

# Detección de eventos del Volcán Cotopaxi empleando clasificación supervisada

*Event detection from Cotopaxi Volcano using supervised classification*

*Detecção de eventos de Cotopaxi Volcano utilizando classificação supervisionada*

**Román Alcides Lara Cueva**

Universidad de las Fuerzas Armadas-ESPE, Ecuador

[ralara@espe.edu.ec](mailto:ralara@espe.edu.ec)

**Valeria Paillacho**

Universidad de las Fuerzas Armadas-ESPE, Ecuador

[vapaillacho1@espe.edu.ec](mailto:vapaillacho1@espe.edu.ec)

**Michelle Villalva**

Universidad de las Fuerzas Armadas-ESPE, Ecuador

[mdvillalva@espe.edu.ec](mailto:mdvillalva@espe.edu.ec)

## Resumen

Debido a su ubicación geográfica, el Ecuador se encuentra en una zona de alta sismicidad volcánica, por lo que para el análisis e identificación de las señales provenientes de un volcán son necesarias herramientas automáticas de gran eficacia. En esta investigación se presenta una detección basada en clasificación de eventos sismo-volcánicos y no volcánicos registrados en el lapso de 2009 a 2010. Para la extracción de características más significativas de las señales se plantean dos técnicas: en el dominio frecuencial por medio de la densidad espectral de los eventos y en el dominio de escala mediante la energía de los coeficientes de aproximación y detalle. Para la clasificación supervisada se emplea el algoritmo de árboles de decisión. El rendimiento más alto fue obtenido a partir del análisis energético de los coeficientes con una precisión del 98 % aproximadamente.

**Palabras clave:** aprendizaje de máquina, detección de eventos sísmicos, árboles de decisión.

## Abstract

Ecuador by its geographic location is a zone of volcanic seismicity, therefore for in order to analyze and identify signals from a volcano. It is required of automated tools with highly effacing. In this paper, we present a classification-based detector of seismic-volcanic events and not volcanic from volcano Cotopaxi registered in the period from 2009 to 2010. In order to feature extraction of the signals we propose two techniques, in the frequency domain using the power spectral density of events and in the scale domain employing energy of approximation and detail coefficients. For supervised classification we use the decision trees algorithm. The highest performance was obtained from the energy analysis of the coefficients with an accuracy of 98 %.

**Key words:** machine learning, seismic event detection, decision tree.

## Resumo

Devido à sua localização geográfica, o Equador está em uma área vulcânica de alta atividade sísmica, assim, para análise e identificação de sinais de um vulcão são necessárias ferramentas automatizadas altamente eficientes. Neste detecção baseada classificação de pesquisa de eventos sísmicos-vulcânicas e não-vulcânicos registrados no período de 2009 a 2010. Para a extração da maior parte das características significativas dos sinais levantou duas técnicas é apresentado: no domínio da frequência pela densidade espectral de eventos e domínio de escala por coeficientes de aproximação de energia e detalhes. Para algoritmo de classificação supervisionada ele usa árvores de decisão. O rendimento mais elevado foi obtido a partir da análise de energia dos coeficientes com uma precisão de cerca de 98%.

**Palavras-chave:** aprendizagem de máquina, detectando eventos sísmicos, árvores de decisão.

**Fecha recepción:** Junio 2015

**Fecha aceptación:** Diciembre 2015

## Introducción

Debido a la ubicación del Ecuador en el llamado Cinturón de Fuego del Pacífico en Sudamérica, dicho país un sinnúmero de volcanes activos que pueden variar en grado de peligrosidad (Segovia y Alvarado, 2009). Uno de los principales volcanes del territorio ecuatoriano en esta zona geográfica es el Cotopaxi, un gran estratovolcán que despierta interés en la comunidad científica por los peligros con que se asocia (Andrade et al., 2005) y por haber presentado en los últimos años un significativo aumento en sismicidad volcánica.

El Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN) efectúa una monitorización ininterrumpida (Viracucha y de la Bastida, 2014) por medio de la red de sensores de alta precisión instalada en diferentes estaciones, encargadas principalmente de proporcionar información de las señales sísmicas. Las señales captadas por estos instrumentos son: 1) eventos de largo periodo (LP), 2) eventos volcano-tectónicos (VT), 3) eventos híbridos (HYB) y 4) tremores (TRE). Adicionalmente en el registro sísmico se pueden reportar con una gran probabilidad de ocurrencia señales no volcánicas conocidas como los Truenos (LGH), las cuales generalmente acompañan a los eventos anteriormente mencionados. Una vez obtenidos los registros sísmicos, los especialistas examinan cada uno con el fin de establecer el tipo de señal al que pertenecen; sin embargo, dicho estudio es susceptible de cometer errores al depender de la falibilidad de la persona que efectúa esta tarea (Vargas Jiménez, Rodríguez Espinosa, y Otero Foliaco, 2013).

En general, las investigaciones en este contexto realizan una caracterización y detección mediante varias metodologías, un ejemplo es lo realizado por Gabarda y Cristóbal (2010), donde se plantea una detección de señales sísmicas empleando la Distribución de Wigner a partir de series en el dominio temporal y frecuencial. Del mismo modo, Jaramillo, León, Lara-Cueva, Benítez y Ruiz (2014), establecen una detección basada en estimadores espectrales clásicos y de máxima entropía, en el cual se definen claramente cuatro zonas de concentración de la densidad de potencia de los eventos. Por otro lado, Prasad, Reddy y Nirjogi (2014), realizan una detección mediante un análisis espectral al aplicar una Wavelet madre de tipo Haar con la identificación de la frecuencia de operación de las señales, así como también la presencia o no del evento. Por último, Lois, Psarakis y Pikoulis (2013), instauran un sistema de detección en dos etapas, la

primera de ellas realiza la estimación de las funciones empíricas del proceso de ruido, mientras que la segunda propone la umbralización de las señales.

Sin embargo, tras revisar estos trabajos previos la mayoría de ellos no eran sistemas que pudieran ser implementados en un ambiente de tiempo real, es por ello que este estudio se enmarca en establecer una detección basada en clasificación por medio del algoritmo de aprendizaje de máquina conocido como árboles de decisión (DT, del inglés *Decision Trees*), a través de características obtenidas en el dominio de la frecuencia y escala.

Al depender la detección y clasificación de eventos del criterio y destreza del analista, el objetivo principal de esta investigación es aportar al IGEPN con una herramienta automática para detectar señales sismo-volcánicas y no volcánicas, que permita informar de manera oportuna a los organismos competentes para una mitigación de los riesgos existentes frente a un posible proceso eruptivo. Por esta razón se plantea una extracción de características mediante las Transformadas de Fourier y Wavelet, en el primer caso con un análisis de la densidad espectral de potencia empleando el método de Welch, mientras que en el segundo caso el análisis se basa en la energía de los coeficientes de aproximación y detalle de las mismas. Enseguida se realiza una clasificación supervisada a través de DT para la creación de modelos predictivos, la cual finalmente pasa a la etapa de detección donde luego de un post-procesamiento se evalúa el rendimiento a través de métricas establecidas.

El artículo está organizado de la siguiente manera. En la sección II se describe la metodología para el pre-procesamiento, extracción de características, el algoritmo de clasificación DT y la detección de las señales en análisis. En la sección III se muestran los resultados experimentales e interpretación de cada una de las etapas del detector. Por último se realizan las conclusiones y discusión de la investigación en la sección IV.

## **Metodología**

El algoritmo de detección automático planteado se enfoca en tres tipos de señales, los LP, VT y LGH; los dos primeros eventos permiten verificar si existe un aumento en la actividad tectónica que contribuya a la aparición de una erupción volcánica (Lahr, Chouet, Stephens, Power, & Page, 1994), al contrario de los LGH, que es preciso discriminar de las otras señales de origen sismo-volcánico. La metodología propuesta consta de cuatro etapas principales mostradas en la

figura 1: 1) pre-procesamiento junto con la segmentación, 2) caracterización en el dominio frecuencia y de escala, 3) clasificador de árboles de decisión y 4) post-procesamiento con la detección.

