



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES

Carrera de Ingeniería en Electrónica y Telecomunicaciones

OPTIMIZACIÓN DE LOS PARÁMETROS LIBRES DE UN CLASIFICADOR DE MICRO SISMOS BASADO EN AUTOENCODERS, EMPLEANDO ESPACIOS LATENTES

Autor: Mejía Martínez Angie Nicole

Director: Ing. Carlos Paúl Bernal



- ▶ **01. Introducción**
- ▶ **02. Materiales**
- ▶ **03. Metodología**
- ▶ **04. Resultados**
- ▶ **05. Conclusiones y Recomendaciones**





01. Introducción



02. Materiales



03. Metodología



04. Resultados



05. Conclusiones y Recomendaciones

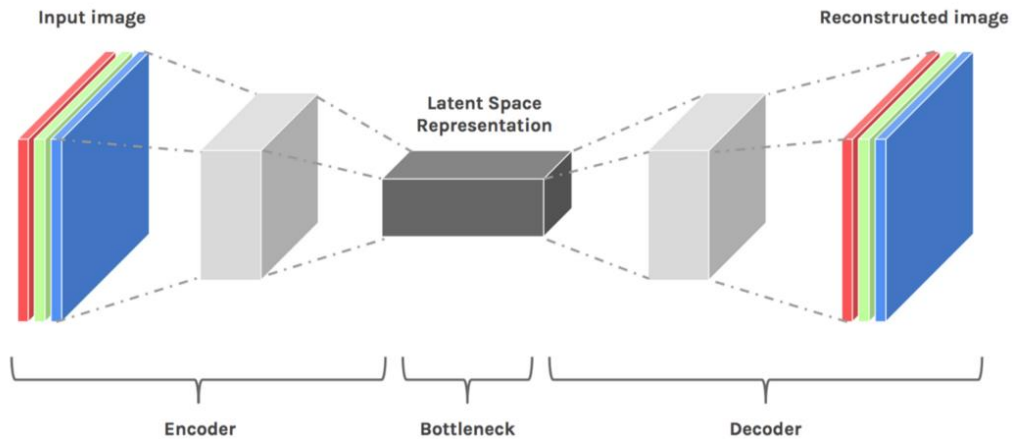


“Generación de señales sintéticas de eventos sismo-volcánicos del volcán Cotopaxi a través de un modelo de red neuronal adversario generativo condicional” (Galarza Cruz & Vega Vergara, 2022)

“Clasificación de eventos sismo volcánicos usando características psicoacústicas mediante técnicas de aprendizaje automático supervisado y no supervisado” (Minango Guatumillo, 2022)

Implementación de un sistema clasificador de microterremotos del volcán Cotopaxi basado en técnicas de Deep Learning (Iglesias, Rosero, 2022)





- Prevenir fallas
- Mejora de procesos
- Funcionalidad efectiva del modelo
- Análisis de Desempeño

Objetivo General

- Optimizar los parámetros libres de un clasificador de micro terremotos basado en autoencoders, empleando espacios latentes.



Objetivos Específicos

- Analizar la correlación de los modelos de entrenamiento de la clasificación sísmica utilizando espacios latentes de las características espectrales.
- Evaluar el desempeño de los modelos de entrenamiento al trabajar con autoencoders.
- Mejorar el porcentaje de desempeño al optimizar los parámetros libres de un clasificador.
- Identificar los parámetros que maximicen el porcentaje de acierto de un evento sísmico.





01. Introducción



02. Materiales



03. Metodología



04. Resultados



05. Conclusiones y Recomendaciones



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

MATLAB® Software de Desarrollo



- Análisis de Datos
 - Gráficas
 - Programación
- Creación de apps
- Cálculo en Paralelo

Dell G15 Ryzen Edition Características



- AMD Ryzen 7 5800H
- 16 GB DDR4 3200 MHz
- 500GB (SSD M.2 NVme)
- Nvidia GeForce RTX 3060

Base de datos de Instituto Geofísico de la Politécnica Nacional



- Reunidas en el 2012 – 2019
- Frecuencias de 50 y 100 Hz
- Conformada por VT, LP, HB, RG y IQ

