

CARACTERIZACIÓN DE SEÑALES VOLCÁNICAS PARA LA CLASIFICACIÓN
SUPERVISADA DE EVENTOS SÍSMICOS



ANGELA PATRICIA CABRERA RIOBAMBA
EDISON GERARDO ROSERO NARVAEZ

UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA ELECTRÓNICA
SAN JUAN DE PASTO
2019

**CARACTERIZACIÓN DE SEÑALES VOLCÁNICAS PARA LA CLASIFICACIÓN
SUPERVISADA DE EVENTOS SÍSMICOS**

ANGELA PATRICIA CABRERA RIOBAMBA

EDISON GERARDO ROSERO NARVAEZ

**TRABAJO DE GRADO PARA OPTAR POR EL TÍTULO DE INGENIERO
ELECTRÓNICO**

ASESOR

PhD. DIEGO HERNÁN PELUFFO ORDÓÑEZ

INGENIERO ELECTRÓNICO

ASESORES EXTERNOS

PAOLA ALEXANDRA CASTRO CABRERA

ESTUDIANTE DOCTORADO DE LA UNIVERSIDAD NACIONAL DE COLOMBIA

JOHN MAKARIO LONDOÑO BONILLA

**FUNCIONARIO DEL OBSERVATORIO VULCANOLÓGICO Y SISMOLÓGICO
DE MANIZALES**

**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA ELECTRÓNICA
SAN JUAN DE PASTO
2019**

NOTA DE RESPONSABILIDAD

“La Universidad de Nariño no se hace responsable por las opiniones o resultados obtenidos en el presente trabajo y para su publicación priman las normas sobre el derecho de autor.”

Acuerdo 1. Artículo 324. Octubre 11 de 1966, emanado del honorable Consejo Directivo de la Universidad de Nariño.

NOTA DE ACEPTACIÓN:

Firma del presidente del jurado

Firma del jurado

Firma del jurado

San Juan de Pasto, 14/05/19

AGRADECIMIENTOS

“Agradezco primero a Dios por todas las bendiciones que me ha dado, agradezco a mi madre, a mi compañero de vida. A mi madre por apoyarme en cada uno de los retos que me propongo y ser el ejemplo de lucha incansable y perseverancia. A mi compañero por brindarme esa fortaleza y apoyo tan importante y necesario en este proceso”.

Angela Patricia Cabrera Riobamba.

“Agradezco a mi madre porque gracias al apoyo incondicional alcanzar esta meta hoy es posible”.

Edison Gerardo Rosero Narváez.

“Agradecemos al observatorio vulcanológico y sismológico de Manizales por brindarnos los datos a trabajar y la asesoría brindada por Paola Castro”.

Edison Rosero y Angela Cabrera.

DEDICATORIA

“A mi madre, ella es la persona con más tenacidad que conozco. Cada acto de su vida está enfocado en llevar adelante a sus hijos. Tan fuerte e insistente, una gran inspiración para mi vida. El reflejo de un amor infinito e incondicional”.

Angela Patricia Cabrera Riobamba.

“A mi madre y a mi hermano, son la muestra perfecta de sabiduría y paciencia, la energía perfecta que recarga mis ánimos para alcanzar cada una de mis metas”.

Edison Gerardo Rosero Narváez.

RESUMEN

Tradicionalmente, la vigilancia de un volcán está a cargo de los observatorios vulcanológicos donde el estudio de la sismicidad volcánica es una herramienta importante, junto con otras ciencias pueden describir la mecánica interna de la edificación volcánica. Este análisis requiere detección y registro de las señales en las estaciones vulcanológicas, transmisión a los observatorios, en donde se realiza un preprocesamiento a la señal en busca de características distintivas para su clasificación, una interpretación de la tasa de sismos, la ubicación y otros aspectos. Generalmente, esta tarea en su mayoría es ejecutada por operadores humanos expertos quienes, en temporadas de alta actividad volcánica ralentizan el tratamiento de las señales y en ocasiones resulta no ser efectivo.

En la actualidad, se han desarrollado investigaciones que proponen la clasificación automática de los eventos sismo-volcánicos, basadas en técnicas de aprendizaje de maquina (machine learning), con el fin de determinar el tipo evento registrado. No obstante, aún no existen métodos concretos de caracterización y clasificación. Diversos enfoques muestran grandes avances, pero debido a la variabilidad que presentan estas señales no hay un método definido de caracterización, lo que representa un problema abierto y acertado para la investigación.

Por lo anterior, en este trabajo se desarrolló una metodología de comparación de técnicas de extracción de características. Se utilizó representación de las señales sísmicas en el dominio del tiempo, frecuencia, tiempo-frecuencia y Cepstral. Se optimizó el conjunto de atributos realizando selección de características por asignación de ponderados. Se ejecutó una clasificación supervisada haciendo uso de registros conocidos, obtenidos de una base de datos de eventos sismo-volcánicos, facilitada por el Observatorio Vulcanológico y Sismológico de Manizales (OVSM). Finalmente, se obtuvo medidas de desempeño de la clasificación para determinar el subconjunto de características que mejor representan y discriminan las señales.

ABSTRACT

Traditionally, the monitoring of a volcano is in charge by study of volcanic seismicity, based on recognizing the types of seismic events, which, together with other sciences, can describe the internal mechanics of volcanic edifice. This analysis requires of the detection and recording of the signals in the vulcanological stations, transmission to the observatories, where the signal is preprocessed, in search of distinctive characteristics for its classification, and an interpretation of the earthquake rate, the location and other aspects. Generally, this task is mostly carried out by expert human operators, who, in seasons of high volcanic activity, often slow down the treatment of signals and sometimes it turns out not to be very effective.

At present, researches have been developed that proposes the automatic classification of volcanic earthquake events, based on machine learning techniques, in order to determine the type of event recorded. However, there are still no specific methods of characterization and classification, several approaches show great advances, but due to the variability that these signals present, there is no defined method of characterization, which represents an open and successful problem for research.

Therefore, in this work we developed a comparison methodology technique for extracting characteristics, where we used representation of seismic signals, in the domain of time, frequency, time-frequency and cepstral. The set of attributes was optimized, with methods of selection of characteristics by weighted assignment, a supervised classification was made, making use of known records, obtained from a database of volcanic earthquake events, facilitated by the Manizales seismological observatory (OVSM). Finally, performance measurements of the classification were obtained, to determine the subset of characteristics that best represent and discriminate the signals.

TABLA DE CONTENIDO

INTRODUCCIÓN	15
Descripción del problema	17
Justificación	18
Contribuciones de este trabajo	18
Objetivo general	19
Objetivos específicos	19
Marco Teórico	19
Monitoreo Volcánico	19
Sismología Volcánica	20
Observatorio vulcanológico	21
Redes sísmicas volcánicas	22
Sistemas de registro de señales sísmicas	23
Tipos de sismos volcánicos	23
Señales sísmicas propias de actividad volcánica	24
Eventos externos del volcán	26
Representación de las señales sísmo-volcánicas	27
Frecuencia – Transformada de Fourier	27
Tiempo frecuencia – Espectrograma	27
Cepstral – Coeficientes Cepstrales en las frecuencias de Mel	29
Sistemas computacionales para reconocimiento de señales sísmo- volcánicas	30
Reconocimiento de patrones	31
Caracterización	31
Selección de características	32
Clasificación	33
1. DESARROLLO DE LA INVESTIGACIÓN	34
1.1. REGISTRO SEÑALES SÍSMO-VOLCANICA	34
1.2. PRE-PROCESAMIENTO	35

1.3. REPRESENTACIÓN SEÑALES VOLCÁNICAS	36
1.3.1. Análisis frecuencia.....	36
1.3.2. Análisis Tiempo-Frecuencia.....	36
1.3.3. Análisis Cepstral	38
1.4. CARACTERIZACIÓN SEÑALES SÍSMO-VOLCÁNICAS.....	38
1.5. SELECCIÓN DE CARACTERÍSTICAS.....	43
1.6. CLASIFICACIÓN.....	44
1.6.1. Clasificadores polinomiales de grado lineal y superior	44
➤ Clasificación basada en la densidad normal.....	44
➤ Clasificadores no lineales	45
2. MARCO EXPERIMENTAL.....	47
2.1. MEDIDAS DE CLASIFICACIÓN	47
2.1.1. Curvas ROC	47
2.2. PRUEBAS APLICADAS SOBRE LA BASE DE DATOS.....	48
3. RESULTADOS.....	50
3.1. INTERFAZ DE VISUALIZACIÓN DE RESULTADOS.....	61
4. CONCLUSIONES Y TRABAJO FUTURO	63
REFERENCIAS.....	64
ANEXOS.....	70

LISTA DE FIGURAS

Figura 1. Estructura de un sismómetro. Fuente: [19].....	22
Figura 2. Evento sísmico vulcano-tectónico VT. Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz	24
Figura 3. Evento sísmico largo periodo LP. Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.....	25
Figura 4. Evento sísmico Tremor TR. Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.....	26
Figura 5. Sismograma y espectrograma de un evento VT. Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.....	29
Figura 6. Diagrama de bloques extracción parámetros MFCC. Fuente: [32].....	30
Figura 7. Metodología para el reconocimiento de patrones en señales sísmico-volcánicas. Fuente: Esta investigación.	34
Figura 8. Espectro de evento sísmico vulcano-tectónico Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.....	36
Figura 9. Espectrograma evento sísmico vulcano-tectónico Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.	37
Figura 10. a) Diagrama cajas y bigotes clasificación de eventos VT y LP utilizando el conjunto de características completo. b) curva ROC para el clasificador LDC.	51
Figura 11. a) Diagrama cajas y bigotes clasificación de eventos VT - IC utilizando el conjunto de características de tiempo. b) curva ROC para el clasificador PARZEN.....	53
Figura 12. a) Diagrama cajas y bigotes clasificación de eventos VT - TR utilizando el conjunto de características de tiempo. b) curva ROC para el clasificador SVM.	54
Figura 13. a) Diagrama cajas y bigotes clasificación de eventos LP-IC utilizando la matriz de características completa. b) curva ROC para el clasificador SVM.	56
Figura 14. a) Diagrama cajas y bigotes clasificación de eventos LP-TR utilizando el grupo de características de tiempo. b) curva ROC para el clasificador Parzen.	57
Figura 15. Diagrama cajas y bigotes clasificación de eventos IC-TR utilizando las características de tiempo, b) Curva ROC para el clasificador PARZEN.....	59
Figura 16 Grafica promedio medidas de desempeño para cada grupo de características, clasificación multiclase con KNN.	60
Figura 17. Interfaz gráfica de clasificación de señales sismo volcanicas.....	62

LISTA DE TABLAS

Tabla 1. Volcanes con enjambres de eventos sísmicos. Fuente [9].....	21
Tabla 2. Características dominio temporal.....	38
Tabla 3. Características dominio frecuencia.....	40
Tabla 4. Características dominio Tiempo- Frecuencia.....	41
Tabla 5. Medidas de desempeño <i>Se</i> , <i>Sp</i> y <i>Exactitud</i> de clasificación de eventos VT-LP. 50	50
Tabla 6. Medidas de desempeño <i>Se</i> , <i>Sp</i> y <i>Exactitud</i> de clasificación de eventos VT-IC.. 52	52
Tabla 7. Medidas de desempeño <i>Se</i> , <i>Sp</i> y <i>Exactitud</i> de clasificación de eventos VT-TR. 53	53
Tabla 8. Medidas de desempeño <i>Se</i> , <i>Sp</i> y <i>Exactitud</i> de clasificación de eventos LP-IC.. 55	55
Tabla 9. Medidas de desempeño <i>Se</i> , <i>Sp</i> y <i>Exactitud</i> de clasificación de eventos LP-TR. 56	56
Tabla 10. Medidas de desempeño <i>Se</i> , <i>Sp</i> y <i>Exactitud</i> de clasificación de eventos IC-TR.58	58
Tabla 11. Matriz de confusión y medidas de desempeño para clasificación multiclase con K- NN con diferentes grupos de características.....	60
Tabla 12. Matriz de confusión y medidas de desempeño para clasificación multiclase con K- NN con matriz de características concatenando los grupos anteriores.....	61

LISTA DE ANEXOS

ANEXO 1. Página web.....	70
---------------------------------	-----------

GLOSARIO

Frecuencia de muestreo: Se refiere a la cantidad de muestras por unidad de tiempo tomadas de una señal continua para producir una señal discreta.

Clase: Agrupación de objetos que tiene características comunes.

Características: Se refiere a los atributos que describen las instancias del conjunto de datos.

Algoritmo: Conjunto definido de reglas o procesos que llevan a la solución de un problema en un número determinado de pasos.

Preprocesamiento: Es un proceso donde se filtra información de las bases de datos usando diferentes mecanismos como el filtraje, normalización, entre otros.

Selección de características: Hace referencia al proceso de elegir información relevante de un conjunto de datos.

INTRODUCCIÓN

Los volcanes se forman de lava y gases que emergen desde el interior de la tierra. Un volcán activo representa un riesgo potencial para las poblaciones cercanas, por esta razón, es necesario e importante una vigilancia continua y casi en tiempo real con el propósito de declarar alertas tempranas de riesgo para la humanidad. Generalmente, el estudio de la sismicidad volcánica es la principal herramienta de monitoreo, dado que el análisis de las diferentes señales sismo-volcánicas aporta información significativa y de gran volumen, tanto de los acontecimientos dentro como fuera de la edificación volcánica, teniendo información del movimiento del magma, fluidos termales, fractura de las paredes internas, caída de rocas, lahares y fractura de hielos en aquellos volcanes con cubierta glacial, entre otros.

El monitoreo volcánico, a menudo, se lleva a término con la implementación de estaciones sísmicas y observatorios, siendo las primeras las encargadas en detectar, registrar y posteriormente transmitir los sismos. En la actualidad, la tecnología de herramientas de instrumentación y telemetría permiten obtener gran cantidad de información sísmica del volcán, donde se realiza un preprocesamiento, seguido de una caracterización y clasificación de cada evento sismo-volcánico. Una estación sísmica registra cientos de enjambres sísmicos al día, de un volcán que este presentando una alteración en su actividad.