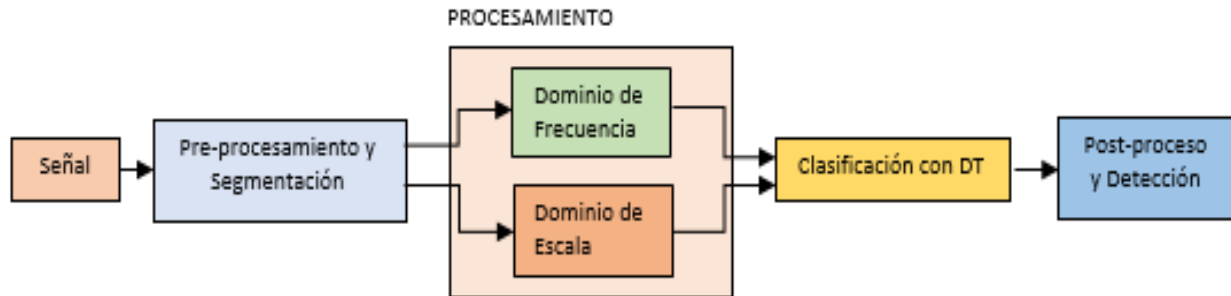


Figura 1. Diagrama de bloques del detector de eventos sismo-volcánicos.

### Pre-procesamiento y segmentación

El pre-procesamiento contribuye a conseguir las señales en aproximadamente su forma original, ya que estas al ser resultado del continuo contacto entre las placas tectónicas pueden provocar errores en la medición sísmica. Por ese motivo en esta etapa inicialmente las señales pasan por un filtro de respuesta finita al impulso (FIR, del inglés *Finite Impulse Response*) de tipo pasa-banda con un orden de 128, definido en el rango de 0.5 Hz a 50 Hz. Dicho rango de frecuencia se establece para considerar las bandas de operación de frecuencia de todas las señales, además de descartar a los microsismos en la frecuencia de 0.2 Hz (Akram y Eaton, 2012). A continuación se realiza la normalización de las señales para nivelar los valores en el registro sísmico con una media cero y varianza uno ( $\mu = 0$ ,  $\nu = 1$ ).

Adicionalmente para satisfacer los requerimientos de tiempo real donde influyen varios parámetros como: el número de características, la dimensión de las matrices de entrenamiento que se emplean para determinar el modelo predictivo, entre otros factores, debe existir una estrecha relación entre la complejidad del clasificador y el tamaño de la base de datos (Cortés, Benítez, García, y Álvarez, 2015). De acuerdo a esto se realiza la segmentación de las señales, donde se considera que el tamaño de las matrices características es dependiente del valor de la ventana, un tamaño menor de ventana aumenta la dimensión de la matriz y viceversa, de ahí que el valor de la ventana que se eligió fue de 15s como se muestra en la figura 2, en base a los rendimientos obtenidos por Lara-Cueva, Benítez, Carrera, Ruiz, y Rojo-Álvarez, donde cada

señal segmentada se guarda en la matriz  $\mathbf{L}$  constituida por  $\mathbf{L} = [\mathbf{I}_1^T, \mathbf{I}_2^T, \dots, \mathbf{I}_m^T]^T$ , con una dimensión  $m = r/\omega$ , y donde  $r$  es el tiempo total de un registro  $\mathbf{I}_i$  capturado por el sismógrafo de valor igual a 120s, definido por  $\mathbf{I}_i = [\mathbf{I}_{i,1}^T, \mathbf{I}_{i,2}^T, \dots, \mathbf{I}_{i,j}^T]^T$ , en el cual  $\mathbf{I}_{i,j}$  representa un segmento de 15s.

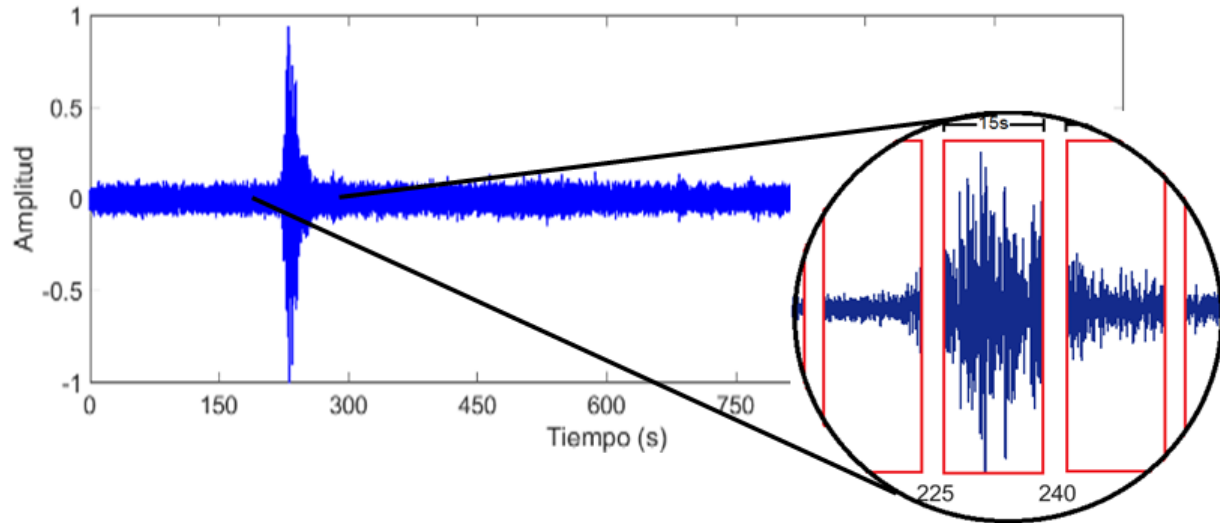


Figura 2. Ejemplo de la segmentación de una señal LP con una ventana de 15s.

Finalmente en concordancia con las etiquetas de inicio y fin de los eventos establecidos de forma visual, se asigna a cada segmento una etiqueta. Dicha etiqueta sirve para la identificación de las diversas clases de señales sísmicas, que se utilizan para entrenamiento en el algoritmo de DT. De esta manera se fija el valor de: -1 al ruido de fondo (BN, del inglés *Background Noise*), 1 a los eventos LP, 2 a los VT y 3 a los LGH.

### Extracción de características

Esta etapa es fundamental en la detección porque sirve para identificar las características representativas de las señales sismo-volcánicas y no volcánicas, ya sea de forma frecuencial o de escala, y para realizar una comparativa entre ambas técnicas.

### Dominio de la frecuencia

La Transformada Rápida de Fourier (FFT, del inglés *Fast Fourier Transform*), se empleó para tomar las características más relevantes de las señales en el dominio de la frecuencia, por lo que se aplicó la densidad espectral de potencia (PSD, del inglés *Power Spectral Density*) con el

método Welch. Este método consiste en seccionar la señal en el dominio del tiempo con la finalidad de que se traslapen para que no exista pérdida de información en los extremos, posteriormente se estima y promedian los periodogramas de cada una de estas partes (Undurraga y Uribarri, 2006). Los parámetros que se consideraron son: cada sección emplea una ventana de Hamming de longitud igual al segmento, además se tomaron 512 puntos para la resolución de representación de frecuencia y un solapamiento del 50 %, lo que originó  $g_{i,j} = g_n(\mathbf{I}_{i,j})$ , siendo  $g_n$  el operador que realiza la PSD con la resolución respectiva para cada grabación, creando la matriz  $\mathbf{G}$ , dada por  $\mathbf{G} = [\mathbf{g}_1^T, \mathbf{g}_2^T, \dots, \mathbf{g}_i^T]^T$ , donde  $\mathbf{g}_i$  está definida por  $\mathbf{g}_i = [\mathbf{g}_{i,1}^T, \mathbf{g}_{i,2}^T, \dots, \mathbf{g}_{i,j}^T]^T$ , con una dimensión  $m = N \times j$  casos, donde  $j$  representa el número de características dada por  $n = \frac{n_f}{2} + 1$ , siendo  $n_f$  los puntos de la FFT.

