



# ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

## DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y TELECOMUNICACIONES CARRERA DE TELECOMUNICACIONES

**“Identificación de los intervalos de confianza,  
anchos de banda y evolución temporal de las  
principales señales sísmicas del volcán Llaima”**

***Autores:*** Arequipa Moreta Xavier Alexander,  
Villacrés Figueroa Esteban Eduardo

***Director del Proyecto:*** Ing. Román Alcides Lara Cueva, PhD.

*07 de septiembre de 2023*

VERSIÓN: 1.1



# Agenda

- 1.- **Introducción**
- 2.- **Desarrollo**
- 3.- **Pruebas y Resultados**
- 4.- **Conclusiones y Trabajos futuros**



# 1. Introducción

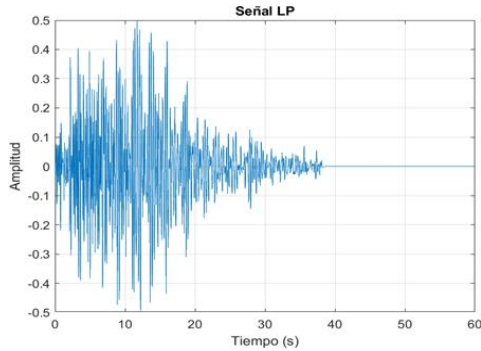
## Motivación e Importancia



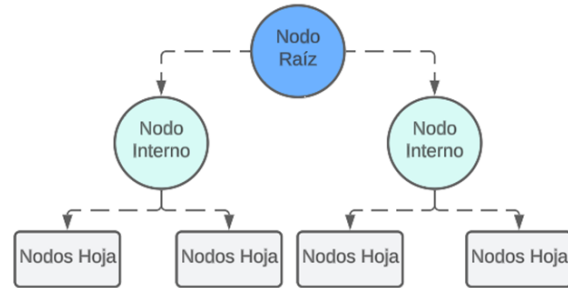
# 1. Introducción

## Antecedentes

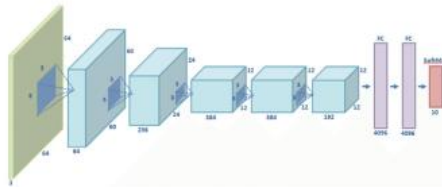
Análisis de Espectro de Frecuencia



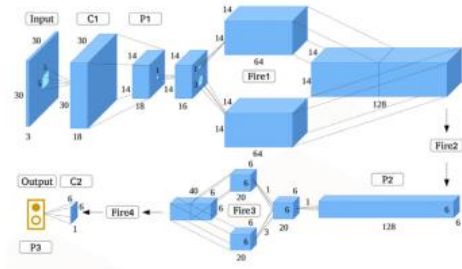
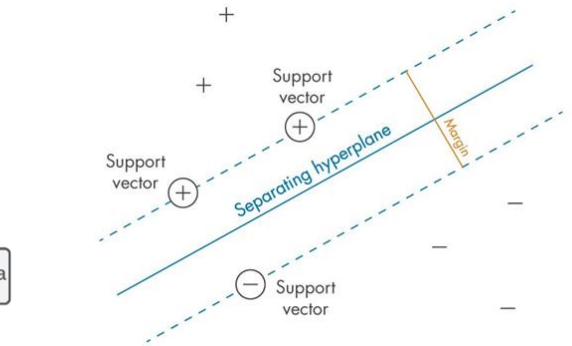
Machine Learning Tradicional



Deep Learning



(a) AlexNet [26]



(b) SqueezeNet [27]

# 1. Introducción

## ***General***

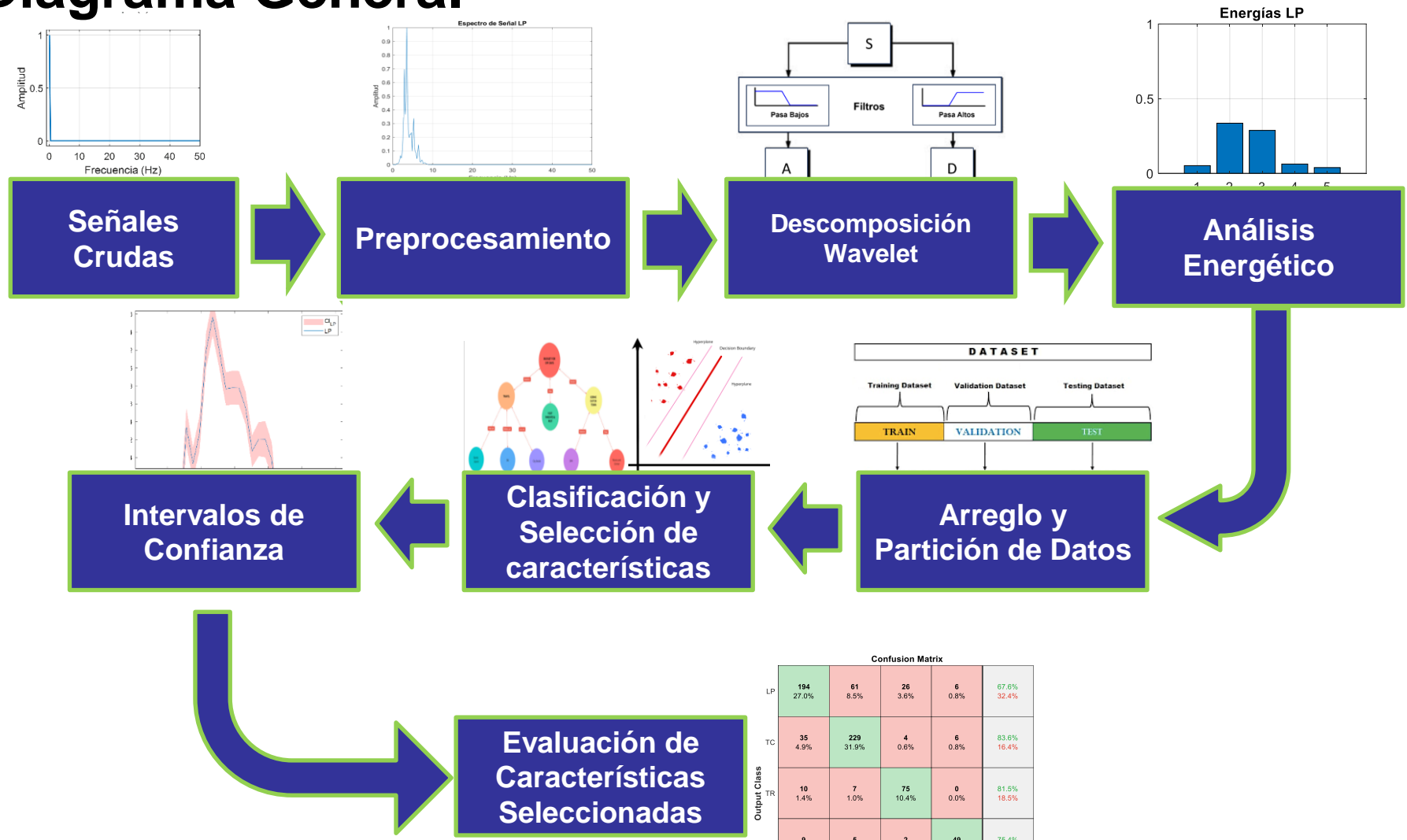
Identificar de los intervalos de confianza, anchos de banda y evolución temporal de las principales señales sísmicas del volcán Llaima.

## ***Actividades***

1. Identificación de las propiedades y variables del problema a tener en cuenta.
2. Reunión de la base de datos representativa, con datos provistos por el observatorio Observatorio Volcanológico de los Andes del Sur.
3. Creación de una estructura de datos (temporales, espaciales, otros.) con un soporte común.
4. Pruebas con estimadores espectrales de series temporales y técnicas de re-muestreo.
5. Pruebas y evaluación del desempeño.



# 2. Desarrollo Diagrama General



Confusion Matrix

	LP	TC	TR	VT	
LP	194 27.0%	61 8.5%	26 3.6%	6 0.8%	67.6% 32.4%
TC	35 4.9%	229 31.9%	4 0.6%	6 0.8%	83.6% 16.4%
TR	10 1.4%	7 1.0%	75 10.4%	0 0.0%	81.5% 18.5%
VT	9 1.3%	5 0.7%	2 0.3%	49 6.8%	75.4% 24.6%
	78.2% 21.8%	75.6% 24.2%	70.1% 29.9%	80.3% 19.7%	76.2% 23.8%
	Target Class				