La cantidad de información de la actividad sísmica de un volcán hace que su procesamiento en los observatorios no vaya tan rápido, debido a que frecuentemente, esto es realizado por operadores humanos [1]. Basados en su experiencia buscan aspectos como: cambio en la velocidad, tamaño, forma de onda, duración, ubicación, entre otros, con el propósito de describir el evento volcánico para tomar decisiones y pronósticos de una posible alerta de riesgo. Siendo así de gran importancia el aprovechamiento de la información obtenida del volcán, se ve necesario implementar métodos automáticos con objeto de procesar y clasificar las señales sísmicas.

En la literatura se evidencia varias investigaciones, que contribuyen en la automatización de la identificación y clasificación de sismos volcánicos con rendimiento variable, que pueden facilitar y agilizar el procesamiento de los datos. Autores muestran diversos enfoques en aprendizaje automático usando técnicas como redes neuronales artificiales (ANN, Artificial Neural Networks). La referencia [4] usa una ANN simple, consiste en una entrada, una capa oculta y una salida que devuelve un vector de pertenencia a una clase. Los parámetros que describen las características de cada evento fueron interpretados desde un operador experto, donde se encuentran características como la duración de cada evento, las frecuencias dominantes, la impulsividad de la señal entre otras, obteniendo un rendimiento de clasificación entre el 80% y 95%. Así mismo las referencias [5] y [6] utilizan una red neuronal perceptrón multicapa alcan

zando sensibilidades de 93% y 89%, respectivamente.

Por otra parte, la referencia [1] valida el rendimiento del algoritmo de aprendizaje automático árboles de decisión (RF, Random Forest), para clasificar dos tipos de eventos, volcano-tectónicos (VT) y caída de rocas (Rock-fall) en el volcán Pitón de la Fournaise en Francia. El clasificador se entrena a partir de 60 atributos en los que se encuentran forma de onda, espectrales, pseudo espectrales y polaridad, donde el rendimiento se mantuvo por encima del 90%. Técnicas como los modelos ocultos de Markov (HMM, Hidden Markov Model) se han utilizado en varios estudios como el realizado en el volcán de la isla Decepción en la Antártida [7], alcanzando rendimientos desde 90% hasta el 95%. La referencia [2] y [5] trabajan técnicas de machine learning bajo el algoritmo máquinas de vectores de soporte (SVM, Support Vector Machines) utilizando métodos supervisados alcanzando 92% y sobre el 80% respectivamente de clasificación.

Aunque los estudios mencionados muestran una aceptación de los algoritmos de clasificación con respecto a las señales sismo-volcánicas, es difícil considerar cual sería el mejor, ya que en cada estudio siguen diferentes metodologías donde tienen en cuenta aspectos como el número de tipo de eventos a clasificar, el entrenamiento de los clasificadores y sin duda la más destacada, la forma de relacionar el vector de características.

En la actualidad, no hay definida una metodología que proporcione una buena caracterización y clasificación de eventos sismo-volcánicos. Algunos autores demuestran que las características espectrales representan mejor los sismos volcánicos como en la referencia [6]. También existen estudios en que la forma de onda son atributos más discriminantes como en [5]. Asimismo, se encuentran investigaciones que muestran una mejor representación usando un conjunto de atributos de diferente tipo como en las referencias [1] y [2]. Lo que demuestra la variabilidad de las señales sísmicas de volcán a volcán, dejando el tema abierto para nuevas investigaciones que puedan acercarse a la búsqueda de las características más representativas en las señales trabajadas con el propósito de obtener una buena clasificación.

En este trabajo se desarrolló un sistema de reconocimiento de patrones mediante técnicas de clasificación supervisada, utilizando una base de datos que contiene señales sismo-volcánicas del volcán Nevado del Ruiz, segmentadas y clasificadas por expertos del observatorio vulcanológico de Manizales (OVSM), con la que se realizó una clasificación biclase, con el objeto de determinar las características más representativas de cada tipo de evento. Se implementó técnicas de extracción de características utilizando representaciones en el dominio del tiempo, frecuencia,

tiempo-frecuencia y Cepstral. La representación adecuada está dada por medidas de desempeño de clasificación.

Modalidad: Investigación.

Linea: Procesamiento de datos.

Descripción del problema

El estudio de la sismicidad volcánica es clave y sigue siendo una de las herramientas más confiables para el monitoreo del volcán, suministrando grandes posibilidades de declaración de alertas tempranas en el ámbito de prevención de riesgos y desastres. El movimiento del magma hacia la superficie provoca estrés y movimientos en la edificación volcánica, generando señales sísmicas de diferente naturalidad, provocando regulares patrones sísmicos conocidos como eventos sismo-volcánicos. Los observatorios reciben datos de la actividad sísmica en los volcanes a partir de las estaciones sísmicas mediante técnicas de telemetría, el procesado y análisis de los datos en la mayoría de casos es realizado por operadores humanos, quienes entre otras actividades se encargan de caracterizar las señales y asignar una clase al evento vulcanológico de acuerdo a su experiencia y conocimiento.

La tecnología de la actualidad provee sistemas de detección y transmisión que enriquecen la información que recibe el observatorio. Cuando el volcán presenta mayor actividad los eventos sismo-volcánicos aumentan considerablemente, llegando a registrar cientos de sismos al día [1]. Los Datos obtenidos necesitan ser analizados casi en tiempo real, para interpretar los movimientos y comportamiento de la edificación volcánica, lo que representa un trabajo indispensable para los científicos, sobrecargándose de trabajo y sin el aprovechamiento total de la información. Siendo así de gran importancia implementar métodos automáticos con objeto de procesar y clasificar las señales sísmicas.

De acuerdo a lo anterior, se planteó un sistema automático de clasificación de eventos vulcanológicos que ayude a reducir y soportar el trabajo de los científicos en los observatorios. Se han desarrollado varias investigaciones que se basan en técnicas de aprendizaje automático con métodos de reconocimiento de patrones supervisados, con el propósito de discriminar las clases de eventos vulcanológicos. Sin embargo, no existe un acuerdo científico que sistematice la forma de representar las señales sísmicas obtenidas de la actividad volcánica, presenta un problema abierto y oportuno de investigar para aportar un conjunto de atributos basados en las investigaciones de la literatura volcánica, que logre la mejor discriminación entre las clases de eventos volcánicos.

Justificación

La sismología volcánica es una de las técnicas más antiguas, que permite conocer el estado de un volcán. La interpretación de las señales sísmicas adquiridas de la actividad volcánica, es una tarea fundamental para el aprovechamiento total de la información que brindan. Se puede obtener información de posibles intrusiones magmáticas, alteraciones de una cámara magmática preexistente, ascenso de magma ante una inminente erupción, etcétera [8].

Los observatorios vulcanológicos cumplen una difícil tarea, al interpretar toda la información sísmica del volcán recibida de las estaciones que, en conjunto con diferentes técnicas de monitoreo volcánico como la geodesia, geoquímica, entre otras, ayudan a pronosticar posibles erupciones y tomar medidas de precaución ante una alerta de riesgo. Por tanto, fue necesario automatizar los procesos de discriminación de señales sísmicos volcánicos, que permitieron asignar una clase a las señales obtenidas con eficiencia.

Las investigaciones realizadas por científicos, muestran grandes avances en este aspecto, sin embargo, no existe un determinado procedimiento a seguir para encontrar un sistema totalmente automático que clasifique las señales sísmicas de los volcanes. En virtud de ello este trabajo se desarrolla con el fin de encontrar las mejores técnicas de caracterización, basada en investigaciones de varios autores, para lograr una clasificación con alto desempeño, que proporcione asistencia a los observatorios que monitorean las edificaciones volcánicas.

Contribuciones de este trabajo

El análisis de las señales sísmicas de un volcán activo fue una herramienta importante en el monitoreo volcánico. Aportó al entendimiento de estas señales, partiendo de su caracterización fue una contribución para los observatorios vulcanológicos, dando pie al desarrollo de sofisticadas herramientas que permitieron la completa automatización de una clasificación con alta eficiencia, con el propósito de declarar alertas tempranas y posibles riesgos hacia la humanidad con los volcanes.

Partiendo de la importancia de identificar los tipos de eventos que se está presentando en la edificación volcánica, este trabajo representó un aporte a la comunidad científica y académica, en el área de procesamiento de señales sismo-volcánicas, específicamente, en el análisis de caracterización, buscando un modelo que logró discriminar cada clase de sismo.

Objetivo general

Desarrollar una metodología de caracterización y clasificación de señales sismo-volcánicas que permita identificar parámetros relevantes del tipo de evento volcánico para el reconocimiento de patrones.

Objetivos específicos

- Implementar técnicas de extracción y selección de características de señales sismo-volcánicas a fin de obtener las más discriminantes que permitan la identificación de clases de sismos-volcánicos.
- Analizar las características obtenidas a partir del entrenamiento de clasificadores y determinar el más adecuado para el tipo de señal trabajada.
- Evaluar los resultados obtenidos por el clasificador determinado, realizando análisis de los errores por la clasificación automática.

Marco Teórico

Una onda mecánica que se transmite por el interior de la tierra se conoce como onda sísmica [8], y a través de su estudio y análisis, se conoce el epicentro de acuerdo a su hipocentro o foco, y la causa de su origen.

El monitoreo de las señales sísmicas propias de la actividad volcánica ayuda a comprender los movimientos internos del volcán. Existen varias técnicas de monitoreo de volcanes que, a diferencia de la sismicidad requieren trabajo de campo en toma de muestras, no brindan información en tiempo real y a menudo, necesitan de la luz del día para realizar el monitoreo, convirtiendo el estudio de la sismicidad volcánica en la principal herramienta de vigilancia de los volcanes [3] [9] [10].

Monitoreo Volcánico

Una erupción volcánica está catalogada como uno de los desastres naturales más peligrosos que puede atentar contra la población humana. No solo la lava se produce en una erupción, también se encuentran emisiones como nubes de ceniza, gases tóxicos y los que no son propios de un volcán, pero se producen a raíz de una erupción como: lluvia ácida, lahares, caída de rocas, inundaciones, entre otras [11]. Por esto el monitoreo de los volcanes se ha convertido en una tarea necesaria, dado que se puede declarar alertas tempranas de riesgo, para lograr minimizar el daño hacia la humanidad.

En la vigilancia de un volcán se involucran varias ciencias que se complementan con el propósito de determinar los eventos próximos a ocurrir en la edificación volcánica; entre ellas se encuentra la geodesia, la geoquímica, la gravimetría, el geomagnetismo, la geología y la sismología [8], comprendiendo estudios de la corteza del volcán, estudio de la emisión de gases, cambios gravitacionales, campos electromagnéticos y los sismos volcánicos.


El estudio de la sismología se considera como la mejor herramienta en monitoreo volcánico. Las referencias [12] [13] [14] demuestran que, el análisis de los eventos sísmicos puede describir lo que está ocurriendo en el interior del volcán, brindando información acerca de su origen.

Sismología Volcánica

El estudio de los registros sísmicos de un volcán busca describir su naturaleza y origen, al asociarlos con fuentes como el desplazamiento del magma y fluidos hidrotermales en el interior de la edificación volcánica. La geofísica es la mejor herramienta para comprender la mecánica interna del volcán, los episodios eruptivos y declaración de alerta temprana de riesgo [15].

La sismicidad volcánica que precede y acompaña a las erupciones puede ser altamente específica e individual para un volcán. Por lo tanto, es incierto qué parámetros sísmicos de un volcán podrían servir de modelos estándar para el diagnóstico y pronóstico de erupción en otro volcán. Sin embargo, aunque cada volcán tenga sus propias características sísmicas se han logrado conocer parámetros basados en la experiencia y el estudio de la actividad eruptiva [16]. Las referencias [12] y [15] demuestran la relación de los sismos con procesos eruptivos del volcán, mencionan que las señales de baja frecuencia aumentan antes y durante de una erupción real. La **Tabla 1** muestra algunos ejemplos de volcanes que registraron enjambres de eventos sísmicos asociados o no con actividad eruptiva [9], de aquí la importancia de conocer el tipo de señales sísmicas que el volcán está registrando.

Tabla 1. Volcanes con enjambres de eventos sísmicos.

Volcán	Año	Magnitud _{max}	Comentarios
Arenal	1968	4.5	Pre- erupción
Asama	1916	6.3 	No erupción
Aso	1975	6.0	No erupción

Bandai	1988	5	Erupción provocada
Bezimianny	1956	5.0	Pre- erupción
Fernandina	1968	5.7 M_S	Colapso Caldera
Gorely	1985	≤ 6	Post erupción (8 meses)
Ito-oki	1989	5.5	Pre- erupción
Jan Mayen	1954	5.8	No erupción
Katmai	1912	7.0	Post erupción
Long Valley	1980	6.2 M_L	4 terremotos con $M \geq 6$, no erupción
Miyake-jima	1962	5.9	Fuerte enjambre
Miyake-jima	1982	5.9 M_{JMA}	Enjambre 20 km de profundidad
Miyake-jima	1982	5.9	Durante erupción
Mt. St. Helens	1980	5.2	Erupción provocada

Nota: M_S Magnitud de onda superficial, M_L Magnitud Local (Ritcher), M_{JMA} Magnitud Agencia Meteorológica de Japón. **Fuente** [9].

