

UNIVERSIDAD DE LAS FUERZAS ARMADAS - “ESPE”

DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES

TEMA: “GENERACIÓN DE SEÑALES SINTÉTICAS DE EVENTOS SISMO-VOLCÁNICOS DEL VOLCÁN COTOPAXI A TRAVÉS DE UN MODELO DE RED NEURONAL ADVERSARIO GENERATIVO CONDICIONAL”

AUTORES:

GALARZA CRUZ, CHRISTIAN RAMIRO & VEGA VERGARA, JOSSELYN MICHELLE

Director del Proyecto: Ing. Román Lara, PhD.

Docente Evaluador: Ing. Julio Larco

Director de Carrera: Ing. Carlos Daniel Altamirano, PhD.

Secretaria Académica: Abg. María Fernanda Jaramillo

Julio, 2022



Introducción

Diseño e Implementación

Resultados

Conclusiones

Recomendaciones y Trabajos Futuros

Introducción

Trabajos Relacionados

WaveGAN

Arquitectura de red generativa contradictoria capaz de sintetizar audio. Se puede observar una amplia variedad de dominios de audio, como el habla humana, los sonidos de los pájaros y la música, con resultados bastante convincentes (Donahue et al., 2018).

GAN

Li et al. (2018), se utilizó por primera vez una GAN para extraer una representación compacta y eficaz de las formas de onda sísmicas. Se usaron 700,000 formas de onda recolectadas en el sur de California y Japón y se utilizó un clasificador *Random Forest* para determinar si las formas de onda son generadas por terremotos o fuentes de ruido locales.

Generación de señales sintéticas de eventos sismo-volcánicos del volcán Cotopaxi mediante el método de Bootstrap

Por medio de la extracción de muestras de las respuestas en frecuencia de magnitud y fase de los datos originales de las señales LP y VT, se pueden obtener fases y magnitudes sintéticas, aplicando filtros.

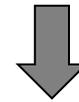
ESeismic-GAN

Modelo de Redes adversarias generativas convolucionales profundas (GAN, del inglés *Generative Adversarial Networks*). Los autores se enfocaron en crear respuestas de frecuencia de magnitud y combinarlas con respuestas de fase genuinas, para obtener nuevas señales volcánicas con la transformada inversa rápida de Fourier. (Grijalva et al., 2020)

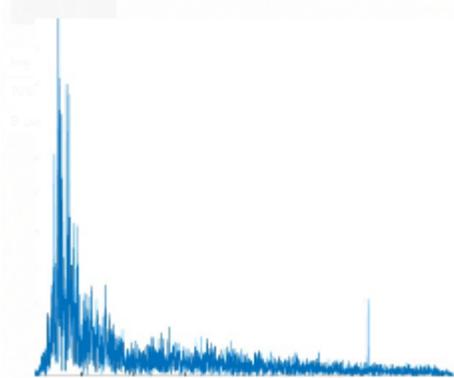
Introducción

Motivación

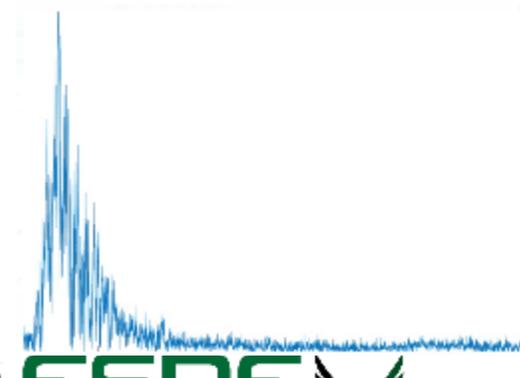
- Dado que cada señal debe interpretarse por separado, la tarea de etiquetar las actividades sismo-volcánicas, que es visual, requiere una cantidad considerable de capital humano y tiempo.
- La tarea de etiquetado es particularmente compleja y los expertos cambian su comprensión y análisis de las señales a lo largo del tiempo.
- Las bases de datos de señales sismo-volcánicas etiquetadas liberadas al público son escasas, debido a la falta de procesamiento y etiquetado de las mismas.



Dominio de la Frec. Real



Dominio de la Frec. Sintética



Objetivo General

Generar señales sintéticas de eventos sismo-volcánicos a través de un modelo de red neuronal generativo adversario condicional (CGAN).

Objetivos Específicos

- Estudiar sobre el estado del arte de la generación de señales sintéticas de una dimensión a través de una CGAN.
- Implementar una arquitectura de un modelo de red neuronal basada en un generador y discriminador.
- Entrenar el modelo con las bases de datos de eventos sismo-volcánicos proporcionados por el Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN).
- Monitorizar el proceso de entrenamiento de la CGAN, de forma que se pueda identificar y corregir posibles fallas de convergencia.
- Generar una base de datos de señales sintéticas de eventos sismo-volcánicos que se encuentren etiquetados.
- Evaluar la base de datos con los expertos del IGEPN.



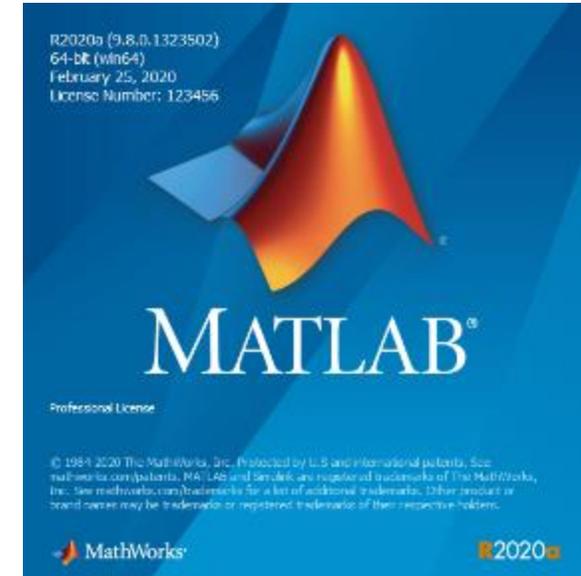
MSI GL63

- Intel Core I7 8th Gen (6 core, 12 threads, 2.2GHz - Turbo 4GHz)
- Nvidia GeForce GTX 1060 (6GB VRAM)
- 16 GB RAM
- 1TB SSD



DELL G3 3579

- Intel Core I7 8th Gen (6 core, 12 threads, 2.2GHz - Turbo 4GHz)
- Nvidia GeForce GTX 1050 Ti (4GB VRAM)
- 16 GB RAM
- 1TB HDD



MATLAB 2020a

- Deep Learning Toolbox
- Implementación de redes generativas.
- Facilidad en el manejo e implementación de capas.

Diseño e Implementación

Modelo generativo GAN



A medida que avanza el entrenamiento, el generador se acerca a producir resultados que pueden engañar al discriminador:



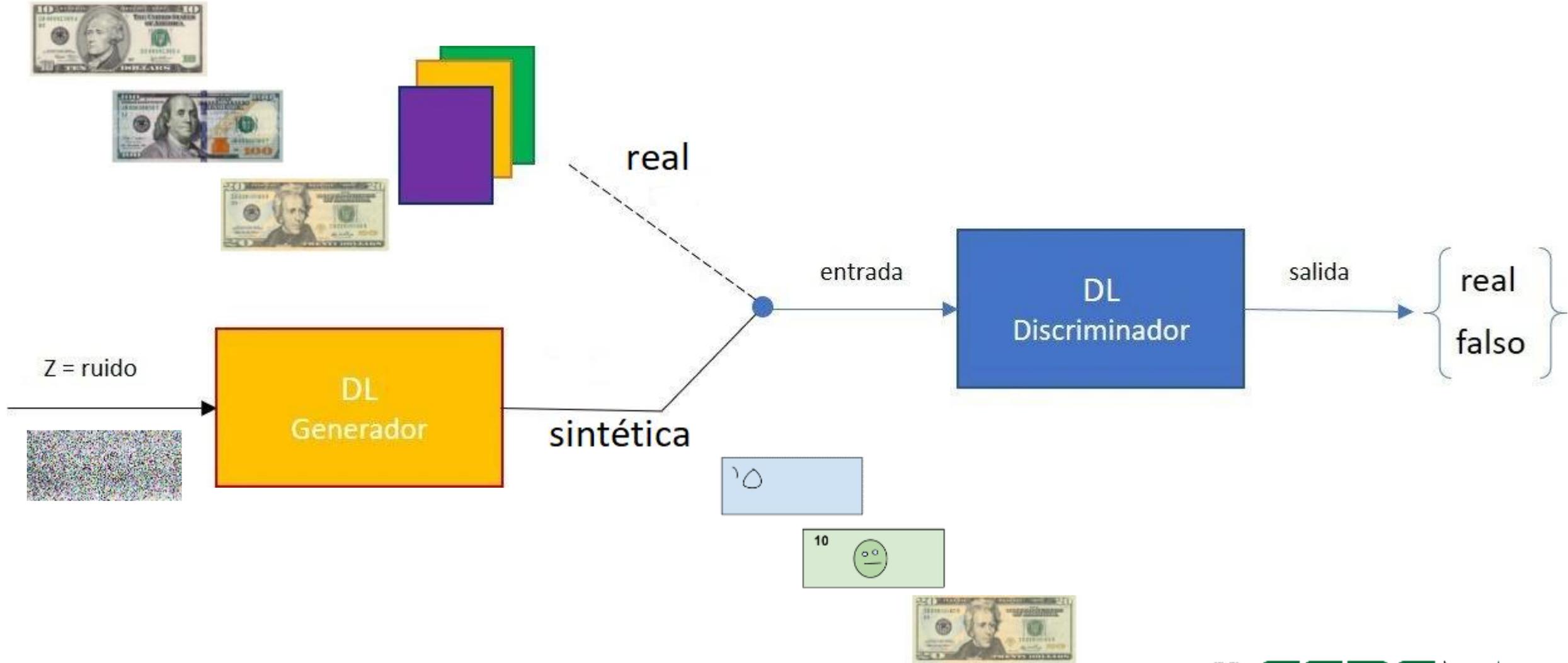
Finalmente, si el entrenamiento del generador va bien, el discriminador empeora al diferenciar entre lo real y lo falso. Comienza a clasificar los datos falsos como reales y su precisión disminuye.



Google Developers (s.f.).