Generador de Señales Sismo volcánicas CGAN



- Generación de VT y LP
- Frecuencia en 100HZ
- Permite agregar ruido



01. Introducción



02. Materiales



03. Metodología



04. Resultados

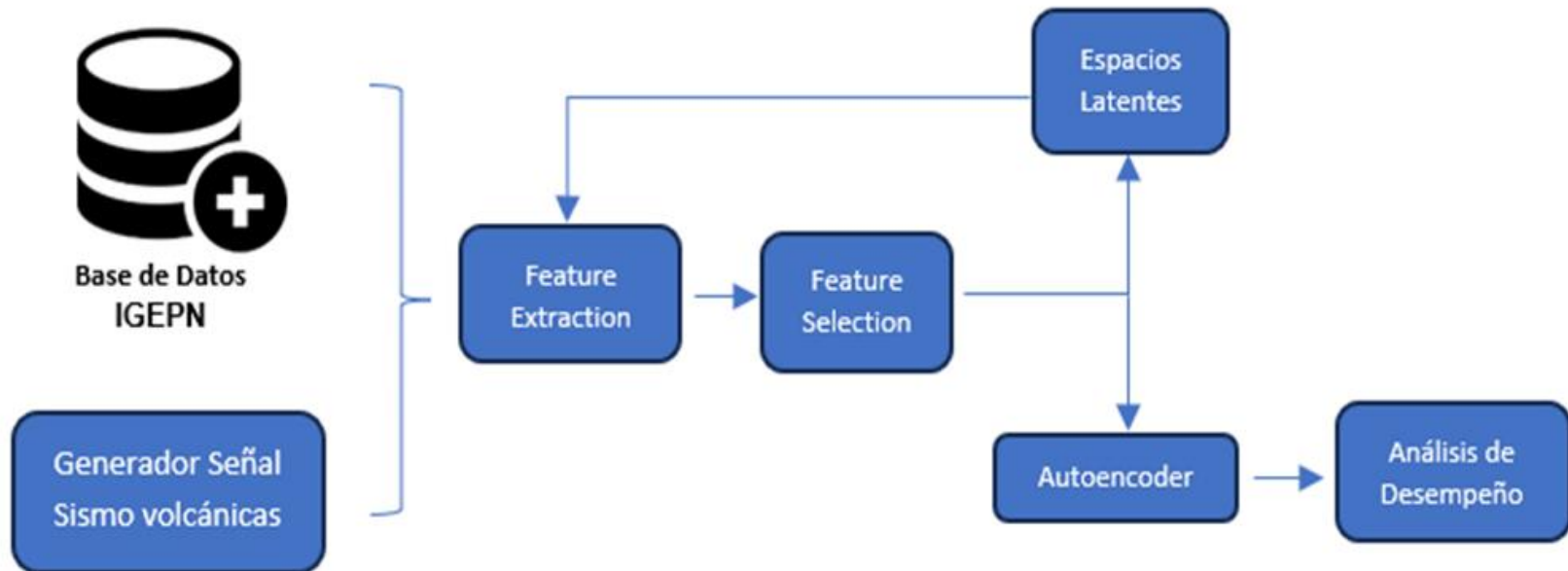


05. Conclusiones y Recomendaciones

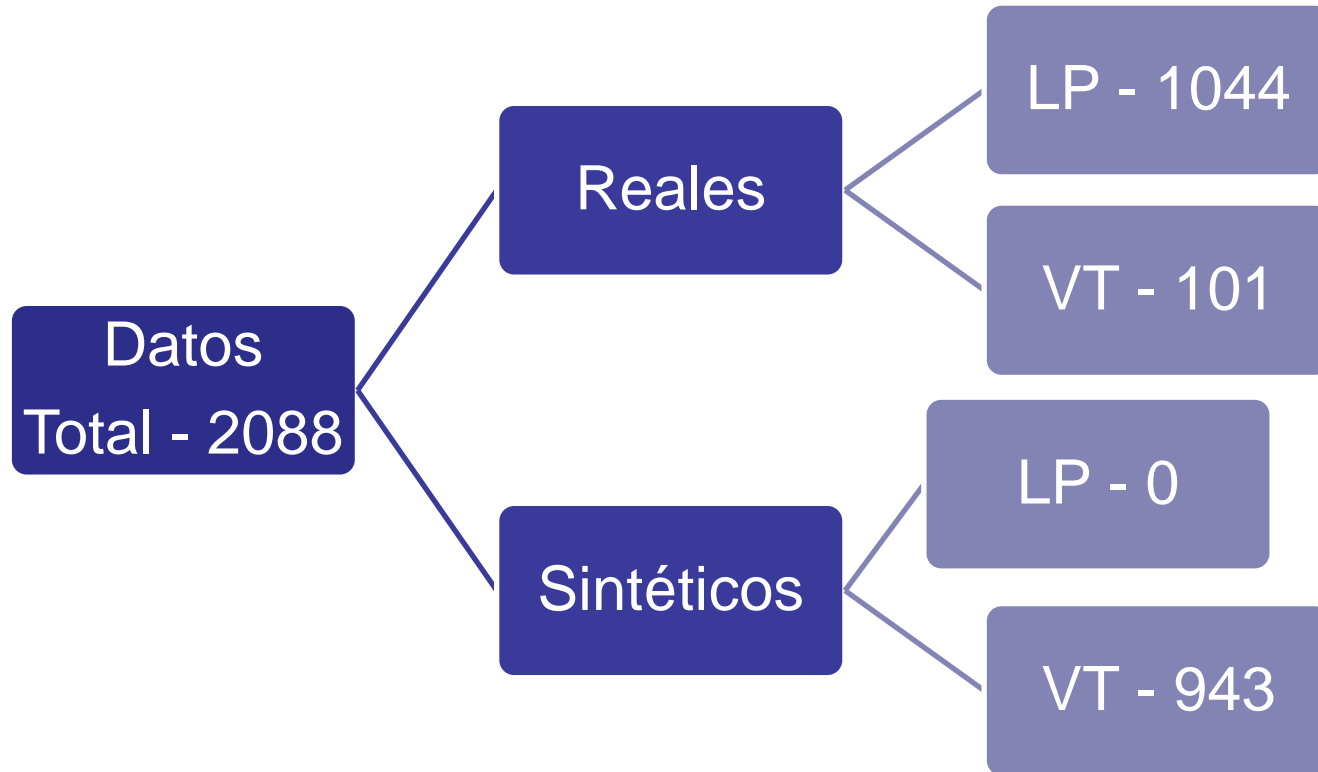


ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

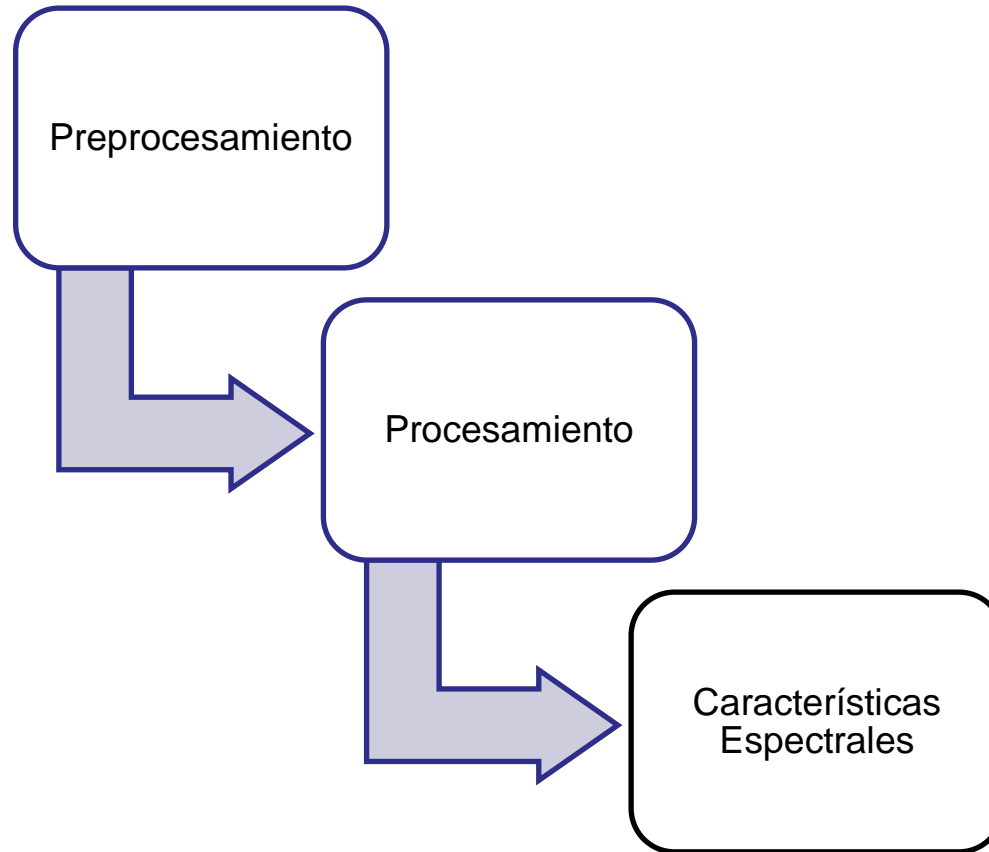
Diagrama de Bloques



Data Augmentation



Feature Extraction



Feature Extraction

Preprocesamiento

- Toma de datos aleatorios para LP reales
- Señales en función de la frecuencia
- Muestro en 100Hz
- Eliminación de la media
- Combinación entre VT y LP
- Datos sintéticos en función de la frecuencia.

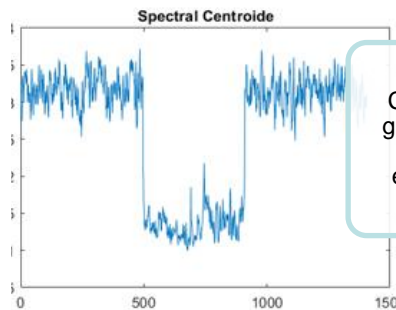
Procesamiento

- Obtención de la Características
- Ventana de tipo rectangular
- Función lineal
- Escala de ERB [0 - 25]Hz
- 11 Características Espectrales