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

# 2. Desarrollo

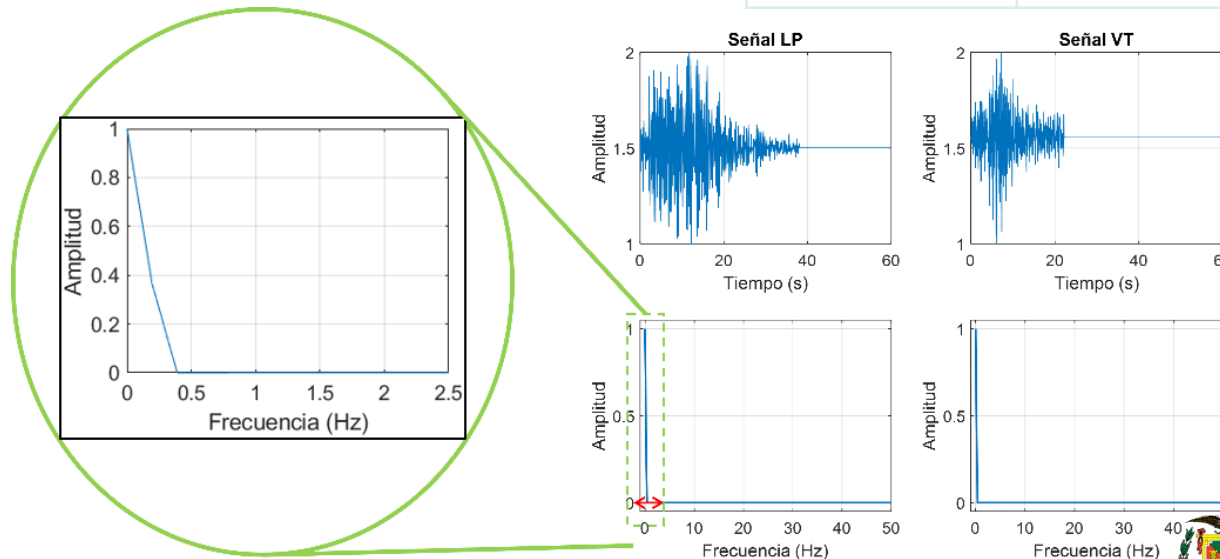
## Señales Crudas

El conjunto de datos crudos consta de un total de 3592 señales representadas por  $M$  y contenidas en la matriz:

$$S = \{s_1^T, s_2^T, \dots, s_M^T\},$$

Cada señal está **muestreada a 100 Hz**.  
Cada señal tiene **una duración de 60 [s]**.  
Cada señal tiene **6 000 muestras**

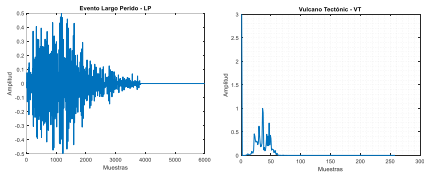
Tipo	Número de señales	Porcentaje
LP	1310	36.47 %
TC	1488	41.42 %
TR	490	13.64 %
VT	304	8.46 %



# 2. Desarrollo

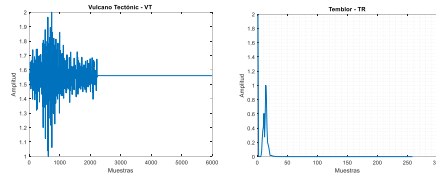
## Preprocesamiento

### Largo Período (LP)



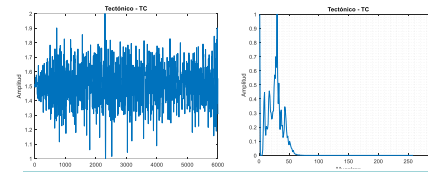
Sismos de baja frecuencia  
Originados cerca de la superficie, a profundidades menores a 1km.  
Duración entre algunos segundos hasta algo más de 1 minuto.  
Por lo general, sus frecuencias se encuentran en el rango de 0.5 a 5 Hz.

### Tremor (TR)



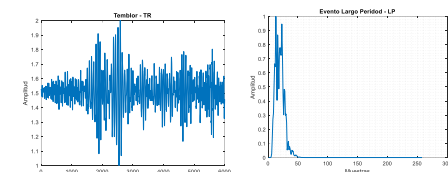
Señales con amplitud constante.  
Periodo de tiempo que puede oscilar entre varios minutos hasta varios días.  
Se deben a deslizamientos de tierra y avalanchas.  
Frecuencias superiores a 6 Hz.

### Tectónico (TC)



Son aquellos que no son producidos por el volcán y están vinculados con fallas geológicas  
La forma en cómo se diferencia los VT de los TC es por el análisis en el tiempo de llegada.  
Frecuencias de 0.1 Hz hasta varios Hz.

### Volcano Tectónico (VT)



Señal de duración variable  
Amplio contenido espectral  
Se presentan en un amplio rango de profundidades, desde las decenas de kilómetros.  
Frecuencias entre 5 y 15 Hz.





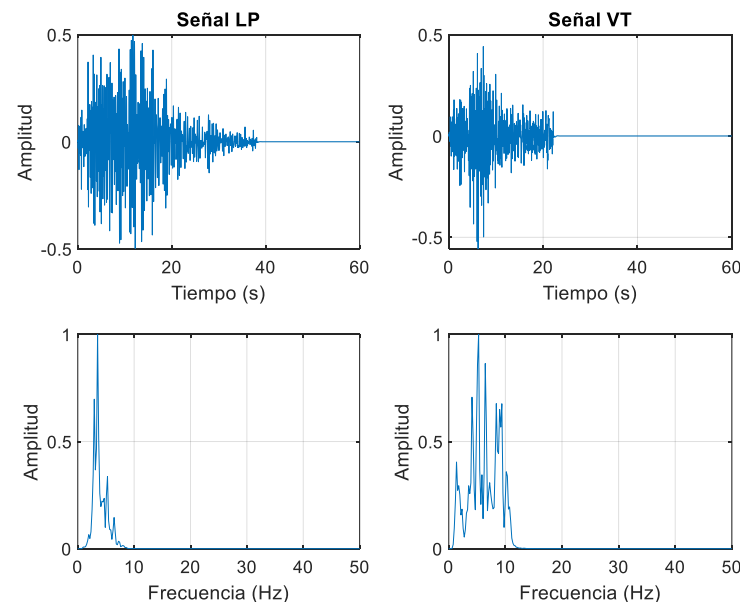
# 2. Desarrollo

## Preprocesamiento

### Filtrado y Normalización

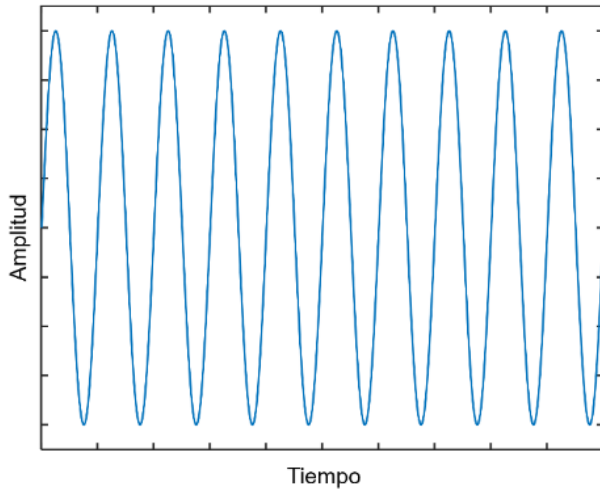
Se utiliza un filtro FIR, para mejorar la calidad de las señales, separar componentes y eliminar ruido, de 0.7 a 49.5 Hz; además de un proceso de normalización en cada señal  $s_M$ . Con ello, se obtienen los vectores  $h_M$  como resultado de aplicar el operador  $h_M\{s_M\}$ , el cual representa la operación matemática de aplicar el proceso anterior.

