

Towards an automatic detection system of signals at cotopaxi volcano

Román Lara-Cueva, Valeria Paillacho-Salazar & Michelle Villalva-Chaluisa

Grupo de Investigación en Sistemas Inteligentes (WICOM-Energy) y Centro de Investigación de Redes Ad Hoc (CIRAD), Departamento de Eléctrica y Electrónica, Universidad de las Fuerzas Armadas-ESPE, Sangolquí, Ecuador. ralara@espe.edu.ec, vapaillacho1@espe.edu.ec, mdvillalva@espe.edu.ec

Received: December 07th, 2015. Received in revised form: July 20th, 2016. Accepted: December 12th, 2016.

Abstract

Currently, Cotopaxi volcano has increased its seismic activity. Efficient and reliable tools are required for volcano monitoring, and early and effective emergency alerts are necessary. This article presents a supervised classification-based detection of seismic-volcanic and non-volcanic events recorded during 2010. Our algorithm use four features acquired using energy of approximation and detail coefficients of the wavelet decomposition analyzed with Daubechies and Symlet families. The classification of events was performed by the algorithm of decision trees with embedded techniques: cross-validation and pruning for reduction in the number features. The best results are obtained by application of the mother wavelet Symlet with accuracy of 99% and precision of 98%.

Keywords: Multilevel decomposition wavelet; decision tree; feature extraction; seismic event detection.

Hacia un sistema de detección automática de señales del volcán Cotopaxi

Resumen

Actualmente el Volcán Cotopaxi ha experimentado un incremento progresivo en su actividad sísmica, por lo que los especialistas requieren herramientas de alta eficiencia y confiabilidad para la monitorización volcánica. Este artículo presenta una detección basada en clasificación supervisada de los eventos sismo-volcánicos y no volcánicos registrados durante el año 2010. Nuestro algoritmo emplea cuatro características adquiridas por medio de la energía de los coeficientes de aproximación y detalle de la descomposición wavelet analizados con las familias Daubechies y Symlet. La clasificación de eventos fue realizada con el algoritmo de árboles de decisión empleando técnicas embebidas: validación cruzada y podamiento para una reducción del número de características. Los mejores resultados son obtenidos con la aplicación de la wavelet madre Symlet con una exactitud del 99% y una precisión de 98%.

Palabras clave: Descomposición multinivel wavelet; árbol de decisión; extracción de características; detección de eventos sísmicos.

1. Introducción

La actividad volcánica constituye un factor predominante en el desencadenamiento de fenómenos como el calentamiento global y el cambio climático, los cuales en los últimos años han tenido trascendencia por el impacto que ocasionan principalmente a los seres vivos en la Tierra [1]. El aumento de la actividad volcánica a nivel mundial representa un riesgo prominente para la población y una desmedida pérdida de recursos [2], por lo que para disminuir los efectos ocasionados por estos sucesos, se han venido desarrollando y perfeccionando varios tipos de sistemas de

monitorización [3,4], que entregan información detallada del comportamiento que presenta un volcán. Estos sistemas reportan eventos tales como: eventos de largo periodo (LP), eventos volcánico - tectónicos (VT), eventos híbridos (HYB) y tremores (TRE) [5]. Al realizar una vigilancia de manera ininterrumpida se genera una gran cantidad de información para ser analizada, generalmente de forma visual, por lo que su determinación y clasificación puede resultar una tarea compleja, de gran consumo de tiempo y sensible a errores.

Para apoyar el trabajo de los especialistas en esta área se han desarrollado varias investigaciones sobre el tema que se describen en detalle más adelante en la Sección II, en las que

se incluyen la caracterización temporal y frecuencial de los eventos sísmicos [6], así como también técnicas para la detección y clasificación de los mismos, empleando algoritmos de aprendizaje de máquina, tales como: redes neuronales artificiales (ANN, del inglés *Artificial Neural Networks*) [7,8], máquina de vectores soporte vectorial (SVM, del inglés *Support Vector Machines*) [9], entre otros; sin embargo la mayoría de estas propuestas no han garantizado un enfoque en tiempo real.

Considerando que la identificación de los eventos de forma visual depende del grado de expertise del especialista, el objetivo principal de esta investigación es realizar una detección automática basada en técnicas de clasificación de señales volcánicas y no volcánicas que permita dictaminar medidas preventivas ante un posible proceso eruptivo, lo que contribuye a precautelar la integridad de los habitantes directamente afectados por un volcán. Para ello se propone utilizar las características representativas de las señales que se obtienen a partir de la extracción de los coeficientes de interés de la Transformada Wavelet, en contexto con las frecuencias de operación de los eventos LP, VT y Truenos (LGH, del inglés *Lightnings*). Asimismo por medio de la aplicación de la Transformada Discreta de Wavelet (DWT, del inglés *Discrete Wavelet Transform*) se busca solucionar el problema de la resolución temporal-frecuencial que habitualmente aparece al aplicar la Transformada de Fourier, al ser dependiente del inventariado, de acuerdo al principio de Heisenberg [10]. Para la elección de la Wavelet Madre (MW, del inglés *Mother Wavelet*) entre las familias se emplea el criterio de similaridad con las señales en análisis [11]. Posteriormente la clasificación se efectúa a través de árboles de decisión (DT, del inglés *Decision Trees*), mediante matrices características divididas en bancos de entrenamiento y prueba. Las técnicas embebidas de DT que se emplean son: podamiento y validación cruzada, las cuales seleccionan las características que permitan discriminar de mejor manera los eventos y evitar el problema del sobreajuste en el clasificador. Por último la detección de las señales se obtiene por medio de un pos-procesamiento, en el cual se desarrolla un algoritmo de ajuste de los eventos precedidos por el clasificador.

El artículo está organizado de la siguiente manera. La Sección II realiza una revisión de los trabajos relacionados en detección y clasificación de eventos vulcanológicos. La Sección III describe la base de datos disponible del volcán Cotopaxi. La metodología propuesta para la caracterización, extracción, selección y detección de las señales provenientes de un volcán se expone en la Sección IV. En la Sección V se muestran los resultados de forma experimental de cada una de las etapas implementadas y los rendimientos obtenidos empleando técnicas de clasificación para la detección. Finalmente en la Sección VI se presenta la discusión y conclusiones de la investigación, así como los trabajos futuros.

2. Trabajos relacionados

Las investigaciones que se relacionan con la detección de las señales de origen sismo-volcánico establecen diferentes métodos basados en análisis de tiempo, frecuencia y escala, dentro de estos difieren el tipo de procesamiento de las señales para la extracción de patrones, considerando varios parámetros.