En la figura 3 se observan las bandas de frecuencia ( $f$ ) de las señales, de donde los LP tienen un contenido espectral en el rango entre  $0.5 \text{ Hz} < f < 5 \text{ Hz}$ , mientras que los VT presentan una  $f < 10 \text{ Hz}$  y finalmente los LGH con valores más altos entre  $17 \text{ Hz} < f < 20 \text{ Hz}$ .

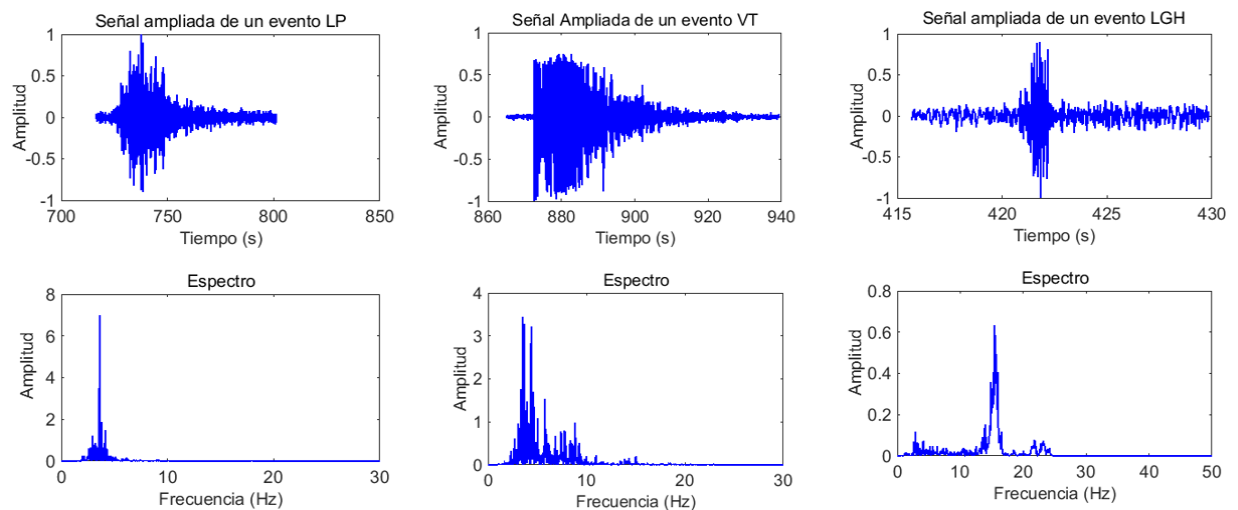


Figura 3. Ejemplos de eventos registrados en el volcán Cotopaxi con sus respectivos espectros.

### Dominio de la escala

La herramienta matemática para el escalamiento es la Transformada Discreta de Wavelet (DWT, del inglés *Discrete Wavelet Transform*), la cual efectúa una descomposición multinivel mediante la aplicación de varios filtros donde en cada nivel se separa la porción de bajas frecuencias (coeficientes de aproximación  $cA$ ) y la porción de altas frecuencias (coeficientes de detalle  $cD$ ) para considerar ciertos componentes de frecuencia de la señal (Gareis, Gentiletti, Acevedo, y

Rufiner, 2009). En la figura 4 se muestra este procedimiento de acuerdo a la frecuencia de muestreo empleada de 100 Hz, con un nivel de descomposición 5 para establecer las bandas de frecuencia que trabajan los eventos, de ahí que resultaron los coeficientes de detalle 2, 3 y 4 (cD2, cD3 y cD4) y el coeficiente de aproximación (cA5). La Wavelet Madre que se aplica es de la familia Daubechies en un orden de 10, por el concepto de similitud de su forma de onda con los diferentes tipos de señales (Saltos Torres, 2014).

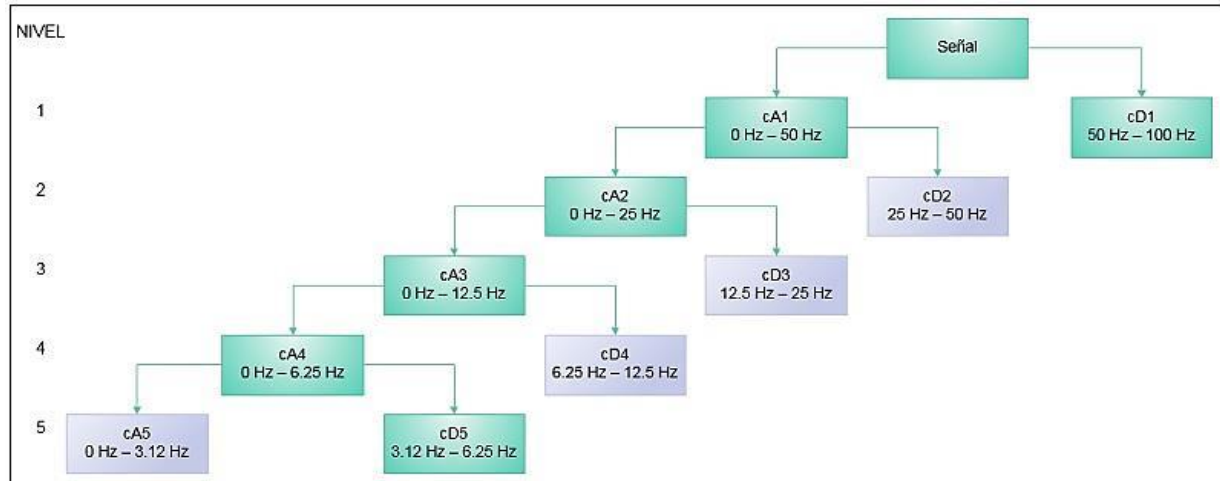


Figura 4. Descomposición Multinivel Wavelet basado en las frecuencias de operación de las señales.

En concordancia con la descomposición multinivel de Wavelet se establece  $\mathbf{d}_{i,j} = g_{\omega}(\mathbf{l}_{i,j})$ , donde  $g_{\omega}$  es el operador de este procedimiento representado por los coeficientes cA y cD, lo que permite originar  $\mathbf{d}_i$ , definida como  $\mathbf{d}_i = [\mathbf{d}_{i,1}^T, \mathbf{d}_{i,2}^T, \dots, \mathbf{d}_{i,j}^T]^T$  para finalmente generar la matriz  $\mathbf{D}$  dada por  $\mathbf{D} = [\mathbf{d}_1^T, \mathbf{d}_2^T, \dots, \mathbf{d}_i^T]^T$ .

Posteriormente se determina la matriz  $\mathbf{E}$  formada por  $\mathbf{E} = [\mathbf{e}_1^T, \mathbf{e}_2^T, \dots, \mathbf{e}_i^T]^T$ , donde cada  $\mathbf{e}_i$  está compuesto por  $\mathbf{e}_i = [e_{i,1}, e_{i,2}, \dots, e_{i,j}]$  con  $j = 1, 2, \dots, 4$ . Esta matriz contiene la energía de cada uno de los 4 coeficientes seleccionados calculados de acuerdo a la ecuación 1.

$$\text{Energía} = \sum_{i=1}^N x^2(i), \tag{1}$$

Aquí  $\mathbf{x}$  representa cada uno de las componentes de la señal.



