3bb628c21eaeae2b8bba31c57'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'93412b65ce389cdbb59a33d264e5eaac'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'6a07ec8281f0b298ef7f530f16665c30'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'417611f5c0f61b9c8ac1166f86cd45af'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'29d26044b7c57ffea877f08d3a0b5fd4'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'508ddfc088c06502d2216cc5d68e989c'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'e3a5617cbb9ec799dca07b606fc9aa89'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'b8bb616dcec433bdbe7d817dc1195804'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'6bed9247020728899a39af3a86cb1631'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'c1cd5bac40598e4162aeccc5c4e44d30'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'7a192ed803ebc1d11a0b608743646951'
1x7 table	1x7 table	1x7 table	1x7 table	Confirmed	'd1c806c6366f131440238e5594abb081'
1x7 table	1x7 table	1x7 table	1x7 table	Discarded	'4d8390a0e507bec95158319c60f951ed'

El proceso inicial implicó la conversión de los archivos .mat a archivos .json, que eran más compatibles con el lenguaje de programación utilizado en este proyecto, en este caso, Python. Posteriormente, se diseñó una estructura de datos que facilitaba el manejo de la información. Esta estructura se basaba en una matriz con diccionarios clave-valor, donde se almacenaban los eventos junto con detalles sobre la estación de registro, el canal de los datos, los

datos recolectados en ese canal y se agregaba una nueva clave que contenía el nombre del archivo *.wav* que se generaría a partir de ese evento, como se muestra en la *Figura 2.*

Figura 2. Ejemplos archivo .json generado para la nueva base de datos

	WAV	Event_Identifier	Station	Channel	Data	StartPoint	EndPoint
0	7e7ea43c735bbfe9e9c1fe73ed97a7e3_BHZ_BREF_EXPL...	7e7ea43c735bbfe9e9c1fe73ed97a7e3	BREF	BHZ	[-64, 127, -144, -150, -190, -166, 285, 39, -9...	686	1744
1	d1c806c6366f131440238e5594abb081_BHZ_BREF_EXPL...	d1c806c6366f131440238e5594abb081	BREF	BHZ	[-228, -224, -282, -205, -128, -232, -169, -14...	749	1743
2	06248a480e26c5bca9b7aa4e57d91402_BHZ_BREF_EXPL...	06248a480e26c5bca9b7aa4e57d91402	BREF	BHZ	[-251, -282, -280, -214, -234, -179, -231, -23...	810	1845
3	60866d09059d17568777c8469e79c68b_BHZ_BREF_EXPL...	60866d09059d17568777c8469e79c68b	BREF	BHZ	[-201, -163, -177, -177, -161, -189, -186, -18...	799	1868
4	80b8a62a7ee952159dfd1db907f7c26e_BHZ_BREF_EXPL...	80b8a62a7ee952159dfd1db907f7c26e	BREF	BHZ	[-286, -242, -244, -265, -244, -309, -328, -26...	701	1779
...	...	...	...	...	...	...	...
247	737253a9183ec4bb53a53081bc7351ec_BHE_BTAM_EXPL...	737253a9183ec4bb53a53081bc7351ec	BTAM	BHE	[420, 440, 428, 404, 399, 387, 369, 369, 382, ...	924.0	2116.0
248	c92208e1b6581443b5d10673009ca17e_BHE_BTAM_EXPL...	c92208e1b6581443b5d10673009ca17e	BTAM	BHE	[150, 161, 157, 158, 181, 182, 177, 172, 167, ...	519.0	1551.0
249	d4cf4eb102f4f87d8fe99e521fcc85d0_BHE_BTAM_EXPL...	d4cf4eb102f4f87d8fe99e521fcc85d0	BTAM	BHE	[343, 355, 350, 345, 363, 412, 474, 484, 452, ...	1040.0	2180.0

Cabe destacar que solo se conservaron los eventos confirmados por expertos, es decir, aquellos eventos sísmicos que fueron validados. Esto se determinó mediante una etiqueta en el dataset original del SeisBenchV1, que confirmaba si el evento había sido validado o descartado (ver Figura 1). De esta manera, el dataset para la generación de audio se redujo significativamente, creando un conjunto de datos más confiable del cual se pudieron generar audios representativos de los eventos sísmicos registrados.

2.2 Generación de los audios de los eventos sísmicos

Una vez generadas las bases de datos para cada uno de los eventos recolectados, se siguieron varios pasos para preparar los audios que se utilizarían en el modelo de entrenamiento.