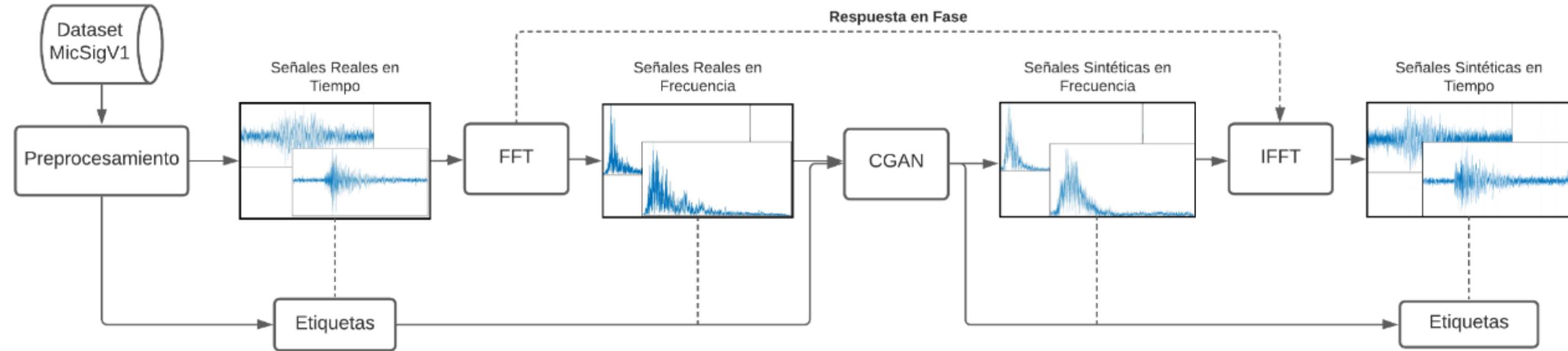
Diseño e Implementación

Modelo generativo GAN



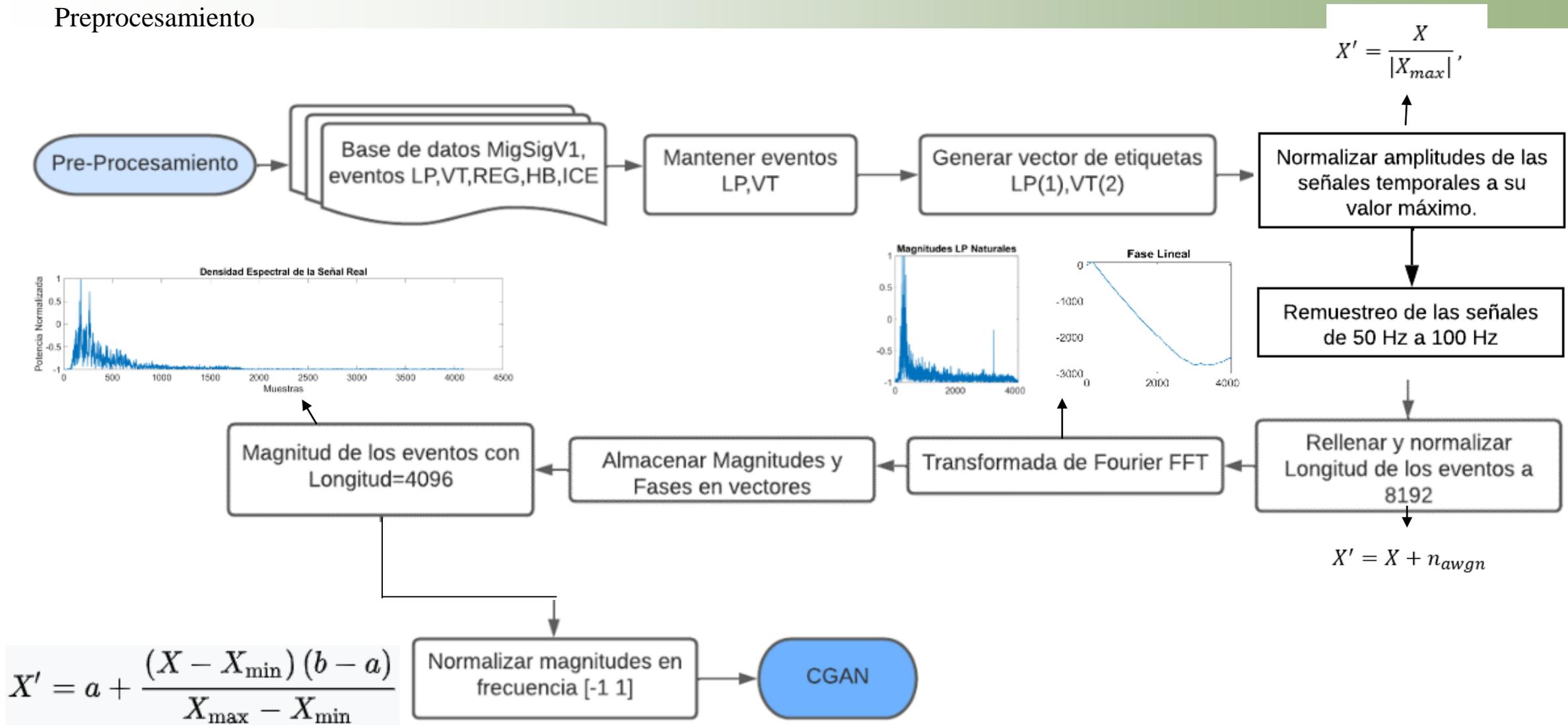
Diseño e Implementación

Diagrama de Bloques del Proyecto



Diseño e Implementación

Preprocesamiento

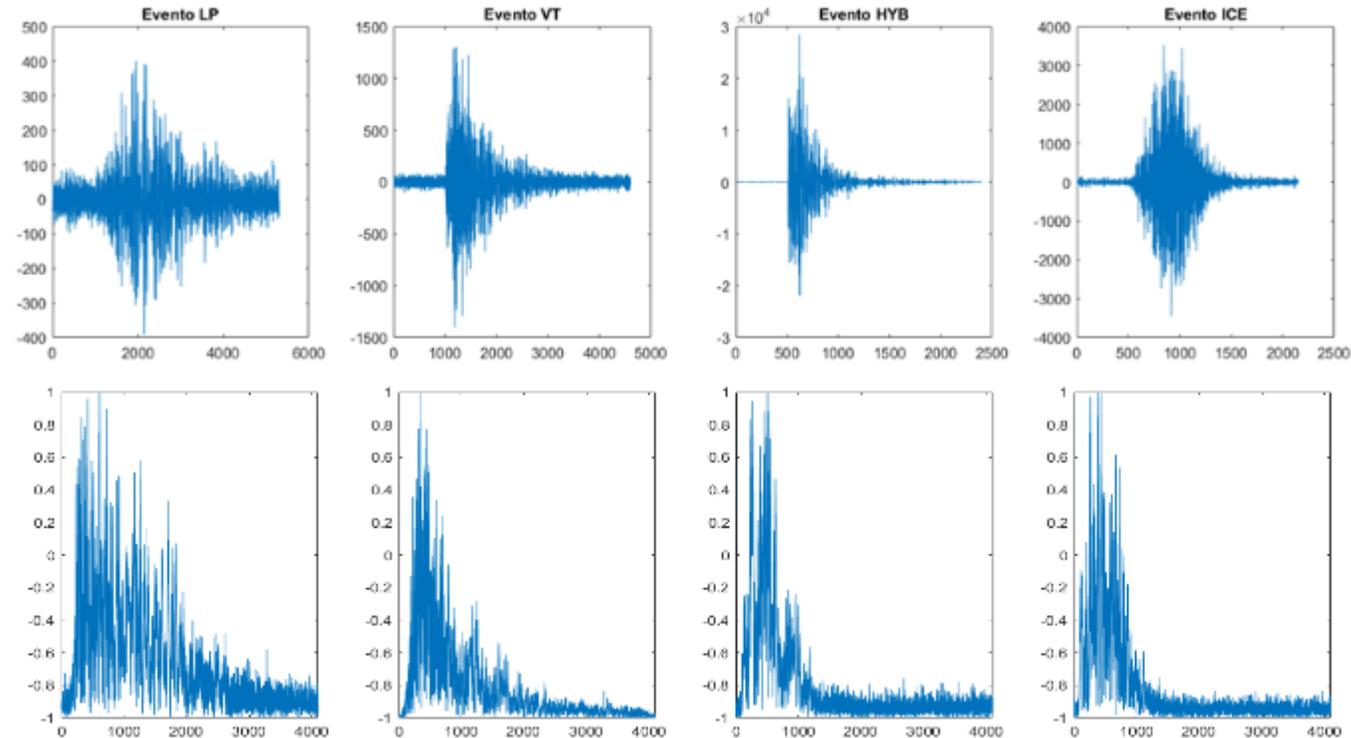


Dataset MicSigV1

- Conjunto de datos de señales sismo-volcánicas discretas del Volcán Cotopaxi por el IGEPN.
- 1187 eventos sísmicos.
 - **1044 LP**
 - **101 VT**
 - 27 REG, 8 HB y 7 ICE
- Los eventos LP y VT son los que poseen mayor ocurrencia frente a los otros.

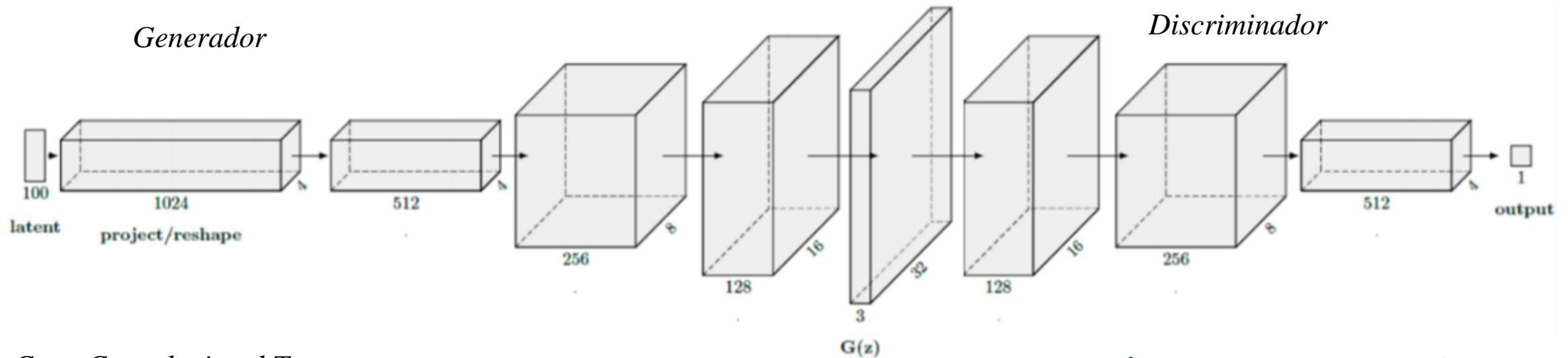


ESeismic Repository



Diseño e Implementación

Diagrama de Bloques de un modelo de CGAN



Capa Convolutiva Transpuesta

$$W_0 = (W_i - 1) \times s + k - 2p$$

Capa Convolutiva

$$W_0 = \frac{W_i - k + 2p}{s} + 1$$

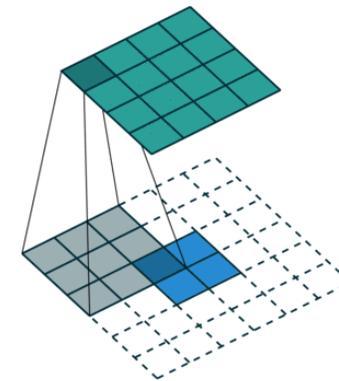
Entrada

$$W_i \times H_i \times C_i \xrightarrow{\text{Salida}} W_o \times H_o \times C_o$$

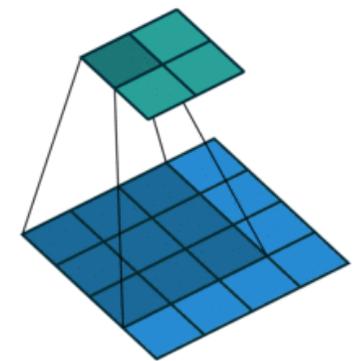
$s = \text{stride}$
 $k = \text{kernel}$
 $p = \text{padding}$
 $W_i = \text{ancho datos de entrada}$
 $W_o = \text{ancho datos de salida}$

$$W_o = 2^n$$

$C_o = \text{cantidad kernels propuestos}$
 en cada capa de convolución.



Convolución Transpuesta



Convolución

Diseño e Implementación

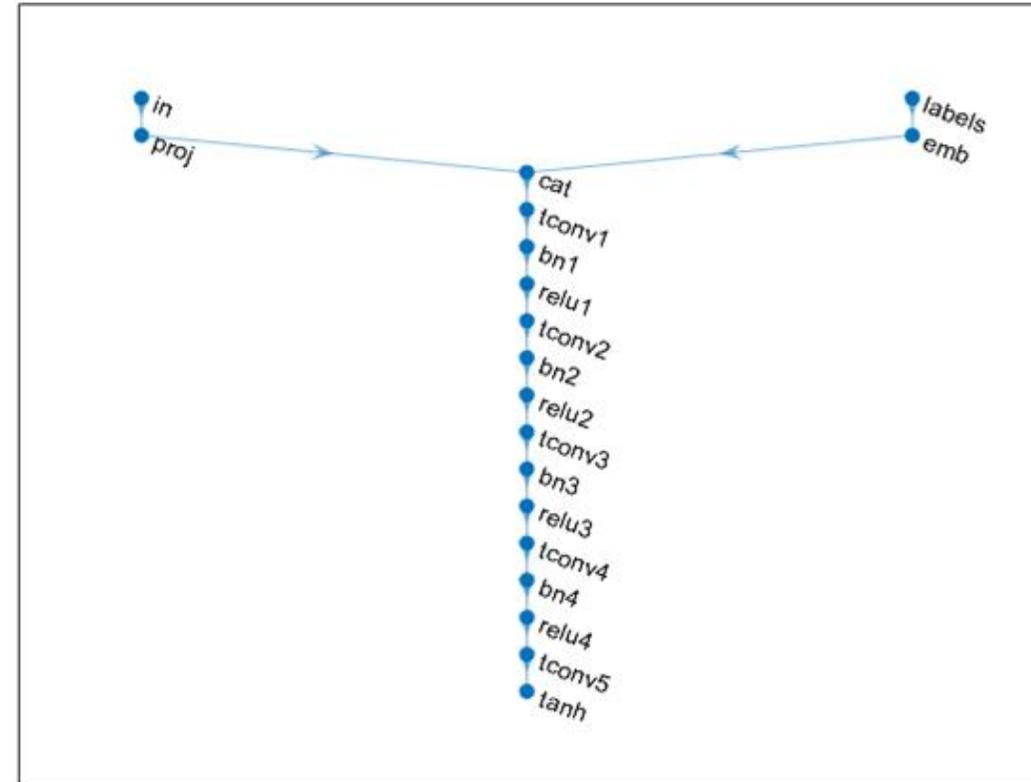
Capas del Modelo CGAN Final

Operación		Salida
Entrada z	100	
Etiquetas	(1x1x1)	
projectAndReshape	(4x1x1024)	(4x1x1025)
ReLU		
Trans Conv1D (Stride=1)	(16 * d, kernel_size = 13)	(16x1x1024)
batchNormalization		
ReLU		
Trans Conv1D (Stride=4)	(8 * d, kernel_size = 6)	(64x1x512)
batchNormalization		
ReLU		
Trans Conv1D (Stride=4)	(4 * d, kernel_size = 6)	(256x1x256)
batchNormalization		
ReLU		
Trans Conv1D (Stride=4)	(2 * d, kernel_size = 6)	(1024x1x128)
batchNormalization		
ReLU		
Trans Conv1D (Stride=4)	(1, kernel_size = 6)	(4096x1x1)
batchNormalization		
Tanh		