### Clasificación con árboles de decisión

Para que se desarrolle la clasificación de los eventos se emplea el algoritmo de aprendizaje DT, el mismo que realiza un aprendizaje supervisado. Esta técnica efectúa de manera clara una representación de la relación entre las variables y las etiquetas reales, por ello los datos de entrada que se le atribuyeron son: las matrices características y los tipos de eventos. El árbol se forma con un nodo raíz, nodos internos y nodos terminales; asimismo en cada nodo se establece una regla que produce la decisión binaria, hasta llegar a un nodo terminal denominado hoja, que representa a una clase. La cantidad de ramas es dependiente de los valores que pueda tomar cada nodo. De esta forma, el modelo se genera con el fin de predecir valores de acuerdo a decisiones secuenciales, a partir de las características representativas de los datos. Este algoritmo realiza un enfoque basado en la teoría de información, donde la elección de un atributo se relaciona directamente con la entropía. La entropía es una medida de incertidumbre en un sistema, la cual permite conocer la cantidad promedio necesaria de bits para poder corregir la salida de un modelo. Este parámetro se representa mediante la ecuación 2.

$$E(Y) = \sum_{i \in C} -p_i \log_2 p_i, \quad (2)$$

Aquí  $C$  es el conjunto de clases a los que puede pertenecer dicho ejemplo y  $p_i$  es la probabilidad de que un ejemplo dado pertenezca a la  $i$ -ésima clase.

Adicionalmente se puede obtener la ganancia de información por medio de la reducción de la entropía, medida que permite establecer el nivel de clasificación de un atributo con los datos de entrenamiento. Está definida mediante la ecuación 3.

$$G(Y, A) = E(Y) - \sum_{v_i \in V} \frac{|Y_{v_i}|}{|Y|} E(Y_{v_i}), \quad (3)$$

Aquí  $V$  representa el conjunto de valores que puede adquirir un atributo  $A$ , mientras que  $Y_{v_i}$  es una parte del conjunto de ejemplos  $Y$ , que tiene como atributo a  $A$  con un valor  $v_i$ . Además se realiza la sumatoria de las entropías de los conjuntos resultantes, incrementadas por los tamaños de dichos conjuntos con respecto al inicial (Galipienso, 2003) (Hall, 1999).

Por último, es preciso mencionar que DT posee una técnica de reducción del tamaño del árbol denominada podamiento, la misma que se encarga de seleccionar las características más

relevantes de las señales, sin afectar el rendimiento en la clasificación. De esta manera se consigue un árbol con una estructura más simple, evitando el problema de sobreajuste (Wei, y otros, 2009).

### Post-procesamiento y rendimiento

Con la obtención de los modelos predictivos mediante la evaluación de las distintas características de la etapa anterior, en esta fase se describe el pos-procesamiento que permite la detección de eventos sismo-volcánicos y no volcánicos. Al determinar las clases de eventos asignados con los valores 1, 2, 3 ó -1 que pertenecen a los LP, VT, LGH o BN respectivamente; se procede a cotejar la información de etiquetas predecidas y reales de los mismos, de acuerdo al concepto de clasificación supervisada por medio de un algoritmo de ajuste. Dicho algoritmo determina la presencia de algún evento con un valor de 1 y en caso contrario con 0. De esta manera se establece un valor de tolerancia, cuyo valor determina un margen donde los parámetros de rendimiento favorezcan la detección.

Las métricas establecidas para calcular el desempeño del detector basada en clasificación DT son: la Exactitud (A), Precisión (P), Sensibilidad (R) y Especificidad (S), las cuales se especifican a continuación en las ecuaciones 4 a la 7.

$$A(\%) = \frac{N_c}{N_T} \times 100, \quad (4)$$

$$P(\%) = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100, \quad (5)$$

$$R(\%) = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100, \quad (6)$$

$$S(\%) = \frac{N_{TN}}{N_{TN} + N_{FP}} \times 100, \quad (7)$$

Aquí  $N_c$  es el número de patrones clasificados correctamente,  $N_T$  es el número total de los patrones usados en la clasificación,  $N_{TP}$  es el número de verdaderos positivos,  $N_{FP}$  es el número de falsos positivos,  $N_{TN}$  es el número de verdaderos negativos y  $N_{FN}$  es el número de falsos negativos.

## Resultados

En esta sección se describen los resultados obtenidos con la metodología planteada. La base de datos para la realización de este trabajo fue provista por el IGEPN, la cual está conformada con 100 señales de cada tipo de evento. Esta base de datos fue dividida en matrices de igual tamaño para el entrenamiento (*train*) y prueba (*test*) garantizando que exista independencia entre las mismas. El software de simulación que se empleó para el desarrollo de los experimentos es Matlab® R2015a, mediante una PC Core(TM) I5 con 2.27 GHz y 8 GB de RAM.

En la fase de extracción de características, en el dominio de frecuencia se obtuvieron 257 características aplicando PSD con el uso de su periodograma de Welch, mientras que en el dominio de la escala al calcular la energía de los coeficientes de aproximación y detalle se lograron solamente 4 características.

En la fase de clasificación de los eventos al aplicar DT, a través de las características obtenidas con los dos métodos de extracción se consiguieron las representaciones de los árboles que se observan en las figuras 5 y 6. El número de nodos para la clasificación con FFT resultó en 40, mientras que con DWT fue una cantidad de 41 nodos. Asimismo, la primera figura muestra que las principales características en FFT son:  $X_{97} \geq 0.45$   $X_{16} \geq 0.82$  y  $X_{22} \geq 3.34$  correspondientes a los valores en amplitud a 18.87 Hz, 3.11 Hz y 4.28 Hz, respectivamente. Además, la característica  $X_{97}$  constituye el nodo raíz del árbol con un valor umbral de 0.45 que permite la identificación entre las 4 posibles clases. En cuanto a la gráfica de clasificación con DWT, las características que elige el algoritmo son la  $cD_4 \geq 1.31$  correspondiente al rango de frecuencia entre  $6.25 \text{ Hz} \leq f \leq 12.5 \text{ Hz}$ , en cambio que para  $cD_2 \geq 3.82$  su frecuencia comprende el intervalo de  $25 \text{ Hz} \leq f \leq 50 \text{ Hz}$ .

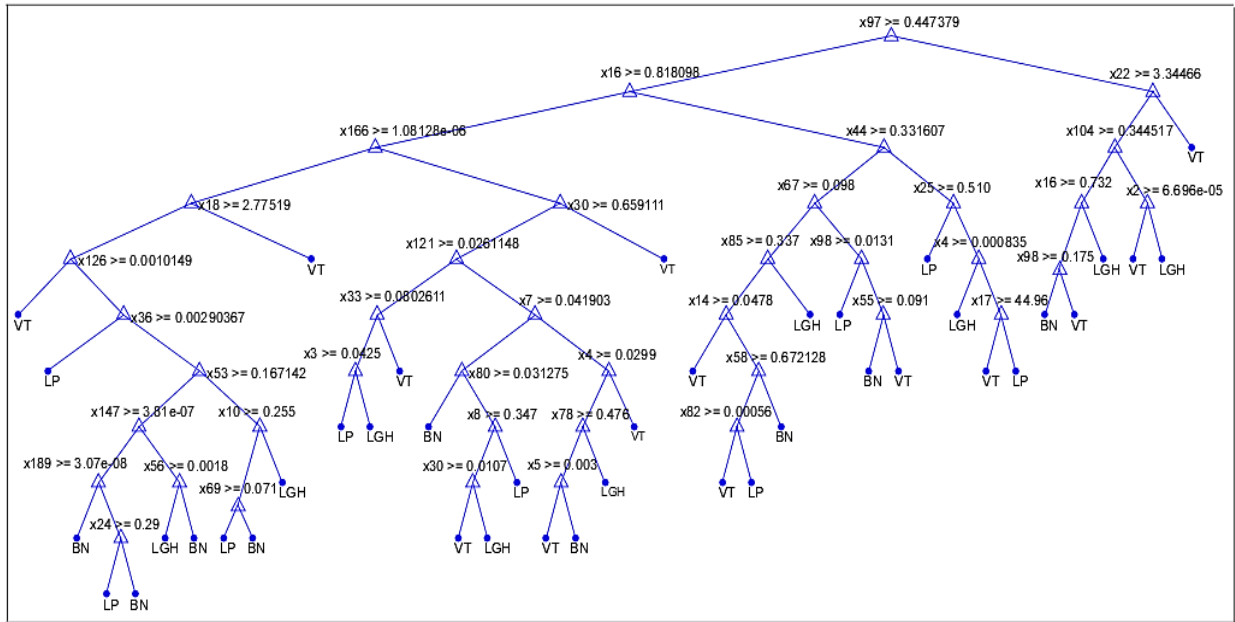


Figura 5. Representación del árbol de decisión obtenido con las características de la FFT.