En primer lugar, se llevó a cabo un análisis de los datos recolectados, para determinar la tasa de muestreo sobre las estaciones y los canales que registraron las lecturas. Inicialmente, varios de estos datos se registraron a 50Hz y se remuestrearon a 100Hz para obtener una representación más detallada de la señal y para estandarizar la tasa de muestreo en todos los datos.

$$\textit{(Ecuación 1)} \quad normalized\_data = \frac{data}{max(abs(data))}$$

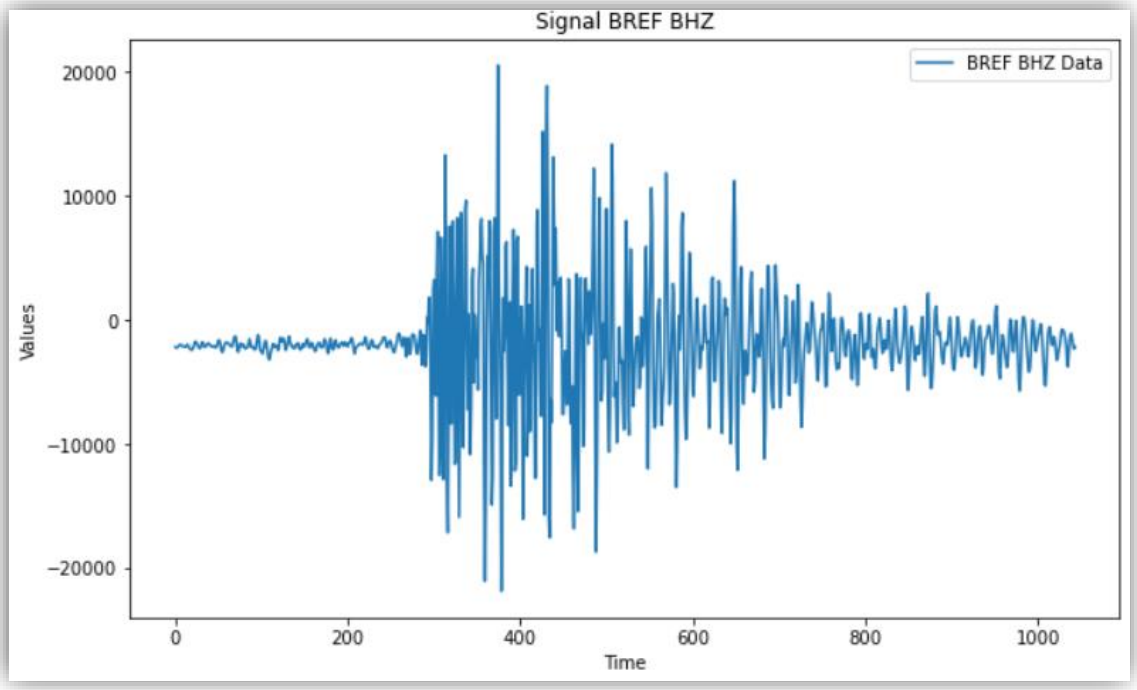


Figura 3. Gráfico de la señal de onda de una muestra tomada por la estación BREF en el canal BHZ

Con los datos remuestrados, se ajustó la frecuencia de las muestras a 16000Hz. Posteriormente, se normalizaron los datos utilizando la *Ecuación 1*, lo que implicó escalar los valores para que quedaran dentro del rango de -1 y 1. Este proceso de normalización aseguró que los datos estuvieran en una escala consistente, lo que es esencial para el entrenamiento efectivo de modelos de aprendizaje automático. La normalización permitió mantener la forma de onda de la señal sin alteraciones, como se muestra en la *Figura 3* y *Figura 4*.