$d=64$, $padding=[1\ 0]$

Ultima funcion de activacion tanh [-1 1]

Estructura Generador



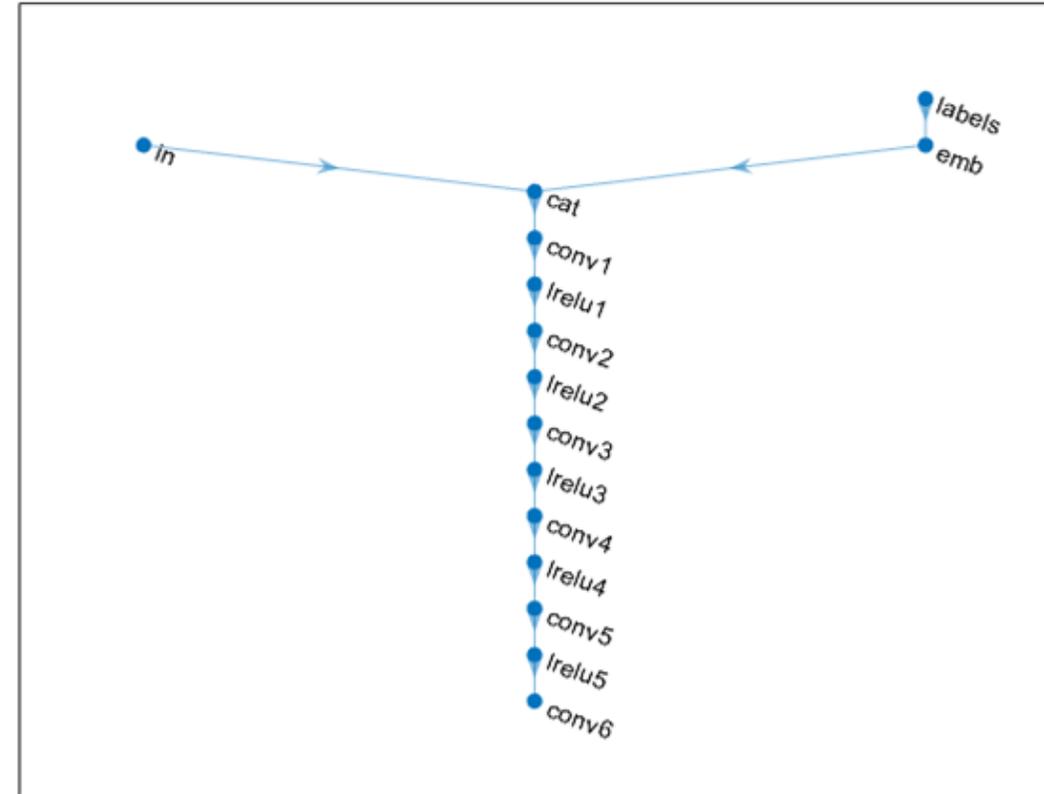
Diseño e Implementación

Capas del Modelo CGAN Final

Operación Entrada x G(z) etiquetas		Salida
	(4096x1x1) (1x1x1)	(4096x1x2)
Conv1D (Stride=4) LReLU ($\alpha = 0.2$)	(d, kernel_size = 13)	(1024x1x64)
Conv1D (Stride=4) LReLU ($\alpha = 0.2$)	(2 * d, kernel_size = 6)	(256x1x128)
Conv1D (Stride=4) LReLU ($\alpha = 0.2$)	(4 * d, kernel_size = 6)	(64x1x256)
Conv1D (Stride=4) LReLU ($\alpha = 0.2$)	(8 * d, kernel_size = 6)	(16x1x512)
Conv1D (Stride=4) LReLU ($\alpha = 0.2$)	(16, kernel_size = 6)	(4x1x1024)
Conv1D (Stride=1)	(1, kernel_size = 4)	(1x1x1)

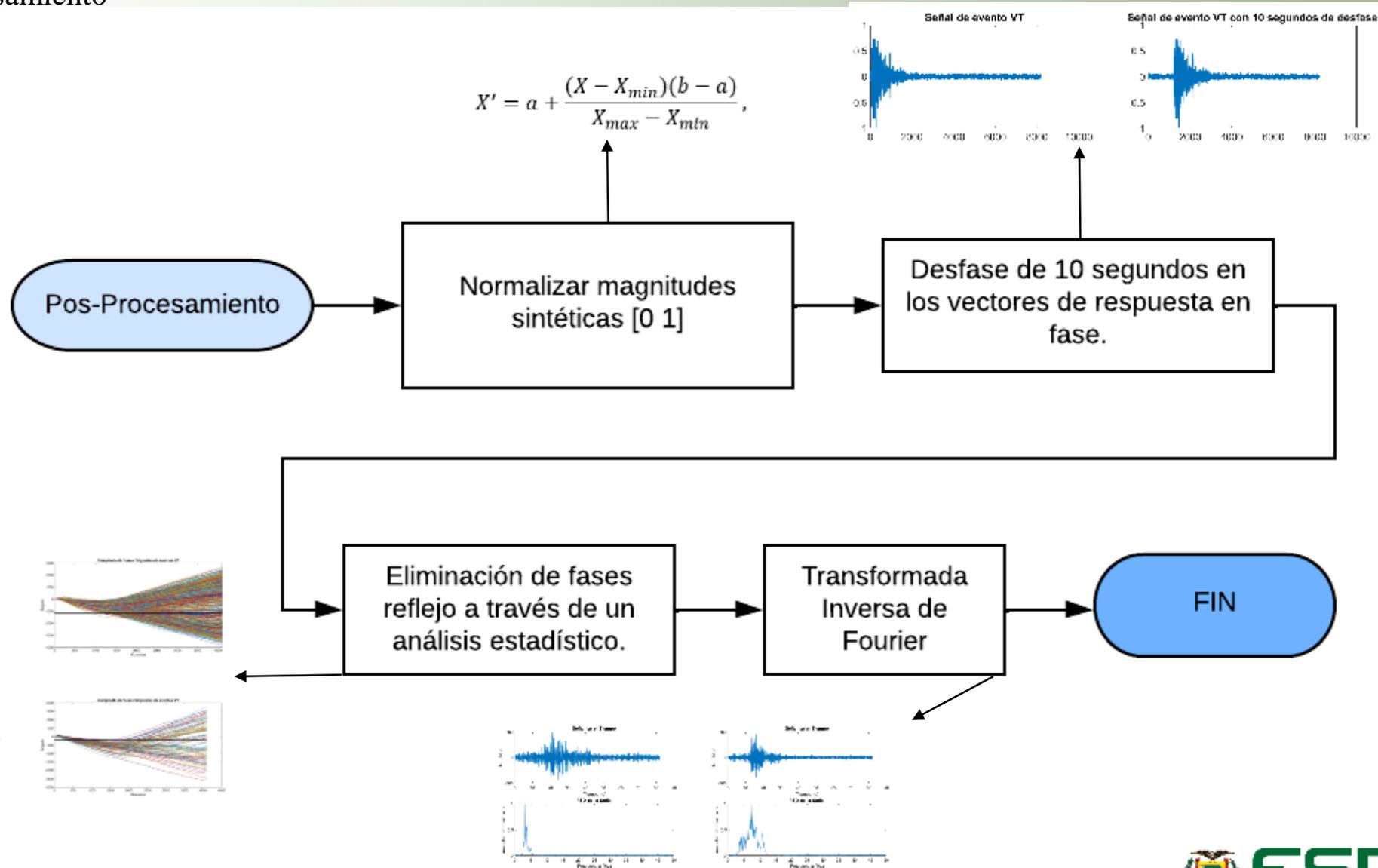
$d=64$, padding=[1 0]

Estructura Discriminador



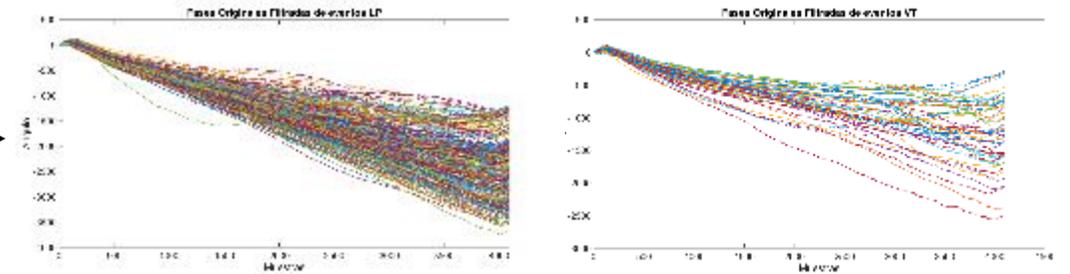
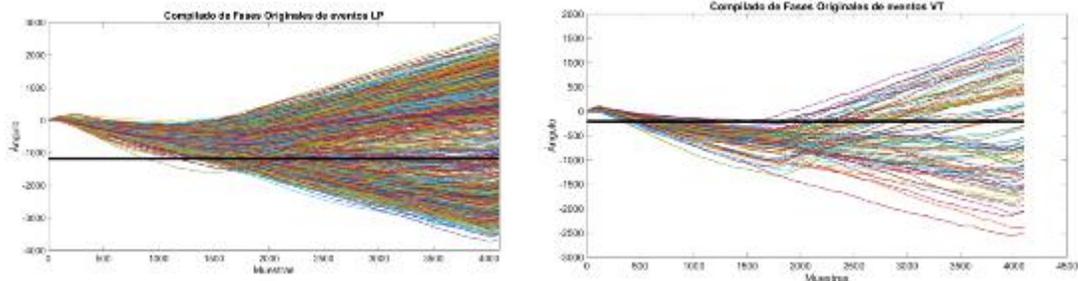
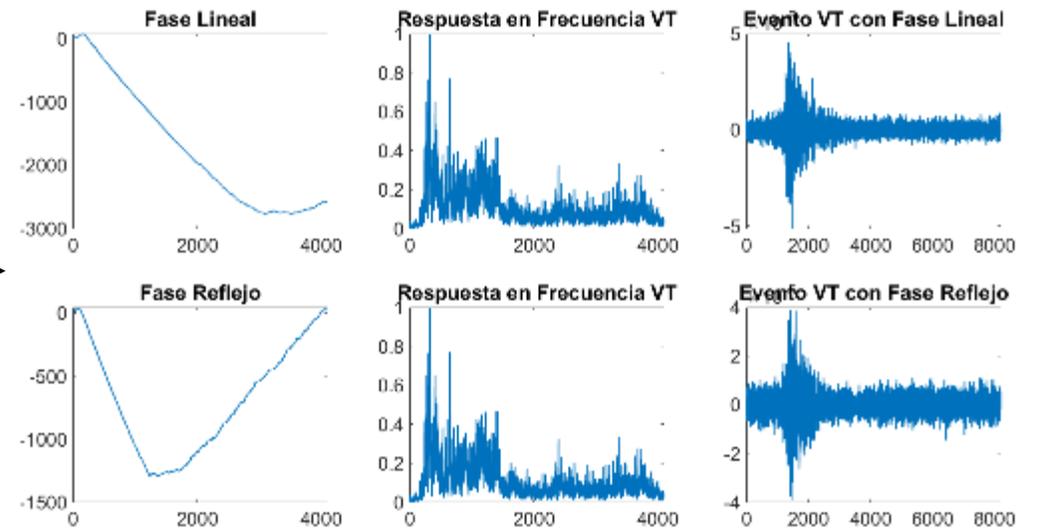
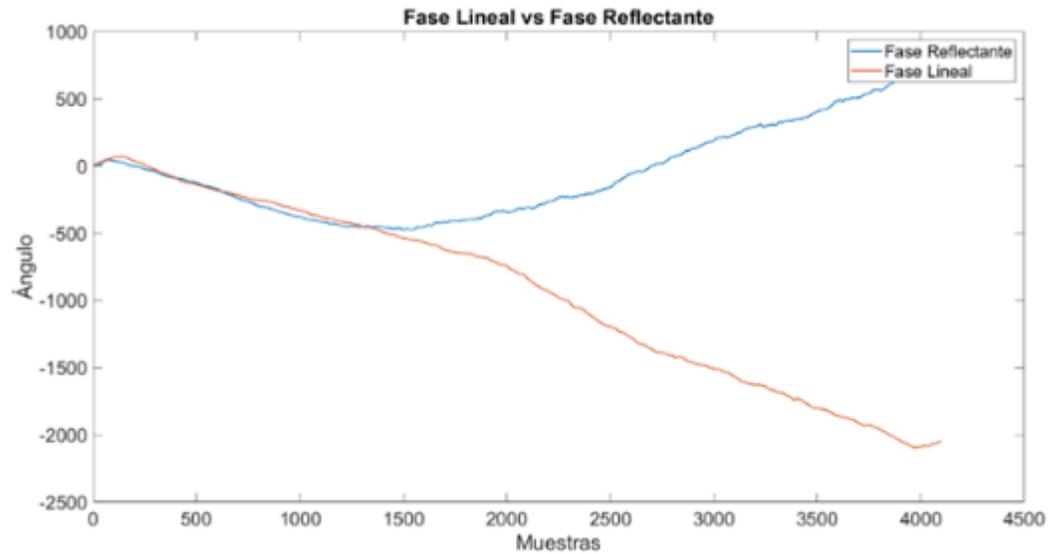
Diseño e Implementación