En [12] el parámetro de detección se calcula como una función de los coeficientes de autocorrelación de 3 canales de datos. Mientras que en [13] describen un método de detección de eventos basados en un análisis tiempo-frecuencia, a través de la distribución de Wigner. Asimismo en [14] se desarrolla un detector de eventos aplicando conceptos de umbral, donde se emplea Transformada Wavelet Haar obteniendo el 1% de detecciones perdidas. Por otra parte en [15] se describe un detector de eventos LP basado en clasificación con algoritmos DT y K-vecinos más cercanos (k-NN, del inglés *k-Nearest Neighbors*), con un número de 49 y 257 características producto de un análisis en escala y frecuencia, de ahí que este último alcanza el 99% de precisión y sensibilidad.

Adicionalmente, se desarrollan clasificadores de eventos como en [16], que realiza un análisis individual de 7 señales del volcán Nevado del Ruiz de Colombia con la extracción de 12 características obtenidas mediante la descomposición Wavelet por medio de técnicas de clasificación como DT, RNA y SVM; particularmente para los LP y VT se presenta errores del 15% y 4% respectivamente. El sistema descrito en [17], muestra una identificación de tres señales, del ruido de fondo (BN, del inglés *Background Noise*), a través de un modelado oculto de Markov (HMM, del inglés *Hidden Markov Model*), con una tasa de clasificación del 80%. Por último en [18] se aplica las transformadas Wavelet discreta, continua y estacionaria, de lo cual se elige la última para la clasificación con RNA con resultados del 89.9%.

Acorde a la revisión de la literatura la mayor parte de las investigaciones utilizan técnicas de aprendizaje de máquina para la clasificación; no obstante pocos realizan una detección basada en clasificación. Además se puede evidenciar que en los trabajos previos se aplican varios métodos para la extracción de características con un número considerable de las mismas, por esta razón se resolvió trabajar con la energía de los coeficientes de aproximación y detalle de Wavelet.

3. Descripción de la base de datos

Esta investigación se enfoca en el estudio de señales del volcán Cotopaxi, debido a que presenta un alto grado de peligrosidad, por su forma eruptiva y sobre todo por su ubicación [19]. Este es uno de los volcanes al cual se le asigna gran parte de los recursos disponibles para su monitorización, lo que asegura un seguimiento continuo del mismo [20].

En el Ecuador, el Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN) es la institución encargada de la monitorización e investigación de eventos de origen tectónico y volcánico, para ello cuenta con una Red de Sismógrafos y Acelerógrafos de alta calidad que cubren el 70% del territorio nacional, las cuales entregan información de manera permanente, lo que permite a los expertos el análisis de datos hipocentrales, mecanismos focales, entre otros [21]. Específicamente el volcán Cotopaxi posee una red compuesta por seis estaciones de corto periodo (SP, del inglés *Short Period*) con una respuesta en frecuencia de 1 a 50 Hz y seis estaciones de banda ancha (BB, del inglés *Broadband*) con una respuesta en frecuencia de 0.1 a 50 Hz. Cada estación presenta instrumentos que son capaces de almacenar señales en registros de 1200s de duración.

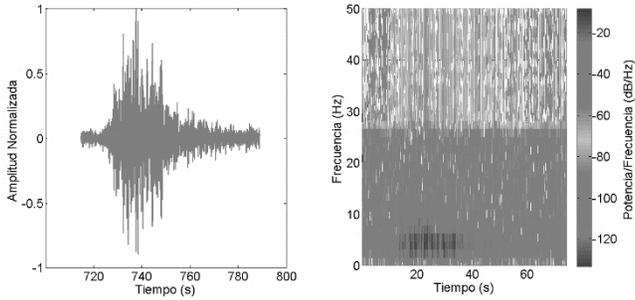


Figura 1. Ejemplo de un evento sismo-volcánico LP en el volcán Cotopaxi con una duración aproximada de 60s y su respectivo espectrograma
Fuente: Elaboración propia con datos del IGEPN

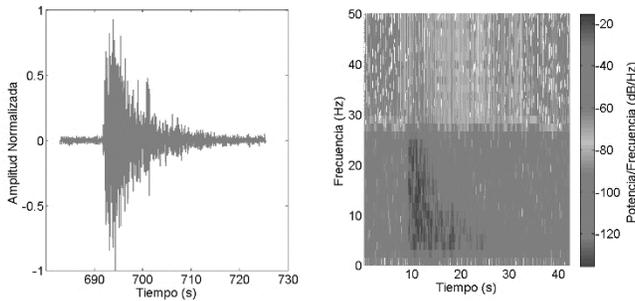


Figura 2. Ejemplo de un evento sismo-volcánico VT en el volcán Cotopaxi con una duración aproximada de 20s y su respectivo espectrograma
Fuente: Elaboración propia con datos del IGEPN

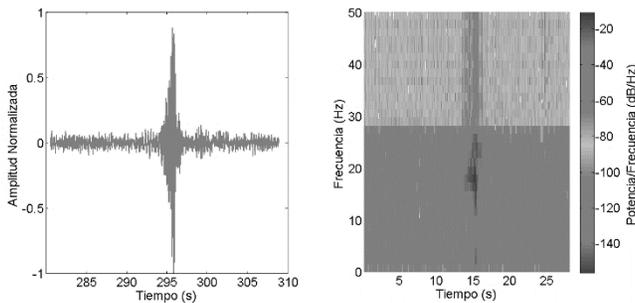


Figura 3. Ejemplo de un evento no volcánico LGH en el volcán Cotopaxi con una duración aproximada de 3s y su respectivo espectrograma
Fuente: Elaboración propia con datos del IGEPN.

Dentro de los eventos para analizar, se buscó aquellos que provean la información adecuada para la construcción de un modelo bien definido por medio de DT, de ahí que se eligió a los LP y VT ya que a partir de sus datos se puede indicar la condición actual del volcán [20] y los LGH necesarios de identificar dentro del registro sísmico por su continua ocurrencia en el mismo. Las Figs.1-3 muestran las bandas de frecuencia (f) de los eventos, de donde en los LP su contenido espectral se encuentra entre $0,5 \text{ Hz} < f < 5 \text{ Hz}$, seguido de los VT que a diferencia del anterior presentan un rango más alto $f < 10 \text{ Hz}$, por último los LGH caracterizado por poseer altas frecuencias entre $17 \text{ Hz} < f < 20 \text{ Hz}$.