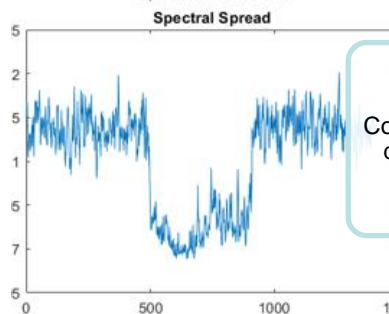
Feature Extraction

Centroide Espectral



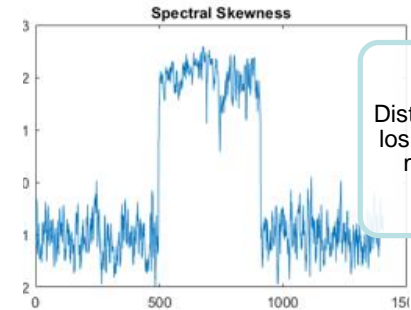
Centro de gravedad y energía espectral

Dispersión Espectral



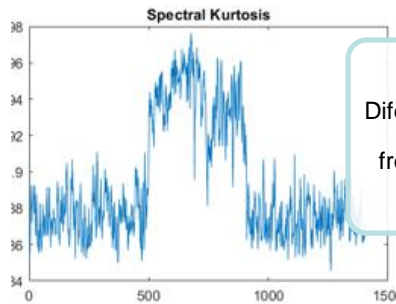
Concentración de energía

Asimetría Espectral



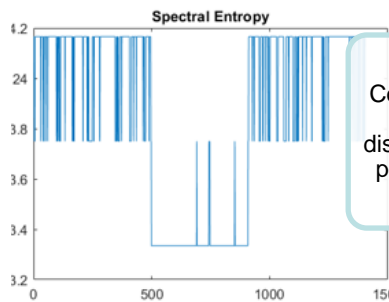
Distribución de los valores de magnitud

Curtosis Espectral



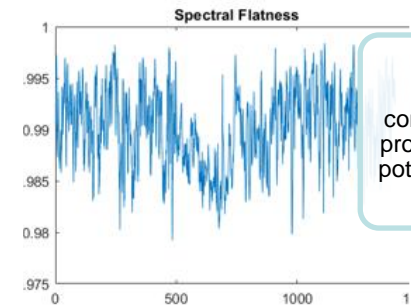
Diferenciación entre frecuencias

Entropía



Comparación de la distribución de probabilidad

Planitud Espectral

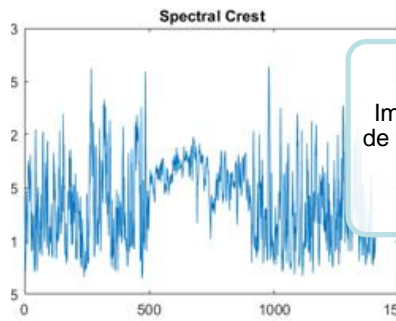


Medida contraria a la proporción de potencia tonal



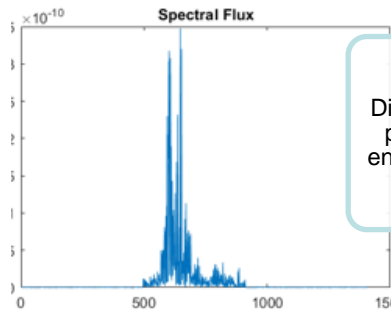
Feature Extraction

Cresta Espectral



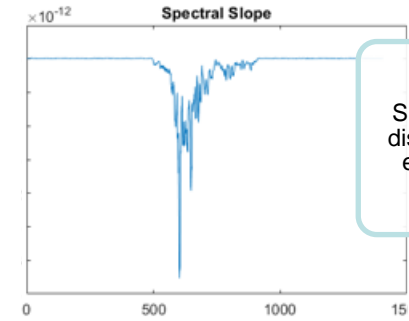
Impulsividad de los eventos sonoros

Flujo Espectral



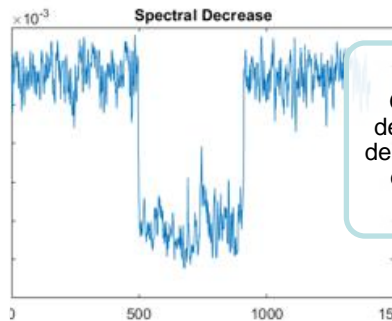
Diferencia de promedios entre cuadros

Pendiente Espectral



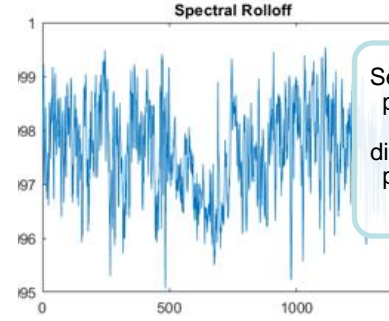
Similar a la disminución espectral

Disminución Espectral



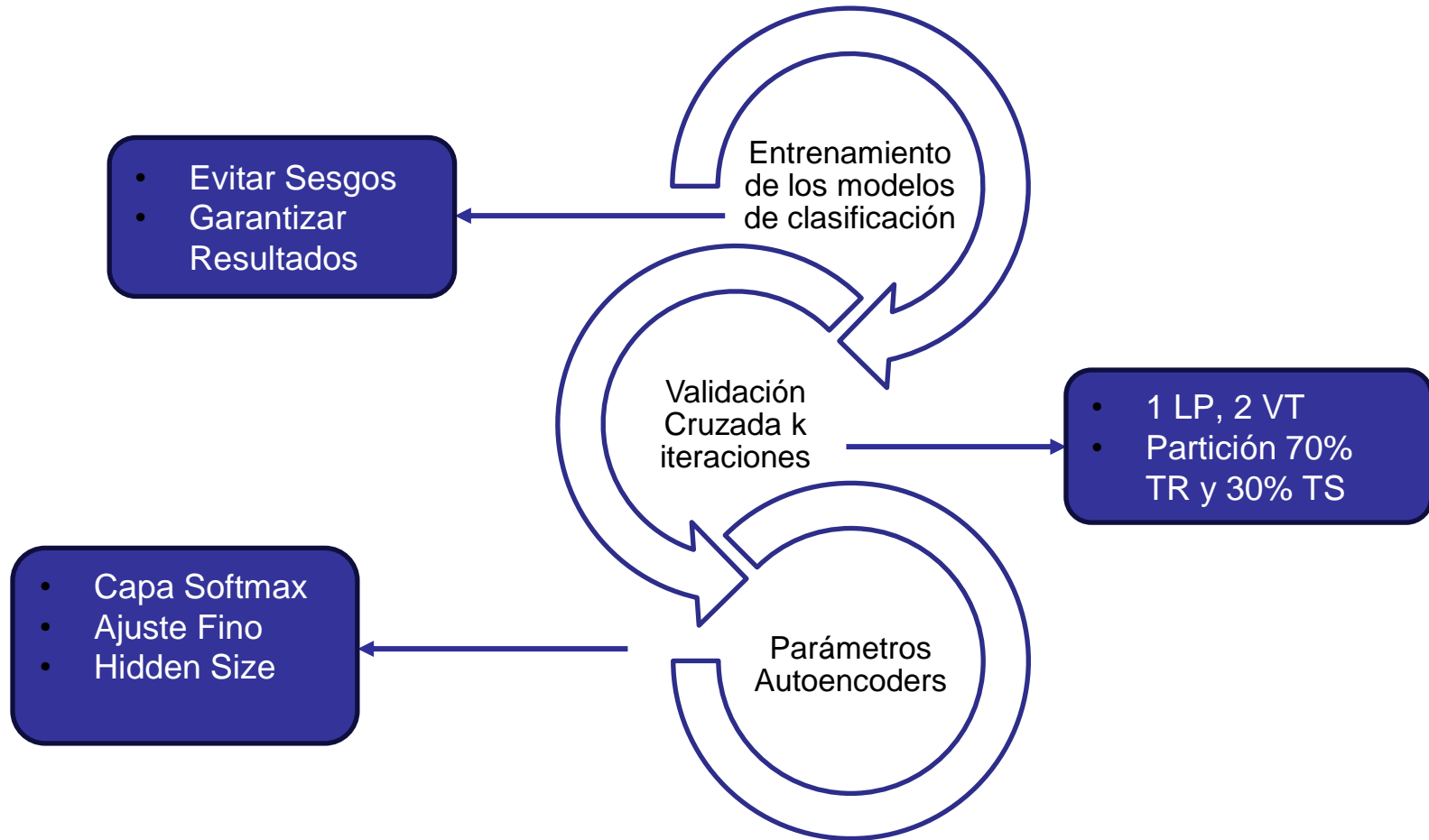
Grado de decaimiento de la amplitud espectral