$$H = \{h_1^T, h_2^T, h_3^T, \dots, h_N^T\}^T.$$

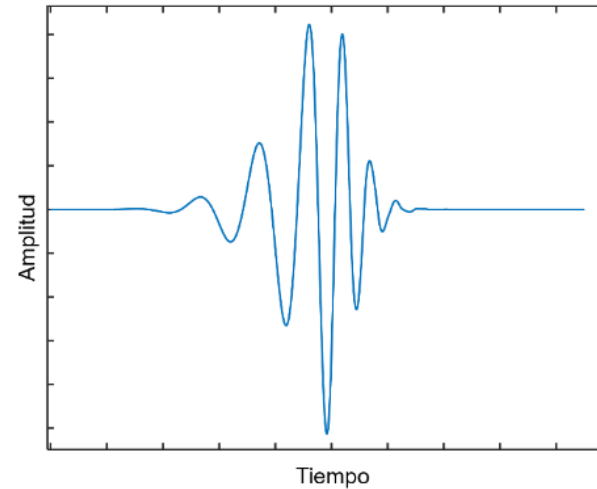


# 2. Desarrollo

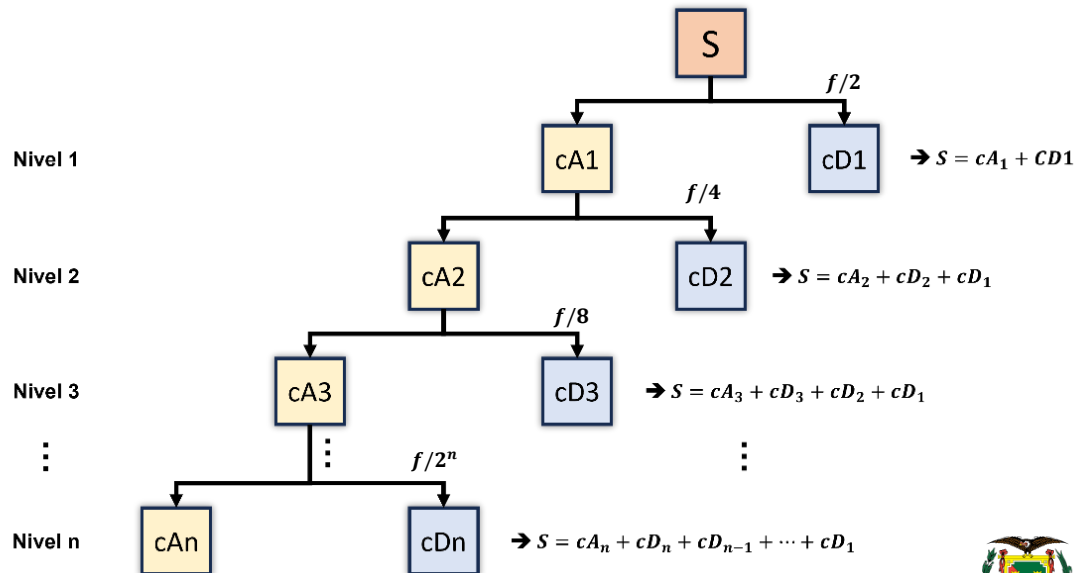
## Descomposición Wavelet



a) Señal sinusoidal



b) Señal wavelet



# 3. Desarrollo

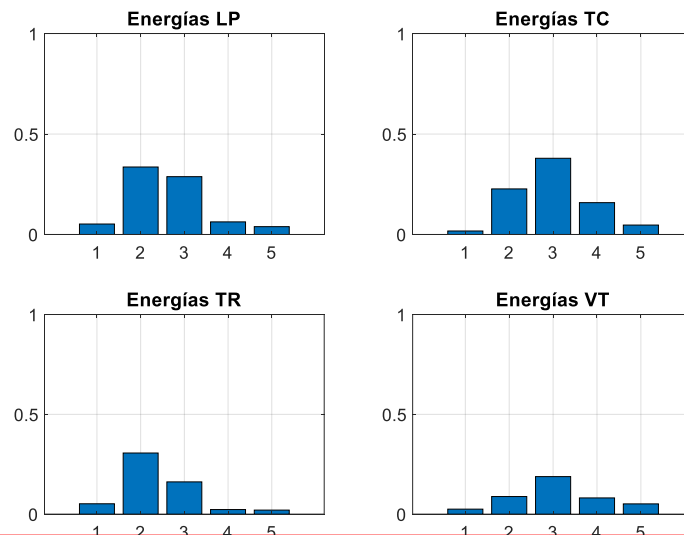
## Análisis Energético

La presente etapa consiste en obtener los valores de energía a partir de los coeficientes wavelet. Para proceder con el cálculo de la energía de la matriz H, a través de la ecuación:

$$E = \sum_{i=1}^M |w_i|^2,$$

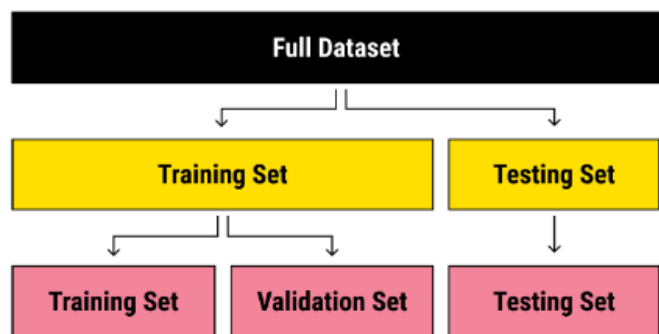
donde  $w$  representa cada coeficiente wavelet de cada señal. De esta forma se genera la matriz  $E$  que contiene las energías:

$$E = [e_1^T, e_2^T, \dots, e_M^T]^T.$$



# 2. Desarrollo

## Arreglo de Datos



DT  
Auto

DT.  
V.C.

DT  
Poda.

SVM

División de datos	
Train: 50% - Test: 50%	
Train: 60% - Test: 40%	
Train: 70% - Test: 30%	
Train: 80% - Test: 20%	
Train: 90% - Test: 10%	

División de datos	
Train: 50% - Test: 50%	
Train: 60% - Test: 40%	
Train: 70% - Test: 30%	
Train: 80% - Test: 20%	
Train: 90% - Test: 10%	

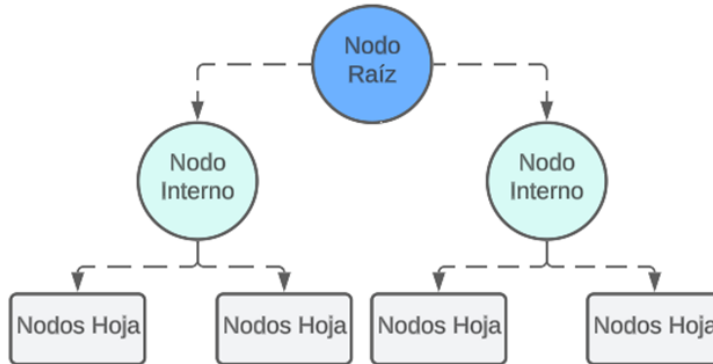
División de datos	
Train: 50% - Test: 50%	
Train: 60% - Test: 40%	
Train: 70% - Test: 30%	
Train: 80% - Test: 20%	
Train: 90% - Test: 10%	

División de datos	
Train: 50% - Test: 50%	
Train: 60% - Test: 40%	
Train: 70% - Test: 30%	
Train: 80% - Test: 20%	
Train: 90% - Test: 10%	

# 2. Desarrollo

## Clasificación y Extracción de Características

- Árboles de Decisión (DT)



$$E(t) = - \sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t)$$

$$G(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2$$

Parámetros Libres DT	
<b>MinLeafSize</b>	Número mínimo de observaciones que se requiere para formar una hoja (nodo terminal) en el árbol.
<b>NumNodes</b>	Indica la cantidad total de nodos en el árbol.
<b>MaxDepth</b>	Indica la profundidad máxima del árbol, es decir, el número máximo de niveles o capas de nodos.
<b>SplitCriterion</b>	Hiperparámetro que controla el criterio de división utilizado, como Gini o entropía.
<b>Prune</b>	Controla si se debe realizar la poda del árbol después de su construcción.
<b>CrossVal</b>	Indica si se debe realizar la validación cruzada para evaluar el rendimiento del modelo.



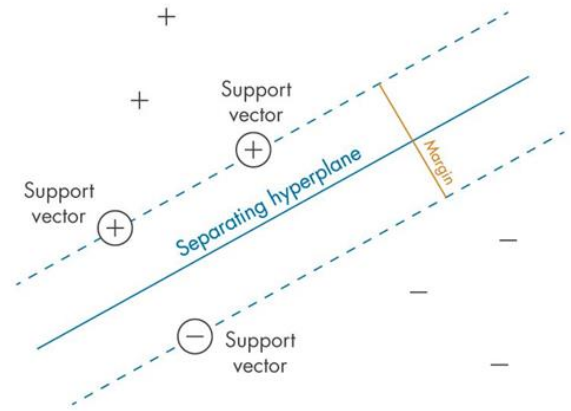
# 2. Desarrollo

## Clasificación y Extracción de Características

- Máquinas de Vector de Soporte (SVM)

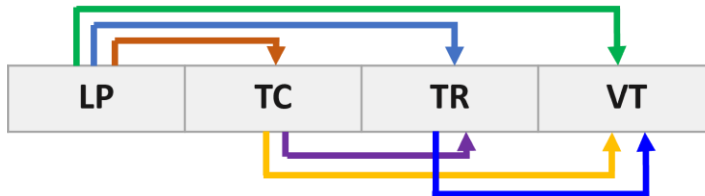
### Parámetros Libres

<b>KernelFunction</b>	Qué función de kernel fue utilizada para transformar los datos.
<b>Cost</b>	Controla la penalización por clasificar erróneamente muestras de entrenamiento
<b>CodingName</b>	Indica la estrategia específica utilizada para clasificar. (onevsone, onevsall)
<b>BinaryLearners</b>	Indica el número de clasificadores binarios utilizados para realizar la clasificación multiclase en función de la estrategia de clasificación.



$$f(x) = w \cdot x + b$$

### Máquinas de Vector de Soporte (SVM)



CLASIFICADOR DE CLASIFICADORES	Clasificadores SVM Binarios
	LP vs TC
	LP vs TR
	LP vs VT
	TC vs TR
	TC vs VT
	TR vs VT

# 2. Desarrollo

## Clasificación y Extracción de Características

### Métricas de Rendimiento

Matriz de confusión, modelo con dos clases.

		Clase Verdadera	
		Positive	Negative
Clase Predicha	Positive	True positives	False positives
	Negative	False negatives	True negatives

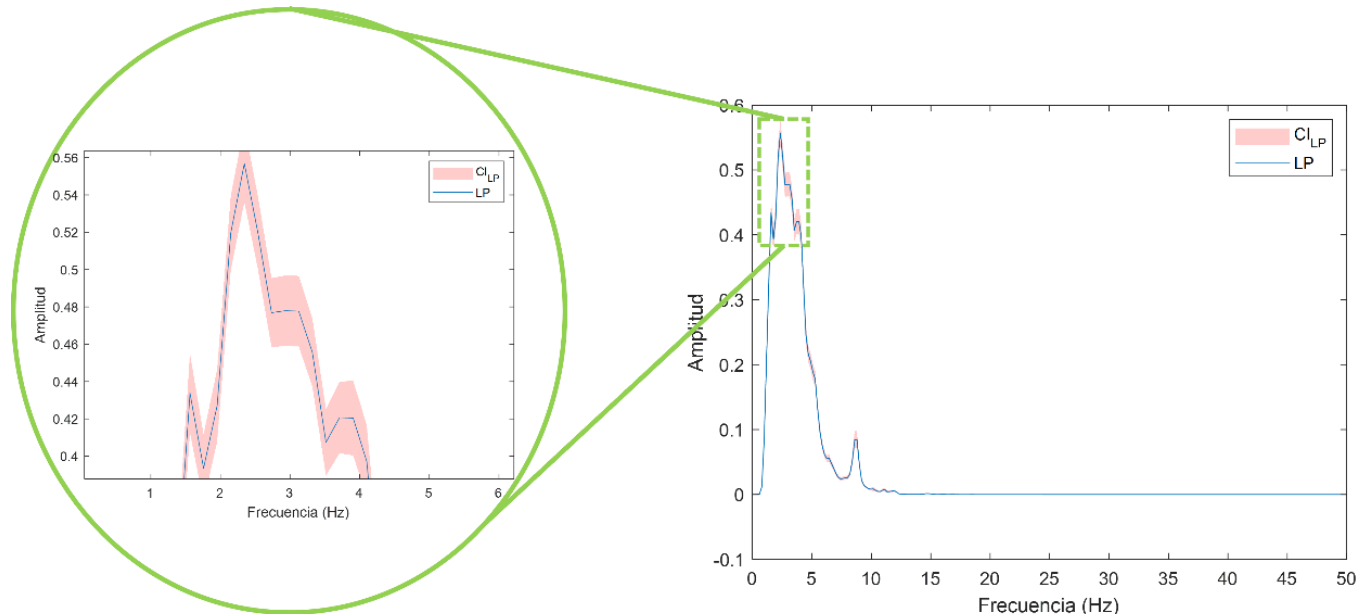
Métrica de Desempeño	Fórmula
Exactitud (%)	$A = \frac{TP + TN}{TP + TN + FP + FN} \times 100$
Precisión (%)	$P = \frac{TP}{TP + FP} \times 100$
Sensibilidad (%)	$R = \frac{TP}{TP + FN} \times 100$
Especificidad (%)	$S = \frac{TN}{TN + FP} \times 100$
BER	$BER = 1 - \frac{R + S}{200}$