Observatorio vulcanológico

Un observatorio vulcanológico es un instituto de carácter gubernamental, cuya función es monitorear volcanes activos y brindar alertas tempranas de riesgo a las autoridades y a la comunidad [17]. El primer observatorio vulcanológico fue fundado en 1845 en Italia (Observatorio Vesubiano) [10]. Ahora la organización mundial de observatorios vulcanológicos (WOVO, World Organization of Volcano Observatories), tiene aproximadamente 90 miembros. Hoy en día se puede encontrar observatorios desde un simple sismógrafo imprimiendo en papel hasta sofisticados observatorios que monitorean varios volcanes a la vez [3][9]. Por ejemplo, el observatorio de Alaska y la agencia meteorológica de Japón tienen a cargo la vigilancia de 32 y 47 volcanes, respectivamente.

Un observatorio está conformado por redes sísmicas y un sistema de software utilizados para monitorear la información. El progreso de la vigilancia volcánica ha involucrado el desarrollo de instrumentos electrónicos cada vez más sofisticados y complejos, como también el aprovechamiento de modelos computacionales para procesar los datos [18]. En los observatorios el análisis de la sismología volcánica es una de las principales herramientas de vigilancia, puesto que el estudio de las señales sísmicas ayuda a entender la estructura y mecánica interna de los volcanes [3]. A partir de estos análisis, los sismólogos comprenden las actividades volcánicas

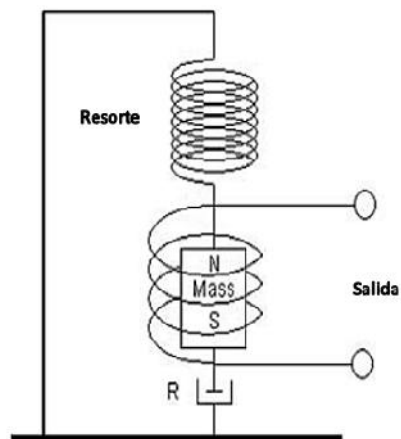
como el movimiento del magma, la fractura de las paredes volcánicas por el estrés de los fluidos termales y gases ascendiendo, entre otras, apoyándose en otras ciencias.

Redes sísmicas volcánicas

La red sísmica encargada de registrar las señales sismo-volcánicas está conformada, generalmente, por un conjunto de estaciones sísmicas distribuidas a lo largo de un territorio en la zona volcánica. Frecuentemente, cuentan con sistemas de instrumentación, adquisición y transmisión de datos. A menudo, las estaciones están ubicadas a distancias de 15 km aproximadamente [9]. Se considera que un volcán activo por lo menos debe tener una estación sísmica para su continuo monitoreo. Las estaciones deben estar distribuidas de tal manera que permitan asegurar la localización y profundidad del sismo. De acuerdo con su ubicación, el sismo puede ser registrado por varias estaciones y, así mismo, estas pueden capturar sus componentes longitudinales y transversales. Las redes modernas se constituyen de 6 o más estaciones permitiendo una mejor detección de sismos y minimizar el error en la búsqueda de su origen [9] [10].

Los instrumentos de las estaciones deben ser de alta sensibilidad debido a que la actividad sísmica genera sismos que van desde 1 – 25 Hz [15]. Los sismómetros utilizados pueden ser de periodo corto, que comprenden una frecuencia de muestreo de hasta 50 Hz y registran datos entre 1 Hz hasta 25 Hz. La figura 1 muestra como está formado un sismómetro, consiste en una masa magnética sostenida por un resorte donde su movimiento es detectado por una bobina, dando de salida un diferencial de voltaje. También se encuentra sismómetros de periodo largo o banda ancha, los cuales alcanzan frecuencias de muestreo hasta de 200 Hz y reconocen datos en una banda de frecuencia mucho más amplia de 0.02 a 100Hz. Su composición es la misma de un sismómetro periodo corto, pero la bobina que detecta el movimiento está conectada a un circuito de realimentación que envía una corriente a través de otra bobina, exactamente para oponerse al movimiento, por lo tanto, la masa se mantiene fija y la corriente de realimentación es proporcional al movimiento en el suelo [9].

Figura 1. Estructura de un sismómetro.



Fuente: [19].

Sistemas de registro de señales sísmicas

Inicialmente, las señales adquiridas de la actividad volcánica son registradas en trenes de impulsos, donde cada impulso representa un evento volcánico. El proceso de segmentación de los registros obtenidos representa una difícil tarea, debido a la dificultad para encontrar los puntos de inicio y final, por la variabilidad que este tipo de señal presenta. Algunos eventos podrían durar desde 10 segundos y otros hasta días [2]. El algoritmo promedio a corto plazo y promedio a largo plazo (STA/LTA, Short Time Average over Long Time Average), presentado por Rex Allen en la referencia [20], sigue siendo el más utilizado en la actualidad. Sin embargo, hay más investigaciones que presentan diferentes métodos basándose en la curtosis y en filtrado óptimo de la señal [21] [22]. Aunque los métodos mencionados anteriormente representan un avance tecnológico en cuanto a la detección de sismos volcánicos, se puede decir que aún no existe un método definitivo que logre detectar los sismos cuando el volcán presenta mayor actividad, pues los algoritmos como el STA/LTA requieren de ajustes manuales como, umbrales y longitudes de ventana.

Tipos de sismos volcánicos

La actividad volcánica presenta gran variedad en sus señales sísmicas, tradicionalmente clasificadas por su apariencia en el sismograma en aspectos como cambio de frecuencia, duración, periodo, entre otros. Inicialmente Minakami definió 4 clases de señales sísmicas, tipo A o frecuencia alta, tipo B o frecuencia baja,

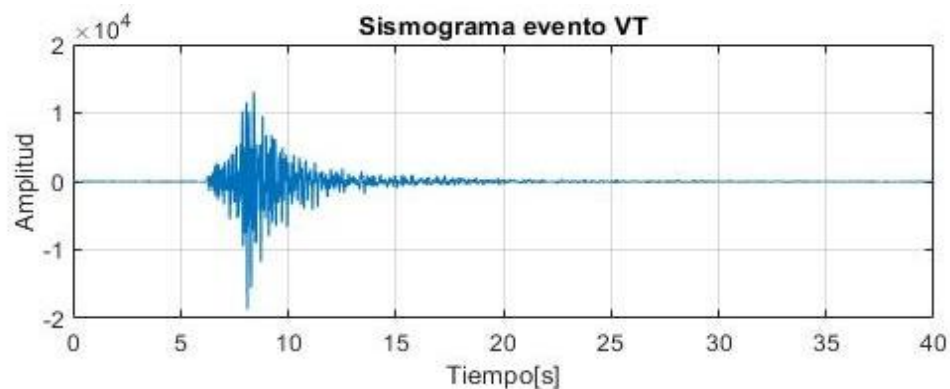
explosiones y tremor volcánico [12]. Sin embargo, estudios más recientes exponen más clases de eventos, por ejemplo, los eventos que son ajenos al volcán como caídas de rocas, lahares, fractura de hielos en aquellos volcanes que poseen capa glacial, entre otros. Para una buena monitorización los sistemas de clasificación deben tener presente tanto la actividad interna como la externa a la edificación volcánica.

Señales sísmicas propias de actividad volcánica

Sismos volcano - tectónicos

Los eventos volcano tectónicos (VT) se generan a partir de la fractura de las paredes volcánicas, por presión del ascenso del magma o fluidos hidrotermales. Los eventos VT son analizados con el propósito de determinar las orientaciones de estrés de las paredes volcánicas, mediante el estudio de los mecanismos focales. Se caracterizan por tener una duración variable, desde los pocos segundos hasta minutos, como se muestra en la figura 2 el inicio del evento es impulsivo y tienen bien definida la onda P y S, su frecuencia dominante se encuentra de 5 – 15 Hz [9]. Estudios demuestran que la sismicidad causada por grandes terremotos distantes al volcán activa los eventos VT [23] [24].

Figura 2. Evento sísmico vulcano-tectónico VT.



Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.

Sismos de largo periodo

En la figura 3 se muestran los eventos de largo periodo (LP) también llamados baja frecuencia (LF, Low Frequency). Se dice que los eventos LP son causados por procesos fluidos que aún no se comprenden bien [25]. Sin embargo, existen estudios que atribuyen a los enjambres de eventos LP a una posible erupción [26]. Existen algunos desacuerdos como el estudio en la referencia [27], quien plantea una posible confusión entre los eventos LP generados a partir de actividad volcánica interna y la actividad externa de los volcanes que tiene capa glacial, ya que la fractura de hielo y el deslizamiento del mismo produce señales sísmicas de las mismas características.

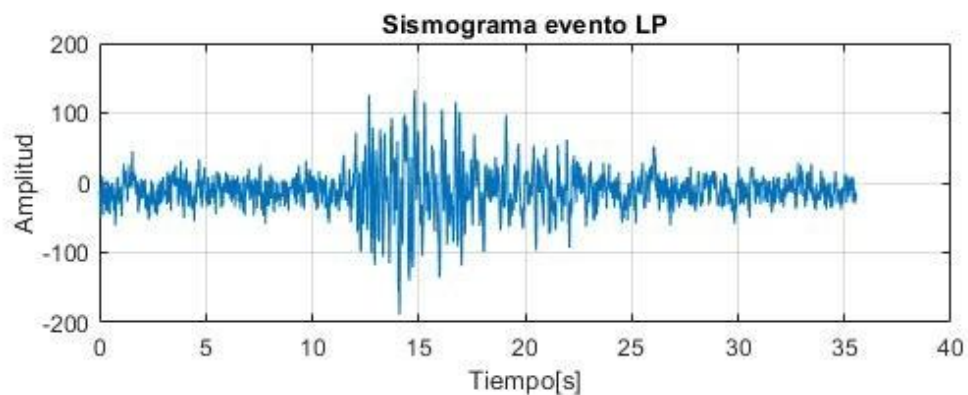


Figura 3. Evento sísmico largo periodo LP. **Fuente:** Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.

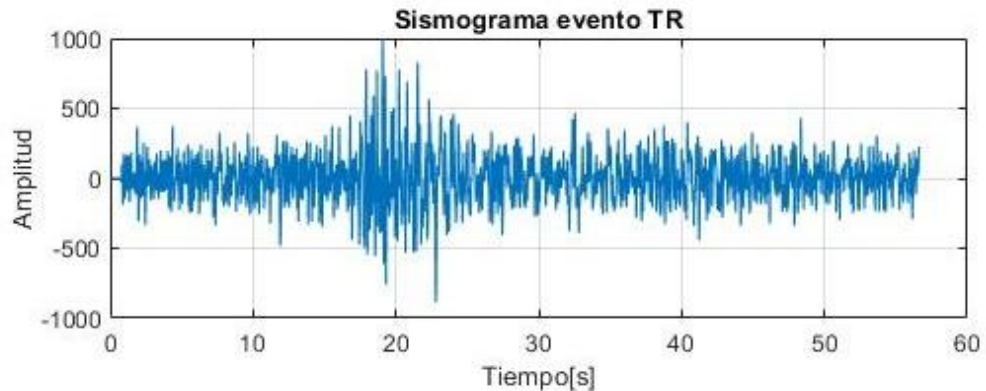
Eventos sísmicos híbridos

Existen algunos eventos que comparten características con eventos de alta frecuencia y de baja frecuencia, estos eventos se los conoce como híbridos (HYB, Hybrid). Se caracterizan por presentar una entrada impulsiva en alta frecuencia con alto contenido espectral, característica de los eventos VT, pero viene acompañada de una señal similar en forma de onda y contenido espectral a un evento [9].

Eventos sísmicos tremor

Este tipo de evento sísmico generalmente se denota por su estrecho rango de frecuencia o sus espectros de pico agudo y su larga duración en comparación con los otros terremotos. La magnitud de este tipo de evento sísmico en presenta continuidad, que en ocasiones cambia de acuerdo a los cambios de energía del fenómeno que lo produce, como el evento TR de la figura 4. Los eventos Tremor se caracterizan por su duración que puede variar desde los minutos llegando a registrarse hasta por meses [10].

Figura 4. Evento sísmico Tremor TR.



Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.

Eventos sísmicos explosión volcánica

Los sismos explosivos (EXP) ocurren junto con las erupciones explosivas y su característica más distinguible es la presencia de ondas de aire en los sismógrafos. La energía expulsada en un sismo explosivo se divide en las fases que viajan por tierra y la que lo hace por aire. Las ondas que viajan por aire vuelven a tierra donde son detectadas por los sismógrafos [3].

Eventos externos del volcán

Además de los tipos propios de la actividad sísmica es importante tener en cuenta los eventos externos del volcán, con el fin de llevar a cabo una mejor clasificación. Los siguientes eventos son tomados de la referencia [9]:

Evento Glacial

Los volcanes de gran altitud frecuentemente están cubiertos por una capa glacial, que a su vez producen eventos sísmicos fractura de hielo (IC, Ice Crack) que a lo largo del límite de la roca con el hielo y dentro del hielo cuando se forman grietas. Los sismogramas de tales eventos se parecen a los eventos volcánicos de largo periodo.

Evento por deslizamiento

Las avalanchas y derrumbes de tierra de diversos tamaños ocurren en los volcanes generalmente, asociados con el derretimiento parcial del hielo de la capa glacial, producidos principalmente en la primavera y el verano. La avalancha principal o deslizamiento genera una señal sísmica de varios minutos de duración, dependiendo del tamaño y la desviación. La amplitud varía con la cantidad de material arrasado. las avalanchas más grandes generan señales sísmicas más fuertes.

Eventos por caída de rocas

La caída de rocas son un tipo de avalancha especial de desprendimiento de tierras en la que uno o unos pocos pedazos de roca se desprenden y caen libremente. Son muy comunes en los bordes de las cúpulas volcánicas y en los volcanes con pendientes muy pronunciadas o acantiladas.

Representación de las señales sísmo-volcánicas

Frecuencia – Transformada de Fourier

La Transformada discreta de Fourier es la principal herramienta que permite diferenciar los registros de las señales de acuerdo a su contenido espectral. Esta transformación matemática permite llevar una señal expresada en el dominio de tiempo a una expresada en el dominio de la frecuencia. Está definida como (3.1) [28]:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}, \quad (3.1)$$

donde $x[n]$ es la señal discreta, N es el número de muestras tomadas de la señal, $e^{-j\frac{2\pi}{N}kn}$ es la N -ésima base de Fourier, con k y n variando entre 0 y $N - 1$.