Reducción Espectral



Se supera una probabilidad en la distribución de probabilidad acumulada

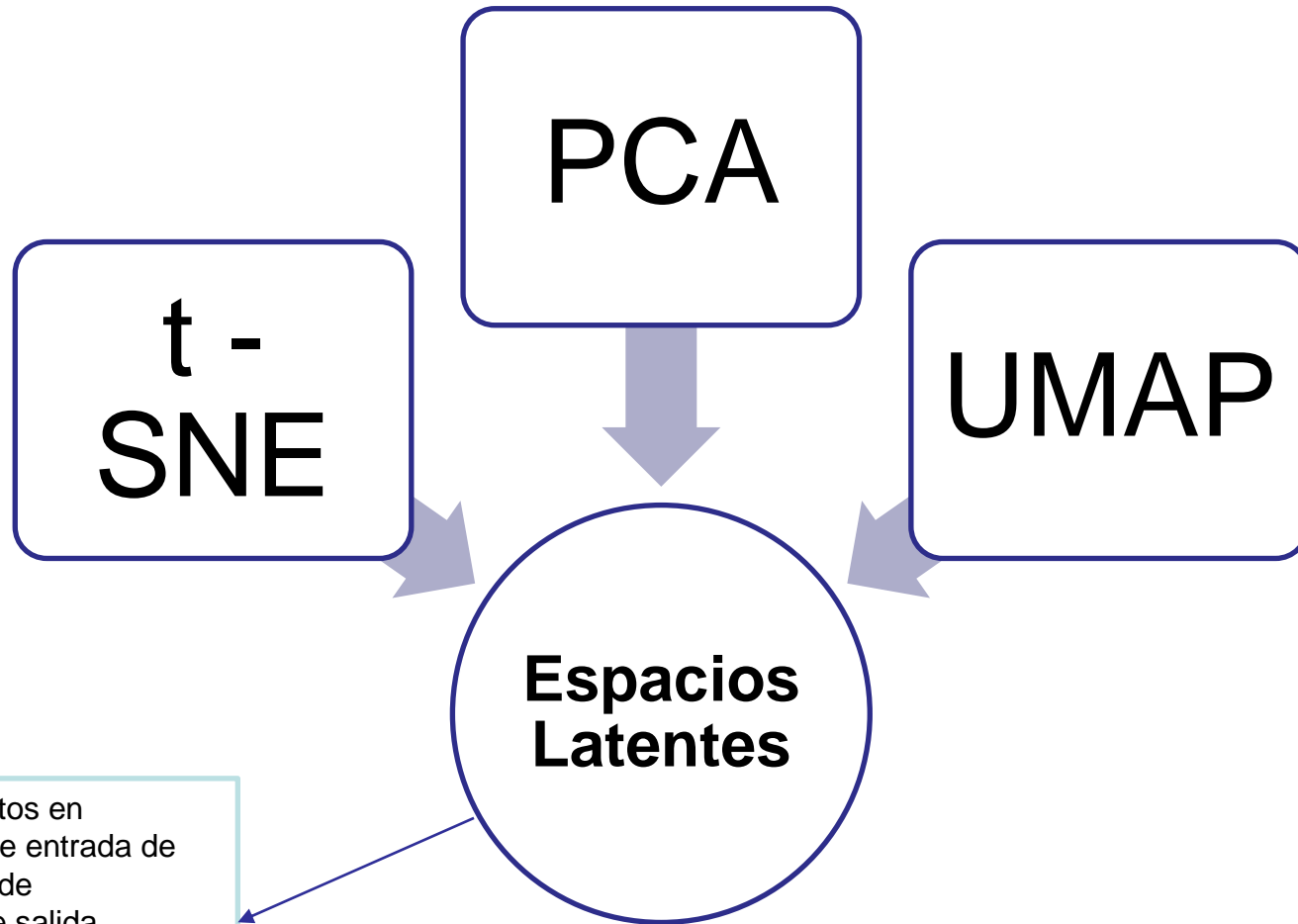
Feature Selection



Feature Selection

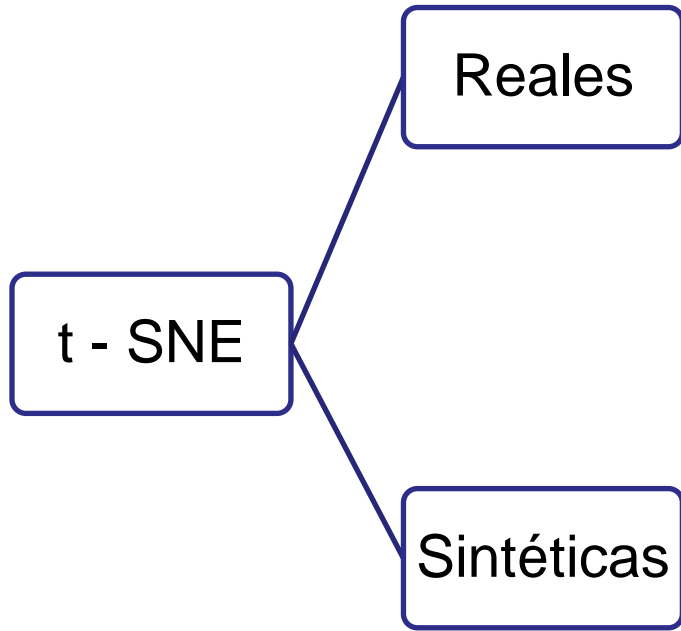
Parámetros del Autoencoder	Descripción	Predeterminado	Ajuste
Encoder Transfer Function	Función de transferencia para el codificador	Sigmoidal	Sigmoidal
Decoder Transfer Function	Función de transferencia para el decodificador	Sigmoidal	Lineal saturada
L_2 Weight Regularization	Coefficiente regulador de peso L_2	0.001	0.004
Sparsity Regularization	Control del impacto del regularizador de dispersión	1	1
Sparsity Proportion	Proporción de entrenamiento, para una red neuronal	0.05	0.05
Parámetros del DT	Descripción	Valor	
Max Num Splits	Número de decisiones de divisiones	10	
Crossval	Validación cruzada con respecto al crecimiento del árbol de decisión	on	



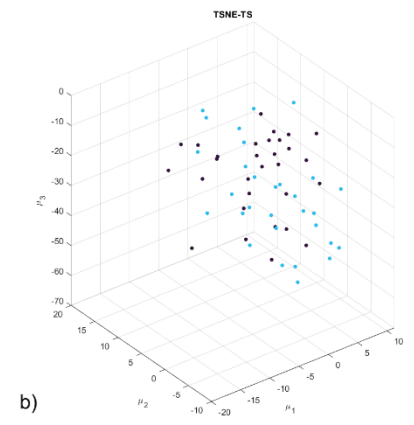
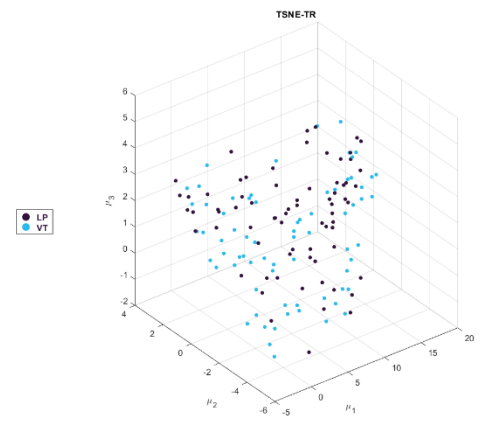
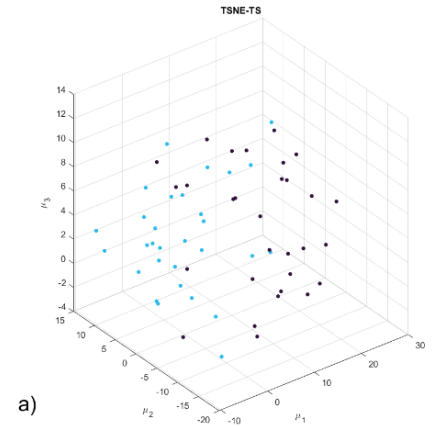
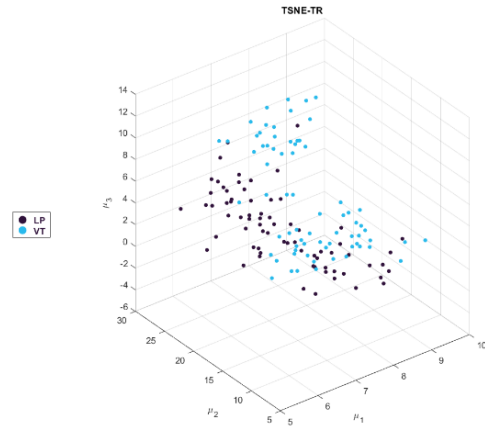


Proyección de datos en n -dimensiones de entrada de nuevos espacios de l -dimensiones de salida ($n > 1$) e intentan mantener gran parte de información posible de las proyecciones.