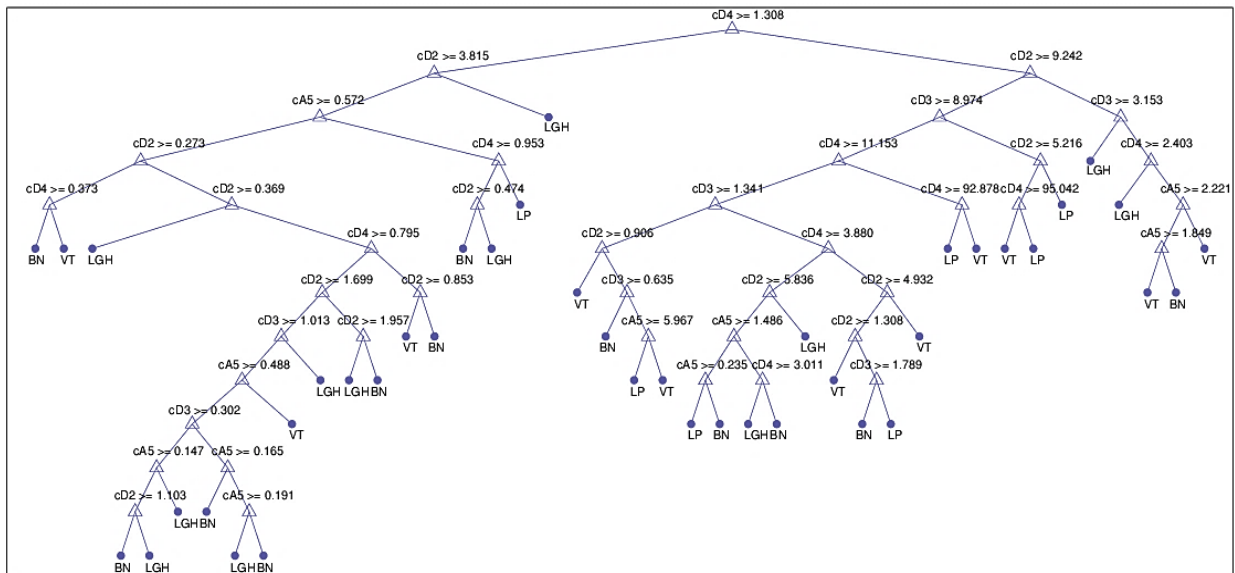


Figura 6. Representación del árbol de decisión obtenido con características de la energía de los coeficientes DWT.

Enseguida en la fase de reducción de características mediante el método de podamiento de DT, el número de ramas se ve disminuido notoriamente en comparación con las gráficas anteriores. En la figura 7 se observa que para la extracción de características con Fourier se necesitan 4 nodos para que se realice una clasificación de los eventos sísmo-volcánicos, estos son  $X_{97} \geq 0.44$ ,  $X_{18} \geq 3.5$ ,  $X_{175} \geq 2.98$  y  $X_{42} \geq 0.47$ , los cuales pertenecen a los valores de amplitudes de 18.87 Hz, 3.11 Hz, 34 Hz y 8.17 Hz, respectivamente. Del mismo modo ocurre en esta etapa con

la técnica de Wavelet, donde existe una repetición del nodo  $cd_2$  dando como resultado una optimización a solo 3 nodos presentada en la figura 8.

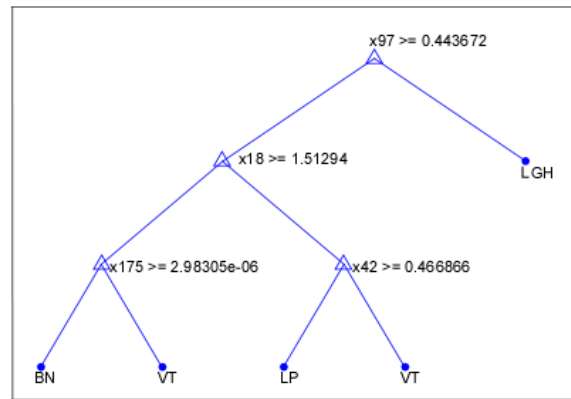


Figura 7. Representación del árbol de decisión empleando la técnica de selección de características podamiento para las características obtenidas con la FFT.

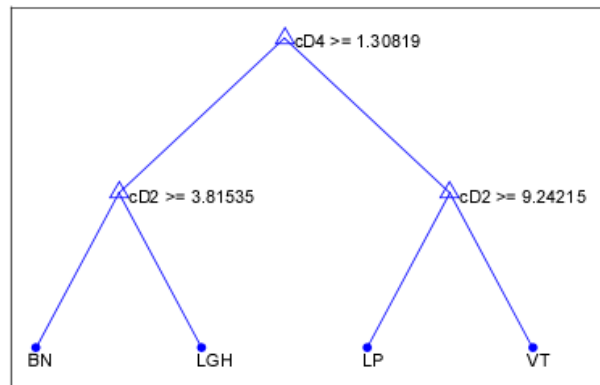


Figura 8. Representación del árbol de decisión empleando la técnica de selección de características podamiento para las características obtenidas con la energía de los coeficientes DWT.

Durante la etapa de post-procesamiento al realizar el ajuste de acuerdo al valor numérico de la tolerancia, las métricas de desempeño del detector se maximizan en un valor de 100. En la figura 9 se muestra la métrica de exactitud, de donde tanto para la FFT como para DWT presentan porcentajes cercanos al 99 %, mientras que cuando se aplica la técnica de selección de características de podamiento existe una pequeña variación en FFT del 1% al contrario de DWT donde las curvas son idénticas.

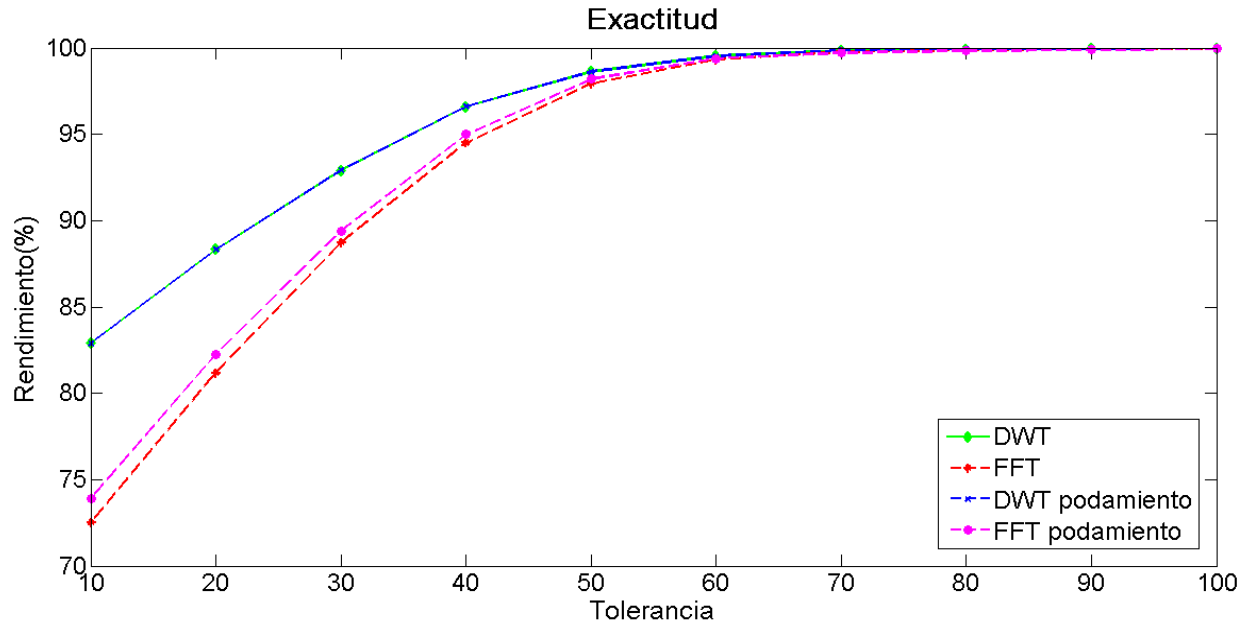


Figura 9. Métrica de exactitud de acuerdo al dominio de la frecuencia (FFT) y dominio de la escala (DWT), con o sin la aplicación de la técnica de selección de características podamiento.