Resulta útil analizar el dominio de la frecuencia, debido que a menudo aspectos como el filtrado selectivo de frecuencias se observan mejor en esta representación.

Tiempo frecuencia – Espectrograma

Con el fin de realizar un análisis más exhaustivo en el análisis frecuencial de las señales sísmo-volcánicas, también se usa la representación en tiempo frecuencia o

espectrograma, resultando de calcular el espectro en ventanas deslizantes sobre la señal. El análisis en Tiempo-Frecuencia utilizado en este trabajo corresponde al que utiliza la Transformada Discreta de Fourier (STFT, Short-Time Fourier Transform) con una ventana deslizante Hamming [29]. Las señales sismo- volcánicas se caracterizan por tener una frecuencia variable a lo largo del tiempo, como se puede ver en la figura 5, con el análisis del espectrograma se puede ver la evolución del espectro con respecto al tiempo. La STFT se define como:

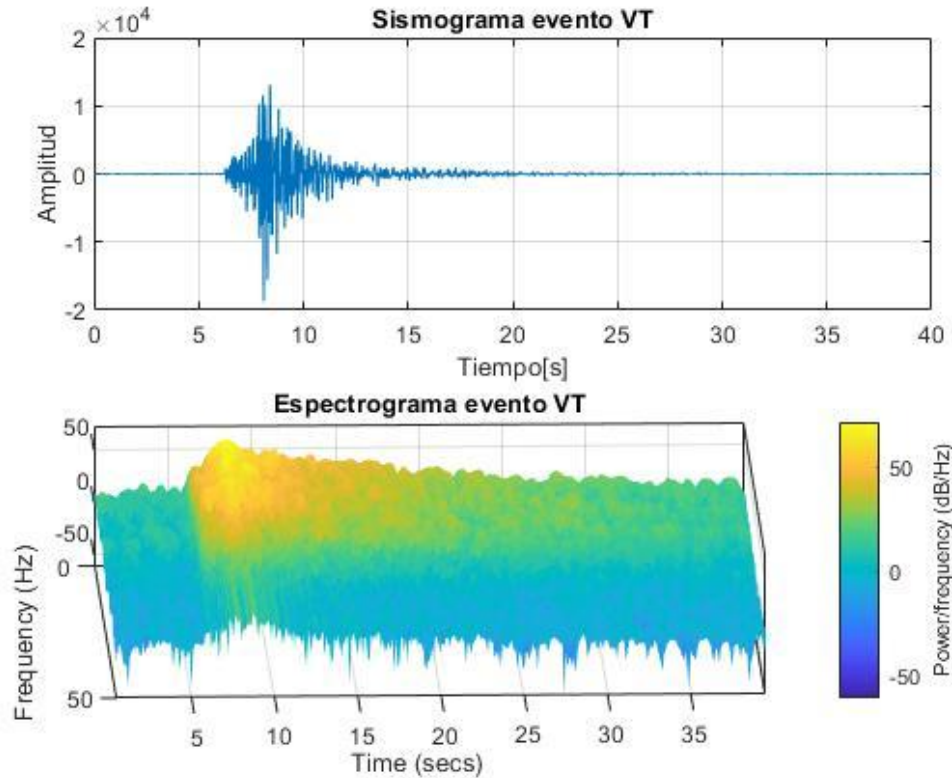
$$STFT\{x[n]\} \equiv X(n, k) = \sum_{m=-\infty}^{+\infty} x[n] w[n-m] e^{-j\frac{2\pi}{N}kn} \quad , \quad (3.2)$$

donde $x[n]$ es la señal y $w[n-m]$ es la ventana. De nuevo, el índice de tiempo discreto m es normalmente considerado como un tiempo "lento" y usualmente no se expresa con tan alta resolución como con el tiempo n [31]. Luego la magnitud cuadrada de la STFT origina el espectrograma así:

$$espectrograma\{x(n)\} \equiv |X(n, k)|^2 \quad , \quad (3.3)$$

El principal propósito de la ventana en la STFT es limitar la extensión de la secuencia que se va a transformar, de forma que las características espectrales sean aproximadamente constantes en el intervalo de duración de la ventana. Cuanto más rápidas sean las variaciones de la señal, más corta debe ser la ventana. La correcta selección de la longitud de la ventana define la precisión en el hallazgo de frecuencias y su ubicación en el tiempo. Una ventana más pequeña da mayor opción a determinar variaciones en el tiempo, pero decrece la resolución en frecuencia. De manera inversa ampliar la longitud de la ventana, amplía la resolución en frecuencia, pero no se permite la resolución en el tiempo.

Figura 5. Sismograma y espectrograma de un evento VT.



Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.

Cepstral – Coeficientes Cepstrales en las frecuencias de Mel

El Cepstrum es el resultado de calcular la transformada inversa de Fourier del espectro de la señal en escala logarítmica.

$$C(s^{(l+1)}) = F^{-1}[\log(|F\{s^{(l)}\}|)] = \sum_{n=0}^{N-1} \log \left(\left| \sum_{k=0}^{N-1} X[k] e^{-j\frac{2\pi}{N}kn} \right| \right) e^{j\frac{2\pi}{N}kn} \quad , (3.4)$$

Donde F es la DFT y F^{-1} es la transformada de Fourier inversa.

Dada la similitud de los registros de los sismos volcánicos con las señales de voz es oportuno utilizar el método paramétrico Cepstral calculando los coeficientes Cepstrales de las frecuencias de Mel (MFCC, Mel Frequency Cepstral Coefficients). Que se han convertido en un estándar en la codificación de la voz [31], y se han

realizado adaptaciones a las señales sísmicas en los estudios realizados en las referencias [31] [32] [33].

La figura 6 muestra el proceso mediante el cual se extraen los coeficientes de Mel:

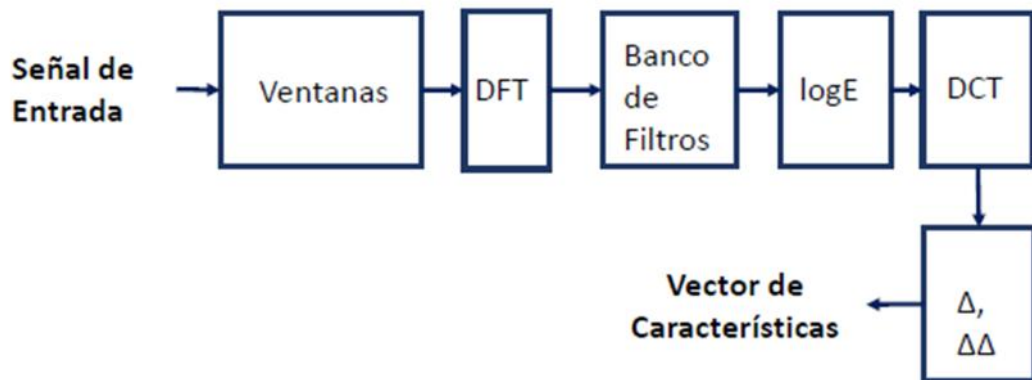


Figura 6. Diagrama de bloques extracción parámetros MFCC. **Fuente:** [32]

Los MFCC se calculan comúnmente de la siguiente forma:

- Separar la señal en ventanas.
- A cada ventana aplicarle la Transformada de Fourier discreta y obtener la potencia espectral.
- Aplicar el banco de filtros correspondientes a la escala Mel al espectro obtenido en el paso anterior y sumar las energías en cada uno de ellos.
- Tomar el logaritmo de todas las energías de cada frecuencia Mel
- Aplicarle la transformada de coseno discreta a estos logaritmos.
- Finalmente se obtienen los coeficientes que buscamos. Aunque opcionalmente se le pueden añadir otros valores como los deltas y/o los delta-deltas [34].

Sistemas computacionales para reconocimiento de señales sismo-volcánicas

El aprendizaje automático o aprendizaje de maquina (machine learning) es un área de alto interés desarrollada en el campo de la inteligencia artificial, que estudia como

las máquinas pueden tomar decisiones a partir de la experiencia acumulada de los casos resueltos anteriormente, para mejorar su actuación [35].

Las técnicas de machine learning son ampliamente utilizadas y en las señales sismo-volcánicas también son de alto interés, una herramienta computarizada capaz de tomar decisiones a partir de la experiencia.

Reconocimiento de patrones

El campo del reconocimiento de patrones se relaciona con el descubrimiento automático de regularidades en los datos y mediante el uso de algoritmos informáticos, realizar acciones tales como clasificar los datos en diferentes categorías [36].

Técnicas de reconocimiento de patrones (RP) han sido utilizada en diferentes áreas de estudio como: procesamiento análisis e interpretación de señales e imágenes, inteligencia artificial, algoritmos genéticos entre otras [37], siendo un campo muy amplio.

Para la mayoría de las aplicaciones prácticas, en un RP las variables de entrada originales normalmente se procesan previamente para transformarlas en un nuevo espacio de variables, donde se espera que el problema de reconocimiento de patrones sea más fácil de resolver. Esto reduce en gran medida la variabilidad dentro de cada clase de datos, ya que la ubicación y la escala de la información es ahora igual, lo que facilita mucho más que un algoritmo de clasificación posterior distinga entre las diferentes clases. Esta etapa de preprocesamiento se denomina extracción de características [36].

Resumiendo, el reconocimiento de patrones tiene como objetivos identificar el patrón o muestra como parte de una clase ya definida basándose en la información ya clasificada (Clasificación supervisada), o asignar una clase todavía no definida (clasificación no supervisada, agrupamiento o clustering).

Caracterización

A medida que el aprendizaje automático aborda la solución de investigaciones de temas importantes los cuales contienen mucha información, es importante centrarse en la información más relevante. La fase inicial del reconocimiento de patrones es la representación en características particulares de cada muestra que faciliten su discriminación [36].

Un algoritmo completo de reconocimiento de patrones basado en información principal puede ser computacionalmente inviable, en su lugar, el objetivo es encontrar características útiles que sean rápidas de computar y también conserven la información discriminativa útil que permite diferenciar las categorías [37].

Las señales sísmico-volcánicas se caracterizan por ser muy variables, sin embargo, investigaciones realizan una caracterización utilizando diferentes técnicas de representación en diferentes dominios, y métodos matemáticos y estadísticos, que logra representar de maneras adecuadas los diferentes tipos de sismos [1] [2] [5] [6]. Teniendo aspectos como: la duración, frecuencias dominantes, la morfología de cada sismo, la energía, entre otros.

Selección de características

A medida que machine learning asume la solución de diversos problemas tecnológicos, la información que representa cada decisión a tomar, usualmente, se ve reflejada en una cantidad abrumadora de información que ralentiza el procesamiento y además puede poseer información redundante. De acuerdo con lo anterior, las técnicas de selección de características buscan reducir la dimensión de la información discriminativa y eliminar información reiterada, potencializando las características más relevantes para la clasificación. A continuación, se generaliza algunas técnicas utilizadas en este campo descritas en [38].

- **Enfoque Embebido, para la selección de características:** Los métodos para inducir descripciones lógicas proporcionan el ejemplo más claro de los métodos de selección de características incorporados dentro de un algoritmo de inducción básico. Para estos métodos el ordenamiento parcial también describe el espacio de las hipótesis y los algoritmos suelen utilizar este ordenamiento para organizar su búsqueda de descripciones de conceptos.
- **Filtro, para la selección de características:** Este método filtra los atributos irrelevantes, utilizando las características generales del conjunto de entrenamiento para seleccionar algunas y excluir otras.
- **Agrupamiento, para selección de características:** El algoritmo genérico de Agrupamiento típico busca el mismo espacio de subconjuntos de características, que los métodos embebido y de filtro, pero evalúa conjuntos alternativos ejecutando algún algoritmo de inducción en los datos de entrenamiento y utilizando la precisión estimada del clasificador resultante como su métrica.

- **Ponderación de características:** Otro enfoque es aplicar una función de ponderación a las características, asignándoles grados de relevancia percibida. Los esquemas de ponderación tienden a ser más fáciles de implementar en configuraciones incrementales en línea y en general, están más motivados por consideraciones de rendimiento.

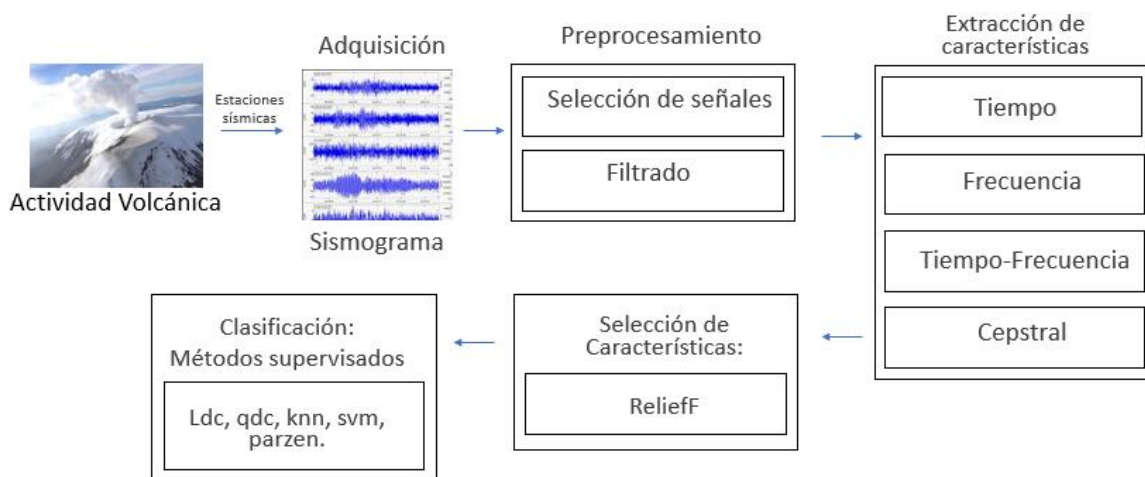
Clasificación

Luego de tener los datos representados en características, viene la etapa de clasificación con el propósito asignar una etiqueta de clase a la información procesada. En este trabajo se tiene en cuenta solo los algoritmos para clasificación supervisada en la cual a partir de un conjunto de datos conocidos por la maquina (conjunto de entrenamiento), se clasifica un segundo conjunto de datos desconocidos (conjunto de prueba).