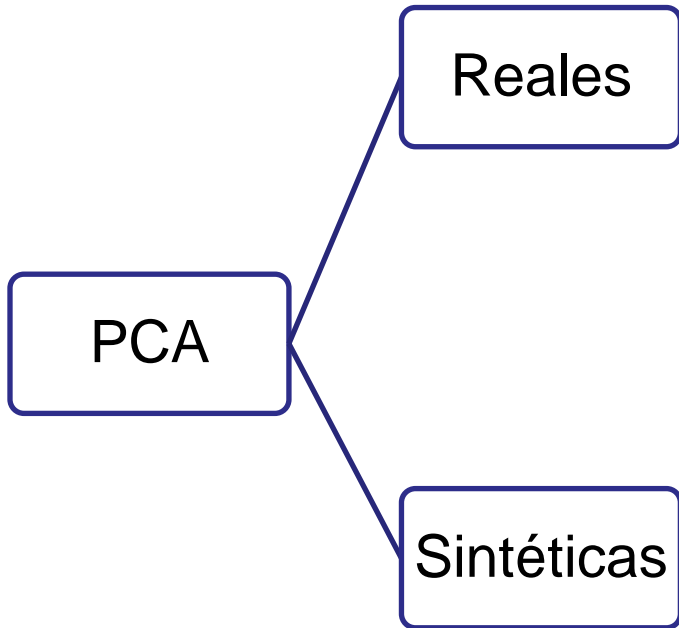
Espacios Latentes



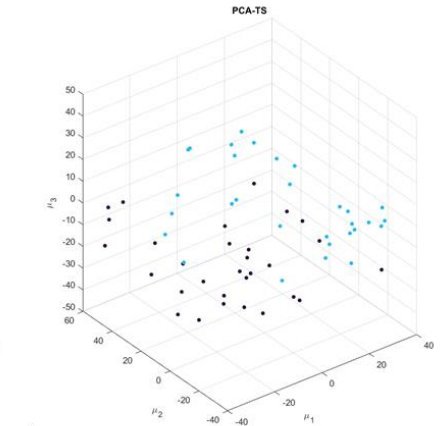
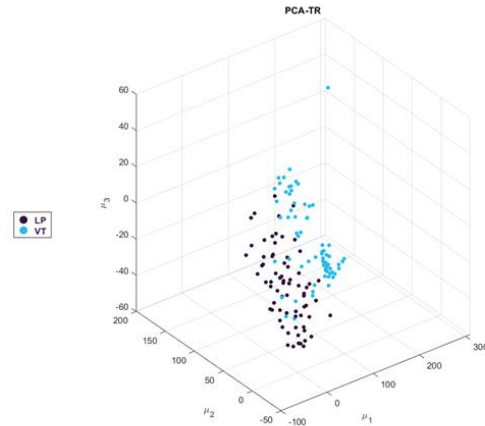
- Optimizar y producir visualizaciones
- Representación de estructuras en diversas escalas
- Enfoque en las estructuras complejas



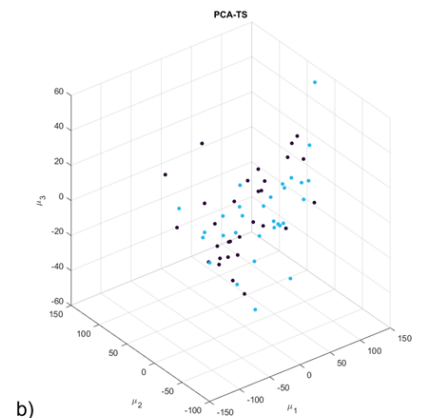
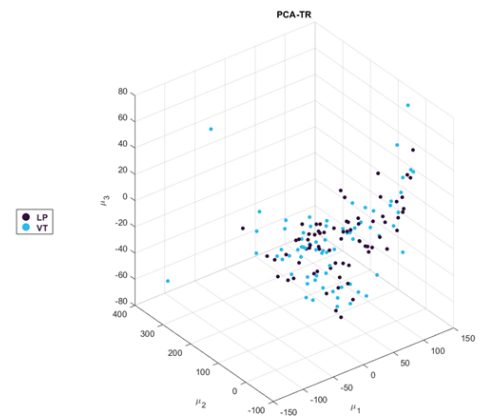
Espacios Latentes



- Dimensionalidad de conjuntos de datos y conserva
- Hallar un conjunto de variables no correlacionadas



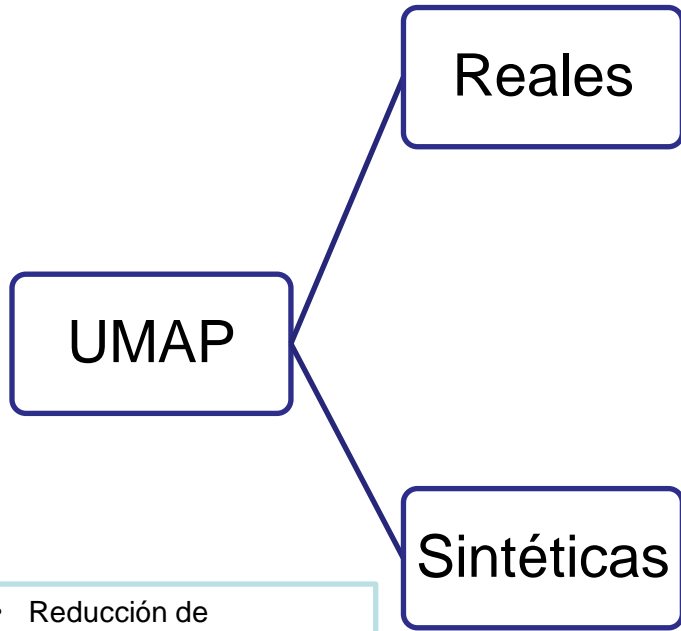
a)



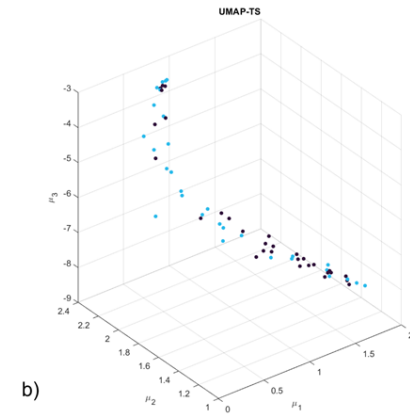
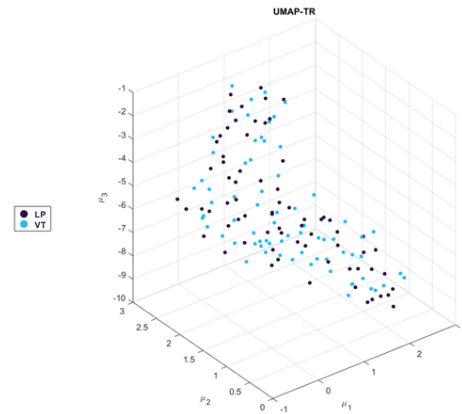
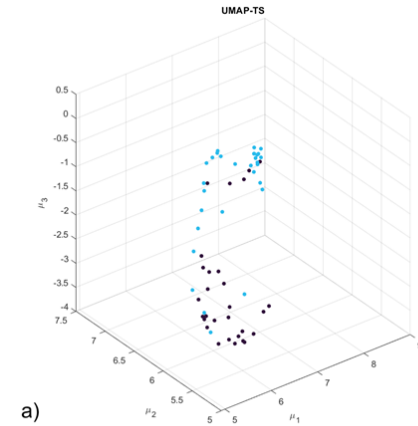
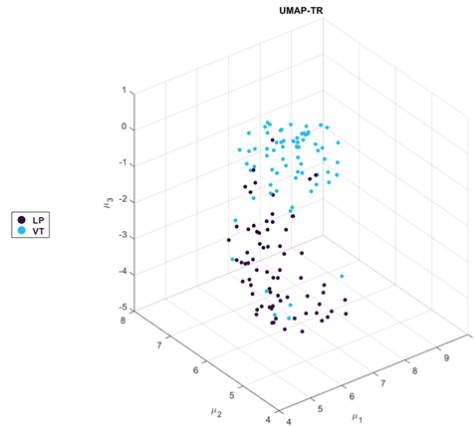
b)