La base de datos fue obtenida a partir de registros proporcionados por el IGEPN durante el año 2010. Dichos registros han sido previamente evaluados por los expertos y

a su vez discriminados en los diferentes tipos de señales sísmicas; con la existencia de 759 LP, 116 VT y 100 LGH. Por consiguiente se definió trabajar con 100 eventos para los LP, 100 VT y 100 LGH, al estimar que la complejidad del detector es dependiente del tamaño de la base de datos [22]. Asimismo la base de datos fue dividida en partes iguales, en el banco de entrenamiento (*train*) y el banco de prueba (*test*), donde cada banco posee señales independientes una del otro.

Seguidamente para la determinación de los tiempos de inicio y fin de los eventos, se realizó una inspección visual de cada una de las señales, de donde se estableció los valores aproximados de los mismos.

4. Materiales y métodos

En este trabajo, se desarrolla una detección automática de tres tipos de señales los LP, VT y LGH, de acuerdo al esquema planteado en la Fig. 4, donde el primer bloque presenta el pre-procesamiento y segmentación, etapa en la cual principalmente se disminuye el ruido existente en cada registro sísmico para ser divididos en ventanas, el segundo bloque considera el procesamiento de la señal donde se realiza procesos para la extracción y selección de características fundamentales para que el algoritmo de aprendizaje DT establezca un modelo adecuado de las señales. Finalmente la etapa de detección es la que garantiza el rendimiento del sistema mediante las métricas de desempeño propuestas.

4.1. Pre-procesamiento

La señales sísmo-volcánicas al ser provenientes de los movimientos de la superficie de la Tierra pueden presentar variaciones por agentes externos; por lo que los instrumentos al poseer una alta sensibilidad pueden registrar en su lectura

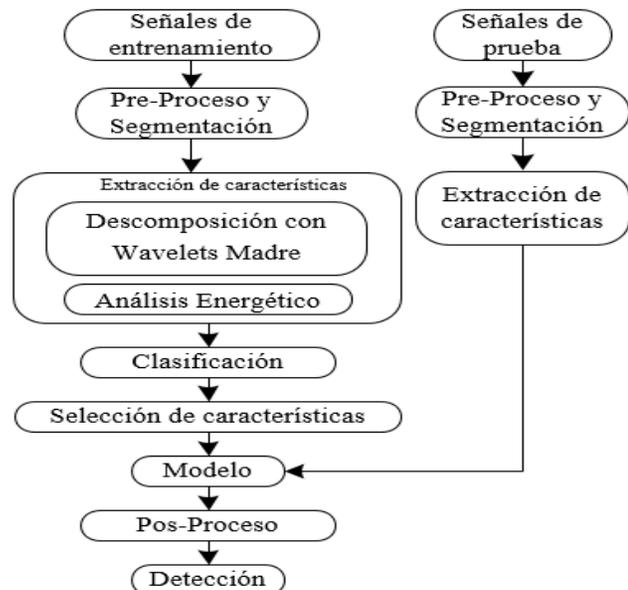


Figura 4. Diagrama de bloques de las etapas implementadas en el detector de eventos sísmo-volcánicos y no volcánicos
Fuente: Elaboración propia

sísmica ruido de fondo, caracterizado por su densidad espectral de potencia plana y por ser una señal que no posee correlación en relación al tiempo; lo que resulta en una adquisición con presencia de errores. El objetivo del pre-procesamiento es la obtención de señales reales, mediante la aplicación de dos procedimientos: el filtraje y la normalización de la información. El filtraje por su parte permite conservar las componentes espectrales principales que contiene la información útil de las señales; al poseer cada evento sísmico una frecuencia de operación, se propuso un filtro de respuesta finita al impulso (*FIR del inglés, Finite Impulse Response*) con un orden de 128, ya que este presenta una amplitud desigual necesaria para las condiciones de las señales. El filtro se encuentra en el rango de $0,5 \text{ Hz} < f < 50 \text{ Hz}$ para evitar la supresión de información importante de las señales y descartar la aparición de un pico prominente en 0.2 Hz producido por sismos de baja intensidad [23]. Finalmente se realiza la normalización de las señales para conseguir una media cero y varianza uno ($\mu = 0, \nu = 1$) de la información contenida en cada registro sísmico.

4.2. Segmentación

Para satisfacer los requerimientos de tiempo real con una carga computacional moderada se establece un proceso de fragmentación de las señales, al ser la dimensión de la matriz de características directamente dependiente del valor de los segmentos, se considera los parámetros de rendimiento alcanzados en [15]; donde se concluye que una ventana $\omega = 15s$ permite detectar los eventos con precisión.

De esta forma se genera una matriz ventaneada \mathbf{s}_i , definida por $\mathbf{s}_i = [\mathbf{s}_{i,1}^T, \mathbf{s}_{i,2}^T, \dots, \mathbf{s}_{i,j}^T]^T$, donde $\mathbf{s}_{i,j}$ es el segmento de una señal con una ventana de 15s, la misma que puede contener algún evento dentro de ella. A partir de aquí, cada señal segmentada se almacena en la matriz \mathbf{S} constituida por $\mathbf{S} = [\mathbf{s}_1^T, \mathbf{s}_2^T, \dots, \mathbf{s}_m^T]^T$, de tamaño $m = r/\omega$ donde r es el tiempo total de un registro capturado por el sismógrafo, este valor es fijo e igual a 1200s.

De igual forma se determinaron los tiempos de inicio y fin, para luego ser etiquetados junto a cada segmento $\mathbf{s}_{i,j}$; con valores de: -1 para BN, 1 para el evento LP, 2 para los VT y 3 para los LGH.

4.3. Extracción de características

Esta etapa constituye la principal dentro del sistema, a partir de las sub-etapas descritas a continuación:

4.3.1. Descomposición multinivel Wavelet

Un sistema Wavelet es un conjunto de bloques de construcción para representar una señal. La expansión Wavelet da una localización tiempo-frecuencia de la señal, esto significa que la mayor parte de energía está bien representada por unos pocos coeficientes de dilatación [24].