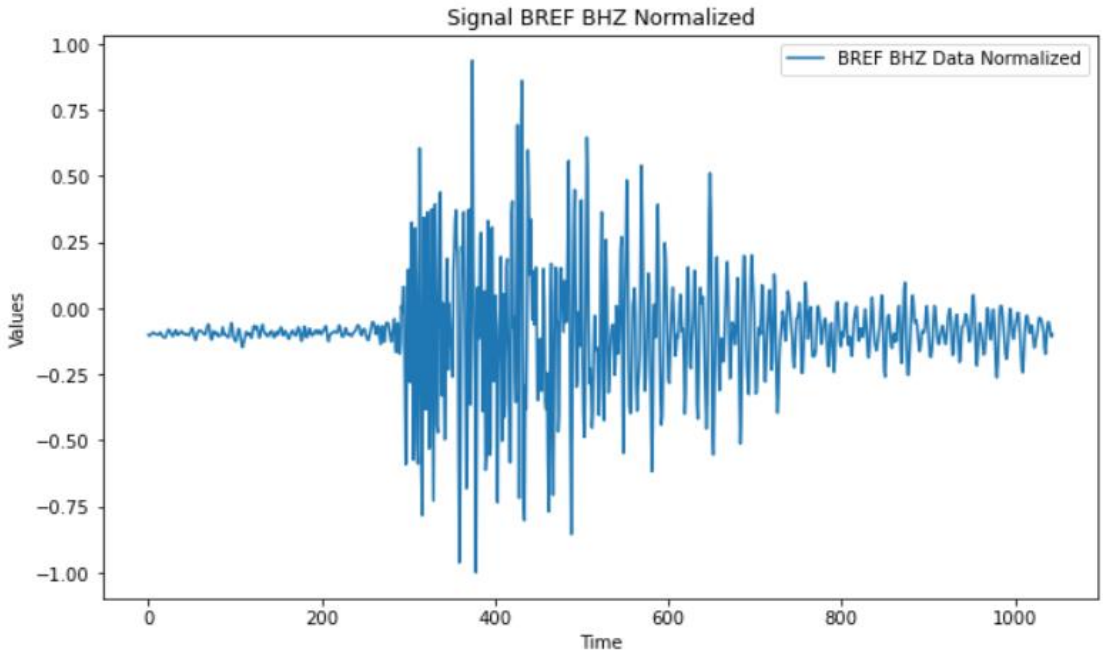


Figura 4. Gráfica de la normalización de la señal de onda de la muestra tomada por la estación BREF y el canal BHZ.

Con estos datos procesados y normalizados, se generaron los archivos de audio en formato *.wav* que serían utilizados en el entrenamiento del modelo de aprendizaje automático. Cada archivo de audio tiene una duración de no más de 2 segundos, pero es representativo del tipo de evento sísmico registrado. Estos archivos de audio, estandarizados y detallados, proporcionaron una base sólida para el entrenamiento del modelo de clasificación de eventos sísmicos.

2.3. Preprocesamiento de los datos

Para el preprocesamiento de los datos de entrenamiento, se desarrollaron diversas clases encargadas de cargar los audios y transformarlos en el tipo de dato correcto aceptado por el modelo de entrenamiento preentrenado que se estaba utilizando para las entradas de datos. Posteriormente, se llevó a cabo un análisis de la distribución de las etiquetas de las clases en todo el conjunto de audios generados del dataset SeisBenchV1, cuyos resultados se presentan en la *Tabla 1*.

Tabla 1. Distribución de las clases de eventos sísmicos

Clase (etiqueta) de Evento sísmico	Número de muestras
LP	35054
VT	23599
HB	1537
TRESP	387
EXPL	245
TRBA	14

Este análisis reveló un desbalance significativo en las clases de eventos sísmicos registrados en el dataset utilizado. Esta disparidad plantea un problema, ya que un conjunto de datos desbalanceado puede conducir al sobreajuste (overfitting) durante el entrenamiento. En otras palabras, el modelo se ajusta demasiado al conjunto de datos de entrenamiento, lo que limita su capacidad para generalizar y puede ocasionar fallos cuando se evalúa con nuevos datos durante las pruebas del modelo (Ying, 2019).

Con esta consideración en mente, se decidió utilizar únicamente los datos de los eventos sísmicos que presentaban una diferencia pequeña con respecto al conjunto de datos proporcionado. Además, se buscó la opinión de un experto en el tema acerca de los canales de las muestras. Tras la consulta, se determinó que las señales sísmicas más seguras y confiables provenían del canal BHZ. Esto resultó en la selección de solo dos clases (eventos sísmicos) para el entrenamiento, el evento sísmico “LT” y “VT”, evitando así el problema del sobreajuste (overfitting). Se utilizaron los audios generados a partir de todas las estaciones, pero los datos se tomaron exclusivamente del canal BHZ, reduciendo así el conjunto de entrenamiento que se emplearía. La distribución de clases utilizada para el entrenamiento se presenta en la *Tabla 2.*

Distribución de las clases a usar en el entrenamiento

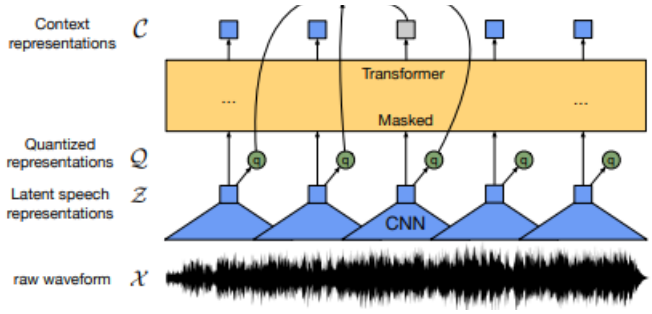
Tabla 2. Distribución de las clases a usar en el entrenamiento

Clase (etiqueta) de Evento sísmico	Número de muestras
LP	11926
VT	8533

2.4. Modelo Preentrenado Wav2Vec2.0

El modelo Wav2Vec2.0, desarrollado por Baevski et al. en 2020, es una arquitectura de aprendizaje profundo diseñada para tareas de procesamiento de audio, con un enfoque particular en la transcripción automática y la representación de características de audio. A diferencia de muchos enfoques previos, Wav2Vec2.0 se destaca por su capacidad para aprender representaciones robustas directamente de datos no etiquetados.