# 2. Desarrollo

## Intervalos de Confianza

El intervalo de confianza para la media  $\mu$ , con desviación estándar desconocida se calcula con la siguiente fórmula:

$$\left( \bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right).$$



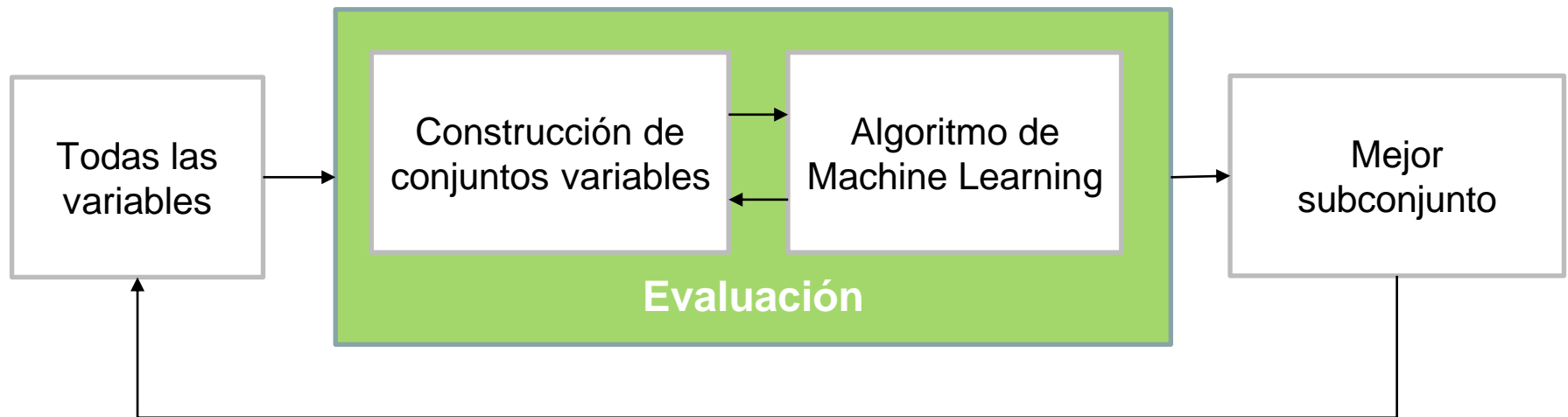
Se utilizó un intervalo de confianza del **99%**.



## 2. Desarrollo

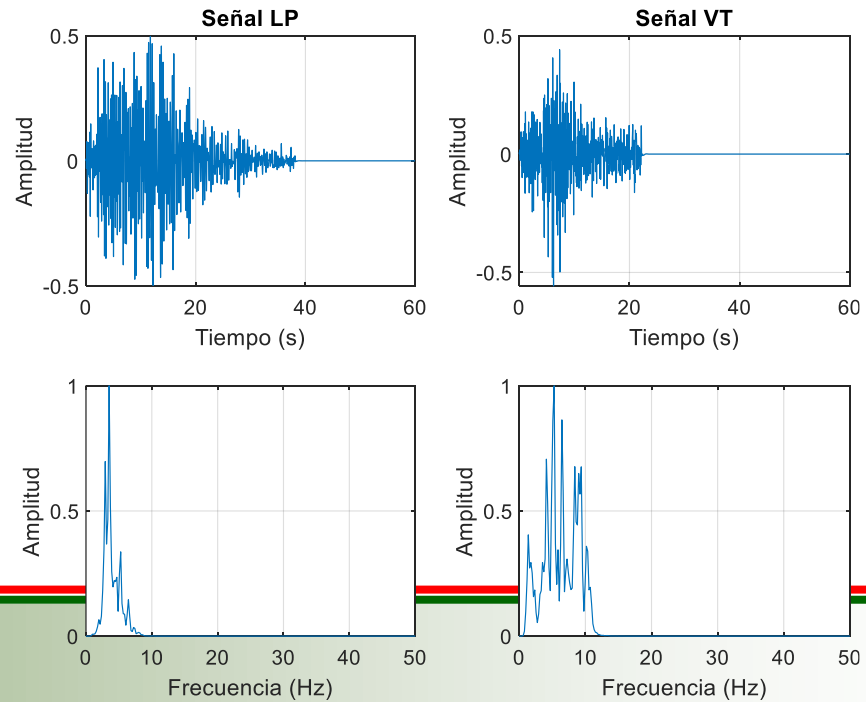
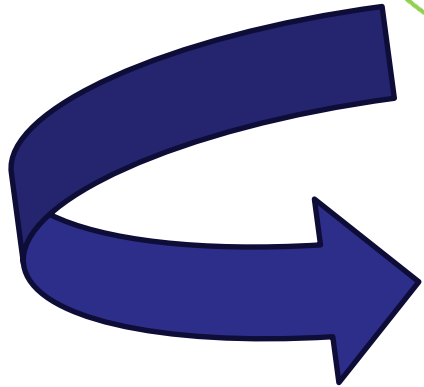
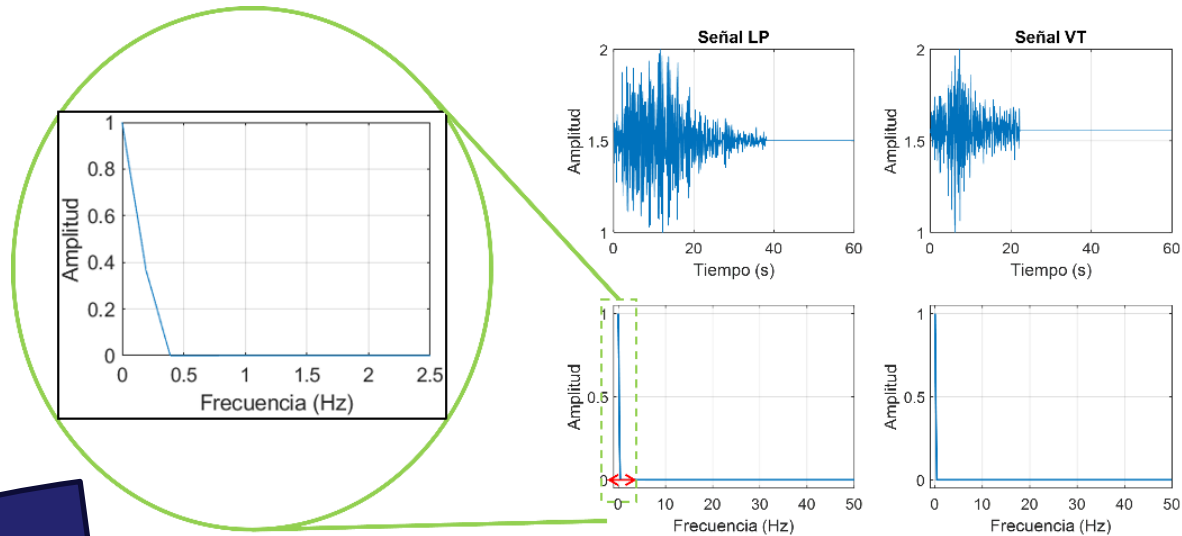
### Evaluación de características seleccionadas

Con el fin de evaluar el rendimiento de las características seleccionadas se procede a clasificar los microsismos nuevamente.