Espacios Latentes



- Reducción de dimensionalidad y la visualización de datos
- Consiste en dos pasos: cálculo de la representación gráfica, descenso de gradiente estocástico al optimizar una incrustación





01. Introducción



02. Materiales



03. Diseño e Implementación



04. Resultados



05. Conclusiones y Recomendaciones



Autoencoders Test	Precisión Test	Exactitud Test	Sensibilidad Test	Especificidad Test	BER Test
90	96,49	93,31	90,65	96,51	0,06
80	96,23	96,28	95,9	96,69	0,04
70	96,49	97,35	97,59	97,45	0,03
60	97,13	96,81	96,26	97,42	0,03
50	96,36	97,17	95,67	98,87	0,03
40	96,68	97,52	96,94	98,15	0,02
30	97,06	93,98	91,29	97,25	0,06
20	97,19	96,11	94,97	97,38	0,04
10	96,81	96,11	94,08	98,47	0,04



Autoencoder	Antes	Después
Precisión	97.619	96.46
Exactitud	98.413	97.35
Sensibilidad	97.619	97.59
Especificidad	98.810	97.45
VER	0.018	0.03

Matriz Combinada Test

Output Class	LP	VT	Target Class
LP	291 46.5%	19 3.0%	93.9% 6.1%
VT	9 1.4%	307 49.0%	97.2% 2.8%
	97.0% 3.0%	94.2% 5.8%	95.5% 4.5%

Matriz Reales Test

Output Class	LP	VT	Target Class
LP	258 45.7%	3 0.5%	98.9% 1.1%
VT	27 4.8%	277 49.0%	91.1% 8.9%
	90.5% 9.5%	98.9% 1.1%	94.7% 5.3%

Matriz Sinteticos Test

Output Class	LP	VT	Target Class
LP	30 50.0%	1 1.7%	96.8% 3.2%
VT	3 5.0%	26 43.3%	89.7% 10.3%
	90.9% 9.1%	96.3% 3.7%	93.3% 6.7%



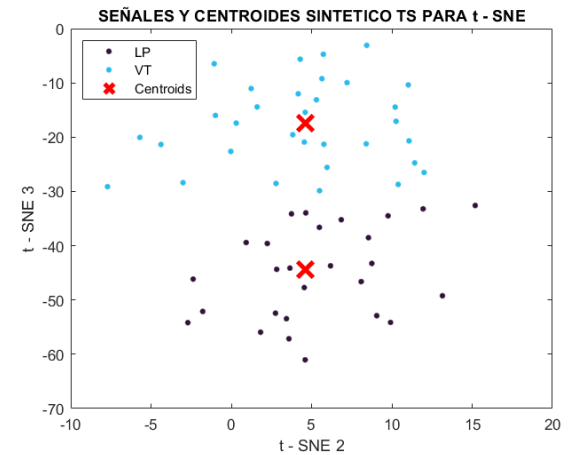
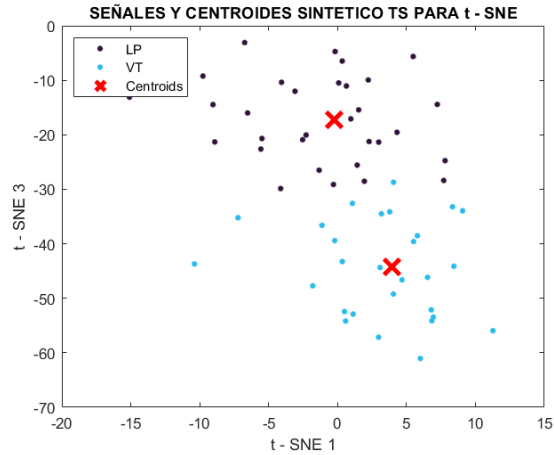
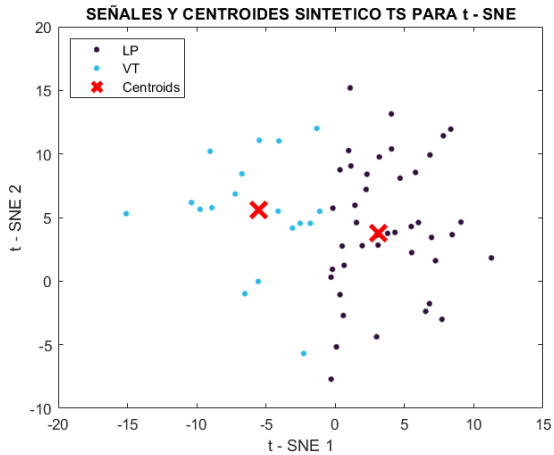
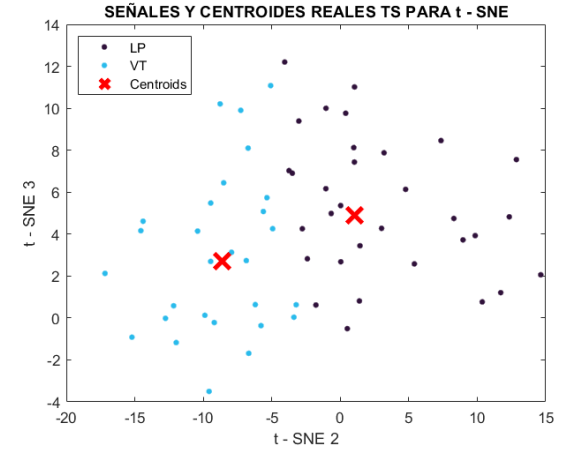
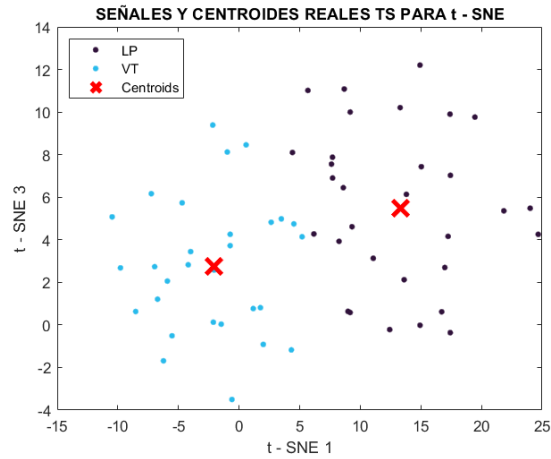
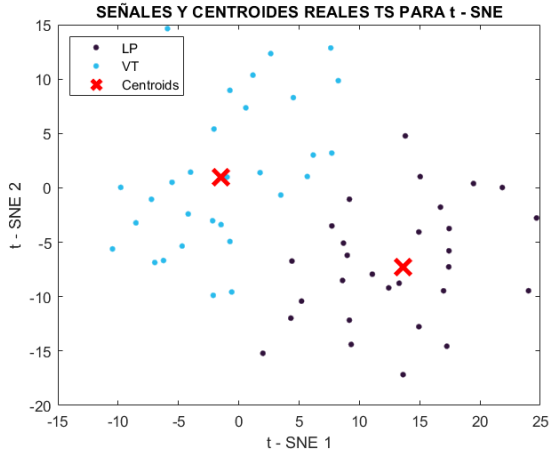
CENTROIDE	MODELO	DATOS REALES $(\mu_{m1}, \mu_{m2}, \mu_{m3})$	DATOS SINTETICOS $(\mu_{m1}, \mu_{m2}, \mu_{m3})$	Dispersión	Distancia Euclidiana TS
LP	PCA	(3.35, -8.24, 10.72)	(22.07, -4.94, 1.24)	21,25	36,95
	TSNE	(0.85, 1.31, 3.62)	(1.60, 3.94, -33.22)	36,95	21,25
	UMAP	(6.92, 6.40, -1.37)	(0.86, 1.74, -5.48)	8,69	8,69
VT	PCA	(-10.21, 10.54, -17.88)	(4.00, -1.70, 5.50)	22,48	33,18
	TSNE	(10.43, -6.41, 4.86)	(0.13, 5.35, -26.66)	35,19	35,19
	UMAP	(6.09, 5.85, -2.65)	(1.09, 1.58, -6.21)	7,48	7,479