La DWT analiza la señal para ser descompuesta en varios niveles de aproximación y detalle mediante la aplicación de un proceso de filtraje tipo pasa-alto y pasa-bajo [25], en

Tabla 1. Coeficientes de aproximación y detalle de los eventos sismo-volcánicos y no volcánicos

Tipo de Evento	Frecuencia de Operación	Coefficientes de aproximación y detalle
LP	0-3.1Hz , 25-50Hz	cA5, cD2
VT	6.2-12.5Hz, 25-50Hz	cD4, cD2
LGH	12.5-25Hz , 25-50Hz	cD3, cD2

Fuente: Elaboración propia

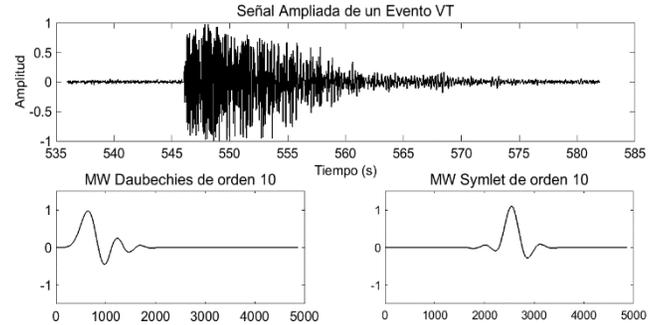


Figura 5. Comparación de la forma de onda de un evento VT con las familias DB10 y Sym10

Fuente: Elaboración propia

efecto se aplica la DWT a la matriz \mathbf{S} lo que origina \mathbf{w}_i dada por $\mathbf{w}_i = [\mathbf{w}_{i,1}^T, \mathbf{w}_{i,2}^T, \dots, \mathbf{w}_{i,j}^T]^T$, donde $\mathbf{w}_{i,j} = g_\omega(\mathbf{s}_{i,j})$, siendo g_ω el operador de la descomposición multinivel Wavelet definido por los coeficientes de aproximación y detalle; para finalmente conformar la matriz \mathbf{W} establecida por $\mathbf{W} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_i^T]^T$.

Los coeficientes de aproximación y detalle se aplican con relación a las bandas de operación descritas anteriormente en la Sección III, además se incluye el análisis en la banda de $25 \text{ Hz} < f < 50 \text{ Hz}$, esto de acuerdo a los resultados en [15], donde se muestra que los eventos LP tienen características importantes en ese rango de frecuencia que pueden aportar a una mejor discriminación.

La Tabla 1 describe los coeficientes de aproximación (cA) y los coeficientes de detalle (cD) dispuestos bajo los valores numéricos de los niveles de descomposición a ser utilizados en la extracción.

Por concepto de similaridad se eligió como MW en orden 10 a las familias *Daubechies* (DB10) y *Symlet* (Sym10) [11] al contrastarlas con las señales de interés, lo que se visualiza en la Fig. 5.

4.3.2. Análisis energético

Una vez identificados los coeficientes de aproximación y detalle (cA5, cD4, cD3 y cD2), se procede al cálculo de la energía de la matriz \mathbf{W} , por medio de la ec. (1).

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

Donde x representa cada uno componentes de la señal.

De esta manera se elabora la matriz \mathbf{E} compuesta por $\mathbf{E} = [\mathbf{e}_1^T, \mathbf{e}_2^T, \dots, \mathbf{e}_i^T]^T$, donde cada \mathbf{e}_i está representado por $\mathbf{e}_i = [\mathbf{e}_{i,1}, \mathbf{e}_{i,2}, \dots, \mathbf{e}_{i,j}]$ con $j = 1, 2, \dots, 4$.

4.4. Algoritmo de clasificación DT

Para esta etapa la técnica para el aprendizaje automático empleada es árboles de decisión de tipo supervisada, de acuerdo a [26-28]. El objetivo de DT es la construcción de un modelo estimado para la predicción de valores, de acuerdo a decisiones de desambiguación a partir de las características más significativas de los datos. El árbol de decisión inicia en el nodo raíz; cada condición en él está representada por un nodo, de donde las ramas son las posibles respuestas que pueden contribuir a la solución de dicha condición. Asimismo existen los nodos hoja que proporcionan el valor de la variable destino.

Para obtener una clasificación con la más alta precisión posible es necesaria una generalización adecuada conseguida por medio de la creación de subárboles, definidos por las particiones recursivas del árbol de lo cual un nuevo nodo cumple la función de nodo raíz.

El parámetro libre en DT es la frondosidad o profundidad (*depth*), el cual está establecido por las métricas de Entropía ó Índice de Gini que se expresan en la ec. (2) y ec. (3) [29], de lo cual el primero mantiene la ganancia de información de los datos mientras que el segundo agrega la función de impureza.

$$E(t) = - \sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t) \quad (2)$$

$$G(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2 \quad (3)$$

Donde $p(i|t)$ denota la fracción de registros pertenecientes a la clase i en un nodo t dado, y c es el número de clases.

4.5. Detección y rendimiento

La detección está basada en algoritmos de clasificación específicamente en DT, el clasificador predice una clase K con valores de; 1, 2, 3 ó -1 correspondiente a los eventos LP, VT, LGH o BN respectivamente. A continuación se realiza una reducción de características mediante las técnicas de selección embebidas de DT (validación cruzada y podamiento [30,31]) para las dos familias de MW. Los modelos obtenidos en la etapa de clasificación se verifican a partir de la matriz \mathbf{E} y las etiquetas del banco de prueba, donde los datos predecidos son sometidos a un procesamiento, el mismo que consisten en lo siguiente; inicialmente se asigna el valor de 1 a la existencia de un evento (1, 2 o 3) y 0 al BN identificando únicamente la presencia del evento sin importar cuál sea. De esta manera se examinó las etiquetas, tomando en cuenta que las señales en

análisis pueden estar contenidas en más de dos segmentos, para lo cual se disminuyó a tan solo una etiqueta por evento considerando un criterio de transición de 1 a 0.

Por último y en relación con la clasificación supervisada se realiza una comparación de las etiquetas predecidas con las etiquetas reales, donde el algoritmo determina automáticamente un umbral que maximiza el rendimiento del detector.

La evaluación de la detección de los eventos se estableció bajo los parámetros de Exactitud (A), Precisión (P), Sensibilidad (R) y Especificidad (S) de acuerdo a las expresiones definidas en ec. (4)-(7).

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

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

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

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

Donde N_c es el número de patrones clasificados correctamente, N_T es el número total de los patrones empleados 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.

5. Resultados

Con la finalidad de evaluar la metodología propuesta, en esta sección se presenta los resultados de cada etapa de la detección basada en clasificación de eventos sismo-volcánicos y no volcánicos. El desarrollo de los experimentos se llevó a cabo en el software de simulación Matlab® R2013a, en una PC Core I7 con 2.80 GHz y 8 GB de RAM.

En la Fig. 6 se muestra un ejemplo del proceso de descomposición basado en frecuencias de un evento LP, en el cual se puede observar que el nivel de descomposición para este tipo de evento es el 5 que comprende el rango de frecuencias entre 0 Hz a 3.1 Hz (cA5). Del mismo modo se examinó el coeficiente cD2 que presenta un intervalo comprendido de 25 Hz a 50 Hz.