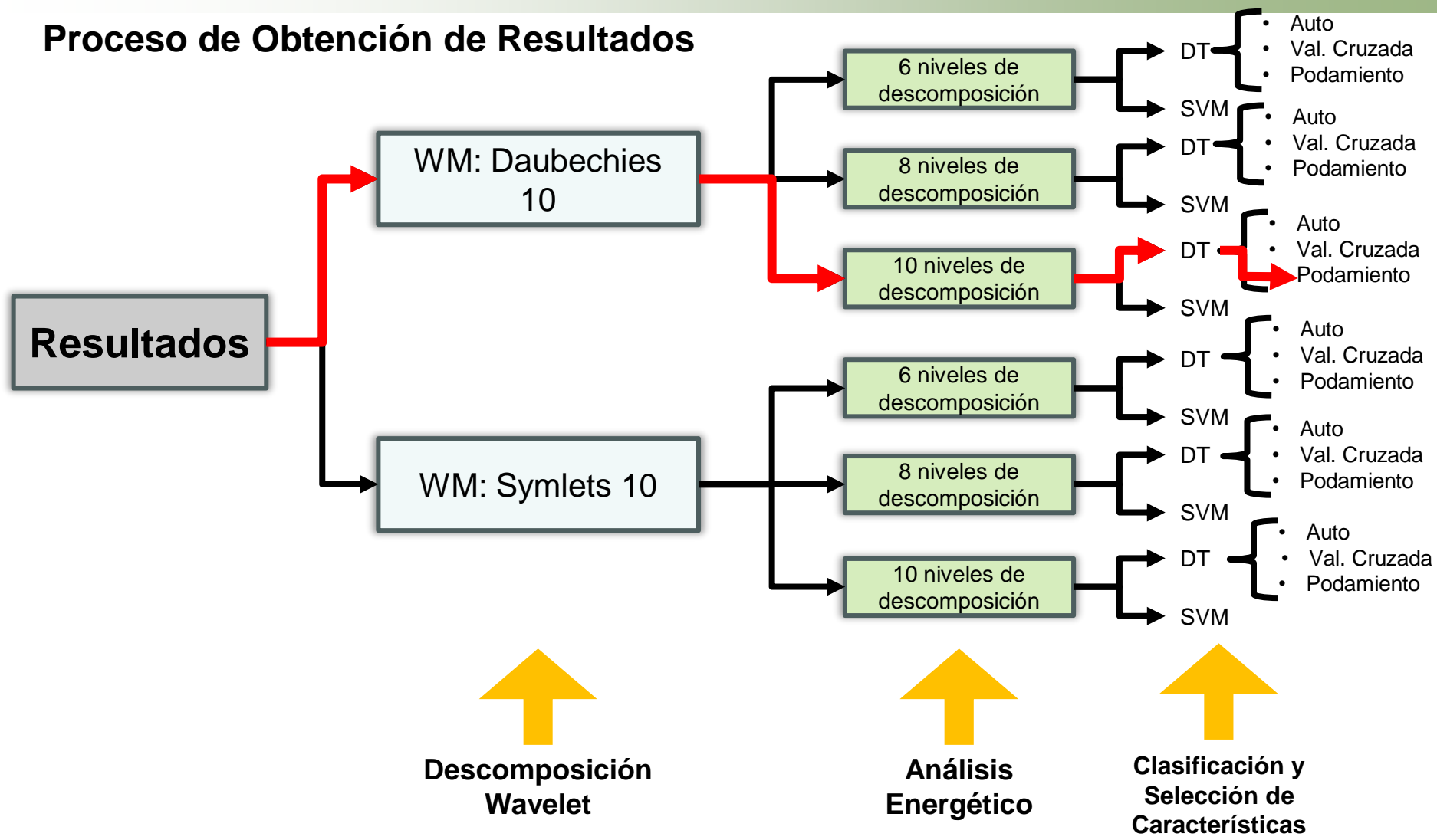
# 3. Pruebas y Resultados

## Preprocesamiento



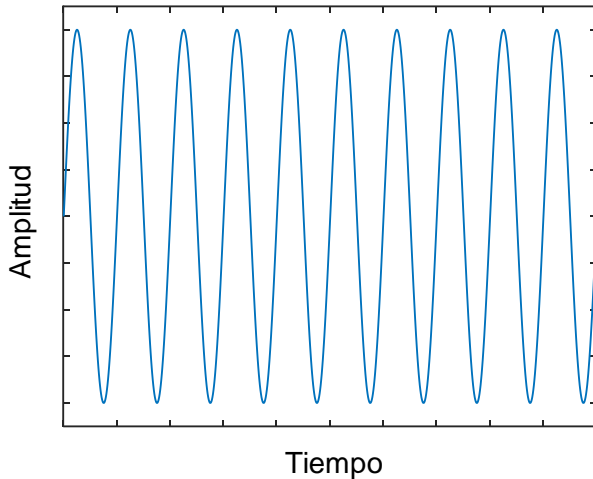
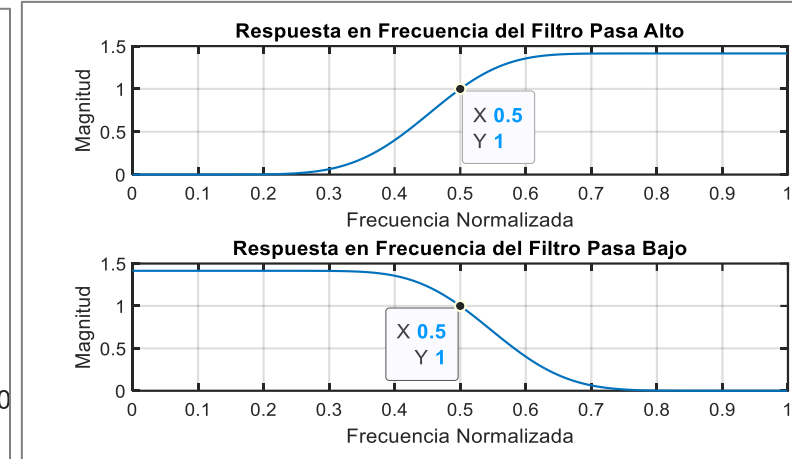
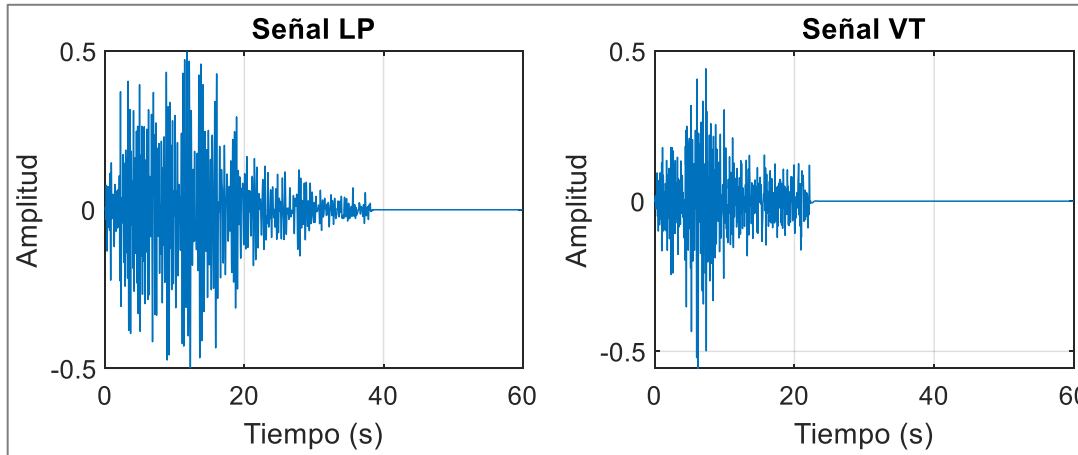
# 3. Pruebas y Resultados