CENTROIDE	MODELO	DATOS REALES (θ, ρ, z)	DATOS SINTETICOS (θ, ρ, z)	ANGULO DE DESFACE (θ)
LP	PCA	(292.13°, 8.89, 10.72)	(347.40°, 22.62, 1.24)	55.27°
	TSNE	(56.91°, 1.56, 3.63)	(67.82°, 4.25, -33.22)	10.91
	UMAP	(42.76°, 9.43, -1.37)	(63.78°, 1.94, -5.49)	21.02°
VT	PCA	(134.09°, 14.67, -17.88)	(203.04°, 4.35, 5.50)	68.95°
	TSNE	(328.44°, 12.24, 4.86)	(88.65°, 5.35, -26.66)	239.79°
	UMAP	(43.84°, 8.45, -2.65)	(55.41°, 1.92, -6.21)	11.57°



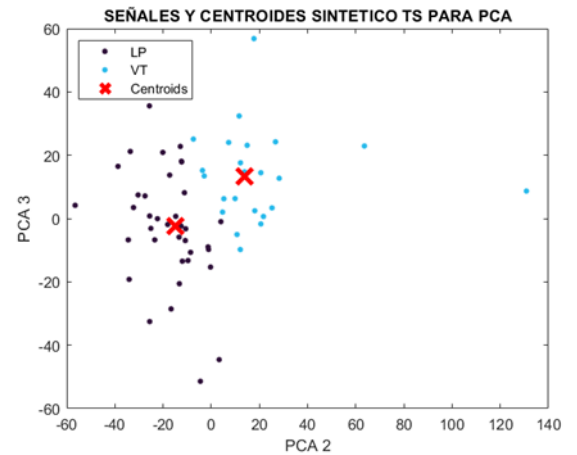
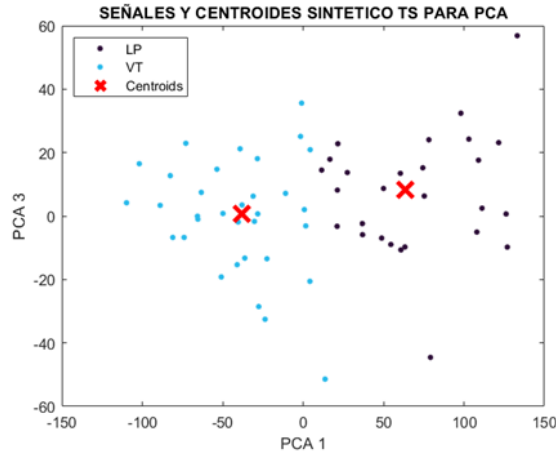
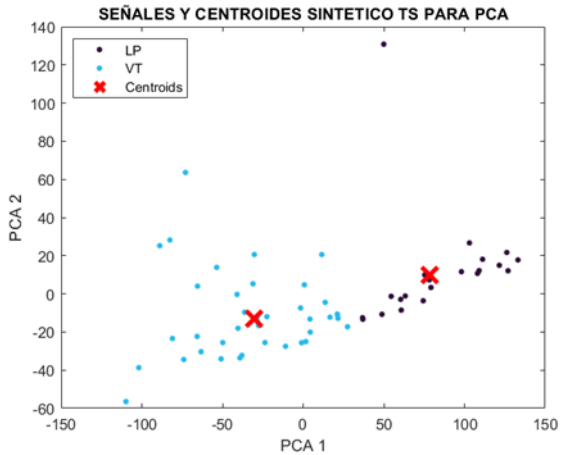
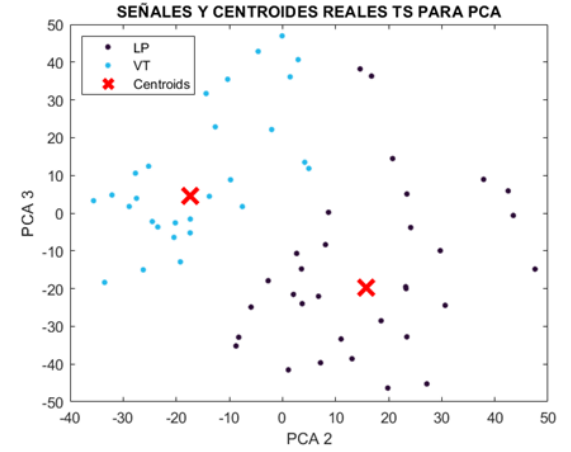
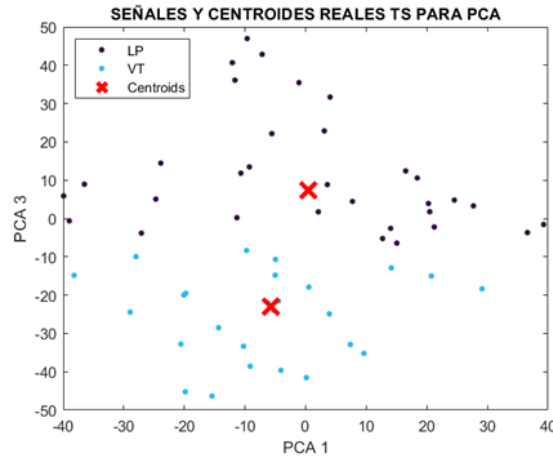
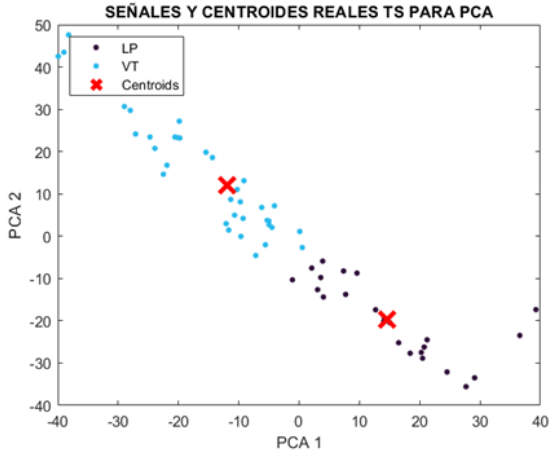
Comparación Gráfica (t - SNE)