Por otra parte en la Tabla 2 se resumen los valores medios de la energía calculada para los 4 coeficientes con las wavelet madre Sym10 y DB10, donde se puede observar los valores representativos para cada tipo de evento. El valor mínimo energía se evidencia en las señales no volcánicas LGH en relación a los eventos de origen volcánico.

La representación del árbol de las principales características para los dos tipos de MW, se consiguen en base al algoritmo de clasificación DT; de donde para DB10 se estructura un árbol de 44 nodos, al contrario de Sym10 que abarca 41 nodos lo cual se ilustra en la Fig. 7. La identificación del nodo raíz significa para

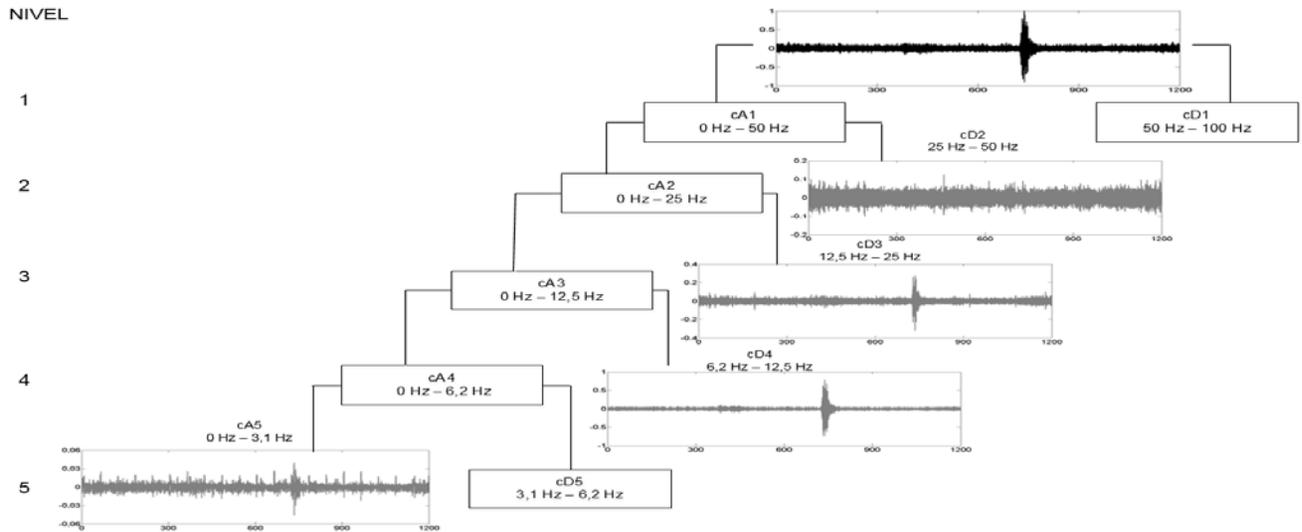


Figura 6. Descomposición multinivel de Wavelet de un evento sísmo-volcánico LP, basado en bandas de frecuencia.
Fuente: Elaboración propia

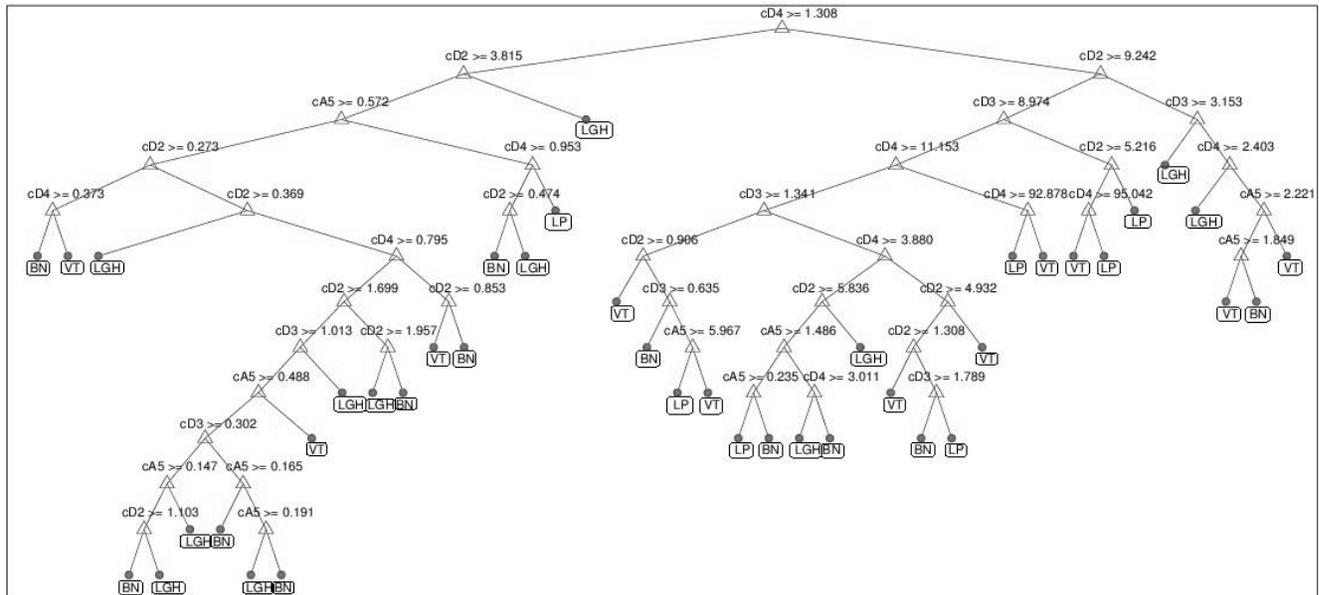


Figura 7. Representación del árbol de decisión de las características de energía de los coeficientes cA y cD de la matriz W aplicando la MW Sym10.
Fuente: Elaboración propia

Tabla 2.
Energía de los coeficientes de aproximación y detalle de los eventos volcánicos y no volcánicos

Evento	cA5		cD4		cD3		cD2	
	Sym	DB	Sym	DB	Sym	DB	Sym	DB
LP	1.60	1.19	3.54	3.64	1.96	1.97	3.02	3.01
VT	0.86	0.66	1.87	1.89	1.90	1.89	2.49	2.49
LGH	0.48	0.42	0.88	0.89	1.07	1.07	1.78	1.78

Fuente: Elaboración propia

el árbol continuar con la formación de las reglas que contribuyen a la determinación de los tres tipos de señales. En ambos casos el coeficiente cD4 funciona como nodo raíz; pero con una diferencia de umbrales, siendo 1.169 para

DB10 y 1.308 para Sym10. En consecuencia la banda de frecuencias de 6.2 Hz a 12.5 Hz, es indispensable para la formación de los árboles.