Posprocesamiento



Diseño e Implementación

IFFT: Magnitudes Sintéticas + Fases originales

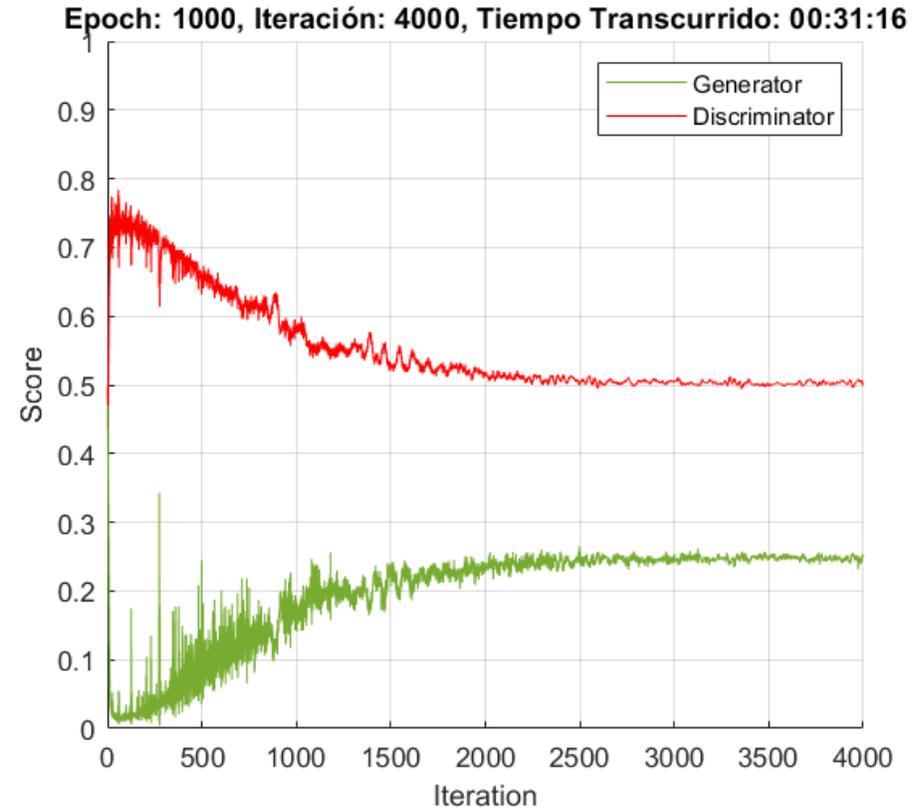
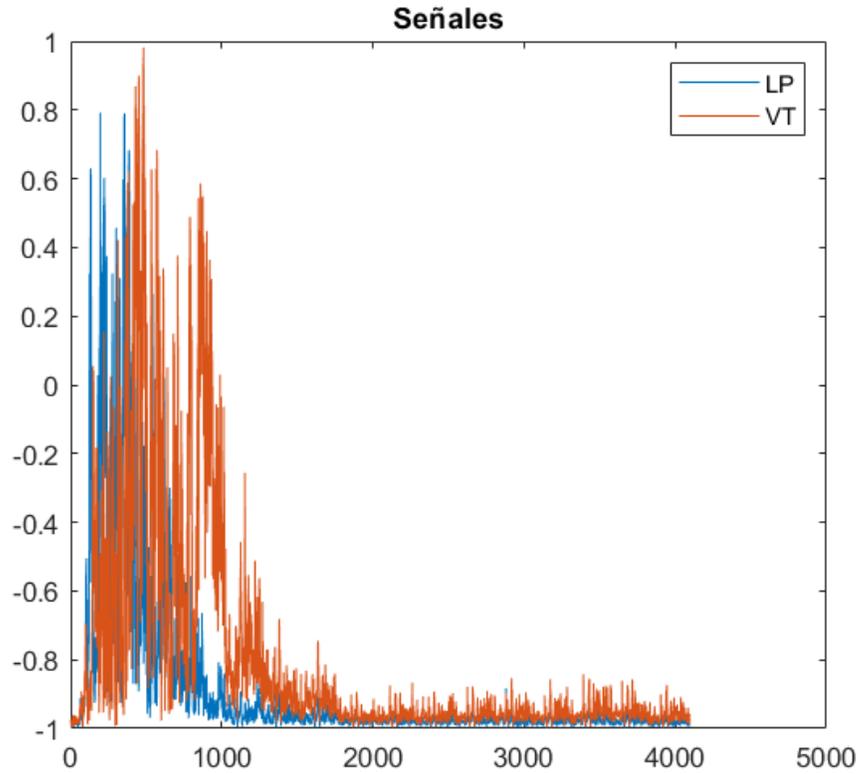


Fase LP=378

Fase VT=55

Resultados

Modelo Experimental Final

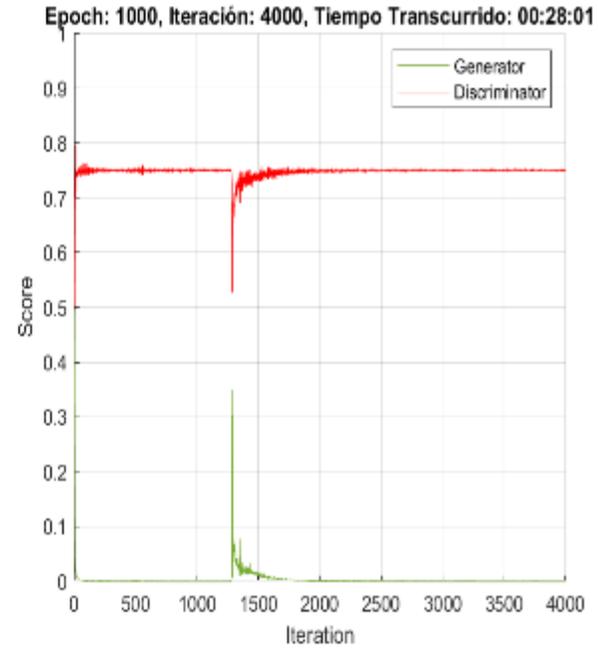
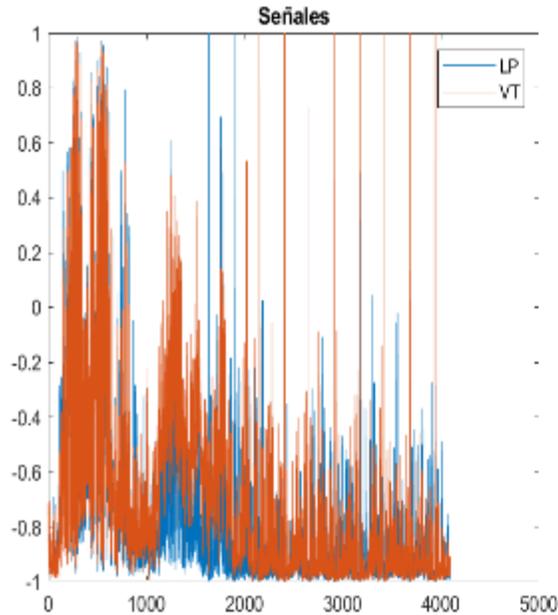


- Tiempo de entrenamiento 31 minutos.
- EPOCH=1000
- Iteraciones=4000
- Se entrenó con la GPU de los equipos.

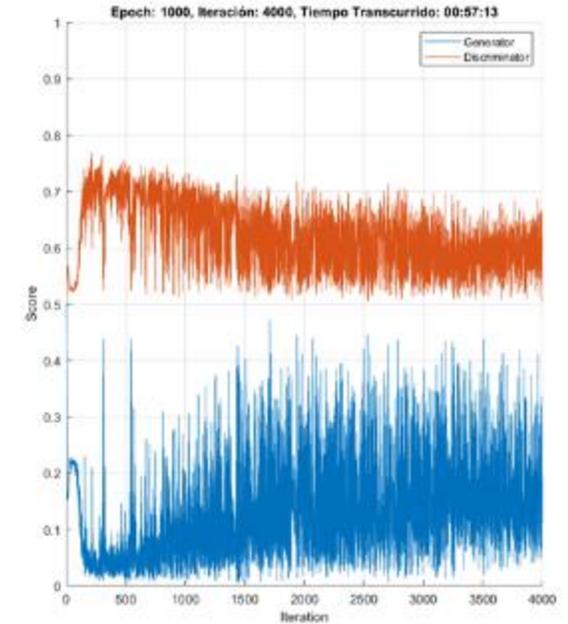
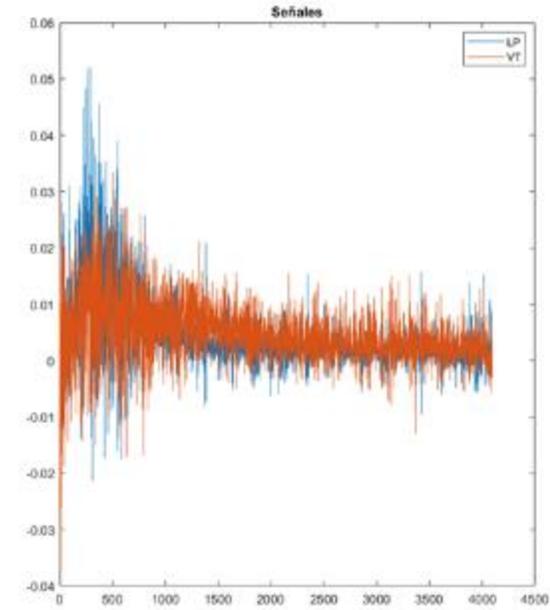
Resultados