Figura 5 Representación general del modelo Wav2Vec2 previo al fine tune



Note: Tomado de wav2vev 2.0: A Framework for Self-Supervised Learning of Speech Representations (p. 2) por Baevski, et al., 2020.

El proceso de entrenamiento de Wav2Vec2.0 se basa en un enfoque auto supervisado. Inicialmente, el modelo se entrena para predecir muestras futuras en un espectrograma de audio a partir de muestras pasadas. Esto se logra mediante la maximización de la similitud coseno entre las representaciones latentes aprendidas por el modelo y las representaciones del espectrograma de audio. Además, se compone de un codificador de características convolucionales multicapa de

$f: X \rightarrow Z$, donde X representa el audio no procesado (*raw audio*) y Z es la salida del modelo, la cuales son representaciones latentes de los audios (Baevski, et al., 2020).

Posteriormente, el modelo se somete a un ajuste fino utilizando un conjunto de datos etiquetado para una tarea específica, como la clasificación de audio. Este proceso de ajuste fino adapta las representaciones latentes aprendidas durante la fase auto supervisada para ser más específicas y útiles para la tarea de interés, en este caso, para la clasificación de señales sísmicas.

2.5. Clasificador

Para el conjunto de datos de audios etiquetados, derivado de la recopilación en la base de datos *SeisBenchV1*, se llevó a cabo el ajuste fino del modelo Wav2Vec2.0 con el propósito de abordar la tarea de clasificación de eventos sísmicos.

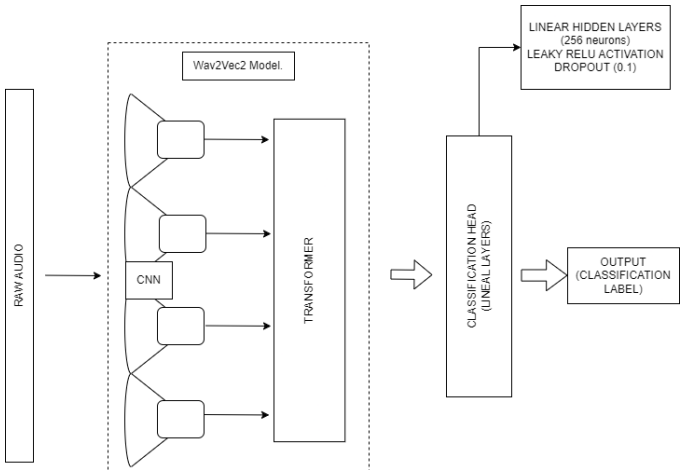


Figura 6. Fine Tune del Modelo Wav2Vec 2.0 para la tarea de clasificación.

La arquitectura del modelo consta de dos partes fundamentales. En primer lugar, se emplea la arquitectura preentrenada de Wav2Vec 2.0 para extraer las características de los *raw audios* generados a partir del conjunto entrenable de datos. Estos datos fueron ajustados a un *sample rate* de 16000Hz para garantizar una representación precisa del contenido acústico. La segunda parte del modelo comprende el entrenamiento supervisado con una primera

optimización de parámetros en *learning rate*, *weight decay y neuronas*. La cabeza de clasificación (*classification head*) se compone con dos capas ocultas con 256 neuronas con funciones de activación Leaky ReLU (Nielsen, 2015), seguidas por una capa de salida con un tamaño equivalente al número de etiquetas necesarias para la tarea de clasificación. En este caso, la capa de salida se dimensiona con dos neuronas, cada una representando la probabilidad asociada a una de las etiquetas de clasificación. Así, se genera un vector de probabilidades para cada etiqueta, lo que facilita la interpretación de la salida del modelo. El optimizador usado fue Adam (Tato, Nkambou, 2018) seteado en un *learning rate* de 1*e-5*, una regularización *L2* con un *weight decay* de 1*e-4* para la penalización de los pesos del conjunto de datos.

2.6. Configuración del conjunto de datos

Se utilizaron 14,206 eventos sísmicos, clasificados como del tipo LP y VT, provenientes de la estación BREF y el canal BHZ para el proceso de entrenamiento. Para este propósito, la técnica de Stratified KFold Cross Validation con cinco folds (5-SKCV) (Berrar, 2018) fue implementada, garantizando una distribución equilibrada de los datos de las diferentes etiquetas en todos los subconjuntos generados. En cada *fold* de entrenamiento se separó un 80% para formar los conjuntos de entrenamiento y 20% para validación.