En las Figs. 8, 9 se observa la aplicación de las técnicas de selección embebidas del algoritmo DT para la MW Sym10. De tal modo que para la validación cruzada se genera un árbol compuesto por 3 nodos, siendo el nodo cD4 el que permite discriminar directamente las señales no volcánicas de los eventos sísmo-volcánicos. De ahí que si el valor de decisión en dicho nodo es inferior al umbral de 1.308 corresponde al ruido o un LGH, de lo contrario se puede identificar entre las clases LP o VT. De la misma manera la estructura obtenida mediante el podamiento refleja una semejanza a la técnica anterior; no obstante en los nodos cD2 difieren en sus umbrales.

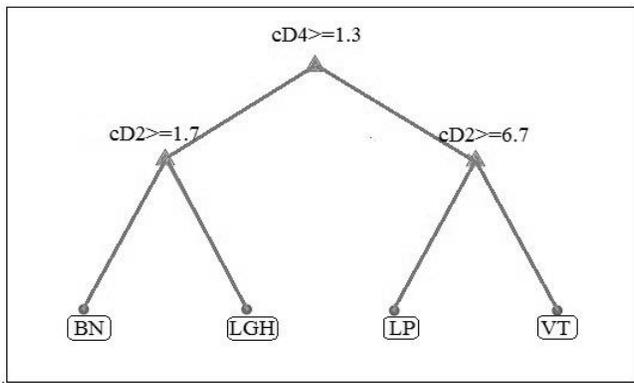


Figura 8. Características con la técnica de selección embebida validación cruzada

Fuente: Elaboración propia

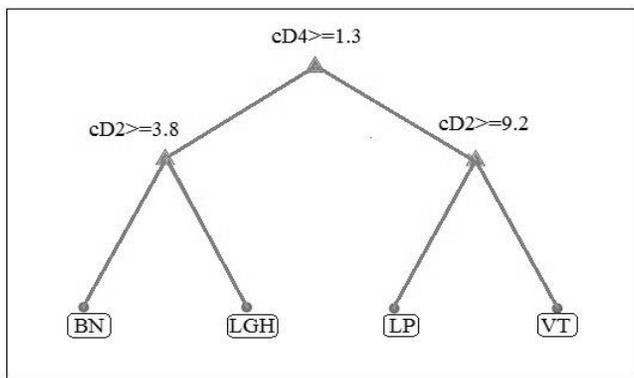


Figura 9. Características con la técnica de selección embebida podamiento

Fuente: Elaboración propia

Tabla 3. Resultados del rendimiento de la matriz W con Symlet 10

Método	A (%)	P (%)	R (%)	S (%)
Defecto	99.98	98.65	99.32	99.98
V. Cruzada	99.88	97.83	91.84	99.97
Podamiento	99.98	98.65	99.32	99.98

Fuente: Elaboración propia

Los parámetros de desempeño de la detección basada en clasificación se adquirieron a partir del banco de características *test*. Mediante la aplicación de los dos tipos de MW se consiguió el modelo por defecto y los modelos mediante las técnicas de selección de características: validación cruzada y podamiento

En la Tabla 3 se describe la aplicación de características por la MW Symlet 10, de lo cual con el modelo por defecto y podamiento se muestra una igualdad en las cuatro métricas de rendimiento. Al contrastar los modelos anteriores con el de validación cruzada existe una mínima variación en A, P y S. En cambio para la R se diferencia en aproximadamente un 7%, lo que demuestra que existe un número menor de verdaderos positivos; además de contribuir a la disminución de la probabilidad de que exista un evento erróneo.

El análisis con MW DB10 se explica en la Tabla 4, de donde se evidencia una coincidencia de rendimientos entre el modelo por defecto y la técnica de podamiento. De igual forma al equiparar

Tabla 4. Resultados del rendimiento de la matriz W con Daubechies 10

Método	A (%)	P (%)	R (%)	S (%)
Defecto	99.83	93.75	91.84	99.93
V. Cruzada	99.88	98.53	91.16	99.98
Podamiento	99.83	93.75	91.84	99.93

Fuente: Elaboración propia

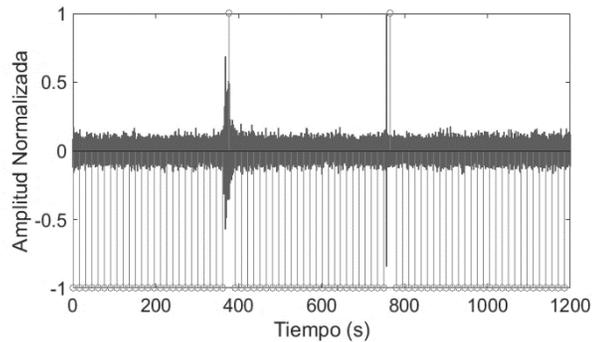


Figura 10. Detección de los eventos sismo-volcánico LP y no volcánico LGH

Fuente: Elaboración propia

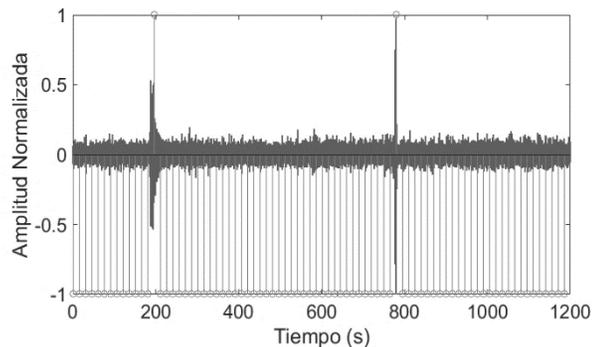


Figura 10. Detección de los eventos sismo-volcánico VT y no volcánico LGH

Fuente: Elaboración propia

los resultados ya mencionados, con el de validación cruzada se observa un valor más alto en el parámetro de P con el 98.53%.

En definitiva al evaluar los 6 modelos existe una analogía de resultados; sin embargo el contraste uno del otro radica en el número de características empleadas en el cálculo del rendimiento final de la detección. Las técnicas de selección de características definen que tan solo con las características *cD2* y *cD4* se puede establecer las cuatro clases de señales.