## Proceso de Obtención de Resultados

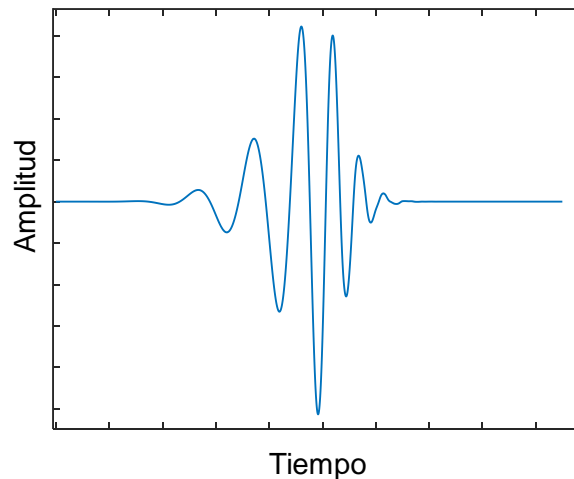


# 3. Pruebas y Resultados

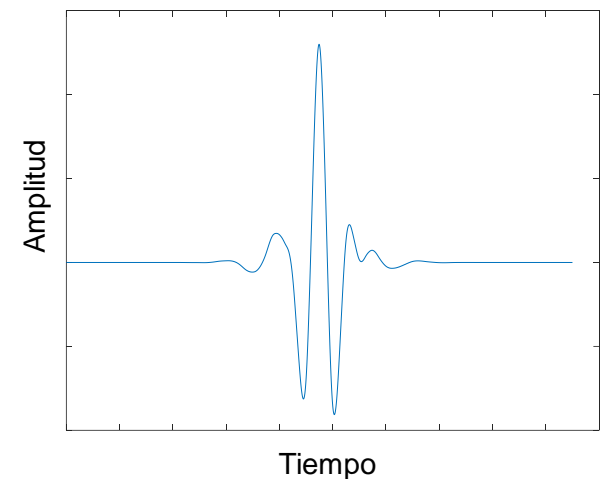
## Elección Wavelet Madre



a) Señal sinusoidal



b) Wavelet Daubechies orden 10



c) Wavelet Symlets orden 10

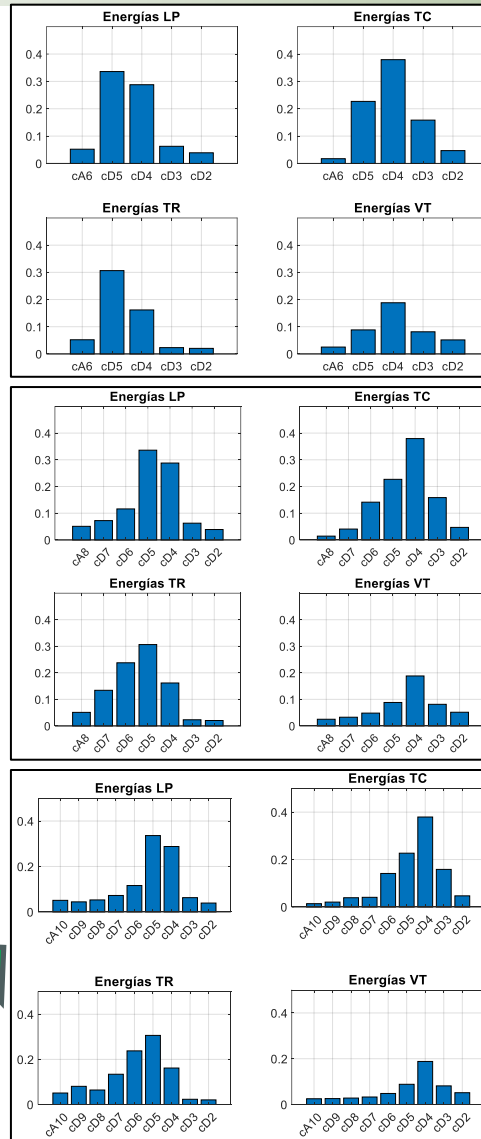
# 3. Pruebas y Resultados

## Energías de Coef. Wavelet

Nivel de Desc.	Etiqu.	Rangos de Frec.	Bandwidth
1	cD1	25 - 50	25
2	cD2	12,5 - 25	12,5
3	cD3	6,25 - 12,5	6,25
4	cD4	3,12 - 6,25	3,12
5	cD5	1,56 - 3,12	1,56
6	cD6	0,78 - 1,56	0,78
7	cD7	0,39 - 0,78	0,39
8	cD8	0,195 - 0,39	0,195
9	cD9	0,0975 - 0,195	0,0975
10	cD10	0,04875 - 0,0975	0,04875
10	cA10	0 - 0,04875	0,04875

Detalle de bandas de Frecuencia de Wavelet

Se emplea como MW: Daubechies y Symlets. Pero a continuación se presentan los mejores resultados:



6 Niveles



10 Niveles

Se determina:  
El nivel ideal de descomposición es de 10 niveles. Por:

- Número de características.
- Permite una mejor identificación.
- Por prueba y error, arroja los mejores resultados.

Descomposición Wavelet Daubechies



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

# 3. Pruebas y Resultados

## Arreglo y partición de datos

DT  
Auto

División de datos	A (%)	P (%)	S (%)	R (%)	BER
Train: 50% - Test: 50%	80,3	80,5	92,5	75	0,2
Train: 60% - Test: 40%	81,5	82,2	93	77,9	0,2
Train: 70% - Test: 30%	83,1	82,9	93,7	79,1	0,1
Train: 80% - Test: 20%	85,6	82,9	92,7	81,3	0.09
Train: 90% - Test: 10%	84,7	81,9	94,4	81,2	0,1

DT.  
V.C.

División de datos	A (%)	P (%)	S (%)	R (%)	BER
Train: 50% - Test: 50%	78,5	77,9	91,9	73,8	0,2
Train: 60% - Test: 40%	80,6	81,2	92,6	76,1	0,2
Train: 70% - Test: 30%	79,5	79,3	92,3	76,3	0,2
Train: 80% - Test: 20%	83,3	80,1	94.1	80,8	0.13
Train: 90% - Test: 10%	83,3	82,5	93,7	79,8	0,15

DT.  
Poda.

División de datos	A (%)	P (%)	S (%)	R (%)	BER
Train: 50% - Test: 50%	80,5	82,2	92,4	74,6	0,2
Train: 60% - Test: 40%	82,7	83,7	93,4	78,1	0,1
Train: 70% - Test: 30%	84,2	83,1	94,2	80	0,1
Train: 80% - Test: 20%	84,5	83,9	94,3	80,9	0.12
Train: 90% - Test: 10%	84,1	82,4	94,2	80,2	0,1

SVM

División de datos	A (%)	P (%)	S (%)	R (%)	BER
Train: 50% - Test: 50%	85,2	84,8	94,4	80,4	0,1
Train: 60% - Test: 40%	86,1	86,4	94,7	80,7	0,1
Train: 70% - Test: 30%	86,3	87,3	94,7	80,9	0,1
Train: 80% - Test: 20%	87,7	88,0	95,4	82,8	0,11
Train: 90% - Test: 10%	84,7	81,9	94,4	81,2	0,1



# 3. Pruebas y Resultados

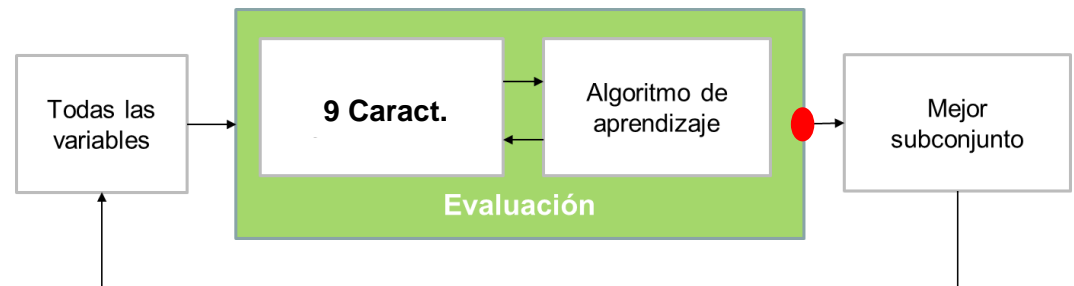
## Métricas de Modelos ML

Métrica MW	A (%)		P (%)		S (%)		R (%)		BER	
	Db	Sym	Db	Sym	Db	Sym	Db	Sym	Db	Sym
DT – Auto.	84	78	81	77	94	92	81	78	0,13	0,15
DT- Val Cruz.	83	79	80	78	94	92	81	79	0,13	0,14
DT – Poda.	85	81	84	81	94	93	81	81	0,12	0,13
<b>SVM</b>	<b>88</b>	<b>82</b>	<b>88</b>	<b>81</b>	<b>95</b>	<b>93</b>	<b>83</b>	<b>79</b>	<b>0.11</b>	<b>0,14</b>

Confusion Matrix

Output Class	LP	TC	TR	VT	
LP	238 33.1%	10 1.4%	27 3.8%	7 1.0%	84.4% 15.6%
TC	15 2.1%	270 37.6%	2 0.3%	8 1.1%	91.5% 8.5%
TR	6 0.8%	9 1.3%	73 10.2%	1 0.1%	82.0% 18.0%
VT	2 0.3%	1 0.1%	0 0.0%	49 6.8%	94.2% 5.8%
	91.2% 8.8%	93.1% 6.9%	71.6% 28.4%	75.4% 24.6%	87.7% 12.3%
	LP	TC	TR	VT	
	Target Class				

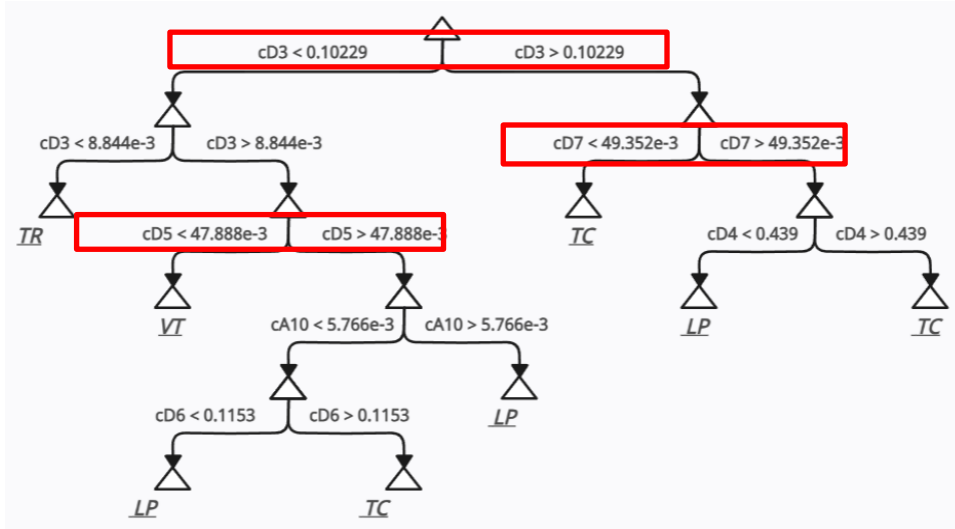
Matriz de Confusión para SVM



# 3. Pruebas y Resultados

## Identificación de Características

- Árboles de Decisión (DT)



Características con DT	
<b>cD3</b>	6,25 – 12,5 Hz
<b>cD5</b>	1,56 – 3,12 Hz
<b>cD7</b>	0,39 – 0,78 Hz

Pesos en modelo DT

Parámetros Libres DT	
<b>MinLeafSize</b>	6
<b>NumNodes</b>	193
<b>MaxDepth</b>	3
<b>SplitCriterion</b>	Gini
<b>Prune</b>	-
<b>CrossVal</b>	-