1. DESARROLLO DE LA INVESTIGACIÓN

La metodología propuesta en este trabajo se orientó a encontrar las características que mejor representen a las señales sísmico-volcánicas de acuerdo a la literatura, para obtener la mejor clasificación. Se incluyó las siguientes etapas preprocesamiento, extracción de características, selección de características y clasificación como se muestra en la figura 7.

Figura 7. Metodología para el reconocimiento de patrones en señales sísmico-volcánicas.



Fuente: Esta investigación.

1.1. REGISTRO SEÑALES SÍSMO-VOLCANICA

Las señales sísmicas utilizadas en esta investigación fueron extraídas de una base de datos de registros sísmicos otorgada por el Observatorio Vulcanológico y Sismológico de Manizales (OVSM), correspondientes al volcán Nevado del Ruiz. La base de datos cuenta con registro de eventos sísmo-volcánicos segmentados y etiquetados, digitalizados a una frecuencia de muestreo de 100 HZ.

Eventos volcánicos de diferentes estaciones sísmicas con su respectiva componente, vertical, este y norte conforman la base de datos. Para el desarrollo de este trabajo se utilizó el registro de los eventos sísmicos para la estación sísmica OLLETA, que cuenta con 1860 registros clasificados de la siguiente manera: 450 datos registrados como eventos Largo Periodo (LP), 450 datos como Vulcano-Tectónicos (VT), 450 registros clasificados como eventos relacionados con la capa

glacial (IC), 450 eventos Tremor (TR), 56 eventos clasificados como eventos Híbridos (HYB) y 4 eventos Explosión (EXP).

En la base de datos de la estación OLLETA se encontró un problema de desbalanceo de clases, que significa diferencia de registros entre las clases de eventos sísmicos, si se lo trabaja en ese estado afectaría en gran magnitud el desempeño del clasificador. Para evitar el problema descrito, en este trabajo se utilizaron las clases balanceadas, es decir las que tienen igual número de registros, consecuentemente, la base de datos quedo conformada por cuatro clases de igual cantidad de información, los eventos analizados son VT, LP, TR e IC.

1.2. PRE-PROCESAMIENTO

Antes de someter las señales sismo-volcánicas a un procesamiento de señales se eliminaron los niveles de offset usando:

$$S = S - \bar{S}, \quad (4.1)$$

Donde S representa cada muestra de la base de datos y \bar{S} su correspondiente media.

Después, se aplicó un filtro Butterworth pasa banda, de 0.5 a 25 Hz para eliminar ruido y frecuencias que no pertenecen a las ondas sísmicas, que por naturaleza de origen están comprendidas entre 1 y 25 Hz.

Después del filtrado se realizó un análisis visual en donde se encontró señales que podrían no pertenecer a una señal sismo volcánica. Debido a esto, se aplicó un análisis de relación señal a ruido (SNR, Signal to Noise Ratio), definido como el margen que hay entre el nivel de referencia y el ruido de fondo de un determinado sistema. La relación señal ruido, es la comparación entre la potencia de la señal sísmica y la potencia del ruido que podría ser causado por el instrumento sísmico u otras fuentes ajenas con la naturaleza del sismo. En términos cuantitativos, la energía de la señal para una frecuencia dada, dividida entre la energía remanente considerada ruido [45]. Con la SNR se logró identificar algunos eventos que no corresponden a sismos volcánicos. Estas muestras fueron retiradas de la base de datos trabajada con el fin de que no estropeen la tarea de caracterización y clasificación. Se redujo la base de datos a 1784 registros con 4 clases VT, LP, IC y TR con 446 registros cada una.

1.3. REPRESENTACIÓN SEÑALES VOLCÁNICAS

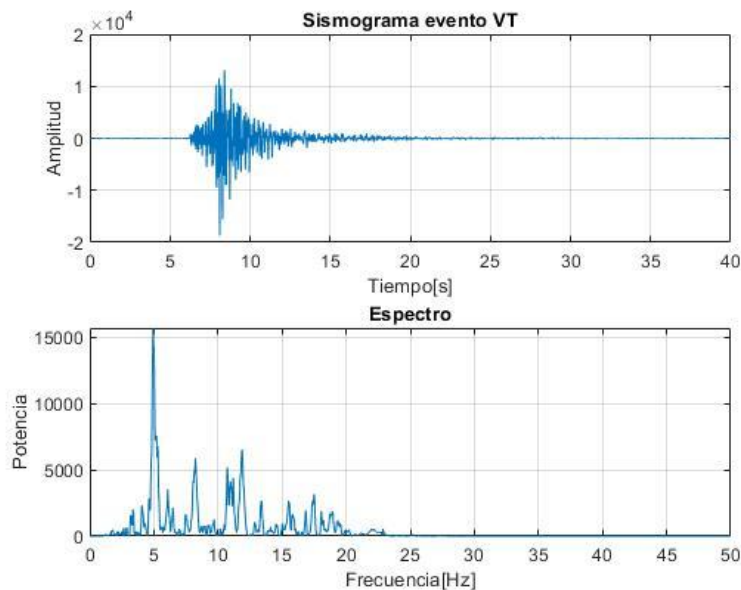
Para comprender mejor las señales sismo-volcánicas, se realizó la descomposición en 3 dominios, frecuencia, tiempo-frecuencia y Cepstral partiendo del dominio tiempo, estado natural en el que se encuentran las muestras.

1.3.1. Análisis frecuencia

El análisis en el dominio frecuencia es usado en el tipo de señales de estudio puesto que se caracterizan por su variabilidad en el contenido frecuencial. Es por eso que el análisis de su espectro en el dominio de la frecuencia, resulta conveniente.

Se calculo la transformada discreta de Fourier y se obtuvo el contenido espectral de las señales sísmicas, con el que se estudió las diferentes características.

Figura 8. Espectro de evento sísmico vulcano-tectónico.



Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.

1.3.2. Análisis Tiempo-Frecuencia

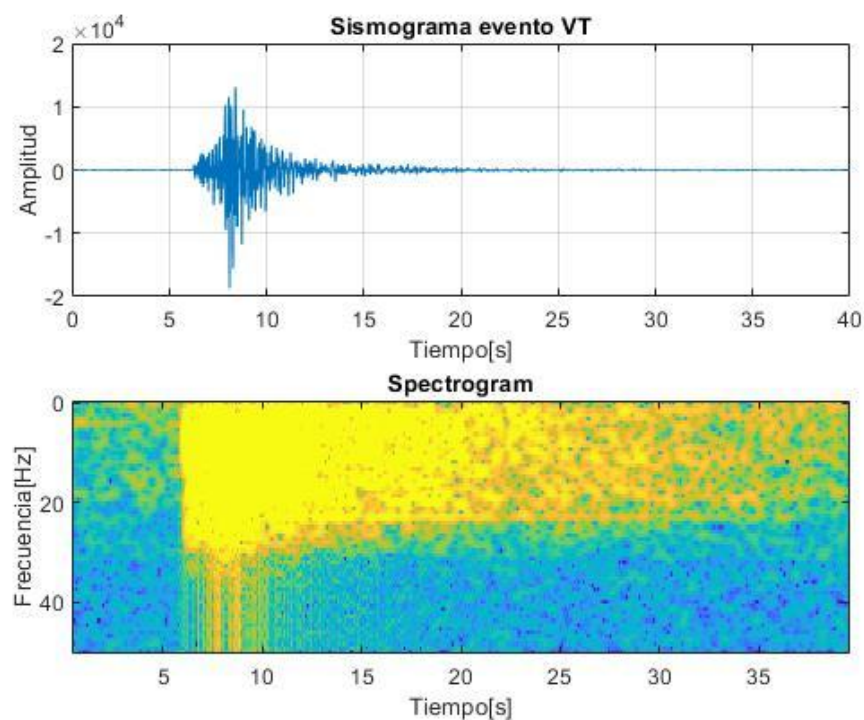
El estudio en el dominio tiempo-frecuencia, aportó mejor contenido a la representación de la señal, donde se observó el comportamiento de la frecuencia con respecto al tiempo. Este análisis es conocido como espectrograma. Los

espectrogramas facilitan un mejor análisis de los fenómenos sísmicos, tanto visual como computarizado [46] [47].

Para construir los espectrogramas de las señales en este trabajo, se calculó la transformada discreta de Fourier, con una ventana deslizante de un segmento de la señal sísmica. Este análisis está basado en Hilbert et al 2017, donde plantea los parámetros a tener en cuenta al calcular el espectrograma para las señales sismo-volcánicas. Se implementa valores indicados por los autores y se corrobora la funcionalidad para nuestras señales, como se observa en la figura 4.

Se calculó la transformada de Fourier discreta con una ventana móvil de 1 segundo (100 muestras) y un 90% de superposición.

Figura 9. Espectrograma evento sísmico vulcano-tectónico.



Fuente: Base de datos sísmicos del OVSM en Volcán Nevado del Ruiz.

1.3.3. Análisis Cepstral

De acuerdo a la literatura se tiene varias características de similitud entre los sismos volcánicos y las señales de audio, cuya caracterización más importante en ese campo es del cepstrum a través de los coeficientes cepstrales de frecuencias de Mel MFCC, para nuestro trabajo se aplicó esta representación a las señales.

1.4. CARACTERIZACIÓN SEÑALES SÍSMO-VOLCÁNICAS

Las señales sísmicas son señales no estacionarias y transitorias, presentan frecuencias y amplitudes variantes a lo largo del registro. Lo que hace necesario trabajar las características en diferentes representaciones con el propósito de encontrar un grupo de atributos que discrimine al máximo estas señales. En este trabajo se realizó un grupo de características en diferentes dominios, tiempo, frecuencia, tiempo-frecuencia y Cepstral, recopiladas de diferentes investigaciones desarrolladas en el campo de caracterización de eventos sísmo-volcánicos. En las siguientes tres tablas se presenta las características calculadas a cada evento, de acuerdo a la representación aplicada, Tiempo, Frecuencia, Tiempo-Frecuencia, y se describe la forma de calcular los coeficientes de frecuencias de Mel.

Tabla 2. Características dominio temporal.

Características dominio Temporal			
N°	Descripción	Fórmula matemática	Referencia
1	Media	$\mu = \frac{1}{N} \sum_{n=0}^{N-1} s[n]$	[5]
2	Varianza	$\sigma^2 = \frac{1}{N-1} \sum_{n=0}^{N-1} (s[n] - \mu)^2$	[2]
3	Desviación estándar	$\sigma = \sqrt{\frac{1}{N-1} \sum_{n=0}^{N-1} (s[n] - \mu)^2}$	[5]
4	Mediana	$\text{mediana}(s[n])$	[5]
5	Valor máximo	$\max(s[n])$	[5]

6	Valor RMS	$r_{rms} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} s[n]^2}$	[2]
7	Energía de la señal	$E = \sum_{n=0}^{N-1} s[n] ^2$	[2]
8	Tiempo de la energía central	$\bar{t} = \frac{1}{E} \sum_{n=0}^{N-1} n s[n] ^2$	[2]
9	Entropía de Shannon	$-\sum_{n=0}^{N-1} p(s[n]) \log_2(p(s[n]))$	[2]
10	Contraste de la señal	$\frac{\max(s[n]) - \min(s[n])}{\mu}$	[2]
11	Duración	$T = t_e - t_s$	[51] [5]
12	Contraste de la señal envolvente	$\frac{M_{QX}(e[n])}{Median(e[n])}$	[51]
13	Contraste envolvente	$\frac{M_{QX}(e[n])}{Median(e[n])}$	[1]
14	Tiempo Ascendente	$\frac{t_{max} - t_s}{t_e - t_{max}}$	[5]
15	Curtosis señal	$\frac{1}{N} \sum_{n=0}^{N-1} \left[\frac{s[n] - \mu}{\sigma} \right]^4$	[52]
16	Curtosis envolvente	$\frac{1}{N} \sum_{n=0}^{N-1} \left[\frac{e[n] - \mu}{\sigma} \right]^4$	[51]
17	Sesgo señal	$\frac{1}{N} \sum_{n=0}^{N-1} \left[\frac{s[n] - \mu}{\sigma} \right]^3$	[52]
18	Sesgo envolvente	$\frac{1}{N} \sum_{n=0}^{N-1} \left[\frac{e[n] - \mu}{\sigma} \right]^3$	[51]
19	Numero de picos en la autocorrelación	$a[\tau] = \frac{1}{N} \sum_{n=0}^{N-\tau} s[n]s[n + \tau]$	[6]

20	Energía 1/3 de la autocorrelación	$\sum_{n=0}^T a[n] ^2$	[6]
21	Energía 2/3 de la autocorrelación	$\sum_{n=3}^T a[n] ^2$	[6]
22	Relación energías	$\frac{\sum_{n=0}^T a[n] ^2}{\sum_{n=3}^T a[n] ^2}$	[6]
23-27	Energía señal envolvente usando filtros	$ES^i = \sum_{n=0}^{N-1} e[n] ^2$	[5]
28-32	Curtosis señal usando filtros	$\frac{1}{N} \sum_{n=0}^{N-1} \left[\frac{e[n] - \mu}{\sigma} \right]^4$	[1]

En la tabla 2 se establecen características estadísticas, forma de onda y análisis de la energía. N es el número de muestras de la señal, T es duración, t_s es tiempo de inicio de la señal, t_e es tiempo final de la señal, t_{max} es el tiempo de la máxima amplitud, $e[n]$ es la envolvente, $p(s[n])$ es la probabilidad del conjunto de registros. Para las características numero 23 a 27 y 28 a 32 se utilizó 5 filtros pasa banda en diferentes frecuencias, aplicados a la señal original y su envolvente de la siguiente manera $F1(0.1-1 \text{ Hz})$, $F2(1-3 \text{ Hz})$, $F3(3-10 \text{ Hz})$, $F4(10-20 \text{ Hz})$, $F5(20-50 \text{ Hz})$.