Por último se observa ejemplos de la salida del detector. En la Fig. 10 se determina un evento sismo-volcánico LP y un evento no volcánico LGH del BN; mientras que en la Fig. 11 se identifica un evento VT y LGH.

6. Discusión y conclusiones

La detección y clasificación de eventos sismo-volcánicos constituye una línea de investigación relevante en el desarrollo de sistemas automáticos que permitan indicar la

ocurrencia de algún proceso eruptivo o sísmico con un factor de riesgo. Por tal motivo en el presente estudio se desarrolló un sistema detector de eventos sismo-volcánicos y no volcánicos, planteado en varias etapas de manera sistemática y descriptiva, obteniendo una mejora en el parámetro de exactitud A del 11% aproximadamente, en comparación con el trabajo propuesto en [17], al emplear ambos bancos de entrenamiento y prueba independientes.

Las principales características de las señales para el algoritmo de clasificación DT, se extrajeron a través de un análisis de energía de los coeficientes cA5, cD4, cD3 y cD2 con relación a las frecuencias de operación de las señales; de donde se obtuvo menor número de características que los estudios realizados en [15] y [16], en los cuales se emplea 49 y 12 características respectivamente, con técnicas asociadas a la Transformada Wavelet, sin embargo en el primer caso se realiza una descomposición de nivel 4 mientras que en el segundo caso el análisis de los coeficientes de aproximación y detalle se realiza en función a la autocorrelación en frecuencia y momentos estadísticos.

Por medio de las técnicas embebidas del algoritmo de clasificación DT, validación cruzada y podamiento se eliminaron los nodos innecesarios a través del parámetro de frondosidad, de donde su número se reduce de 41 a 3 características resultantes con la MW Sym10. Además mediante el valor de los umbrales del nodo cD4 fue posible discriminar las señales de origen no volcánicas de los eventos sismo-volcánicos si necesidad que evaluar cD2.

Para la evaluación del rendimiento del detector basado en clasificación se puso en consideración 6 modelos, de lo cual el modelo de podamiento correspondiente a la extracción de características con la MW Sym10 arrojó los mejores resultados con un 99.98% en A, 98.65% en P, 99.32% en S y 99.98% en R. Los resultados alcanzan estos valores debido a que la detección se realiza bajo un aprendizaje supervisado.

En el futuro se complementará este trabajo con la detección y clasificación de los eventos HYB y TRE aplicando la técnica de aprendizaje SVM en su modo multi-clase, al no ser considerados en esta investigación por presentar un número reducido de los mismos en la base de datos disponible.

Agradecimientos

Los autores desean expresar su agradecimiento a la Universidad de las Fuerzas Armadas – ESPE por el apoyo económico en el desarrollo de este trabajo a través del Proyecto 2013-PIT-014 y 2015-PIC-004, así como también al IGEPN por proveer la base de datos empleada en esta investigación.

Referencias

- [1] González, M., Jurado, E., González, S., Aguirre, O., Jiménez, J. and Nívar-Cháidez, J.D.J., Cambio climático mundial: Origen y consecuencias. *Ciencia UANL*, 6(3), pp. 377-385, 2003.
- [2] Tilling, R. y Beate, B., Los peligros volcánicos. *Apuntes breves sobre un curso breve*. Organización Mundial de Observatorios Vulcanológicos, 1993. pp. 2-3.
- [3] Nugroho, A., and Winarko, E., Geographical information system and Web Service implementation for volcanic eruptions geologic disasters surveillance, monitoring, and mitigation system. *International Conference on Electrical Engineering and Informatics (ICEEI)*, 2011. pp. 17-19. DOI: 10.1109/ICEEI.2011.6021766
- [4] Wener-Allen, G., Johnson J., Ruiz, M. and Lees, J., Monitoring volcanic eruptions with a wireless sensor network, *Wireless Sensor Networks*, 2005. Workshop on Proceedings of the Second European, 2005. pp. 108-120. DOI: 10.1109/EWSN.2005.1462003
- [5] Wassermann, J., *Volcano seismology*. Fürstenfeldbruck, Geophysikalisches Observatorium der Ludwig-Maximilians Universität München, 2011. pp. 1-67.
- [6] Arciniega-Ceballos, A., Chouet, B. and Dawson, P., Long-period events and tremor at Popocatepetl volcano (1994-2000) and their broadband characteristics. *Bulletin of Volcanology*, 65(2), pp. 124-135, 2003. DOI: 10.1007/s00445-002-0248-8
- [7] Ibs-von Seht, M., Detection and identification of seismic signals recorded at Krakatau volcano (Indonesia) using artificial neural networks, *Journal of Volcanology and Geothermal Research*, 176(4), pp. 448-456, 2008. DOI: 10.1016/j.jvolgeores.2008.04.015
- [8] Curilem, G., Vergara, J., Fuentealba, G., Acuña, G. and Chacón, M., Classification of seismic signals al Villarrica volcano (Chile) using neural networks and genetic algorithms. *Journal of Volcanology and Geothermal Research*, 180(1), pp. 1-8, 2009. DOI: 10.1016/j.jvolgeores.2008.12.002
- [9] Celik, E., Atalay, M. and Bayer, H. Earthquake prediction using seismic bumps with Artificial Neural Networks and Support Vector Machines. *Signal Processing and Communications Applications Conference (SIU)*, 2014. pp. 730-733. DOI: 10.1109/SIU.2014.6830333
- [10] Doroslovacki, M., Discrete-time signals and uncertainty relations involving ordinary second moments in time and frequency. *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1994. pp. 186-189. DOI: 10.1109/TFSA.1994.467262
- [11] Ahuja, N., Lertrattanapanich, S. and Bose, N., Properties determining choice of mother wavelet. *IEE Proceedings-Vision, Image and Signal Processing*, 152(5), pp. 659-664, 2005. DOI: 10.1049/ip-vis:20045034
- [12] Magotra, N., Nalley, D. and Brozek, J., Seismic event detection using three-component data. *IEEE Transactions on Geoscience and Remote Sensing*, 30(3), pp. 642-644, 1992. DOI: 10.1109/36.142949
- [13] Gabarda, S. and Cristóbal, G., Detection of events in seismic time series by time-frequency methods. *Signal Processing, IET*, 4(4), pp. 413-420, 2010. DOI: 10.1049/iet-spr.2009.0125
- [14] Thanasopoulos, I.A. and Avaritsiotis, J.N., Seismic detection and time of arrival estimation in noisy environments based on the haar wavelet transform. *Sensor Array and Multichannel Signal Processing Workshop*, 2008. SAM 2008. 5th IEEE, Darmstadt, 2008. pp. 433-436. DOI: 10.1109/SAM.2008.4606906
- [15] Lara-Cueva, R.A., Benítez, D.S., Carrera, E.V., Ruiz, M. and Rojo-Álvarez, J.L., Automatic recognition of long period events from volcano tectonic earthquakes at Cotopaxi volcano. *IEEE Transactions on Geoscience and Remote Sensing*, 54(9), pp. 5247-5257, 2016. DOI: 10.1109/TGRS.2016.2559440
- [16] Hurtado, J.E., Henao, R. y Castellanos G. Clasificación de señales sísmicas por medio de onditas y máquinas de soporte vectorial. *Universidad Nacional de Colombia*. Colombia. Observatorio Sismológico de Quindío. Colombia. INGEOMINAS. sfsf CO, 2002.
- [17] Gutiérrez, L., Ibáñez, J., Cortes, G., Ramírez, J., Benítez, C., Tenorio, V. and Isaac, A., Volcano-seismic signal detection and classification processing using hidden Markov models. Application to San Cristóbal volcano, Nicaragua. *Geoscience and Remote Sensing Symposium*, 2009 IEEE International, IGARSS 2009, pp. 12-17. DOI: 10.1109/IGARSS.2009.5417428
- [18] Hoffman, A., Hoogenboezem, C., van der Merwe, N. and Tollig, J.A., Seismic buffer recognition using wavelet based features. *Geoscience and Remote Sensing Symposium Proceedings*, 1998 IEEE International, IGARSS 1998, pp. 1333-1335. DOI: 10.1109/IGARSS.1998.691399
- [19] Andrade, D., Hall, M., Mothes, P., Troncoso, L., Eissen, J., Samaniego, P., Egred, J., Ramón, P., Rivero, D. y Yepes, H., Los peligros volcánicos asociados con el Cotopaxi. *Corporación Editorial Nacional*, 2005.