Modelos con Fallos

No convergencia



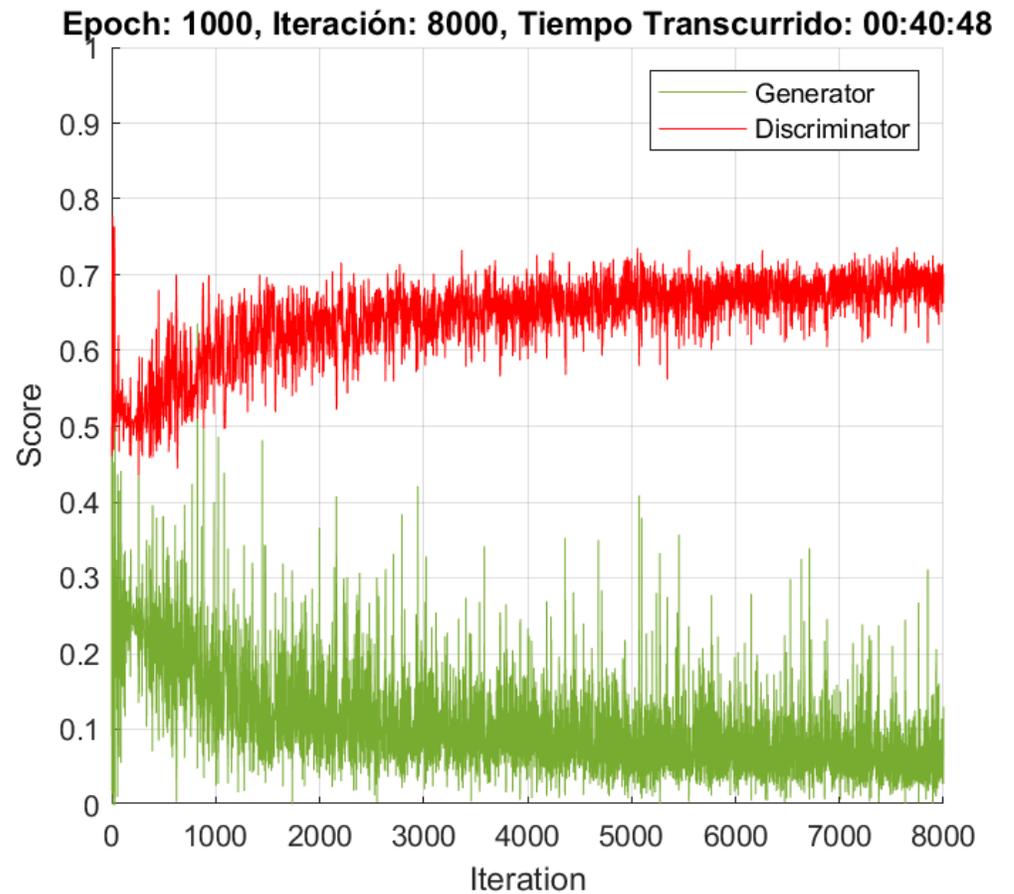
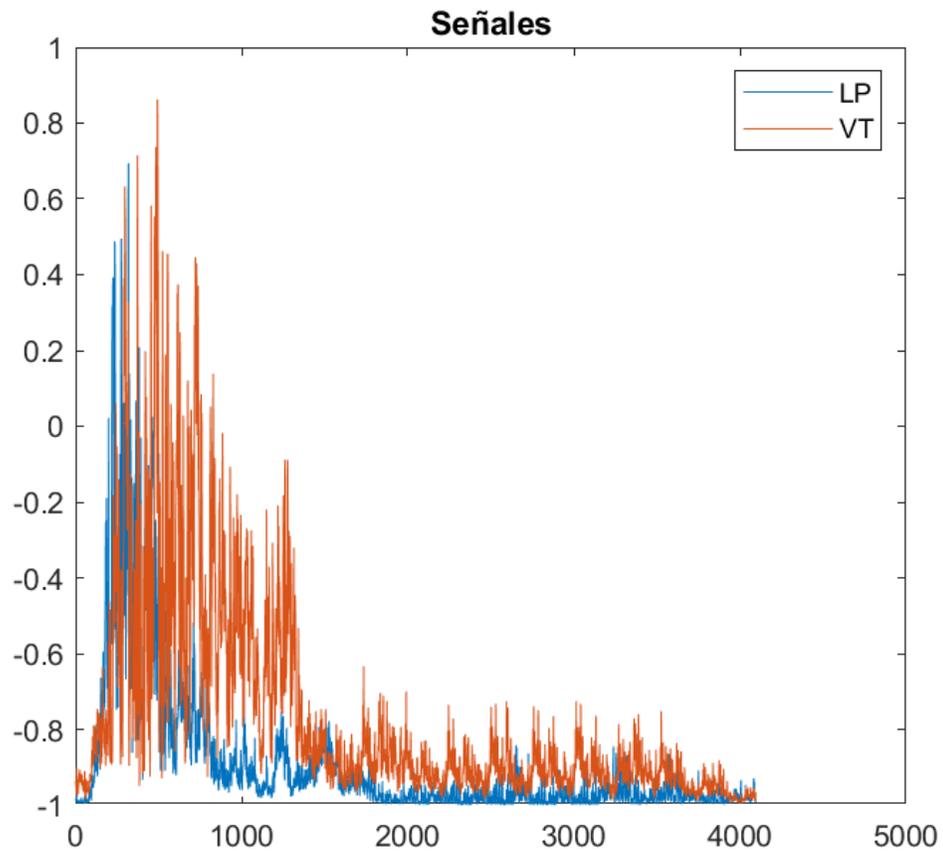
Colapso de modo