Para el conjunto de prueba se destinó los índices generados por el SKCV específicamente para evaluar el rendimiento del modelo entrenado en la tarea de clasificación, utilizando la arquitectura Wav2Vec 2.0.

2.7. Métricas de Evaluación

Durante la fase de entrenamiento y prueba, se calcularon métricas clave para evaluar el rendimiento del modelo. Estas métricas incluyeron el Accuracy (ACC), el F1 Score y la pérdida (*CrossEntropyLoss*) en los conjuntos de datos de entrenamiento y validación obtenidos a través

de la estrategia de Stratified KFold Cross Validation (SKCV). Además, estas métricas se evaluaron en el conjunto de prueba para obtener una visión integral del desempeño del modelo en datos no vistos.

En particular, se generó la matriz de confusión a partir del conjunto de prueba, proporcionando una representación detallada de las predicciones del modelo. Adicional, se calculó el Área Bajo la Curva (Area Under Curve, AUC) mediante la curva de Características Operativas del Receptor (Receiver Operating Characteristics, ROC), ofreciendo una medida adicional de la capacidad del modelo para distinguir entre las clases de interés.

Estas métricas combinadas ofrecen una evaluación completa y detallada del rendimiento del modelo, abordando tanto la precisión global como la capacidad de discriminación en el conjunto de prueba, lo que proporciona información valiosa sobre la efectividad del modelo en la clasificación de eventos sísmicos, en este caso del tipo LP y VT.

También se utilizaron monitores sobre el valor del F1 score y loss sobre el conjunto de validación para generar callbacks que permitan detener el entrenamiento en caso de que se haya llegado a la convergencia del modelo antes de las épocas establecidas.

3. Resultados y Discusión

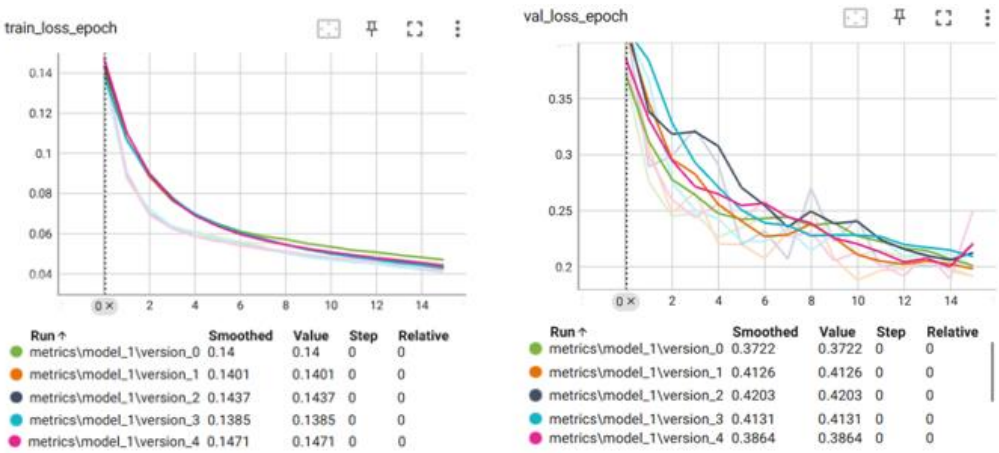


Figura 7. Gráficas de Loss vs Epochs para el train y validation.

En la *Figura 7* se muestran las curvas de *loss vs epochs* en los pasos de entrenamiento y validación. Esta gráfica nos brinda una herramienta visual para analizar si el entrenamiento está resultando en un sobreajuste (overfitting) en el conjunto de entrenamiento. En el lado izquierdo, se pueden observar las curvas correspondientes a los folds de entrenamiento, las cuales son suaves y disminuyen de manera constante hasta alcanzar una pérdida mínima de 0.04. Por otro lado, la gráfica del conjunto de validación, aunque no sigue la misma tendencia de disminución que las curvas de entrenamiento, alcanza un valor de pérdida de 0.2.

Aunque la diferencia entre las pérdidas en ambos conjuntos es de aproximadamente 0.16 puntos, las gráficas indican que el modelo generado evita el sobreajuste en los datos de entrenamiento. Este comportamiento es crucial, ya que sugiere que el modelo es capaz de generalizar de manera efectiva a nuevos conjuntos de datos, lo cual se refleja en su capacidad para clasificar de manera adecuada al realizar experimentos con cambios en los subsets de validación y entrenamiento.