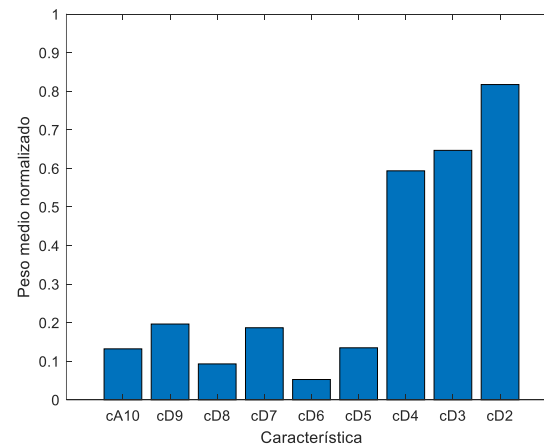
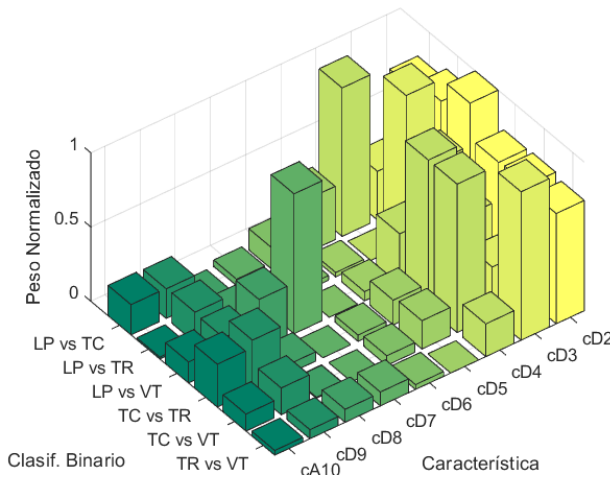


# 3. Pruebas y Resultados

## Identificación de Características

- Máquinas de Vector de Soporte (SVM)

Parámetros Libres	
KernelFunction	Linear
Cost	1 cuando se equivoca y 0 cuando no
CodingName	onevsone
BinaryLearners	6



Características con SVM	
cD4	3,12 - 6,25 Hz
cD3	6,25 - 12,5 Hz
cD2	3,12 - 6,25 Hz

Pesos en modelo SVM

# 3. Pruebas y Resultados

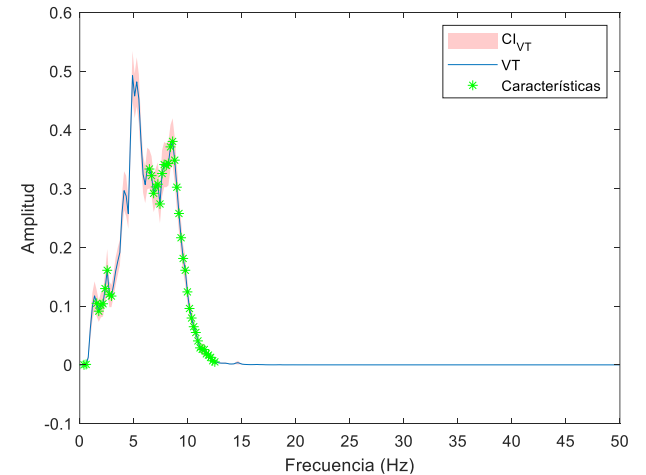
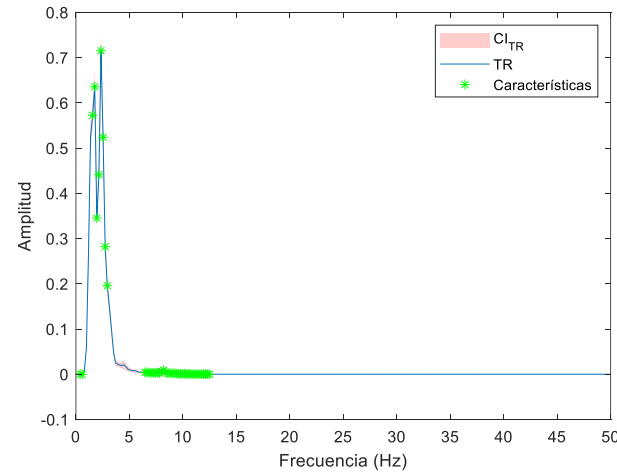
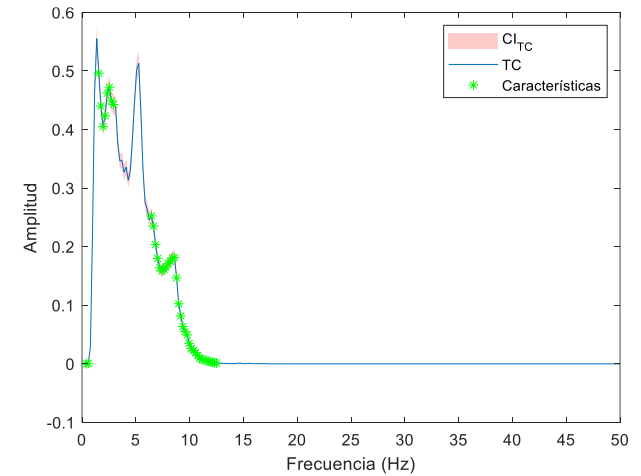
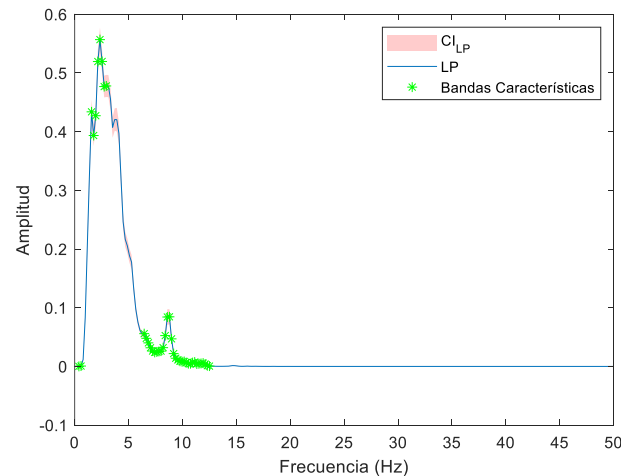
## PSD Welch Promedio e Intervalos de Confianza – Modelo DT

### Características con DT

**cD3**      6,25 – 12,5 Hz

**cD5**      1,56 – 3,12 Hz

**cD7**      0,39 – 0,78 Hz



Grado de Confianza: 99%  
Exactitud de clasificación: 84.5%



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

# 3. Pruebas y Resultados

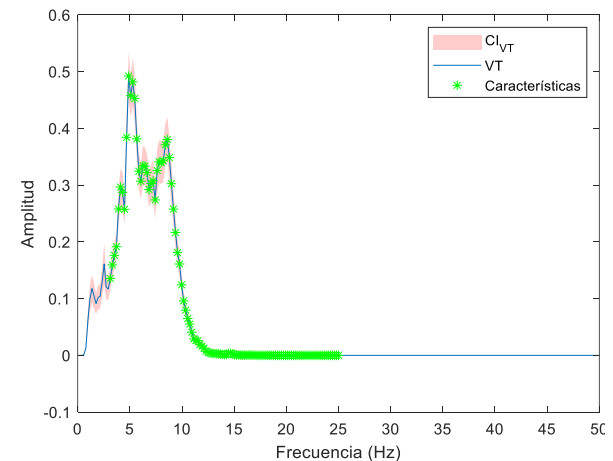
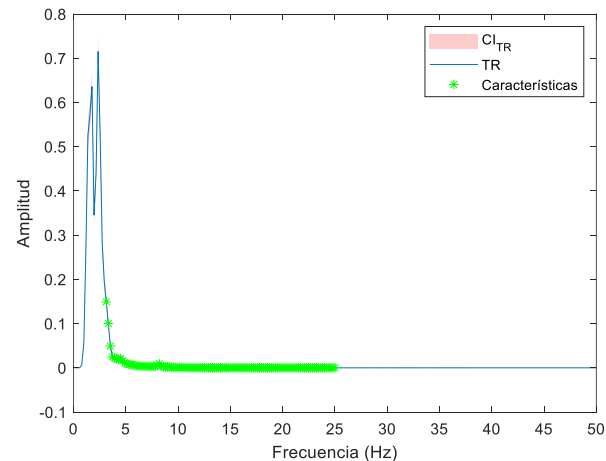
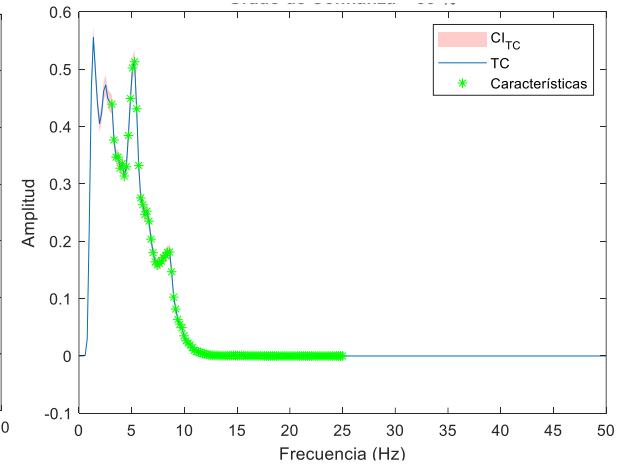
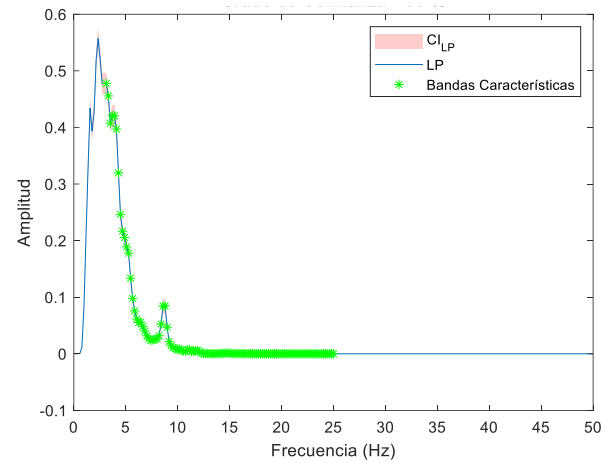
## PSD Welch Promedio e Intervalos de Confianza – Modelo SVM