En la figura 10 se observa la métrica de precisión, donde las metodologías de extracción frecuencial y de escala se diferencian en 2 %, con 96 % para la técnica de Fourier y 98 % para la Wavelet. Al emplear el podamiento en DWT se mantiene en los mismos valores, sin embargo, en FFT al igual que la métrica anterior cambia en 1 %.

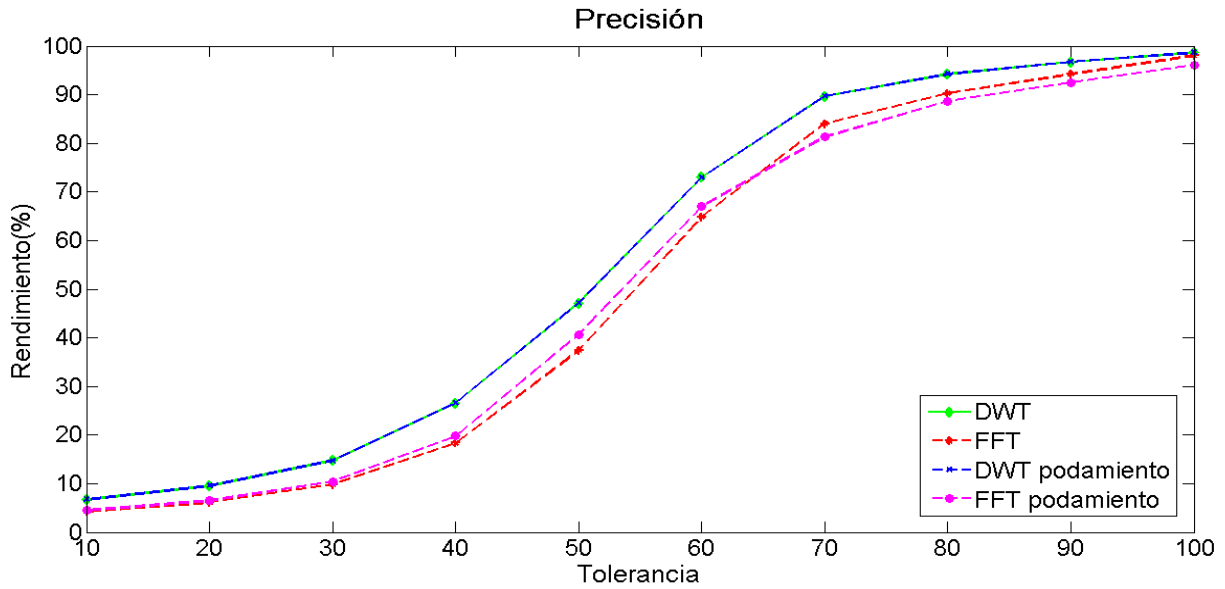


Figura 10. Métrica de precisión de acuerdo al dominio de la frecuencia (FFT) y dominio de la escala (DWT), con o sin la aplicación de la técnica de selección de características podamiento.

Por otro lado, en la figura 11 se muestra la métrica de sensibilidad, donde se distingue un gran contraste de las técnicas utilizadas en los valores que presentan, de lo cual sin emplear el podamiento la FFT y DWT resultan en un valor del 99.3 %, no obstante al considerar la selección de características se tiene en FFT el valor del 100 % a diferencia de DWT que se mantiene en 99.3 %.

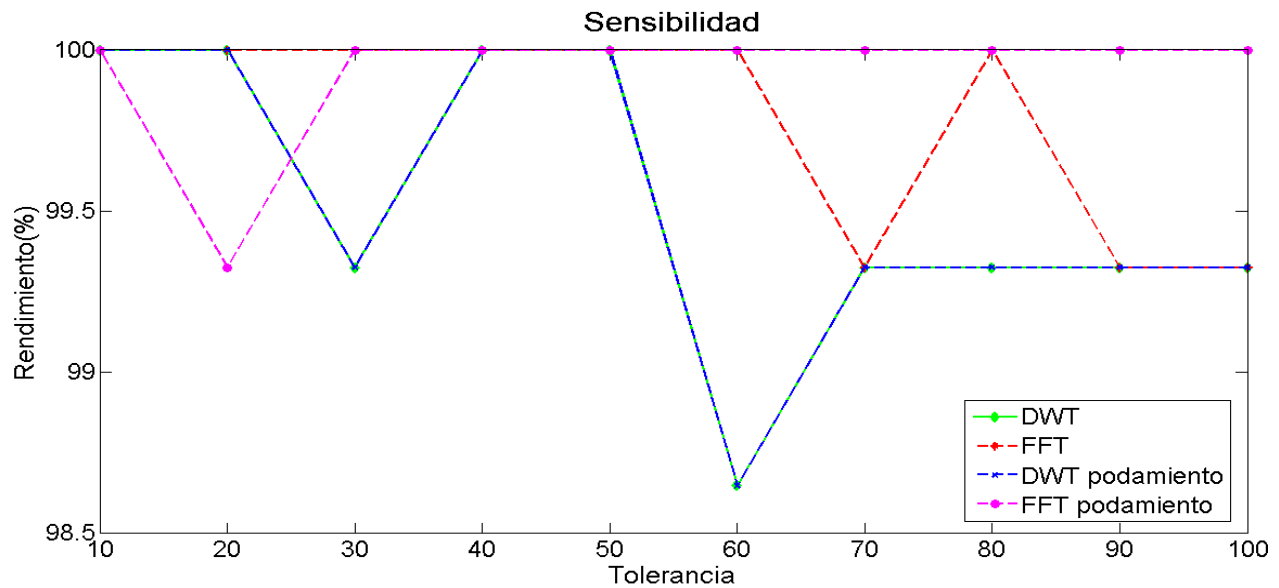


Figura 11. Métrica de sensibilidad de acuerdo al dominio de la frecuencia (FFT) y dominio de la escala (DWT), con o sin la aplicación de la técnica de selección de características podamiento.

Finalmente, la métrica de especificidad se presenta en la figura 12, al igual que la exactitud presenta valores cercanos al 99 % para FFT y DWT. Al utilizar el podamiento, en DWT la curvas se superponen, pero en FFT se inicia con una insignificante diferencia hasta alcanzar el mismo valor en la tolerancia establecida.