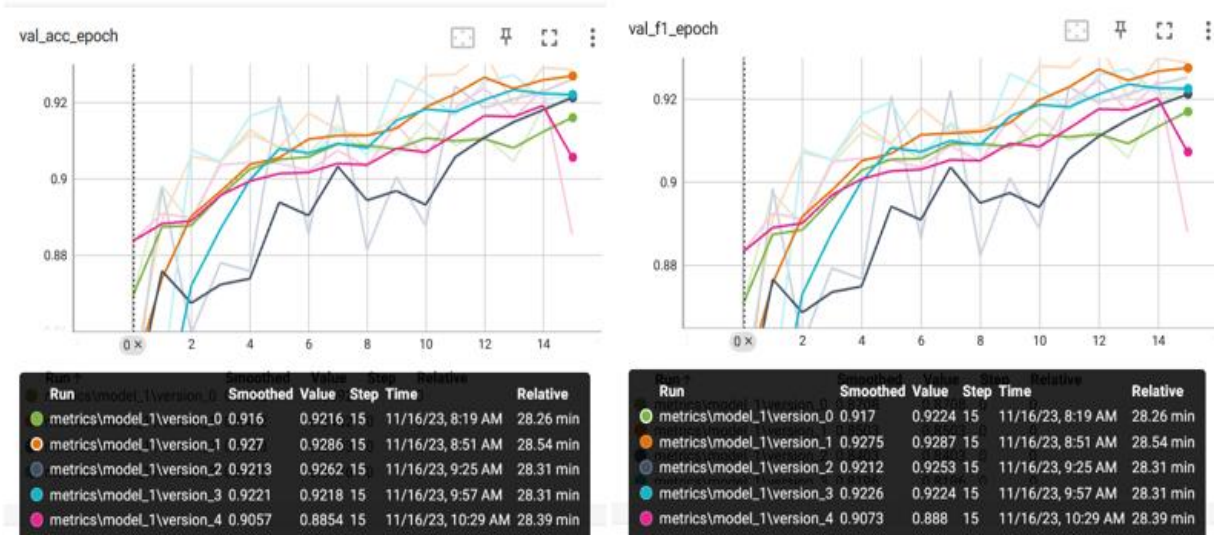


Figura 8. Resultados accuracy y f1 score para el conjunto de validación.

En la *Figura 8*, se presentan las gráficas de *accuracy vs epochs*, y *f1 vs epochs*, la cuales muestran como el modelo se fue mejorando la clasificación con el conjunto de entrenamiento sin llegar al overfitting sobre ese conjunto de datos. El accuracy y el F1 score en cada uno de los *folds* rondan alrededor de 0.921 y 0.928 en la clasificación sobre el conjunto de validación.

Tabla 3. Resultados de la evaluación de las métricas sobre cada fold de prueba.

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Accuracy	0.9288	0.9245	0.9238	0.9329	0.9292
F1 Score	0.9296	0.9244	0.9232	0.9329	0.9298
Loss	0.1889	0.2149	0.2128	0.2012	0.1898

La *Tabla 3* detalla los resultados de la evaluación de métricas (ACC, F1 y Loss) sobre cada subconjunto de prueba generado en los cinco folds proporcionados por el SKCV. Al aplicar el conjunto de prueba al modelo entrenado, se observa que las métricas no difieren significativamente de los valores obtenidos en el conjunto de validación. Esto sugiere que el modelo, entrenado con los hiperparámetros definidos, es capaz de realizar una clasificación adecuada en un nuevo conjunto de datos de entrada. Los consistentes resultados de accuracy, F1 score y pérdida en diferentes folds respaldan la afirmación de que el modelo puede generalizar efectivamente, manteniendo un rendimiento sólido en datos no vistos.

En la *Figura 9*, se muestra la matriz de confusión resultante del conjunto de prueba total obtenida de la concatenación de cada una de las predicciones generadas en los cada fold con su respectivo conjunto de prueba.

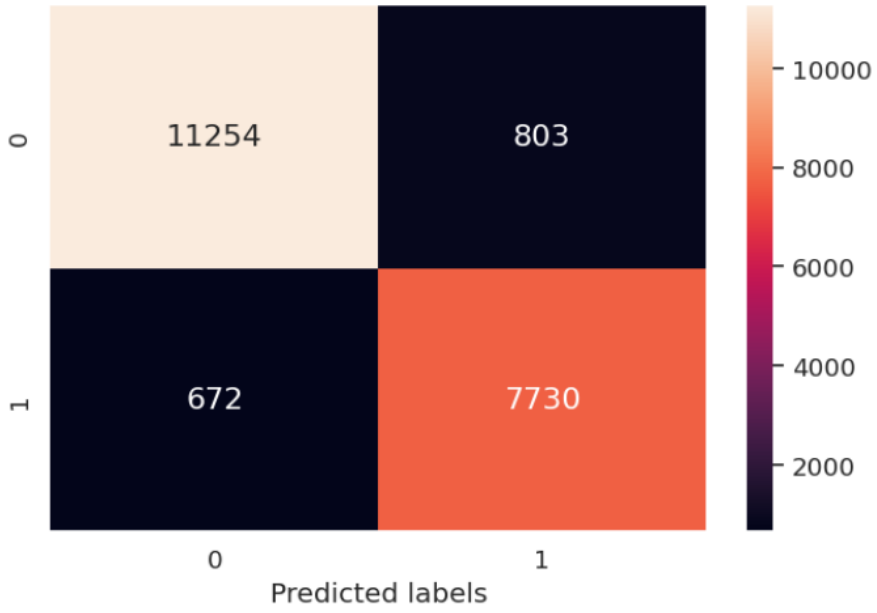


Figura 9. Matriz de Confusión generada a partir del conjunto de prueba.