Tabla 3. Características dominio frecuencia.

Características dominio Frecuencia			
No	Descripción	Fórmula Matemática	Referencia
1	Media espectral	$\mu = \frac{1}{N} \sum_{n=0}^{N-1} S[n]$	[1]
2	Valor máximo potencia espectral	$\max(\sum_{n=0}^{N-1} S[n])$	[1]
3	Valor frecuencia máxima	f_{max}	[5]
4	Mediana espectral	$median(S[n])$	[1]
5	Varianza	$\sigma^2 = \frac{1}{N-1} \sum_{n=0}^{N-1} (S[n] - \mu)^2$	[1]

6	Riqueza espectral	$\frac{\max S[n] }{\text{median}\{ S[n] \}}$	[2]
7	Frecuencia centroide	$\frac{\sum_n f[n] S[n] }{\sum_n S[n] }$	[1]
8	F centroide cuartil 1	$\frac{\sum_{f^1} f[n] S[n] }{\sum_{f^1} S[n] }$	[1]
9	F Centroide cuartil 3	$\frac{\sum_{f^3} f[n] S[n] }{\sum_{f^3} S[n] }$	[1]
10	Numero de picos	$S[n] > 0.75 f_{max}$	[1]
11	Media de picos de frecuencia	$\text{media}(S[n] > 0.75 f_{max})$	[1]
12-16	Energía espectral utilizando cinco filtros a la señal	$E^l f = \sum_{n=0}^{N-1} S[n] ^2$	[1]
17	Gamma1	$\gamma_1 = \frac{\sum_n f[n] S[n] }{\sum_n S[n] ^2}$	[1]
18	Gamma2	$\gamma_2 = \sqrt{\frac{\sum_n f[n]^2 S[n] ^2}{\sum_n S[n] ^2}}$	[1]
19	Relación entre gamma1 y gamma2	$\sqrt{ \gamma_1 ^2 - \gamma_2^2}$	[1]

En la tabla 3 se describe la representación matemática de las características espectrales utilizando la DFT. $f[n]$ es el vector de frecuencia. En la característica 11 a 15 se calculó la energía del espectro usando los filtros descritos en la Tabla 2.

Tabla 4. Características dominio Tiempo- Frecuencia.

Características Dominio Tiempo Frecuencia			
No	Descripción	Fórmula Matemática	Referencia
1	Curtosis máxima	$\text{Kurt}(\max(DFT(n, k)))$	[1]
2	Curtosis mediana	$\text{Kurt}(\text{median}(DFT(n, k)))$	[1]
3	Media contraste, máximo sobre media	$\text{mean}\left(\frac{\max(DFT(n, k))}{\text{mean}(DFT(n, k))}\right)$	[1]
4	Media contraste, máximo sobre mediana	$\text{mean}\left(\frac{\max(DFT(n, k))}{\text{median}(DFT(n, k))}\right)$	[1]

5	Número de picos en Máxima DFT	Numero de picos (Max DFT(n, k))	[1]
6	Numero de picos en Media DFT	Numero de picos (Mean DFT(n, k))	[1]
7	Numero de picos en Mediana DFT	Numero de picos (Median DFT(n, k))	[1]
8	Relación entre 5 y 6	$\frac{\text{Numero de picos (Max DFT(n, k))}}{\text{Numero de picos (Mean DFT(n, k))}}$	[1]
9	Relación entre 5 y 7	$\frac{\text{Numero de picos (Max DFT(n, k))}}{\text{Numero de picos (Median DFT(n, k))}}$	[1]
10	Distribución MaxMeanFreqDTF	$\text{Mean}[\max[DFT(n, k)] - \text{mean}[DFT(n, k)]]$	[1]
11	Distribución MaxMedianFreqDTF	$\text{Mean}[\max[DFT(n, k)] - \text{median}[DFT(n, k)]]$	[1]

En la tabla 4 se establecen características a partir del espectrograma. Se realizó estudio estadístico con el propósito de describir la distribución del espectrograma, la frecuencia con respecto al tiempo.

Además de los atributos relacionados en la tabla 4. Se obtuvo características adicionales al calcular el valor medio de cada banda de frecuencia de la representación en tiempo frecuencia, la magnitud en decibelios se calcula como $20 \log_{10}|S|$, donde S es la matriz que contiene el espectrograma [47] [48]. Siguiendo la investigación realizada en P. Castro et al y R. Avessani et al se tomó los parámetros para calcular el espectrograma, de la siguiente manera FFT de 128 muestras, una ventana de Hamming de 64 muestras y un traslape de 50%. Con lo que se obtuvo una matriz de 65 características.

Dominio Cepstral: finalmente, se calculó los coeficientes de Mel MFCC con el propósito de representar las señales sísmicas, esta sección se desarrolló en base al enfoque propuesto por Álvarez et al. Para calcular los coeficientes, se realizó el siguiente procedimiento:

- Se superpuso una ventana de 4 segundos con un desplazamiento de 2 segundos.
- Se utilizó una transformada discreta de Fourier de 512 puntos (DFT) para calcular el espectro de magnitud.
- Se aplicó un banco de filtros emulado que consta de 16 funciones de ponderación triangular que se espaciaron de manera uniforme entre 0 y 25 Hz. La superposición entre filtros adyacentes es del 50%. El analizador de

banco de filtros proporciona una medida de la energía de la señal en una banda de frecuencia determinada.

- Luego, se calculó el logaritmo de las energías al banco de filtros de salida, lo que da como resultado un vector de características de 16 parámetros.
- Dado que las energías del banco de filtros de registro están altamente correlacionadas, fue necesario aplicar una transformación de des correlación. Por lo tanto, la transformada de coseno discreta (DCT) se utilizó para des correlacionar las características y reducir el número de componentes del vector de características de 16 a 13 coeficientes.
- Finalmente, el vector de características se aumentó con regresiones lineales de las características (derivadas y aceleraciones).
- Todas juntas formaron 26 características para este trabajo, 12 coeficientes, la energía y sus derivadas teniendo en cuenta la información del estudio P. Castro et al.

1.5. SELECCIÓN DE CARACTERÍSTICAS

Después de la caracterización de las señales sísmicas se obtuvieron un total de 153 características, entre las cuales hubo atributos de poca relevancia a la hora de discriminar las clases de eventos sismo-volcánico, por ello y con el fin de optimizar el sistema se realizó una selección de las características, con el propósito de evitar que el algoritmo de clasificación pierda eficiencia y velocidad.

Partiendo desde el punto de vista de la minería de datos y machine learning, la etapa de selección y/o extracción de características de un conjunto de datos es importante por la presencia de información irrelevante (características que generan problemas de sobre aprendizaje, porque son datos que no aportan información al sistema), información redundante (características relacionadas linealmente que realizan la misma función) y la dimensionalidad (El número de características es mayor al de datos)

El algoritmo de ReliefF desarrollado por Kira y Rendell en 1992, y modificado por Kononenko en 1994, pondera cada característica de acuerdo con su relevancia para la clase [55]. Este algoritmo selecciona una muestra de los datos e identifican la muestra vecina más cercana que pertenece a la misma clase y la muestra vecina más cercana que pertenece a la clase opuesta. Un cambio en el valor de la característica acompañado de un cambio en la clase conduce a una ponderación del atributo basada en la intuición de que el cambio de atributo podría ser responsable del cambio de clase. Por otro lado, un cambio en el valor del atributo acompañado de ningún cambio en la clase conduce a una ponderación menor del atributo basado en la observación de que el cambio de atributo no tuvo efecto en la

clase. Este procedimiento de actualización del peso se realiza para todos los datos. Las actualizaciones de peso se promedian para que el peso final esté en el rango [-1, 1]. Finalmente, para facilitar el análisis de los resultados, se realiza el respectivo ordenamiento de las características con sus respectivos pesos de mayor a menor. La característica con mayor peso representa la característica con mayor relevancia [56]. Se aplica este algoritmo a la matriz de características reduciéndola en un 50%.

1.6. CLASIFICACIÓN

Los métodos de clasificación planteados son supervisados, partiendo de que la base de datos trabajada, tiene información etiquetada por expertos que servirá para crear un conjunto de entrenamiento y otro de prueba, de acuerdo a la base de datos se tiene 4 clases balanceadas: VT, LP, IC, TR con 446 datos cada una, se realizó una clasificación binaria agrupando las 4 clases en grupo de dos clases, el conjunto de entrenamiento dado por el 80% de los datos, y el 20 % para la evaluación.

1.6.1. Clasificadores polinomiales de grado lineal y superior

➤ Clasificación basada en la densidad normal

Clasificador normal lineal (LDC)

Este clasificador asume que todas las clases se caracterizan por múltiples distribuciones normales con una matriz de covarianza común. El clasificador LDC, es matemáticamente expresado en las ecuaciones (4.61) y (4.62), así [57]:

$$R(x) = \left(\frac{\hat{\mu}_A - \hat{\mu}_B}{\hat{\Sigma}} \right)^T x + b, \quad (4.61)$$

$$b = \frac{1}{2} \left(\frac{p(A)}{p(B)} - 1 \right) \left(\frac{\hat{\mu}_A + \hat{\mu}_B}{2} \right)^T \hat{\Sigma}^{-1} \left(\frac{\hat{\mu}_A + \hat{\mu}_B}{2} \right) - \frac{1}{2} \left(\frac{p(A)}{p(B)} - 1 \right) \hat{\mu}_B^T \hat{\Sigma}^{-1} \hat{\mu}_B, \quad (4.62)$$

Donde $R(x)$ límite de decisión del clasificador, $x = [x_1, \dots, x_p]$, es el vector de interés, $\hat{\mu}_A, \hat{\mu}_B$ son los vectores de medias de las clases A y B y $\hat{\Sigma}^{-1}$ la matriz de covarianza común para ambas, $p(A), p(B)$ son las probabilidades a priori de que un x pertenezca a la clase A y B .

Clasificador normal cuadrático (QDC)

Este clasificador asume que las clases tienen múltiples distribuciones normales, pero cada una es caracterizada por una matriz de covarianza diferente. El

clasificador QDC, se expresa matemáticamente en la ecuación (4.63) y su contante en la ecuación (4.64) así [57]:

$$R(x) = -\frac{1}{2}(x - \bar{\mu}_A)^T \hat{\Sigma}_A^{-1}(x - \bar{\mu}_A) + \frac{1}{2}(x - \bar{\mu}_B)^T \hat{\Sigma}_B^{-1}(x - \bar{\mu}_B) + c, \quad (4.63)$$

$$R(x) = -\frac{1}{2}(x - \bar{\mu}_A)^T \hat{\Sigma}_A^{-1}(x - \bar{\mu}_A) + \frac{1}{2}(x - \bar{\mu}_B)^T \hat{\Sigma}_B^{-1}(x - \bar{\mu}_B) + c, \quad (4.64)$$

$$c = \log \left[\frac{p(\bar{\mu}_B)}{p(\bar{\mu}_A)} \right] - \frac{1}{2} \log \left[\frac{|\hat{\Sigma}_B|}{|\hat{\Sigma}_A|} \right]$$

Donde $R(x)$ límite de decisión del clasificador, $x = [x_1, \dots, x_p]$, es el vector de interés, $\bar{\mu}_A$, $\bar{\mu}_B$ son los vectores de medias de las clases A y B , $\hat{\Sigma}_A^{-1}$ y $\hat{\Sigma}_B^{-1}$ la matriz de covarianza.

➤ Clasificadores no lineales

Clasificador de Parzen

Este clasificador nace del concepto básico de aprendizaje supervisado no paramétrico basado en la estimación de densidades de probabilidad de Parzen. En la implementación se utilizan diferentes Kernels [50]. El objetivo de este clasificador es obtener estimaciones de densidades de probabilidad condicional $p(z|wk)$, el espacio de medida es patrocinado en un número finito de regiones disyuntas R_i llamadas cajas y se cuentan las muestras que caen en ellas, siendo la estimación de la densidad de probabilidad dentro de la caja proporcional a tal número, Además $N_{k,j}$ denota el número de muestras con clase WK , en [49] matemáticamente se expresa en la ecuación (4.66):

$$p(z|wk) = \frac{N_{k,j}}{\text{volumen}(R_i) \cdot N_k} \quad (4.65)$$

Máquina de soporte vectorial (SVM)

Este se basa en una sólida base teórica llamada minimización del riesgo empírico (ERM), cuyo objetivo es maximizar las distancias de los objetos de entrenamiento al clasificador [50]. Gracias a la función Kernel que implementa, busca un hiperplano que separe los puntos de una clase con otra de la forma más óptima, teniendo en cuenta la característica fundamental, es decir este tipo de algoritmos buscan el hiperplano que tenga la máxima distancia con los puntos que estén más cerca de el mismo, siendo previamente proyectado a un espacio de dimensión superior [36] [49]. Dado por su naturaleza, el clasificador SVM es muy funcional en clasificaciones biclase.

K vecinos más cercanos (KNN)

El algoritmo clasifica cada nuevo dato en el grupo que corresponda, según tenga k vecinos más cerca de un grupo o de otro. Es decir, calcula la distancia del elemento nuevo a cada uno de los existentes, y ordena dichas distancias de menor a mayor para ir seleccionando el grupo al que pertenecer [39].