### Características con SVM

**cD4** 3,12 - 6,25 Hz

**cD3** 6,25 – 12,5 Hz

**cD2** 3,12 - 6,25 Hz



Grado de Confianza: 99%  
Exactitud de clasificación: 87.7%



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

# 3. Pruebas y Resultados

## Reentrenamiento

Al reentrenar los modelos ML con las tres características identificadas se obtiene:

Método	A (%)	P (%)	S (%)	R (%)	BER
DT - Defecto	75	75	91	75	0,19
DT - Validación Cruzada	71	72	89	72	0.2
DT - Podamiento	<b>76</b>	<b>77</b>	<b>91</b>	<b>76</b>	<b>0.16</b>
SVM	66	50	87	50	0.31

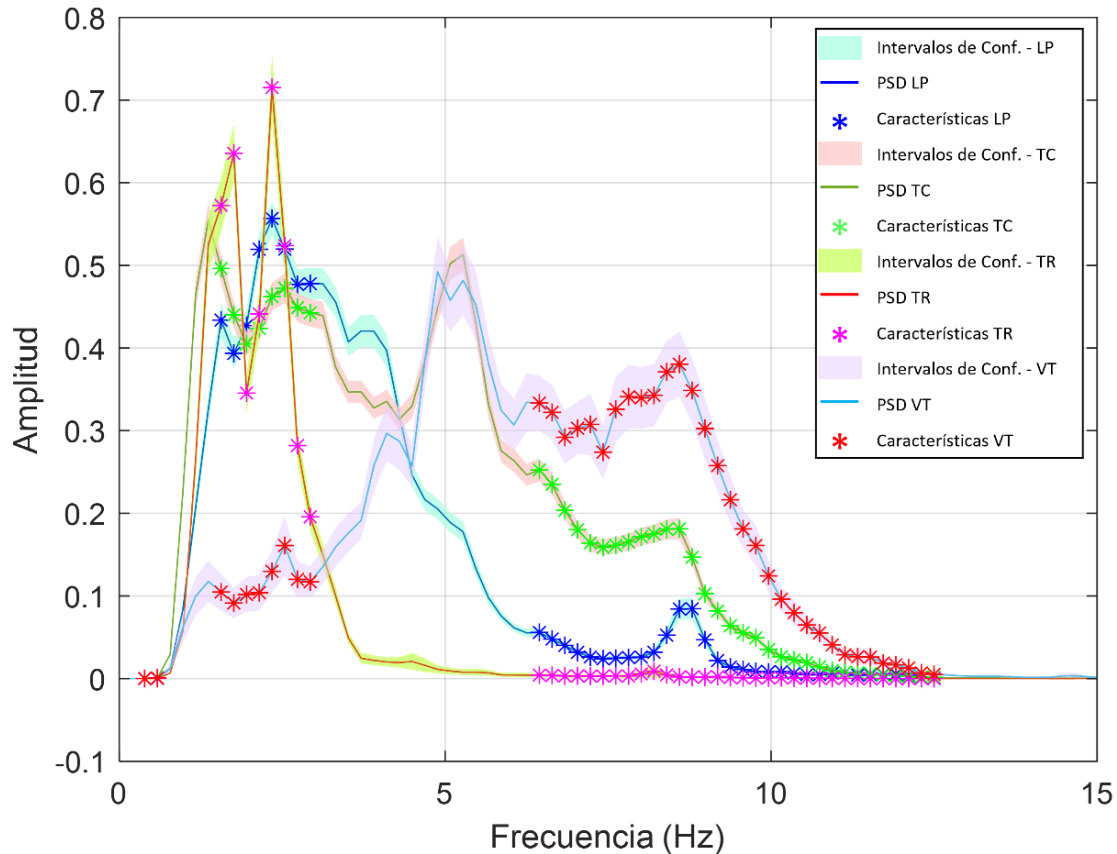
Por lo tanto, las características finales identificadas son:

Orden	Característica	Frecuencia
1	cD3	6.25 - 12.5 Hz
2	cD5	1.56 - 3.12 Hz
3	cD7	0.39 - 0.78 Hz

# 3. Pruebas y Resultados

## Reentrenamiento

Características finales en conjunto con PSD de cada evento y sus intervalos de confianza



Grado de Confianza: 99%  
Exactitud de reentrenamiento: 76%



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

## 4. Conclusiones

- Con el desarrollo del presente trabajo de investigación se ha logrado identificar tres características que permiten definir un microsismo en base a bandas de frecuencia establecidas por la WT, las cuales son: **cD3 (3.12 – 6.25 Hz)**, **cD5 (1.56 - 3.12 Hz)** y **cD7 (0.39 - 0.78 Hz)**. Esto se obtuvo al emplear métodos de selección de características basados en ML, lo que permitió obtener un porcentaje de exactitud de clasificación del **76%** al usar un clasificador DT con podamiento. En consecuencia, se concluye que la relevancia de esto radica en que al usar tan solo tres características de entrada al modelo ML, se obtuvo esta cifra de exactitud y por lo tanto, **sería suficiente con aumentar unas cuantas características más**, ya sea en tiempo o frecuencia, para que el modelo sea capaz de apuntar a una **exactitud de clasificación cercana al 100%**.

## 4. Conclusiones

- Las aplicaciones que tienen las MW son diversas y escoger correctamente la familia de MW adecuada depende del enfoque en el cual se desee aplicar. Es por esto que se considera el principio de similaridad para escoger la MW. En este caso, se han considerado las familias MW Daubechies y MW Symlets ya que son muy similares a las señales de microsismos y de entre estas dos, la **MW que brinda los mejores resultados es la familia de Daubechies**. Pese a que se realizaron pruebas con MW Symlets, se determinó con éxito que mediante la MW Daubechies se obtienen las mejores métricas de rendimiento y, por ende, es el principio mediante el cual se obtuvieron las características principales.
- Los intervalos de confianza que se establecen en el presente trabajo es con un **grado de confianza del 99%** y tras analizar el espectro medio de cada evento se concluye que, en promedio, el espectro de cada señal es muy similar a su media, con lo cual se concluye que la base de datos adquirida **es una base de datos académica** en donde se infiere que muy probablemente existió un proceso previo de elección de las mejores señales con el fin de **eliminar las señales atípicas** para aproximar los datos a una base de datos ideal.



# 4. Conclusiones

- Para el desarrollo se planteó un desarrollo de manera sucesiva en el cual se procede a descartar los métodos y enfoques que no presentan los mejores resultados. Es así que, en un inicio se planteó un análisis de 6, 8 y 10 niveles de descomposición con WT. La razón de escoger estos niveles de análisis se debe a dos factores: la capacidad de diferenciar un evento de otro mediante los niveles de energía de la wavelet con 6, 8 y 10 niveles y segundo, la reducción del número de características. **Con 10 niveles WT se obtienen los mejores resultados.**
- Tal y como se menciona, a medida que se avanzó en el desarrollo del trabajo de investigación, se descartaron varios enfoques de análisis y uno de ellos fue la proporción de división de datos. Mediante procesos de prueba y error, se logró determinar que se obtienen **los mejores resultados con una división de datos del 80% para entrenamiento y 20% para prueba.**





## 4. Conclusiones

- Dentro de todo esto, los mejores resultados obtenidos se resumen en el uso de la descomposición con **WT de 10 niveles** en donde cada coeficiente promedio obtenido es una característica de entrada a un modelo de clasificación basado en ML. La proporción de datos que se emplea es de **80/20** y el modelo ML es **DT con podamiento**. Al seleccionar las tres características más importantes del modelo ML y reentrenar dicho modelo con estas tres características se obtiene un porcentaje de exactitud del **76%**. Bajo estas características es que se obtuvieron los mejores resultados. Al hablar de que se obtuvo un porcentaje de exactitud del 76% al clasificar cuatro eventos de microsismos con un clasificador DT, se destaca la importancia de obtener dicha cifra con apenas tres características de entrada un modelo DT.

## 4. Trabajos Futuros

- Se plantea el desarrollo del mismo proceso bajo **un enfoque de datos balanceado**, dado a que la base de datos empleado durante el desarrollo de este trabajo de investigación consistió en distintos números de señales para cada evento. Sería apropiado desarrollar este procedimiento planteado para la identificación de características e intervalos de confianza bajo un método de análisis con una base de datos balanceada.
- Dentro de ML existen una gran variedad de técnicas que permiten realizar una clasificación. Además de los enfoques en DT y SVM, se pueden explorar **otros algoritmos de ML supervisado a futuro o con Deep Learning**, con el fin de realizar el proceso de selección de características y determinar cuáles son las bandas de frecuencias principales.

## 4. Trabajos Futuros

- Además de los coeficientes wavelet de detalle (cD) y de aproximación (cA) , podría ser interesante **explorar la inclusión de otras características relacionadas al Wavelet**. En la Tabla 10 se encuentran datos estadísticos que se obtienen directamente de la WT mediante el diagrama de caja y bigotes. Con estos datos relacionados netamente con la WT, se puede añadir más características y ampliar la información utilizada por los modelos de ML para intentar mejorar su capacidad de distinción entre diferentes tipos de microsismos.
- Otra sugerencia de trabajo a futuro es utilizar **una mayor cantidad de datos y ampliar el conjunto de señales** de microsismos mediante técnicas de **generación de señales sintéticas** y emplearlas para el entrenamiento y evaluación de los modelos. Esto podría contribuir a una mayor exactitud de clasificación.





**GRACIAS**