Resultados

Modelos con Fallos

Sobre entrenamiento

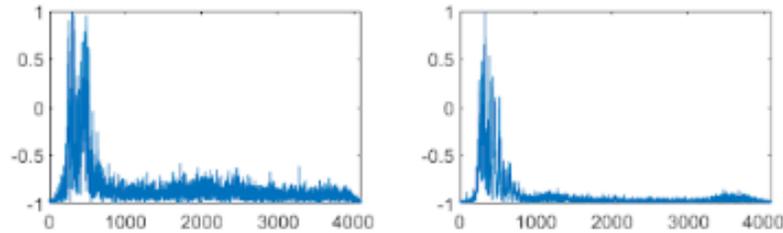


Resultados

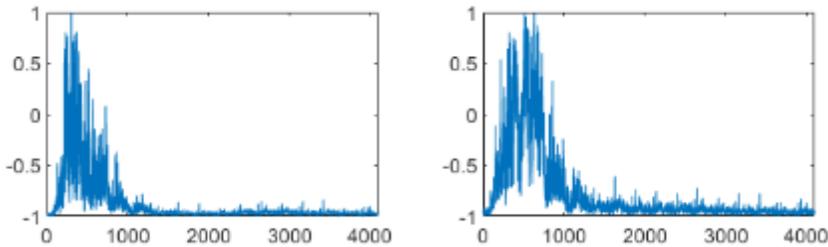
Magnitudes Sintéticas y Reales

Resultados de Magnitudes Reales y Sintéticas LP, señales tiempo LP muestras

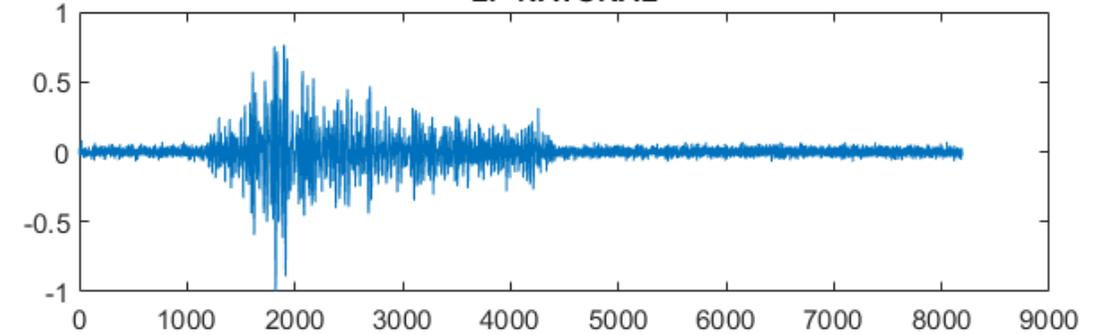
**LP
real**



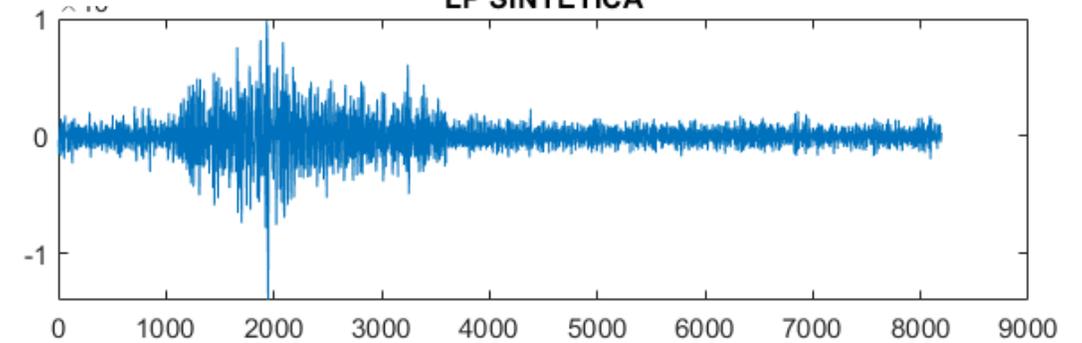
**LP
sintético**



**EVENTO LP TIEMPO
LP NATURAL**



**EVENTO LP TIEMPO
LP SINTETICA**

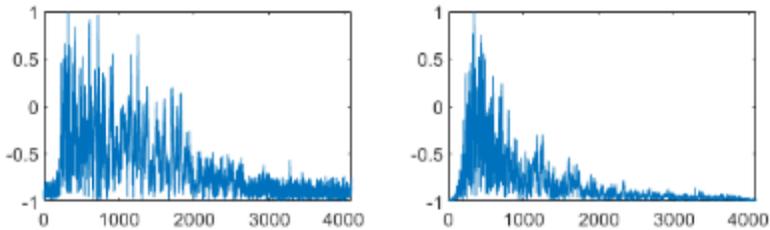


Resultados

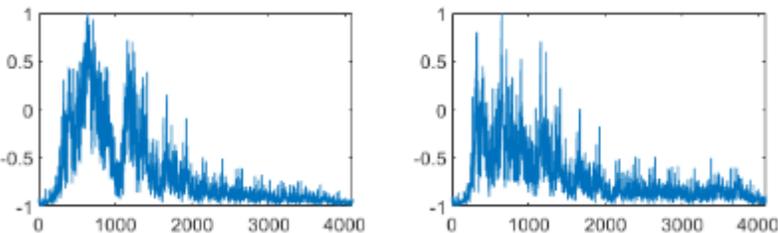
Magnitudes Sintéticas y Reales

*Resultados de Magnitudes Reales y Sintéticas VT
muestras*

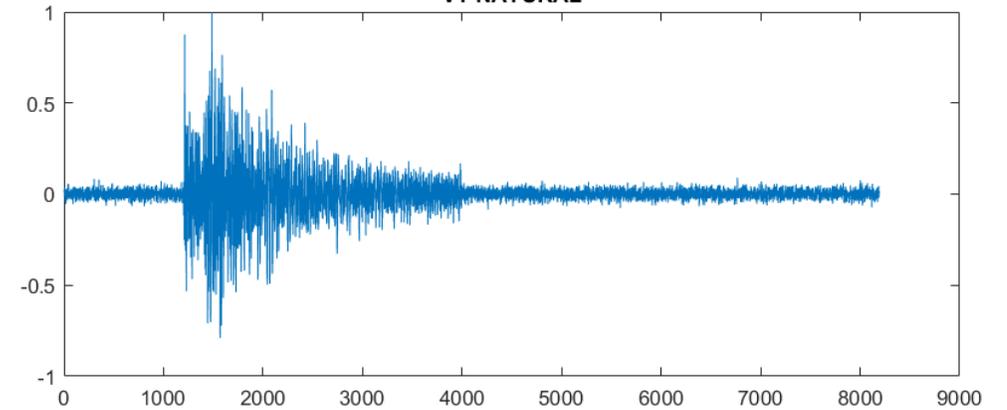
**VT
real**



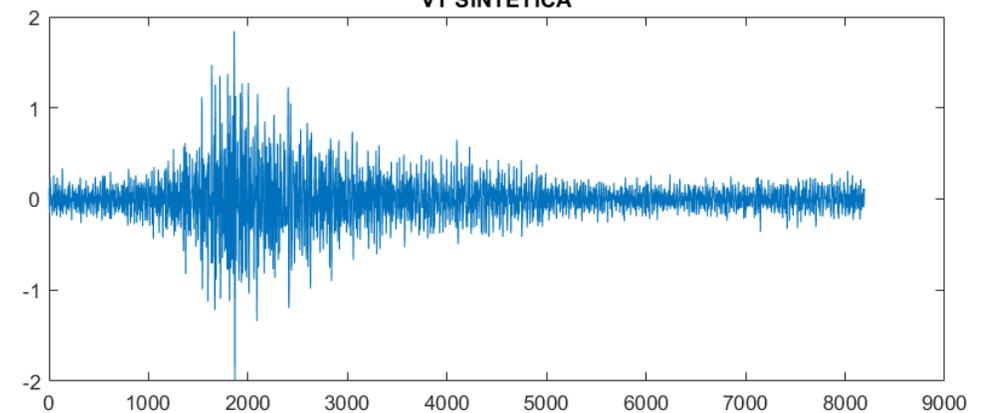
**VT
sintético**



**EVENTO VT TIEMPO
VT NATURAL**



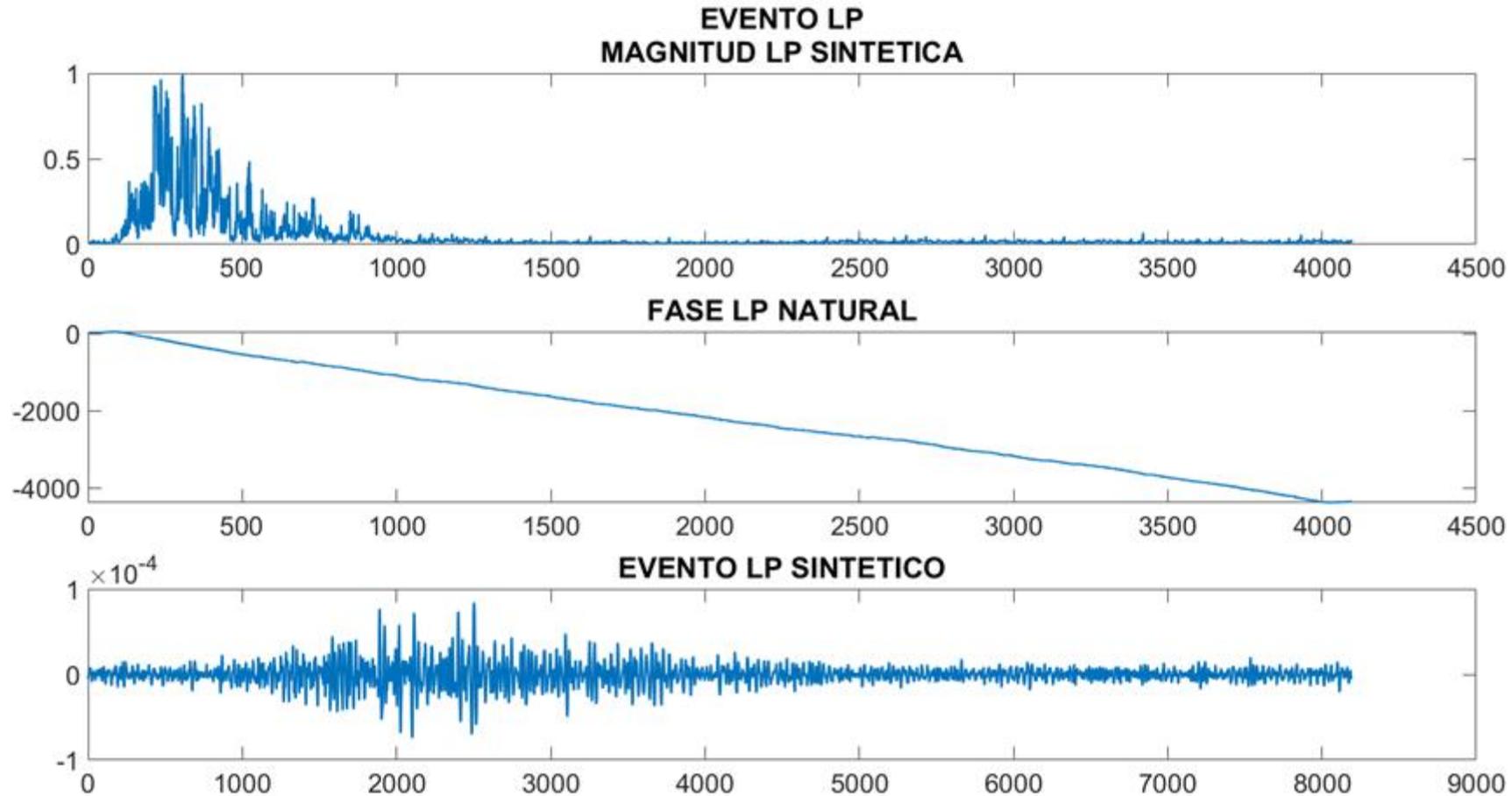
**EVENTO VT TIEMPO
VT SINTETICA**



Resultados

Eventos sintéticos en magnitud, fase y tiempo

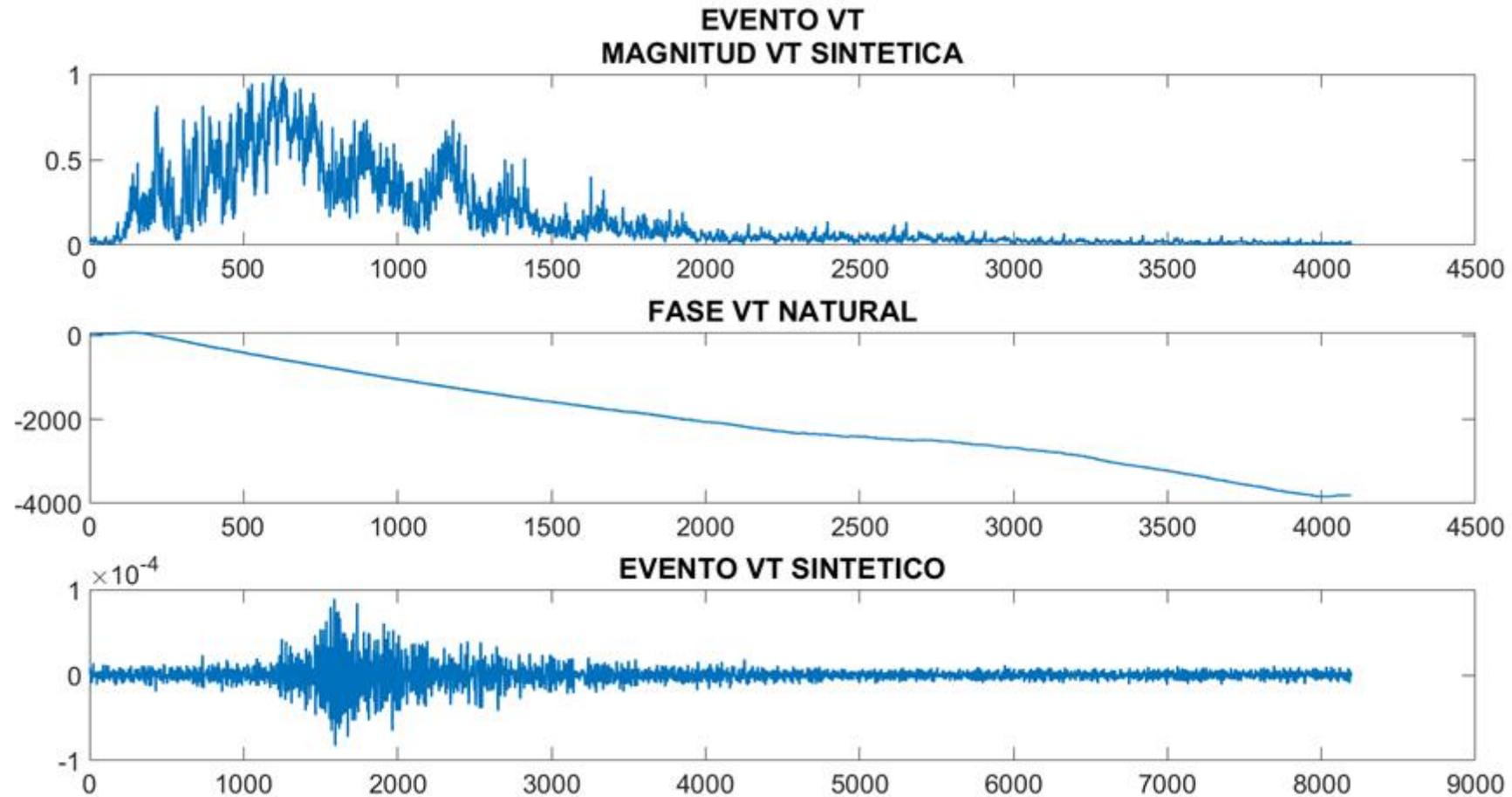
Resultados de Señales sismo-volcánicas LP



Resultados

Eventos sintéticos en magnitud, fase y tiempo

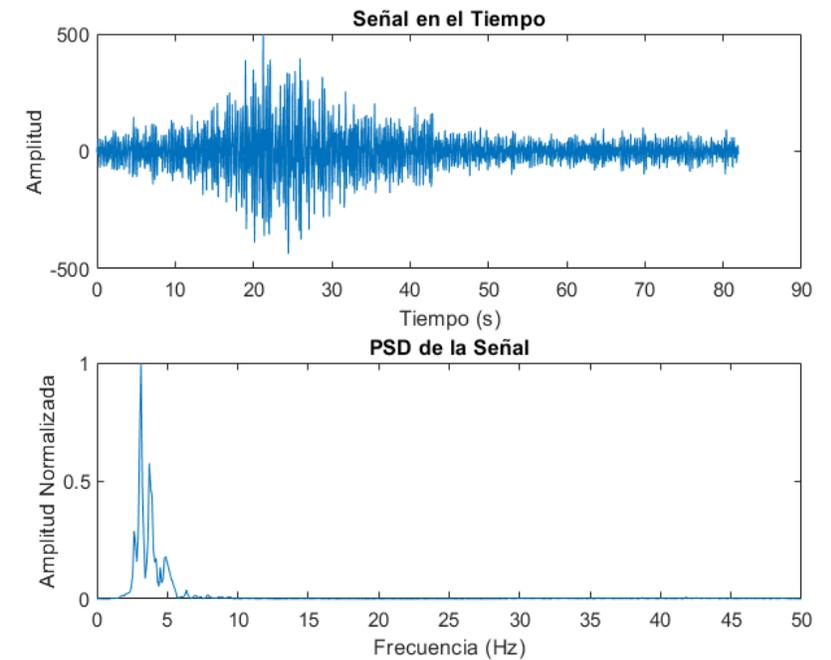
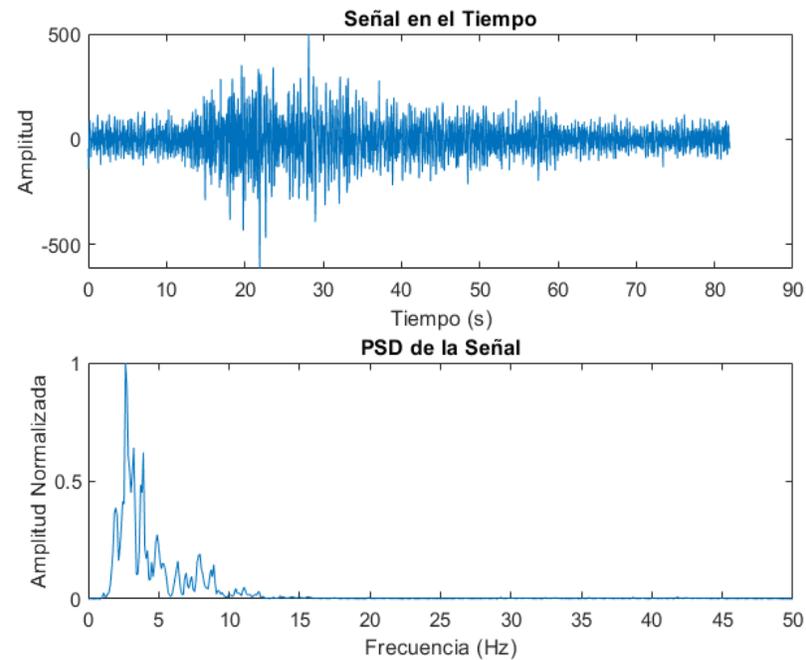
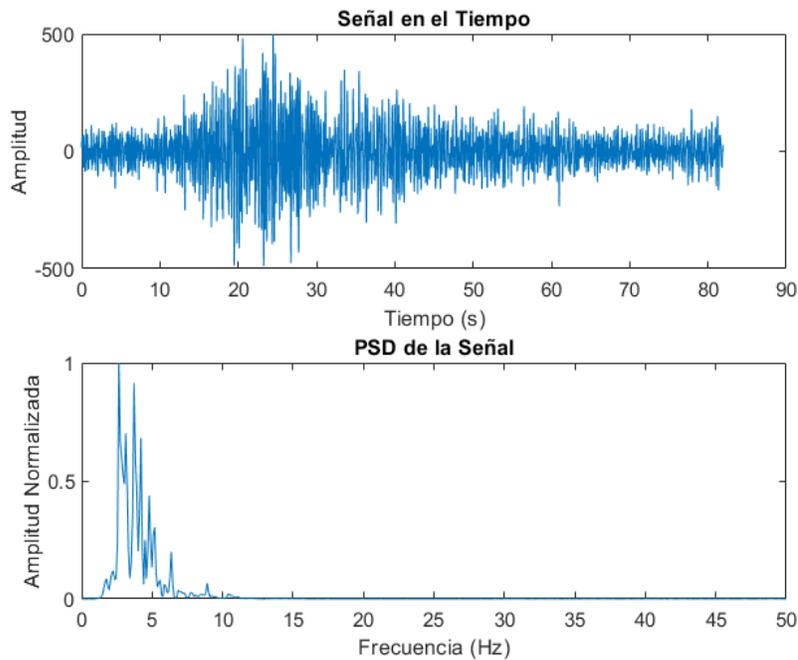
Resultados de Señales sismo-volcánicas VT