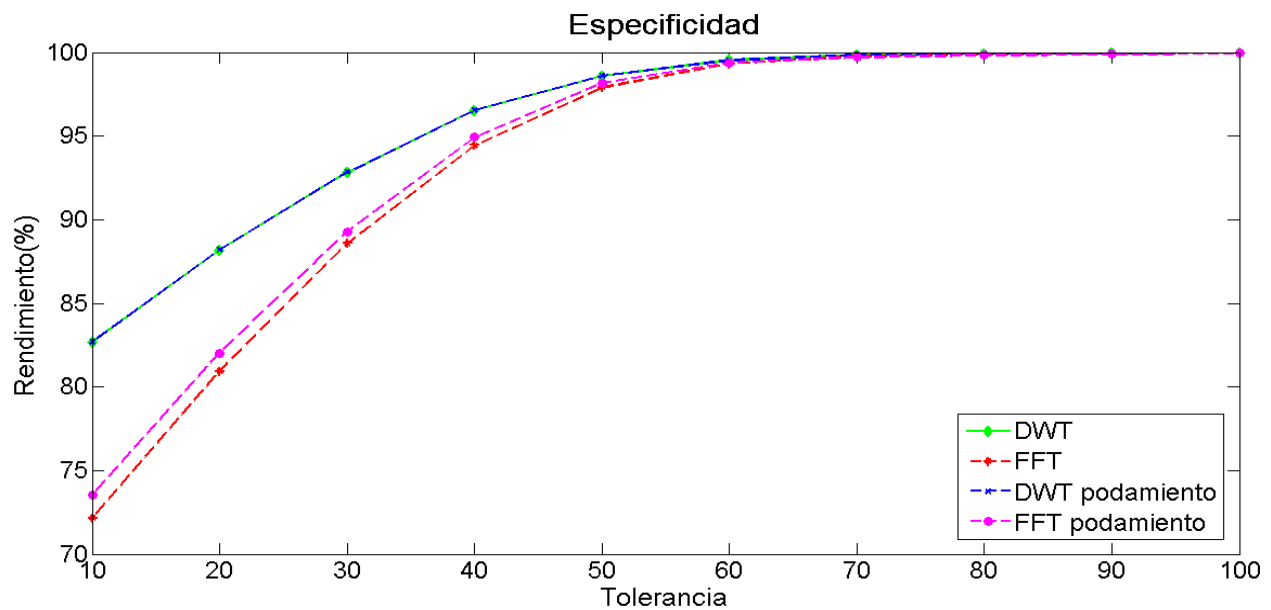


Figura 12. Métrica de especificidad de acuerdo al dominio de la frecuencia (FFT) y dominio de la escala (DWT), con o sin la aplicación de la técnica de selección de características podamiento.



En la figura 13 se visualiza la detección de los eventos, con el modelo generado a partir de las energías de los coeficientes Wavelet. Estos se encuentran enmarcados en un rango de 30s asignando el valor de 1 al inicio y fin, lo que indica la ocurrencia de un evento.

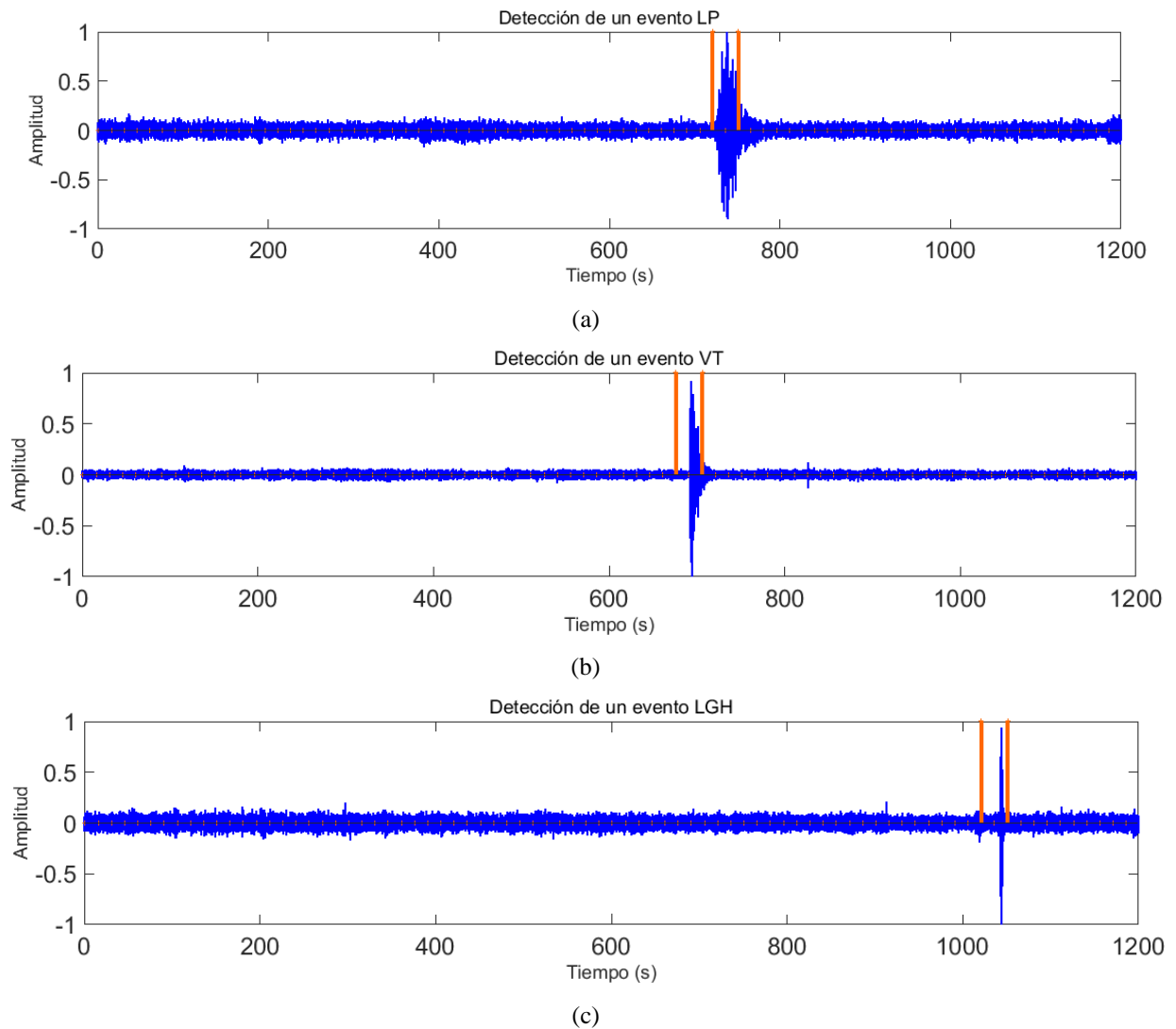


Figura 13. Eventos detectados por el sistema mediante un enmarcado con un rango de 30s, (a) Detección de un evento sismo-volcánico LP, (b) Detección de un evento sismo-volcánico VT, (c) Detección de un evento no

## Discusión y conclusiones

Los trabajos sobre la detección de eventos sismo-volcánicos contribuyen a la prevención y disminución de los daños que pueden provocar un desastre natural de gran magnitud como es una erupción volcánica, motivo por el cual esta investigación se enfoca en aportar al desarrollo de un sistema automático para este tipo de señales.

En referencia a los trabajos relacionados estos presentan distintas técnicas tanto para la obtención de sus características como para la detección de los eventos sísmicos-volcánicos. De acuerdo a Gabarda y Cristóbal (2010), la Distribución de Wigner le permite identificar las señales en un ambiente no supervisado, sin embargo, no se realiza el cálculo del rendimiento en su detección que contribuya a verificar que su método es efectivo. Prasad, Reddy, & Nirjogi (2014) realizan un detector empleando una wavelet madre de tipo Haar para una señal sismo-volcánica, a diferencia del presente trabajo que lo efectúa con dos eventos sismo-volcánicos, una señal no volcánica y el ruido de fondo mediante el análisis energético de la descomposición multinivel wavelet con la familia Daubechies.