Resultados



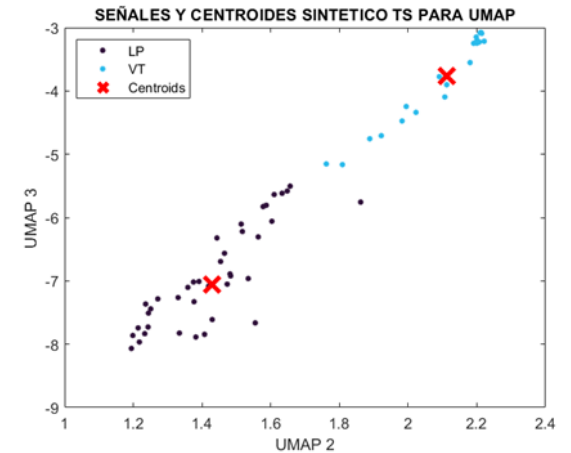
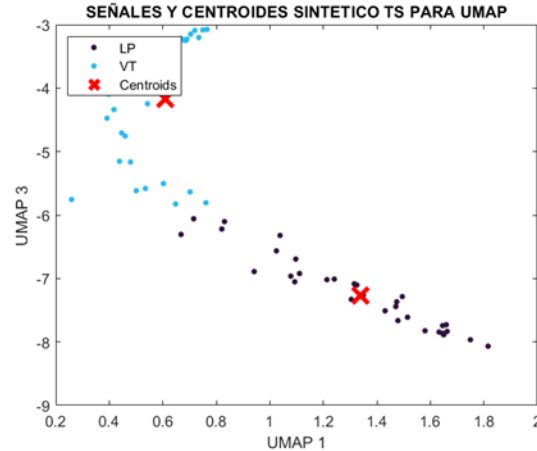
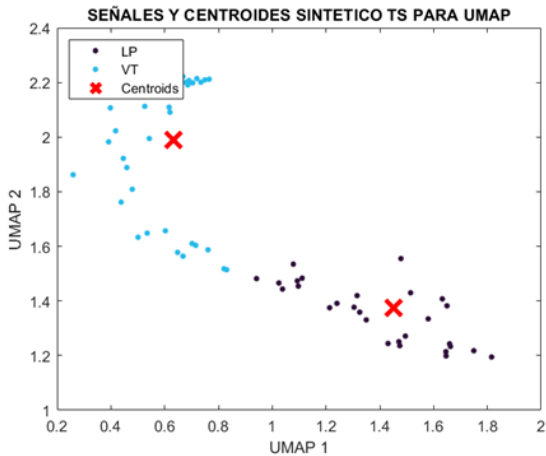
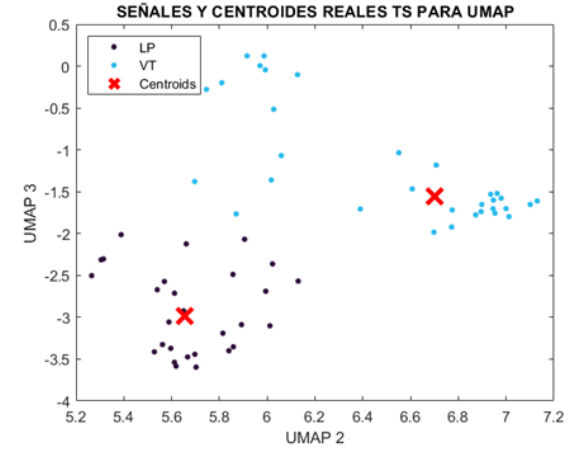
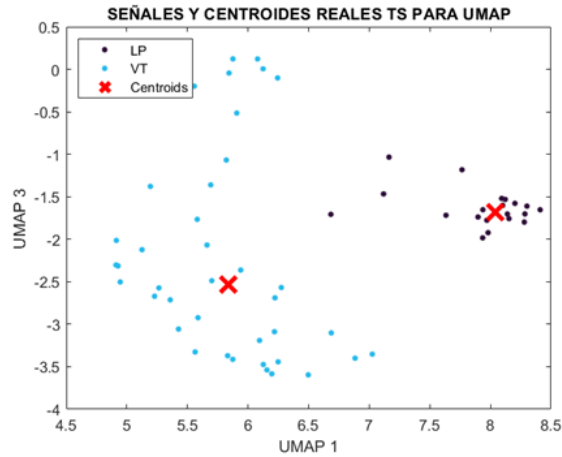
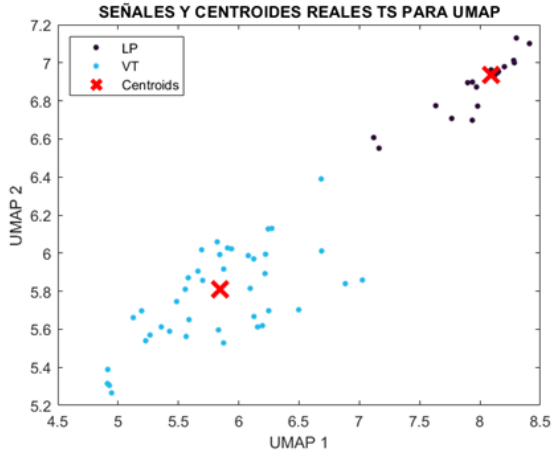
Comparación Gráfica (PCA)

Resultados



Comparación Gráfica (PCA)

Resultados





01. Introducción



02. Materiales



03. Diseño e Implementación



04. Resultados



05. Conclusiones y Recomendaciones



Conclusiones

- Al analizar los modelos de entrenamiento, propuestos en el trabajo y de acuerdo a los resultados obtenidos, a través de los espacios latentes, se pudo observar de forma objetiva y subjetiva que existe separabilidad entre los VT y LP.
- En función del porcentaje de acierto de un evento sísmico, se pudo determinar por medio a de la variación de autoencoders para training y test, que al mantener el 70% para cada uno, se obtiene el mejor desempeño al emplear la base de datos con data augmentation en el modelo de entrenamiento.



Conclusiones

- De acuerdo al trabajo realizado con el generador CGAN al momento de realizar data augmentation, no se logró obtener una mejora en cuanto al desempeño, dado que los porcentajes que reflejaba no superaban al trabajo previo, se presume que las señales sintéticas no contienen ciertas características acústicas, debido a que en el trabajo de (Iglesias Giler & Rosero Arciniega, 2023) si se aprecia una mejora en el desempeño pero al implementar otras características.
- Debido a que no se pudo encontrar una mejora al trabajar con data augmentation, no se encontró parámetros para la maximización del porcentaje de acierto, se procede a comparar los datos sintéticos con los reales, en las medidas objetivas, existen desfasajes, diferencias en las distancias euclidianas y en los centroides, esto se corrobora al emplear o utilizar tres métodos (t – SNE, PCA, UMAP).



Conclusiones

- De los resultados obtenidos de la Tabla 6 a la Tabla 9, se puede concluir que el mejor método al analizar reales como sintéticos fue UMAP, debido al desempeño que presentan en cuanto a la distancia y desfase, en comparación a los otros dos métodos.
- De las técnicas de las manifold learning; con UMAP se obtuvo un mejor resultado en la parte de test, al evaluar de forma cuantitativa a partir de las coordenadas rectangulares y polares de los centroides, con una separación de 8,69 para LP y 7.48 VT, dispersión de 8.69 para LP y 7.48 en VT y desfase de 21.02° en LP y 11.57° en VT.



Conclusiones

- En los parámetros libres del generador se presentaron distintas anomalías en el análisis gráfico de los centroides, como el desfase o posible ensanchamiento en cuanto a la frecuencia, se presume que esto es provocado, debido a que datos sintéticos no fueron trabajados en función del dominio de la frecuencia.



Recomendaciones

- Para llevar a cabo el análisis de los modelos de entrenamiento con espacios latentes, se debe considerar que los datos a emplear se encuentren en el dominio de la frecuencia y con el mismo tamaño, debido a que se pueden presentar diferentes errores en la obtención de las características espectrales, además de influir en la posición del centroide.
- En el momento de evaluar el desempeño de los modelos de entrenamiento al trabajar con autoencoders, se debe tomar en cuenta el porcentaje training y test con la finalidad de adecuar los datos al proceso a realizar, de esta forma se asegura el entrenamiento.



Recomendaciones

- Al mejorar el porcentaje de desempeño en cuanto a la optimización de los parámetros libres de los datos sintéticos, se debe tomar en cuenta que el generador fue realizado en el dominio del tiempo, por lo que se debe considerar adaptar este, en función de la frecuencia para evitar posibles desplazamientos del centroide o ensanchamiento de la frecuencia.
- Se debe tomar en cuenta la funcionalidad de los diferentes modelos al emplear los espacios latentes, dado que estos son aplicados en función del requerimiento, en caso de solo necesitar un análisis cualitativo puede emplearse $t - SNE$, por el contrario, si el estudio es cuantitativo se puede emplear UMAP o PCA.