Aquí, los valores en la diagonal principal representan los casos correctamente clasificados, mientras que los valores fuera de la diagonal indican los casos erróneamente clasificados. Se puede observar que el modelo clasifica positivamente para cada clase una cantidad considerable de muestras, 11254 muestras del evento LP se clasificaron correctamente de 11926 muestras, y 7730 muestras del evento VT se clasificaron como dicho evento de 8533 muestras. Aunque todavía existe un margen de mejora que permita redu cir a un más los datos clasificados incorrectamente como LP y VT.

Finalmente, la evaluación del modelo se complementa con el análisis de la curva ROC presentada en *Figura 10*, que presenta un Área Bajo la Curva (AUC) de 0.98. Esta puntuación en la curva ROC indica una capacidad del modelo para discriminar entre clas.s, evidenciando una alta tasa de verdaderos positivos y una baja tasa de falsos positivos en diversos umbrales de clasificación. Esto permite analizar con más profundidad al modelo para mejorar los resultados iniciales obtenidos para este tipo de tareas sobre el modelo Wav2Vec 2.0 .

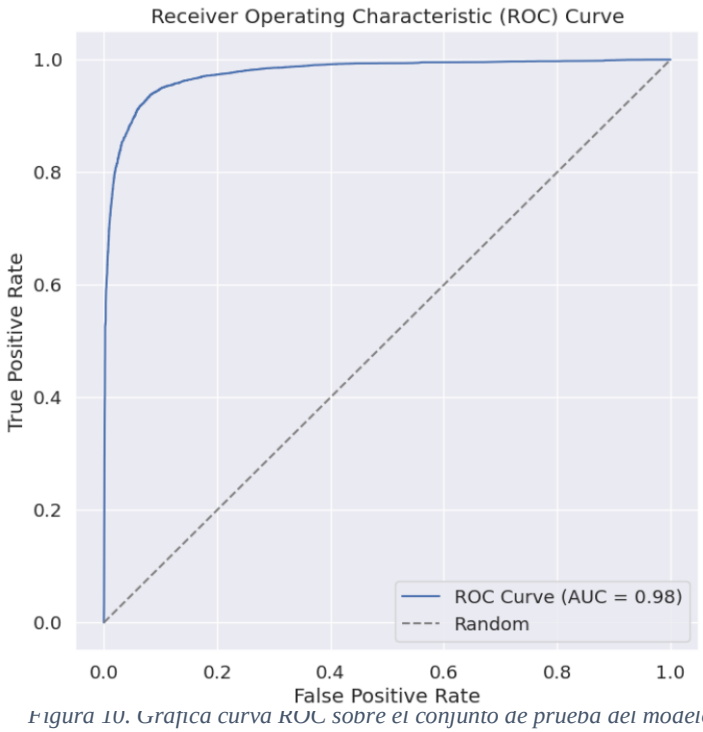


Figura 10. Gráfica curva ROC sobre el conjunto de prueba del modelo

4. Conclusiones

Este trabajo ha permitido ampliar el conocimiento en el ámbito de la sismología al abordar con éxito la aplicación de la tecnología de reconocimiento de voz Wav2Vec en la clasificación de eventos sísmicos. La utilización y adaptación exitosa de este modelo preentrenado ha revelado ser una estrategia viable para la detección y categorización de eventos sísmicos. A pesar de la existencia de enfoques alternativos, como el uso de Redes Neuronales Artificiales (ANN) según Scarpetta et al. (2005) o Procesos Gaussianos según Álvarez, et al. (2007), el objetivo fundamental de este proyecto fue ampliar los métodos de clasificación aplicados a eventos sísmicos mediante una arquitectura focalizada en el reconocimiento de características auditivas fundamentales. Desde una perspectiva de aprendizaje, esta iniciativa ha ofrecido una experiencia valiosa al aplicar técnicas de procesamiento de voz en un contexto no convencional, enriqueciendo así el conocimiento en el campo.