La detección se basa en la clasificación de las señales por medio de dos técnicas en el dominio frecuencial y de la escala. En el primer caso se obtienen 257 características, a partir de la PSD por el método de Welch, al contrario del dominio de la escala con 4 características mediante la energía de los coeficientes de aproximación y detalle. Adicionalmente se empleó la selección de características con la técnica de podamiento, donde se redujo considerablemente la cantidad de las mismas, tal fue el caso del análisis en frecuencia, donde las características principales resultaron ser la  $X_{18}$ ,  $X_{42}$ ,  $X_{97}$  y  $X_{175}$  correspondientes a los valores de amplitudes en las frecuencias de 3.11 Hz, 8.17 Hz, 18.87 Hz, y 34 Hz, respectivamente, mientras que en DWT destacan los coeficientes de detalle,  $cD_2$  que conciernen al rango de frecuencias entre  $6.25 \text{ Hz} \leq f \leq 12.5 \text{ Hz}$  y del mismo modo en  $cD_4$  con frecuencias que varían de  $25 \text{ Hz} \leq f \leq 50 \text{ Hz}$ . De acuerdo a Jaramillo, León, Lara-Cueva, Benítez y Ruiz (2014), las zonas de mayor concentración de frecuencia para cada uno de los eventos son 3.15 Hz para los LP, 6 Hz para los VT y 15 Hz a 35 Hz para los LGH. De donde se evidencia que los valores obtenidos de frecuencia son similares con nuestro trabajo, demostrando de esta manera que el método de extracción de características es adecuado con estas señales.

Las representaciones de los árboles de decisión con el método de podamiento permitieron verificar los umbrales que diferencian una clase de evento de otro; en las características mediante FFT el algoritmo determina que el nodo raíz es  $X_{97}$ , el cual representa a la amplitud de frecuencia de 18.87 Hz, de tal manera que si la señal supera el valor umbral de 0.44 se determina que es un evento LGH. En caso de que no se supere dicho valor, se ingresa a una nueva toma de decisiones, por medio del valor umbral de  $X_{18}$ , perteneciente a la amplitud de frecuencia de 3.11 Hz, el cual a su vez se divide en dos ramificaciones dando origen a los nodos  $X_{175}$  y  $X_{42}$ . Según el  $X_{175}$ , equivalente a la frecuencia de 34 Hz, puede diferenciar entre BN y VT. En cambio, para el nodo  $X_{42}$ , correspondiente a la amplitud de frecuencia de 8.17 Hz, se pueden discriminar dos tipos de eventos, donde si es inferior al valor umbral de 0.47 se identifica que es una clase LP y en caso contrario una VT. De acuerdo a esto se puede evidenciar que la clase VT posee componentes espectrales en la banda de frecuencia de 34 Hz y 8.17 Hz, mientras que para la clase LP su rango se ve limitado hasta 8.17 Hz.

En el dominio de la escala se establece una distribución más simple en su representación del árbol con el nodo raíz  $cD_4$ , que comprende el rango de frecuencias entre 25 Hz a 50 Hz. Dicha característica tiene un valor umbral de 1.31, el cual permite distinguir entre eventos de origen no volcánico y sismo-volcánico sin necesidad de examinar al nodo  $cD_2$ . A continuación se definen dos ramas con la repetición del nodo  $cD_2$ , el mismo que presenta un intervalo de frecuencia de 6.25 Hz a 12.5 Hz. El valor umbral de 3.81 para  $cD_2$  puede identificar entre BN y el evento no volcánico LGH, a diferencia del umbral de 9.24 que en caso de que la energía de la señal sea menor a este valor es un tipo LP, de otra forma resulta ser un VT. En consecuencia se establece que las características de los eventos analizados forman parte de la representación del árbol.

Al comparar la detección de las señales con la extracción de características de FFT y DWT, sin o con el uso del podamiento, se puede observar claramente que los resultados difieren en Fourier, aunque para Wavelet son iguales. Por lo tanto, el mejor modelo evaluado en DT resulta ser DWT con podamiento, al tomar en cuenta una menor cantidad de características y al lograr en los parámetros de rendimiento de A, S y R del 99 % aproximadamente y para P un valor del 98 % con una tolerancia igual a 100.

## Agradecimientos

Los autores agradecen la contribución económica de la Universidad de las Fuerzas Armadas - ESPE en el desarrollo de esta investigación a través de los proyectos 2013-PIT-014 y 2015-PIC-004 y al IGEPN por proporcionar la base de datos utilizada en este trabajo.

## Bibliografía

- Akram, J., & Eaton, D. (2012). Adaptive microseismic event detection and automatic time picking. *GeoConvention*, 1-5.
- Andrade, D., Hall, M., Mothes, P., Troncoso, L., Eissen, J., Samaniego, P.,... Yepes, H. (2005). Los peligros volcánicos asociados con el Cotopaxi. Corporación Editorial Nacional. Obtenido de <http://www.igepn.edu.ec/publicaciones-para-la-comunidad/39-los-peligros-volcanicos-asociados-con-el-cotopaxi/file>
- Cortés, G., Benítez, M., García, L., y Álvarez, I. (2015). A Comparative Study of Dimensionality Reduction Algorithms Applied to Volcano-Seismic Signals. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9, 1-11.
- Gabarda, S., y Cristóbal, G. (2010). Detection of events in seismic time series by time–frequency methods. *IET Signal Processing*, 4(4), 413-420.
- Galipienso, M. I. (2003). *Inteligencia artificial: modelos, técnicas y áreas de aplicación*. Editorial Paraninfo.
- Gareis, I. E., Gentiletti, G., Acevedo, R., y Rufiner, L. (2009). Extracción de características en interfaces cerebro computadoras mediante transformada wavelet discreta: Resultados preliminares. *Memorias del XVII Congreso Argentino de Bioingeniería (SABI 2009)*, 167, 58-62.
- Hall, M. (1999). *Correlation-based feature selection for machine learning*. PhD Thesis. Department of Computer Science, Waikato University, New Zealand.
- Jaramillo, C., León, R., Lara-Cueva, R., Benítez, D. S., y Ruiz, M. (2014). A new structure for sequential detection and maximum entropy spectral estimator for characterization of volcanic seismic signals. *IEEE Latin America Conference on Communications (LATINCOM)*, 1-6.
- Lahr, J. C., Chouet, B. A., Stephens, C. D., Power, J. A., & Page, R. A. (1994). Earthquake classification, location, and error analysis in a volcanic environment: Implications for the

- magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska. *Journal of Volcanology and Geothermal Research*, 62(1-4), 137-151.
- Lara-Cueva, R., Benítez, D., Carrera, E., Ruiz, M., y Rojo-Álvarez, J. (s.f.). Automatic Recognition of Long Period Events from Volcano Tectonic Earthquakes at Cotopaxi Volcano. *Submitted*.
- Lois, A. E., Psarakis, E. Z., & Pikoulis, E. V. (2013). A new thresholding type technique for the detection of seismic events. *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, 1-5.
- Prasad, N., Reddy, K., & Nirjogi, R. (2014). A Novel Approach for Seismic Signal Magnitude Detection Using Haar Wavelet. *5th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, 324-329.
- Salto Torres, M. G. (2014). *Análisis de señales sísmicas del volcán Cotopaxi mediante las transformadas de WAVELET y Fourier (Tesis de Pregrado)*. Universidad de las Fuerzas Armadas ESPE, Sangolquí - Ecuador.
- Segovia, M., y Alvarado, A. (2009). Breve Análisis de la Sismicidad y del Campo de Esfuerzos en el Ecuador. *Geología y Geofísica Marina y Terrestre del Ecuador: desde la costa continental hasta las Islas Galápagos*, 131-149.
- Undurraga, J., y Uribarri, M. (2006). Nuevas Herramientas Para la Medición de Parámetros Acústicos Utilizando Sweeps. *V Congreso Iberoamericano de Acústica*, 25-28.
- Vargas Jiménez, D. S., Rodríguez Espinosa, E. V., y Otero Foliaco, J. E. (2013). Alternativas para la Detección y Monitoreo de Amenazas Sísmicas basadas en Arduino. *Ingenierías USBMed*, 45, 45.
- Viracucha, E., y de la Bastida, J. (2014). Sistema Informático para el Procesamiento y Análisis de Señales Sísmicas de. *Revista Politécnica EPN*, 33(2).
- Wei, J. M., Wang, S. Q., Yu, G., Gu, L., Wang, G. Y., & Yuan, X. J. (2009). A novel method for pruning decision trees. *International Conference on Machine Learning and Cybernetics*, 1, 339 - 343.