Resultados

Señales sintéticas con sus PSD

Resultados de Señales Sintéticas LP

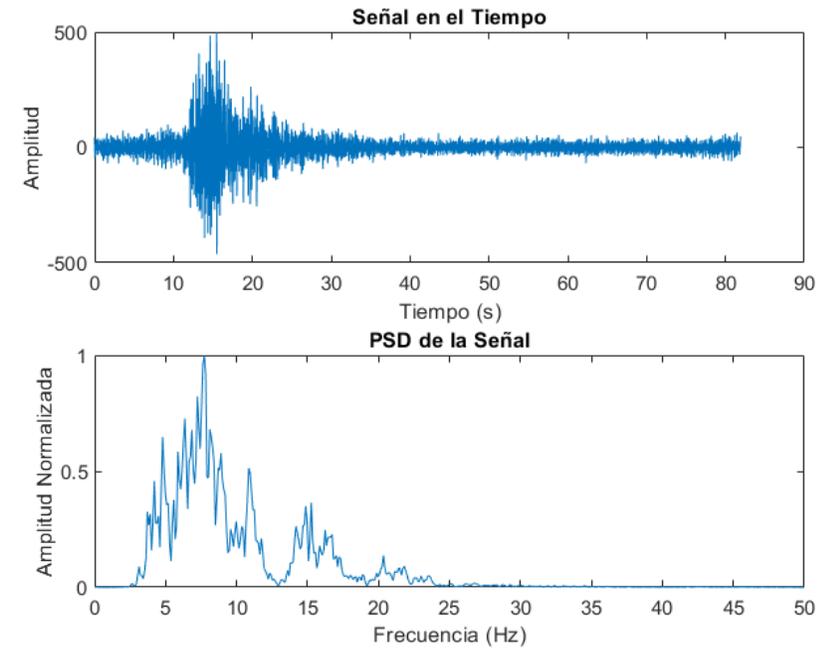
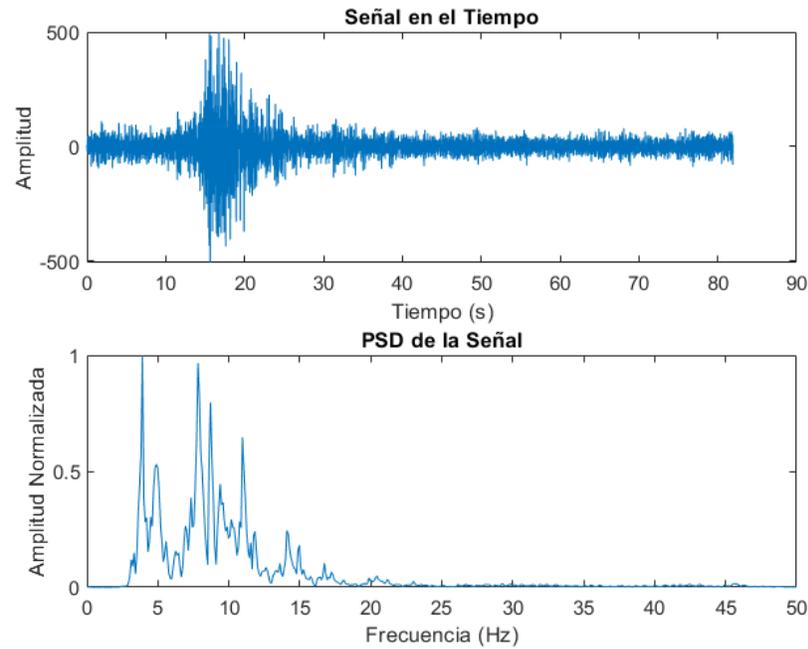
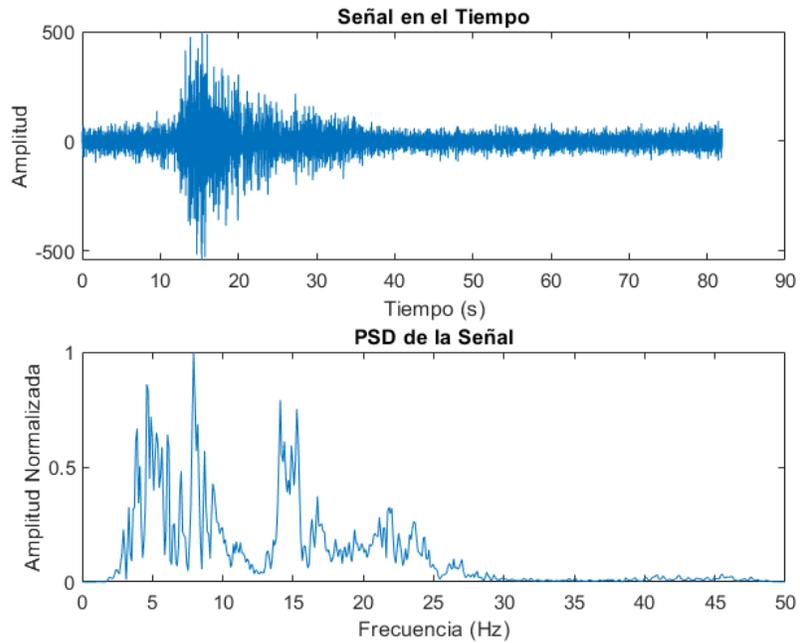


La mayor concentración de potencia en los eventos LP se encuentra en frecuencias de 0 a 5 Hz.

Resultados

Señales sintéticas y sus PSD

Resultados de Señales Sintéticas VT

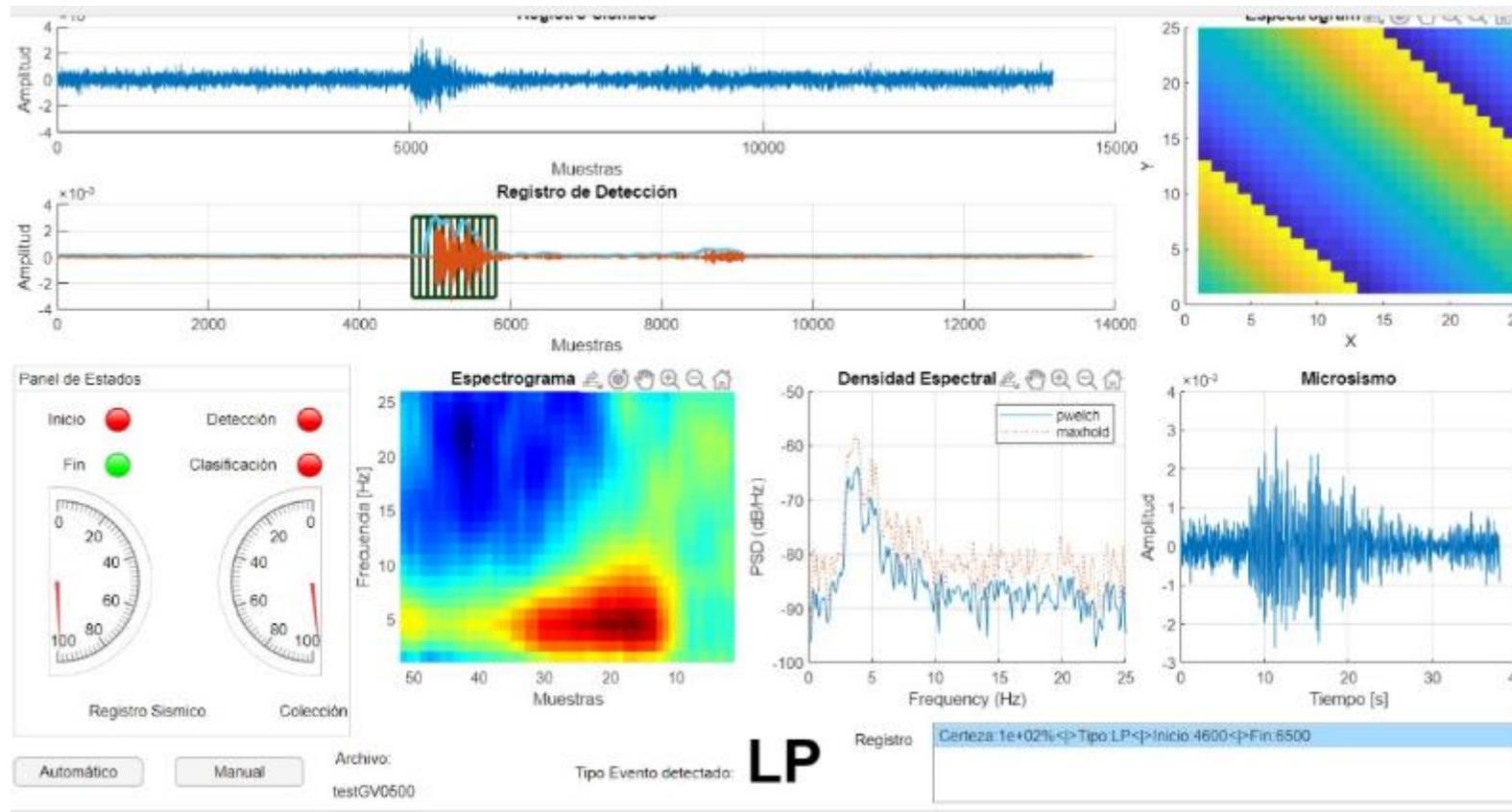


La mayor concentración de potencia en los eventos VT se encuentra en frecuencias de 5 a 15 Hz.

Resultados

Pruebas con un Detector C2N

Detección de una señal LP

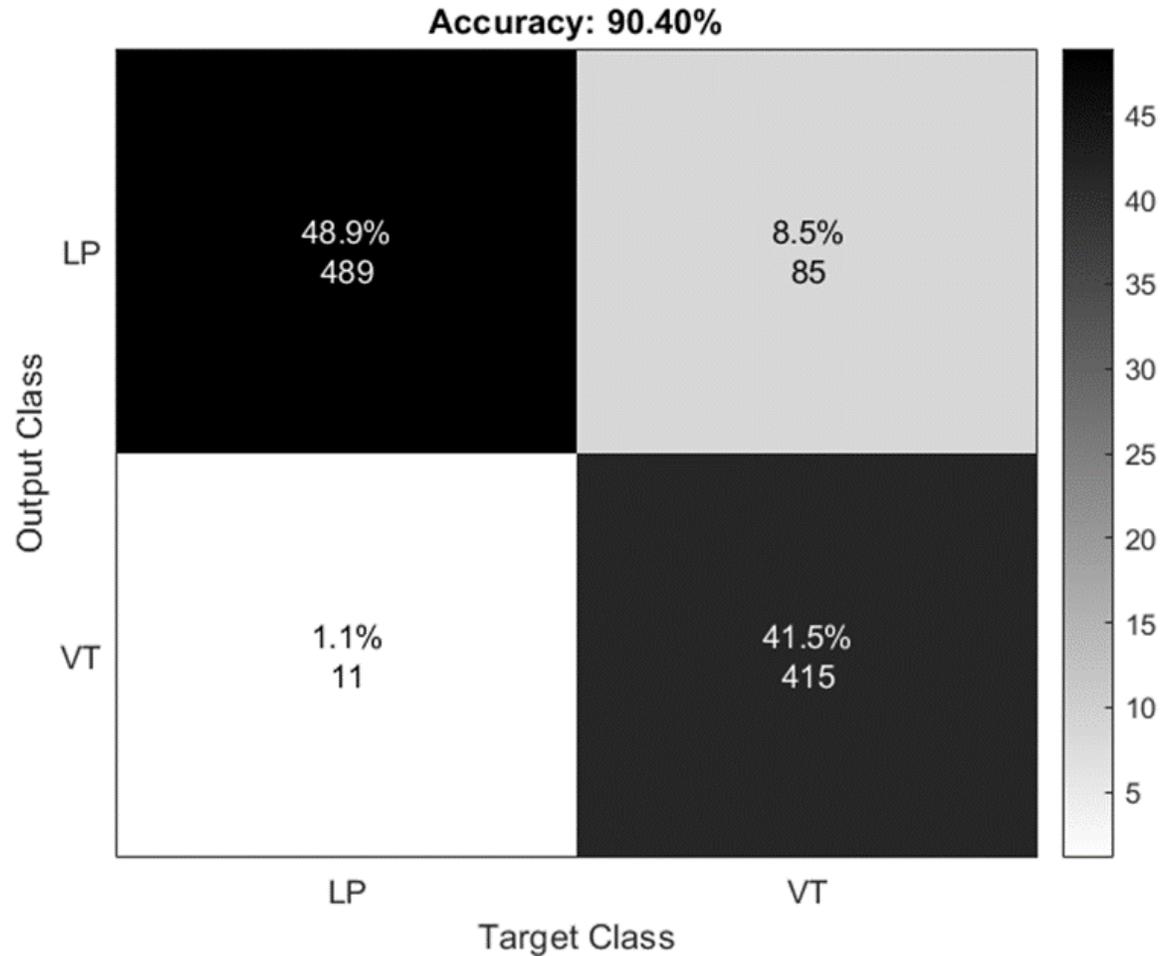


*Lara, F (2021). A deep learning approach for automatic recognition of seismo-volcanic events at the Cotopaxi volcano
Clasificador con una precisión del 97.5%*

Resultados

Pruebas con Sistema de Reconocimiento (Detección + Clasificación):

Se realizaron pruebas con 1000 señales 500 LP y 500 VT con estos datos se obtuvieron los siguientes resultados:



- Se obtuvo un resultado del 100% en la detección.
- Se logró una exactitud del 90.4%.
- Mayor índice de detección de eventos LP sobre los VT.