- [20] Salazar, D. y D'Ercole, D., Percepción del riesgo asociado al volcán Cotopaxi y vulnerabilidad en el Valle de Los Chillos (Ecuador), l'Institut Français d'Études Andines, 38 (3), pp. 849-871, 2009.
- [21] Rivera, K., Mapa de periodos predominantes del suelo del área urbana de la ciudad de Loja (Ecuador), obtenidos a partir de medidas de ruido ambiental, Tesis, Departamento de Geología y Minas, Universidad Técnica Particular de Loja, Loja, Ecuador, 2013.
- [22] Cortes, G., Benitez, M., García, L., Álvarez, I., Ibáñez, J., A comparative study of dimensionality reduction algorithms applied to volcano-seismic signals. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, (99), pp. 1-11, 2015. DOI: 10.1109/JSTARS.2015.2479300
- [23] Chouet, B., Page, R., Stephens, C., Lahr, J. and Power, J., Precursory swarms of long-period events at Redoubt volcano (1989–1990), Alaska: Their origin and use as a forecasting tool. *Journal of Volcanology and Geothermal Research*, 62(1), pp. 95-135, 1994. DOI: 10.1016/0377-0273(94)90030-2
- [24] Burrus, C., Gopinath, R. and Guo, H., Introduction to wavelets and wavelet transforms: A primer, Prentice-Hall International, 1997.
- [25] Mojsilovic, A., Popovic, M. and Rackov, D., On the selection of an optimal wavelet basis for texture characterization. *IEEE Transactions on Image Processing*, 9(12), pp. 2043-2050, 2000. DOI: 10.1109/83.887972
- [26] Bishop, C., *Pattern recognition and machine learning*, Springer, 2006.
- [27] Landgrebe, D., A survey of decision tree classifier methodology. *IEEE transactions on Systems, Man and Cybernetics*, 21(3), pp. 660-674, 1991. DOI: 10.1109/21.97458
- [28] Magerman, D., Statistical decision-tree models for parsing. *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, pp. 276-283, 1995. DOI: 10.3115/981658.981695
- [29] Dougherty, G., *Pattern recognition and classification: An introduction*. Springer Science & Business Media, 2012.
- [30] Esposito, F., Malerba, D., Semeraro, G. and Kay, J., A comparative analysis of methods for pruning decision trees, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), pp. 476-491, 1997. DOI: 10.1109/34.589207
- [31] Dai, Q., A competitive ensemble pruning approach based on cross-validation. *Knowledge-Based Systems*, 37, pp. 394-414, 2013. DOI: 10.1016/j.knsys.2012.08.024

R. Lara-Cueva, es Ing. en Electrónica y Telecomunicaciones en la Escuela Politécnica Nacional, Ecuador, en el año 2001. El grado de MSc. en Sistemas Inalámbricos y Tecnologías Relacionadas del Politécnico di Torino, Italia en el 2005. MSc y Dr en Redes de Telecomunicación para Países en Desarrollo de Universidad Rey Juan Carlos, España en 2010 y 2015, respectivamente. Se unió al Departamento de Ingeniería Eléctrica de la Universidad de las Fuerzas Armadas ESPE en el 2002 y ha sido profesor titular principal de dicha universidad desde el 2005. Ha participado en más de 10 proyectos de investigación con fondos públicos, y dirigido 5 de ellos. Sus principales intereses de investigación incluyen el procesamiento digital de señales, ciudades inteligentes, sistemas inalámbricos y la teoría de aprendizaje automático.

ORCID: 0000-0001-8848-9928

V. Paillacho-Salazar, en mayo del 2016 se graduó de Ing. en Electrónica y Telecomunicaciones en la Universidad de las Fuerzas Armadas ESPE, Ecuador. En el año 2014 se unió al Grupo de Investigación en Sistemas Inteligentes (WICOM-Energy) como ayudante de investigación. Sus áreas de interés incluyen redes de nueva generación, procesamiento digital de señales, comunicaciones inalámbricas e inteligencia artificial.

ORCID: 0000-0002-1776-458X

M. Villalva-Chaluisa, es Ing. Electrónica en Telecomunicaciones en la Universidad de las Fuerzas Armadas ESPE, Ecuador en el año 2016. Actualmente, está vinculada en áreas de investigación relacionadas al procesamiento digital de señales volcánicas. Sus áreas de interés incluyen inteligencia artificial y comunicaciones inalámbricas.

ORCID: 0000-0003-4300-9323



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD DE MINAS

Área Curricular de Ingeniería
Eléctrica e Ingeniería de Control

Oferta de Posgrados

Maestría en Ingeniería - Ingeniería Eléctrica

Mayor información:

E-mail: ingelcontro_med@unal.edu.co
Teléfono: (57-4) 425 52 64