Los ejemplos de entrenamiento son vectores en un espacio característico multidimensional, cada ejemplo está descrito en términos de p atributos considerando q clases para la clasificación. Los valores de los atributos del i –ésimo ejemplo (donde $1 \leq i \leq n$) se representan por el vector p -dimensional

$$x_i = x_{1i}, x_{2i}, \dots, x_{pi} \in X \quad (4.66)$$

El espacio es particionado en regiones por localizaciones y etiquetas de los ejemplos de entrenamiento. Un punto en el espacio es asignado a la clase si esta es la clase más frecuente entre los k ejemplos de entrenamiento más cercano. Generalmente se usa la distancia euclidiana.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^p (x_{ri} - x_{rj})^2} \quad (4.67)$$

2. MARCO EXPERIMENTAL

En esta sección se describe de las medidas de desempeño del clasificador sobre los diferentes conjuntos de características y los experimentos a desarrollar.

2.1. MEDIDAS DE CLASIFICACIÓN

Para calificar el resultado de la clasificación se utilizaron las siguientes medidas de desempeño: *Sensibilidad (Se)*, *Especificidad (Sp)* y *Exactitud*; estas medidas son calculadas con las ecuaciones (5.1), (5.2), (5.3).

$$Se = \frac{Vp}{Vp+Fn} \quad (5.1)$$

$$Sp = \frac{Vn}{Vn+Fp} \quad (5.2)$$

$$Exactitud = \frac{Vn \times Vp}{Vn+Fp+Vp+Fn} \quad (5.3)$$

Y los valores para calcular estas medidas son:

- i. *Vp*: Verdaderos positivos o casos de la clase de interés clasificados correctamente.
- ii. *Vn*: Verdaderos negativos o casos diferentes de la clase de interés clasificados correctamente.
- iii. *Fp*: Falsos positivos o casos diferentes de la clase de interés clasificados como casos de la clase de interés.
- iv. *Fn*: Falsos negativos o casos de la clase de interés clasificados como casos diferentes de la clase de interés.

2.1.1. Curvas ROC

La curva ROC, (Receiver Operating Characteristic), característica operativa del receptor, es una visualización gráfica de la sensibilidad en el eje “y” y 1 – especificidad en el eje x para variar los puntos de corte de los valores de prueba. Esto se representa generalmente en una caja cuadrada para mayor comodidad y sus dos ejes son de 0 a 1. El área bajo la curva ROC (AUC, Area Under the Curve) es una medida efectiva y combinada de sensibilidad y especificidad para evaluar la validez inherente de una prueba de diagnóstico. Máximo AUC = 1 y significa que la prueba es perfecta [53].

AUC es el indicador más utilizado en muchos contextos, se puede interpretar como la probabilidad de que un clasificador ordenará o puntuará una instancia positiva

elegida aleatoriamente más alta que una negativa. A modo de guía para interpretar las curvas ROC se puede decir de los siguientes intervalos de AUC que [0,5] es como lanzar una moneda, [0.5-0.6] test malo, [0,6 – 0,75] test regular, [0.75, 0.9]: test bueno, [0,9 – 0,97] test muy bueno, [0.97, 1] test excelente [54].

2.2. PRUEBAS APLICADAS SOBRE LA BASE DE DATOS

En la metodología propuesta, para las señales sismo-volcánicas se realizó la representación en diferentes dominios consiguiendo varias características temporales, de frecuencia, tiempo-frecuencia y cepstrum, que bien se pueden tomar por separado o combinarse con el propósito de encontrar la mejor clasificación de los diferentes eventos volcánicos. De esta manera se realizan 7 experimentos, con 5 subpruebas para observar la reacción de los clasificadores, lineal (LDC), cuadrático (QDC), K Vecinos más cercanos (K-NN) con $K = 3$ vecinos para no sobre exigir el sistema computacionalmente y minimizar los errores por presencia de la clase más grande, máquina de vectores de soporte (SVM) y clasificador PARZEN, para determinar qué tipo de características se ajustan más a las señales. Se inició con experimentos de 2 clases con el propósito de determinar las características que mejor se ajustan a cada clase. Posteriormente, se experimentó evaluando las características planteadas, clasificando las 4 clases, VT, LP, TR e IC, propuestas en la Sección 4.1. Para esta clasificación se utilizó el clasificador por naturaleza multiclase que es K-NN.

En cada prueba se realizó 20 iteraciones, en cada repetición una se ejecutó los algoritmos de clasificación, los cuales se entrenaron con el 80% de los datos y se desarrolló pruebas con el 20% restante. Se registró el valor de error de clasificación, para encontrar el clasificador con más estabilidad con respecto de su mediana y menor error. Para su visualización se utilizaron diagramas de cajas y bigotes que facilitan el análisis, mostrando en la gráfica la mediana, el punto máximo y mínimo y los cuartiles 2 y 3, logrando identificar el clasificador con mejor desempeño y mayor estabilidad.

Se obtuvieron 3 medidas de desempeño, sensibilidad, especificidad y exactitud. Además, el AUC de la curva ROC y la gráfica del error de clasificación. Como método de selección de mejores características se aplicó el algoritmo de ReliefF.

A continuación, se realizó una clasificación binaria, con selección de características ReliefF, tomando el 50% de los atributos ordenados de acuerdo a la ponderación asignada por el algoritmo, se clasificó con los 5 clasificadores descritos en la sección 4.6. y se realizaron las siguientes pruebas.

- Clasificación utilizando matriz de características temporales.
- Clasificación utilizando matriz de características de frecuencia.

- Clasificación utilizando matriz de características Tiempo-Frecuencia.
- Clasificación utilizando matriz de características de coeficientes cepstrales MFCC.
- Clasificación utilizando matriz completa de características, concatenando cada conjunto descrito anteriormente.

Se combinaron las clases de sismos volcánicos de la siguiente manera.

- **Experimento 1:** Clasificación eventos VT y LP.
- **Experimento 2:** Clasificación eventos VT y IC.
- **Experimento 3:** Clasificación eventos VT y TR.
- **Experimento 4:** Clasificación eventos LP y IC.
- **Experimento 5:** Clasificación eventos LP y TR.
- **Experimento 6:** Clasificación eventos IC y TR.
- **Experimento 7:** Se realizó una clasificación multiclase con selección de características mediante el algoritmo de Relief, se toma el 50% de los atributos ordenados de acuerdo a la ponderación asignada por el algoritmo. Se utilizó K-NN como técnica de clasificación, siguiendo las mismas pruebas.

Finalmente se realizó una interfaz de usuario que permite recrear los resultados obtenidos, donde se permita al usuario interactuar con diferentes eventos, y combinar diferentes grupos de características provenientes de las diferentes representaciones de las señales sísmico volcánicas.

3. RESULTADOS

Después de aplicar la metodología propuesta a los experimentos planteados se obtuvieron los siguientes resultados.

Se utilizo la siguiente connotación para los clasificadores utilizados en las pruebas

C1: Clasificador Lineal (LDC)

C2: Clasificador cuadrático (QDC)

C3: Clasificador K – vecinos más cercanos (KNN)

C4: Maquina de soporte vectorial (SVM)

C5: Clasificador de Parzen (PARZEN)

- **Experimento 1:** Medidas de desempeño para la clasificación de eventos sismo-volcánicos VT- LP

Tabla 5. Medidas de desempeño *Se*, *Sp* y *Exactitud* de clasificación de eventos VT-LP.

SENSIBILIDAD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,73	0,76	0,72	0,74	0,81
Frecuencia	0,74	0,78	0,74	0,73	0,79
Tiempo-frecuencia	0,79	0,70	0,75	0,70	0,78
MFCC	0,60	0,79	0,67	0,71	0,65
Matriz completa	0,84	0,84	0,83	0,90	0,85
ESPECIFICIDAD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,81	0,76	0,81	0,76	0,72
Frecuencia	0,75	0,76	0,73	0,82	0,73
Tiempo-frecuencia	0,60	0,71	0,71	0,71	0,66
MFCC	0,60	0,53	0,65	0,56	0,65
Matriz completa	0,87	0,87	0,75	0,67	0,81
EXACTITUD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,77	0,76	0,76	0,75	0,76
Frecuencia	0,75	0,77	0,74	0,78	0,76
Tiempo-frecuencia	0,69	0,70	0,73	0,70	0,72
MFCC	0,60	0,66	0,66	0,63	0,65

Matriz completa	0,85	0,85	0,79	0,79	0,83
------------------------	------	------	------	------	------

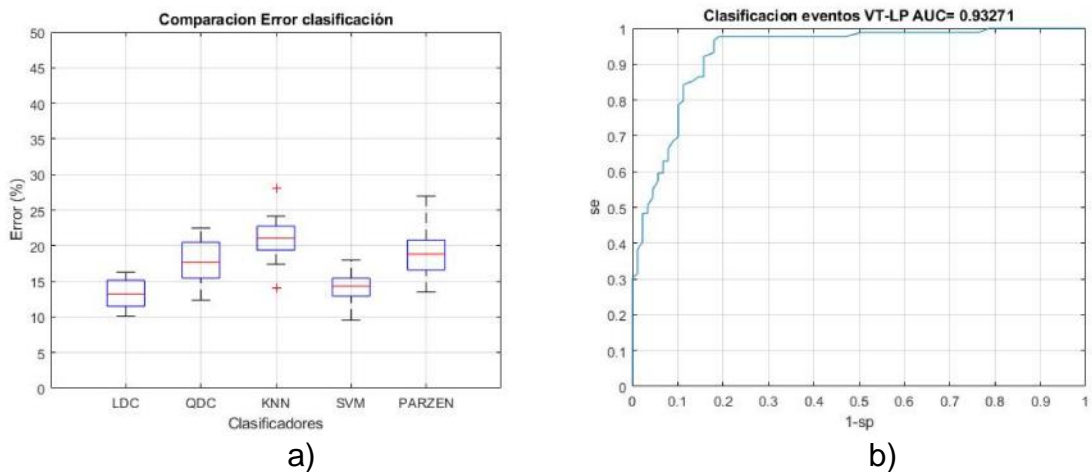
En el experimento 1 se observó mayor sensibilidad en la clasificación realizada con la matriz completa, que contiene concatenados los grupos de características planteados. El valor más alto se dio mediante la clasificación C4, clasificador SVM con el 0,9 de sensibilidad. No obstante, cabe resaltar los valores de la sensibilidad obtenidos por C5 PARZEN, se encuentran en el rango de 0.78 a 0.85 excepto 0.65 que se obtuvo con el grupo de características de MFCC.

En cuanto a la especificidad se obtuvieron los mejores resultados con la matriz completa de características con un valor de 0.87 para los clasificadores C1 y C2. Buenos resultados con las características de tiempo y frecuencia para clasificadores C1, C3 y C4 respectivamente.

La mejor clasificación, de acuerdo a la exactitud se obtuvieron con las características de la matriz completa con valores de 0.85 para el clasificador LDC.

Se muestra en la figura 10 a), el diagrama de cajas y bigotes del error para cada clasificador en las 20 iteraciones. Se verifica que la mejor clasificación se alcanzó utilizando el conjunto de características completo con el clasificador LDC. En la figura 10 b) se muestra la curva ROC para el clasificador LDC con un AUC = 0,93271.

Figura 10. a) Diagrama cajas y bigotes clasificación de eventos VT y LP utilizando el conjunto de características completo. b) curva ROC para el clasificador LDC.



- **Experimento 2:** Medidas de desempeño para la clasificación de eventos sismo-volcánicos VT - IC.

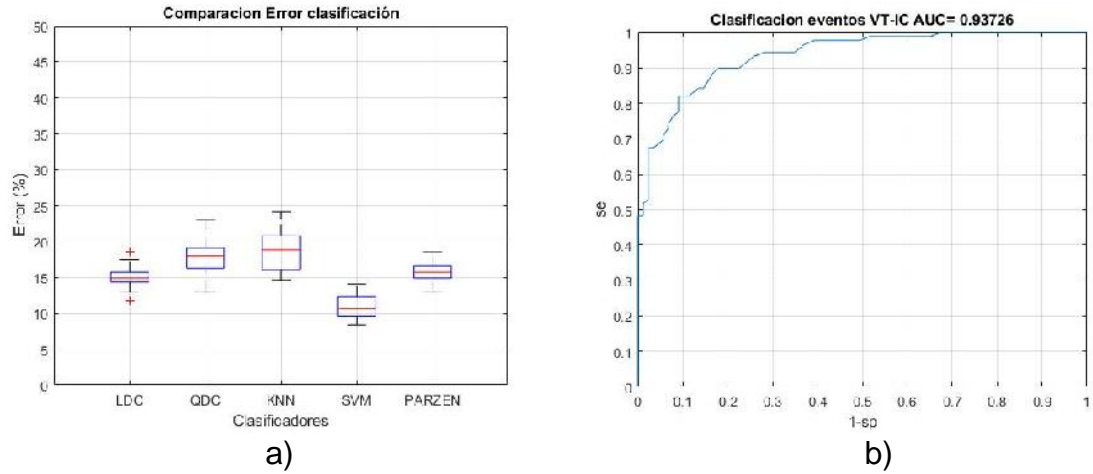
Tabla 6. Medidas de desempeño *Se*, *Sp* y *Exactitud* de clasificación de eventos VT-IC.

SENSIBILIDAD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,92	0,87	0,92	0,90	0,92
Frecuencia	0,75	0,69	0,69	0,73	0,69
Tiempo-frecuencia	0,66	0,63	0,63	0,75	0,76
MFCC	0,67	0,66	0,64	0,69	0,69
Matriz completa	0,87	0,82	0,94	0,74	0,82
ESPECIFICIDAD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,79	0,76	0,81	0,80	0,81
Frecuencia	0,70	0,80	0,80	0,64	0,80
Tiempo-frecuencia	0,72	0,71	0,80	0,71	0,67
MFCC	0,63	0,67	0,67	0,66	0,66
Matriz completa	0,74	0,76	0,65	0,78	0,76
EXACTITUD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,85	0,81	0,87	0,85	0,87
Frecuencia	0,72	0,74	0,74	0,69	0,74
Tiempo-frecuencia	0,69	0,67	0,71	0,73	0,72
MFCC	0,65	0,67	0,66	0,67	0,67
Matriz completa	0,80	0,79	0,80	0,76	0,79

Para la clasificación de eventos VT – IC, de acuerdo al análisis de las medidas de desempeño se encontró que la mejor clasificación está dada por la representación de características temporales, que alcanzo una Sensibilidad mayor a 0,9 para todos los clasificadores excepto C2 con 0,87, y una Especificidad en promedio de 0,79 y una exactitud de clasificación de 0,87 para los clasificadores C3 y C5. Se concluyó que los eventos VT e IC son clasificados de forma adecuada por la representación planteada de características de tiempo.