Durante la ejecución de este proyecto, nos enfrentamos a una dificultad sustancial relacionada con la escasa disponibilidad de datos etiquetados y validados en el contexto específico de eventos sísmicos. Esta limitación representó un obstáculo significativo en la creación de un conjunto de datos más extenso y diversificado, lo cual podría haber mejorado notablemente la capacidad de generalización del modelo. A pesar de este desafío, se adoptó una estrategia cuidadosa que implicó la selección meticulosa de eventos confirmados y la consulta con expertos en el campo. Estas medidas permitieron generar un conjunto de datos aceptable, proporcionando así al modelo las condiciones necesarias para realizar un análisis razonable en relación con la tarea de clasificación para la cual fue adaptado.

Para investigaciones futuras, se recomienda continuar explorando y refinando el modelo presentado en este trabajo con el objetivo de mejorar los resultados obtenidos hasta el momento. Además, se sugiere la investigación de técnicas de oversampling y data augmentation para abordar conjuntos de datos desbalanceados. Este enfoque podría permitir la incorporación de más clases de eventos sísmicos, lo cual tendría un impacto positivo en la capacidad de generalización del modelo. Estas estrategias podrían abrir nuevas oportunidades para fortalecer la robustez y la eficacia del modelo en la clasificación de una gama más amplia de eventos sísmicos, contribuyendo así al avance continuo en este campo de estudio.

En conclusión, este trabajo ha demostrado la viabilidad del modelo Wav2Vec 2.0, como tecnología de reconocimiento de voz, para adaptarse a una tarea de clasificación sobre eventos sísmicos. Los resultados obtenidos respaldan la utilidad de esta aproximación y abren la puerta a futuras investigaciones que exploren aún más la aplicación de modelos de procesamiento de voz en el campo de la sismología.

Bibliografía

Ajmain, E. (2021). Music Genre Classification with Wav2Vec2. Recuperado de https://www.kaggle.com/code/lujar1762/music-genre-classification-with-wav2vec2/notebook

Alvarez, M., Henao, R., Duque E. (2007). Clasificación de Eventos Sísmicos empleando Procesos Gaussianos. Scientia Et Technica, vol. XIII, núm. 35, pp. 145-150, Universidad Tecnológica de Pereira. Recuperado de: https://www.redalyc.org/pdf/849/84903527.pdf

Anónimo. (s.f.). Emotion Recognition in Greek Speech Using Wav2Vec 2.0. Recuperado de https://colab.research.google.com/github/m3hrdadfi/soxan/blob/main/notebooks/Emotion_recognition_in_Greek_speech_using_Wav2Vec2.ipynb#scrollTo=5fbCls1d2yBs

Baevski, A., et al. (2020). Wav2Vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. Recuperado de https://arxiv.org/pdf/1904.05862.pdf

Berrar D., (2018). Cross-validation. Data Science Laboratory, Tokyo Institute of technology. Recuperado de: https://www.researchgate.net/profile/Daniel-Berrar/publication/324701535_Cross-Validation/links/5cb4209c92851c8d22ec4349/Cross-Validation.pdf

Meta. (2020). Wav2vec 2.0: Learning the structure of speech from raw audio. Recuperado de https://ai.meta.com/blog/wav2vec-20-learning-the-structure-of-speech-from-raw-audio/

Nielsen, M. (2015). Neural networks and deep learning. Determination Press.

Peña, S., Perton, M., & Rodríguez, J. L. (2012). Auditory Seismic Signal Processing: A Review. IEEE Transactions on Geoscience and Remote Sensing, 50(12), 4813-4826.

Pérez, N., Benítez, D., Grijalva, F., Lara-Cueva, R., Ruiz, M., & Aguilar, J. (2020). Eseismic: Towards an ecuadorian volcano seismic repository. Journal of Volcanology and Geothermal Research.

Rosero, K. (s.f.). Sound classification and localization using transformers. Recuperado de https://colab.research.google.com/drive/1kGHg2YIfh0d_dvfiykq4b0ScO8zR7mMH

Scarpetta, S., Giudicepetro, F., Ezin, E., Petrosino, S., Pezzo, E., Martini, M., & Marinaro, M. (2005). Automatic Classification of Seismic Signals at Mt. Vesuvius, Italy, using Neural Networks. Bulletin of the Seismological Society of America, 95(1), 185-196.

Stangeland, T. (2021). Seismic Event Classification using Machine Learning.

Tato, A., Nkambou, R. (2018). Improving Adam Optimizer. Quebec, Canadá. Recuperado de: https://openreview.net/pdf?id=HJfpZq1DM

Tibi, R., Linville, L., Young, C., & Brogan, R. (2019). Classification of local seismic events in the Utah region: A comparison of amplitude ratio methods with a spectrogram-based machine learning approach. Bulletin of the Seismological Society of America, 109(6).

Ying, X. (2019, February). An overview of overfitting and its solutions. In *Journal of physics: Conference series* (Vol. 1168, p. 022022). IOP Publishing.