Resultados

Pruebas con un Autoencoder (WiCom)

Se realizaron pruebas con 1000 señales 500 LP y 500 VT con estos datos se obtuvieron los siguientes resultados:

Confusion Matrix

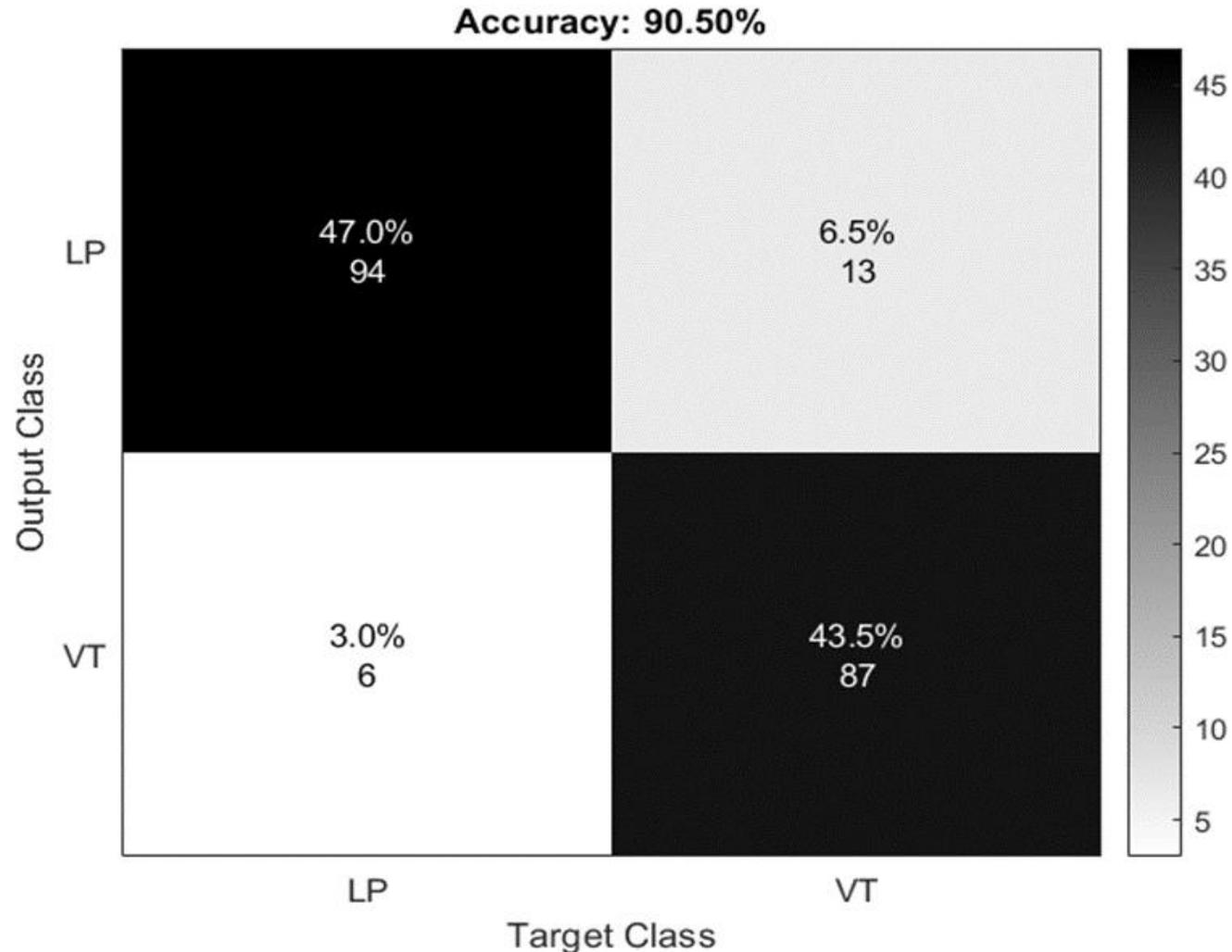
	0	1	
0	500 50.0%	25 2.5%	95.2% 4.8%
1	0 0.0%	475 47.5%	100% 0.0%
	0	1	
	100% 0.0%	95.0% 5.0%	97.5% 2.5%
	Target Class		

- Modelo con Autoencoder WiCom consta de 257 características y tiene una exactitud del 98% para la clasificación de eventos sismo-volcánicos
- Se logró una exactitud del 97.5% con señales sintéticas
- Mayor índice de detección de eventos LP sobre los VT.

Resultados

Formulario Evaluación Visual Eventos Sismo-volcánicos LP y VT Sintéticos

Se realizaron pruebas con 200 señales 100 LP y 100 VT con estos datos se obtuvieron los siguientes resultados:



- Formulario realizado en conjunto con:
 - Nueve expertos del IGEPN y
 - Un experto del SGC.
- Se logró una exactitud del 90.5%
- Las señales en el formulario son naturales y sintéticas.
- Mayor índice de detección de eventos LP sobre los VT.

Resultados

Resultados de las métricas de evaluación en las tres herramientas empleadas

Herramientas de clasificación	Exactitud	Precisión	Sensibilidad	Especificidad	ber
Sistema de Reconocimiento	90.40 %	85.19 %	97.80 %	83.00 %	0.10
Autoencoder	97.50 %	95.20 %	100.0 %	95.00 %	0.30
Clasificación Visual	90.50 %	87.85 %	94.00 %	87.00 %	0.10

- Se obtuvo una mayor exactitud de clasificación con el autoencoder.
- En el Sistema de Reconocimiento y Clasificación Visual se tienen valores de exactitud similares.

Resultados

Resultados de Autoencoder señales naturales, sintéticas y señales naturales aumentados con señales sintéticas

Casos	Exactitud	Precisión	Sensibilidad	Especificidad	ber
Naturales	98.5%	99.0%	98.0%	99.0%	0.02
Sintéticas	96.5%	100%	93.5%	100%	0.04
Real 80%+Sintéticas 20%	98.0%	99.0%	97.1%	99.0%	0.02
Real 60%+Sintéticas 40%	98.0%	100%	96.2%	100%	0.02
Real 40%+Sintéticas 60%	97.0%	100%	94.3%	100%	0.03
Real 20%+Sintéticas 80%	96.5%	100%	93.5%	100%	0.04

- Se ingresaron 100 eventos LP y 100 eventos VT, dependiendo cada caso.
- La exactitud decrece mientras más señales sintéticas se agrega.
- La exactitud se mantiene superior al 95%.
- La precisión se mantiene casi constante.v



Pruebas y Resultados

GUI APP

The screenshot shows the 'CGAN Synthetic Signals' application window. The title bar includes 'Inicio' and 'CGAN' tabs. The main content area is titled 'SISTEMA GENERADOR DE SEÑALES SINTÉTICAS'. It features several input fields and buttons: 'Cantidad de señales sintéticas a generar' (set to 100), 'Etiqueta de los archivos a generar' (set to 'testGV'), 'Directorio donde se guardarán las señales' (with a 'Seleccionar' button), 'Agregar Ruido' (with 'Si' selected), '+ Información', 'Graficar Señal + PSD', and 'Generar'. A 'Status Log' area is at the bottom. Annotations with arrows point to these elements from both sides of the window.

Annotations:

- Quantificar la cantidad de señales sintéticas a generar con el modelo (points to 'Cantidad de señales sintéticas a generar')
- Especificar el directorio de guardado de señales (points to 'Directorio donde se guardarán las señales')
- Información de la APP (points to '+ Información')
- Opción de graficar las señales generadas (points to 'Graficar Señal + PSD')
- Etiqueta de las señales generadas (points to 'Etiqueta de los archivos a generar')
- Posibilidad de agregar ruido a los extremos de la señal para aumentar su tamaño (points to 'Agregar Ruido')
- Botón para finalmente generar las señales bajo los parámetros establecidos (points to 'Generar')

Conclusiones

- Para el presente trabajo de titulación, se utilizó la base de datos MicSigV1 que fue proporcionada por el IGEPN, esta base de datos contiene un número total de 1187 registros de los cuales existen 1044 LP, 101 VT, 27 REG, 8 HB y 7 ICE de las cuales solamente se trabajaron con LP y VT ya que basado en la teoría de la sismica, su incremento desmesurado en un sistema volcánico denota una posible erupción de este.
- En la fase de diseño del modelo CGAN, se implementó un modelo generador capaz de aprender a crear respuestas en frecuencia realistas y un modelo discriminador que aprende a distinguir respuestas en frecuencia falsas. Se configuro para los dos modelos el tamaño de las capas, kernel, stride, la función de activación, cantidad de filtros. Se logra obtener un modelo CGAN definitivo el cual permitió obtener respuestas en frecuencia con las características de eventos sismo volcánicos de tipo LP y VT.
- Durante la fase de entrenamiento en la monitorización de distintos modelos se validó que, al utilizar una mayor cantidad de filtros y capas CGAN, provocaba fallos de entrenamiento propios de los modelos antagónicos adversarios. Pese a tener limitados datos con señales de una naturaleza compleja y haber obtenido distintos modelos fallidos, a través de la variación de los parámetros se identificó un balance en el aprendizaje al utilizar 1000 epoch y con 256 minibatch. Además, gracias a la utilización de GPU se obtuvo una aceleración y optimización del entrenamiento con un tiempo de treinta y un minutos.

Conclusiones

- Se diseñó una aplicación en el software MATLAB 2020 ®, la cual permite generar señales sintéticas de eventos sismo-volcánicos etiquetados. En ella se define la cantidad de señales a generar de las cuales la mitad son del tipo LP y la otra en tipo VT en una relación de 50/50.
- En las herramientas de análisis se obtuvieron las siguientes exactitudes en la clasificación de eventos LP y VT: sistema de reconocimiento, con 90.4%; autoencoder, con 97.5% y clasificación visual, con 90.5%. La herramienta con el mejor resultado de exactitud es el autoencoder, por lo que se realizó una segunda prueba con la mezcla de cierta cantidad de señales sísmicas sintéticas y reales a través de la variación de sus cantidades, se logró una exactitud de entre un 96.5% y a un 98.5%.
- En este proyecto de investigación se implementó un modelo CGAN que genera señales sismo-volcánicas sintéticas para dos tipos de eventos sísmicos: LP y VT. El estudio se basó en el método de generación de magnitudes de respuesta en frecuencia y que, en conjunto con las magnitudes de respuesta en fase de señales sismo-volcánicas reales, se obtuvieron nuevas y únicas señales sintéticas sismo-volcánicas de los dos tipos de eventos en el dominio del tiempo a través de la IFFT. Estas señales sintéticas fueron evaluadas por distintas herramientas de detección y clasificación de eventos sismo-volcánicos de tipo LP y VT donde se obtuvieron porcentajes de exactitud mayores al 90% y en precisión sobre el 99%.

Recomendaciones y Trabajos Futuros

- Al tener una base de datos pequeña, se recomienda enriquecerla con las señales sintéticas como la que se presenta en (B. Perez, 2022). Con ello se tendrá una base de datos más con mayor número de señales sísmo-volcánicas entre naturales y sintéticas, así logrando entrenar la CGAN por más tiempo y tener mejores resultados, en las herramientas de clasificación y detección.
- Se recomienda analizar el uso de las nuevas herramientas que posee el Deep Learning Toolbox de MATLAB® en su versión más actual del año 2022. El entrenamiento de redes generativas de datos unidimensionales se presenta en esta versión en el que se evita procesamiento interno donde se optimiza el código y de tiempos de entrenamiento de los modelos.
- Nuestro grupo de trabajo está interesado en que se generen señales sintéticas de eventos LP, VT, HYB y ICEQUAKE con nuevos métodos de generación de elementos con redes generativas adversarias como son: Auxiliary Classifier Generative Adversarial Network, Progressive Growing Generative Adversarial Network.
- Adicionalmente, con el modelo CGAN diseñado en este trabajo de investigación, se puede llegar a entrenar con bases de datos de diferentes volcanes, como el volcán Chileno Llaima, se debe considerar el preprocesamiento de las señales antes del ingreso al modelo, igualmente se varió el tiempo de entrenamiento de los parámetros de Adam.

UNIVERSIDAD DE LAS FUERZAS ARMADAS - “ESPE”

DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES

TEMA: “GENERACIÓN DE SEÑALES SINTÉTICAS DE EVENTOS SISMO-VOLCÁNICOS DEL VOLCÁN COTOPAXI A TRAVÉS DE UN MODELO DE RED NEURONAL ADVERSARIO GENERATIVO CONDICIONAL”

AUTORES:

GALARZA CRUZ, CHRISTIAN RAMIRO & VEGA VERGARA, JOSSELYN MICHELLE

Director del Proyecto: Dr. Román Lara, PhD.

Docente Evaluador: Ing. Julio Larco

Director de Carrera: Ing. Carlos Daniel Altamirano, PhD.

Secretaria Académica: Abg. María Fernanda Jaramillo

Julio, 2022