Se muestra en la figura 11 a), el diagrama de cajas y bigotes del error para cada clasificador en las 20 iteraciones. En la figura 11 b) se muestra la curva ROC para el clasificador PARZEN con un AUC = 0,932726.

Figura 11. a) Diagrama cajas y bigotes clasificación de eventos VT - IC utilizando el conjunto de características de tiempo. b) curva ROC para el clasificador PARZEN.



➤ **Experimento 3:** Medidas de desempeño para la clasificación de eventos sismo-volcánicos VT - TR.

Tabla 7. Medidas de desempeño Se , Sp y $Exactitud$ de clasificación de eventos VT-TR.

SENSIBILIDAD		CLASIFICADORES				
CARACTERÍSTICAS		C1	C2	C3	C4	C5
Tiempo		0,89	0,80	0,92	0,89	0,89
Frecuencia		0,82	0,81	0,87	0,85	0,82
Tiempo-frecuencia		0,83	0,80	0,88	0,82	0,81
MFCC		0,79	0,71	0,71	0,80	0,83
Matriz completa		0,93	0,88	0,96	0,90	0,96
ESPECIFICIDAD		CLASIFICADORES				
CARACTERÍSTICAS		C1	C2	C3	C4	C5
Tiempo		0,87	0,85	0,79	0,87	0,87
Frecuencia		0,70	0,74	0,73	0,81	0,70
Tiempo-frecuencia		0,88	0,80	0,76	0,84	0,78
MFCC		0,60	0,66	0,66	0,72	0,60
Matriz completa		0,82	0,78	0,79	0,87	0,79
EXACTITUD		CLASIFICADORES				
CARACTERÍSTICAS		C1	C2	C3	C4	C5

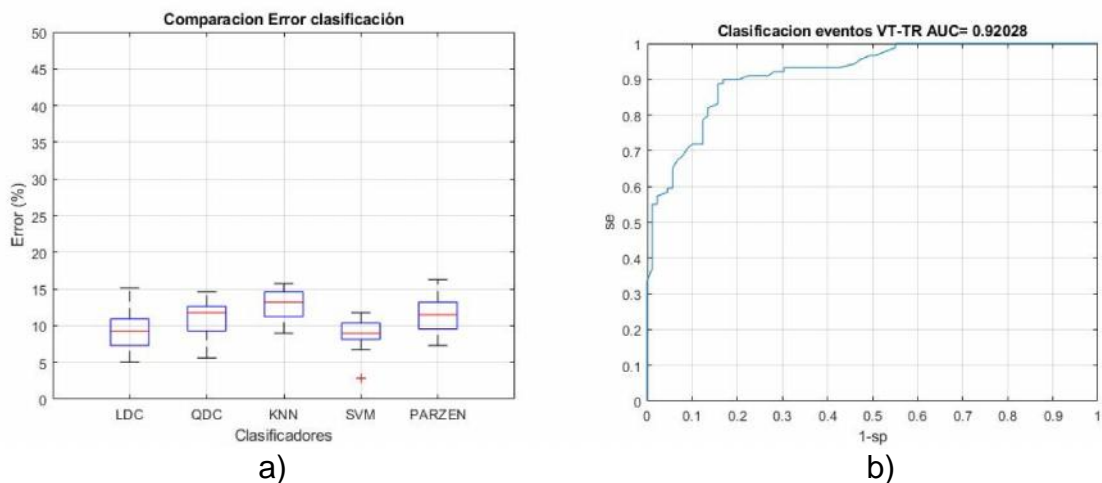
Tiempo	0,88	0,83	0,85	0,88	0,88
Frecuencia	0,76	0,78	0,80	0,83	0,76
Tiempo-frecuencia	0,85	0,80	0,82	0,83	0,79
MFCC	0,69	0,69	0,69	0,76	0,71
Matriz completa	0,88	0,83	0,87	0,88	0,87

En el experimento 3 se encontró que para la medida de Sensibilidad alcanzó valores considerablemente altos utilizando la matriz de tiempo y la matriz completa de características, igualmente para la especificidad y la exactitud, cabe destacar que los resultados fueron buenos en todas las pruebas de clasificación, exceptuando al usar el conjunto de características de MFCC.

De acuerdo al análisis de la tabla 7 las clases de eventos VT- TR bien se pueden clasificar empleando las características de tiempo o la concatenación de todos los grupos. Se logró una exactitud de clasificación de 0.88 utilizando LDC, SVM y PARZEN para la matriz completa de características y el mismo porcentaje para los clasificadores LDC y SVM, con la matriz de características de tiempo. Sin embargo, por coste computacional, elegir las características de tiempo son la mejor opción.

Se muestra en la figura 12 a), el diagrama de cajas y bigotes del error para cada clasificador en las 20 iteraciones. Se verifica que la mejor clasificación se alcanzó utilizando el conjunto de características de tiempo con el clasificador SVM. En b) se muestra la curva ROC para el clasificador SVM con un AUC = 0,92028.

Figura 12. a) Diagrama cajas y bigotes clasificación de eventos VT - TR utilizando el conjunto de características de tiempo. b) curva ROC para el clasificador SVM.



- **Experimento 4:** Medidas de desempeño para la clasificación de eventos sismo- volcánicos LP – IC

Tabla 8. Medidas de desempeño Se , Sp y $Exactitud$ de clasificación de eventos LP-IC.

SENSIBILIDAD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,82	0,82	0,82	0,79	0,83
Frecuencia	0,72	0,71	0,72	0,62	0,71
Tiempo-frecuencia	0,73	0,73	0,64	0,56	0,69
MFCC	0,62	0,65	0,62	0,47	0,57
Matriz completa	0,75	0,75	0,83	0,81	0,82
ESPECIFICIDAD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,71	0,66	0,66	0,75	0,73
Frecuencia	0,48	0,57	0,48	0,47	0,49
Tiempo-frecuencia	0,75	0,70	0,72	0,92	0,75
MFCC	0,61	0,62	0,61	0,82	0,64
Matriz completa	0,81	0,81	0,89	0,79	0,79
EXACTITUD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,76	0,74	0,74	0,77	0,78
Frecuencia	0,60	0,64	0,60	0,54	0,60
Tiempo-frecuencia	0,74	0,71	0,68	0,74	0,72
MFCC	0,61	0,63	0,61	0,65	0,61
Matriz completa	0,78	0,78	0,76	0,80	0,80

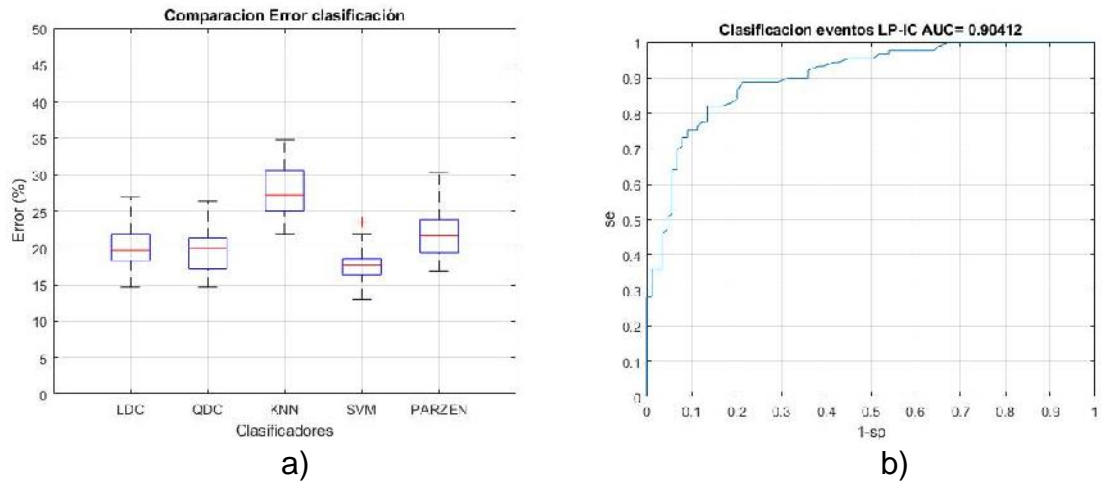
Esta clasificación mostró una mayor Sensibilidad con las características de tiempo y con la matriz completa, en cuanto a la especificidad se logró un valor de 0,92 para las características de tiempo frecuencia usando el clasificador SVM, sin embargo no compensó con la sensibilidad que está en 0,56 para la misma clasificación.

Analizando la tabla 8, se concluye que estas señales fueron mejor representadas por las características de la matriz completa, alcanzando una exactitud de 0.8 para los clasificadores SVM y PARZEN.

A continuación, se muestra en la figura 13 a), el diagrama de cajas y bigotes del error para cada clasificador en las 20 iteraciones. Se evidenció que la mejor clasificación fue alcanzada utilizando el conjunto de características completo con el

clasificador SVM. En b) se muestra la curva ROC para el clasificador SVM con un AUC = 0,93271.

Figura 13. a) Diagrama cajas y bigotes clasificación de eventos LP-IC utilizando la matriz de características completa. b) curva ROC para el clasificador SVM.



➤ **Experimento 5:** Medidas de desempeño para la clasificación de los eventos sismo-volcánicos LP – TR.

Tabla 9. Medidas de desempeño Se , Sp y $Exactitud$ de clasificación de eventos LP-TR.

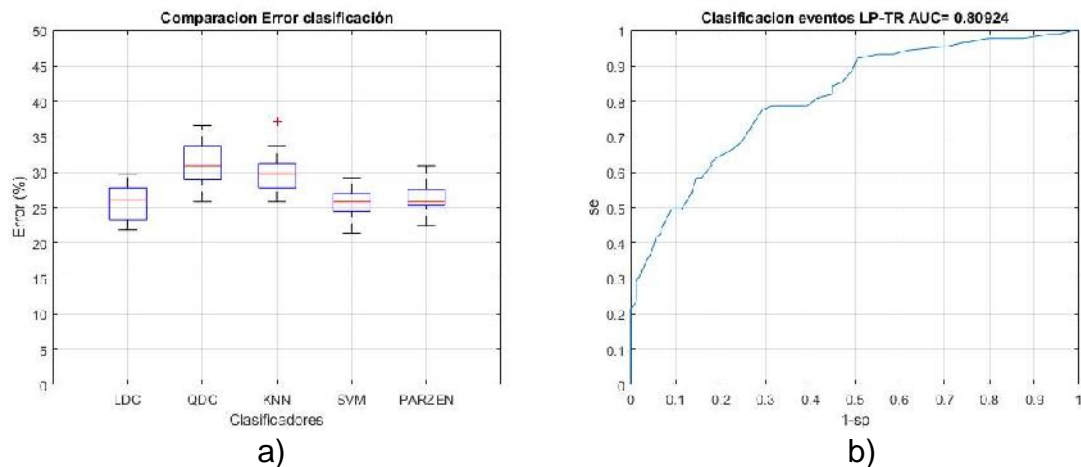
SENSIBILIDAD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,62	0,76	0,73	0,74	0,73
Frecuencia	0,53	0,49	0,55	0,48	0,49
Tiempo-frecuencia	0,47	0,76	0,53	0,64	0,53
MFCC	0,48	0,67	0,63	0,66	0,64
Matriz completa	0,51	0,51	0,46	0,60	0,60
ESPECIFICIDAD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,80	0,72	0,73	0,69	0,73
Frecuencia	0,57	0,63	0,62	0,57	0,63
Tiempo-frecuencia	0,62	0,42	0,73	0,61	0,73
MFCC	0,88	0,71	0,71	0,69	0,71
Matriz completa	0,87	0,87	0,89	0,84	0,84
EXACTITUD	CLASIFICADORES				

CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,71	0,74	0,73	0,71	0,73
Frecuencia	0,55	0,56	0,58	0,53	0,56
Tiempo-frecuencia	0,54	0,59	0,63	0,62	0,63
MFCC	0,68	0,69	0,67	0,67	0,67
Matriz completa	0,69	0,69	0,67	0,72	0,72

En el experimento 5 se observó que los valores más altos de Sensibilidad, corresponde a la clasificación con el grupo de características de tiempo. En cuanto a la especificidad las medidas más altas se obtuvieron cuando se clasificó utilizando la matriz completa de características, pero no se compensan con la sensibilidad de la misma clasificación.

La exactitud fue relativamente baja en comparación con los anteriores experimentos con el valor más alto en 0,74 para el clasificador QDC Y 0,73 para PARZEN utilizando las características de tiempo. De acuerdo a lo anterior la mejor clasificación para estos dos tipos de clase se dio con las características de tiempo. Se muestra en la figura 14 a), el diagrama de cajas y bigotes del error para cada clasificador en las 20 iteraciones. Se verificó que la mejor clasificación se alcanzó utilizando el conjunto de características de tiempo con el clasificador PARZEN. En b) se muestra la curva ROC para el clasificador PARZEN con un AUC = 0,80924.

Figura 14. a) Diagrama cajas y bigotes clasificación de eventos LP-TR utilizando el grupo de características de tiempo. b) curva ROC para el clasificador PARZEN.



- **Experimento 6:** Medidas de desempeño para la clasificación de eventos sismo volcánicos IC - TR.

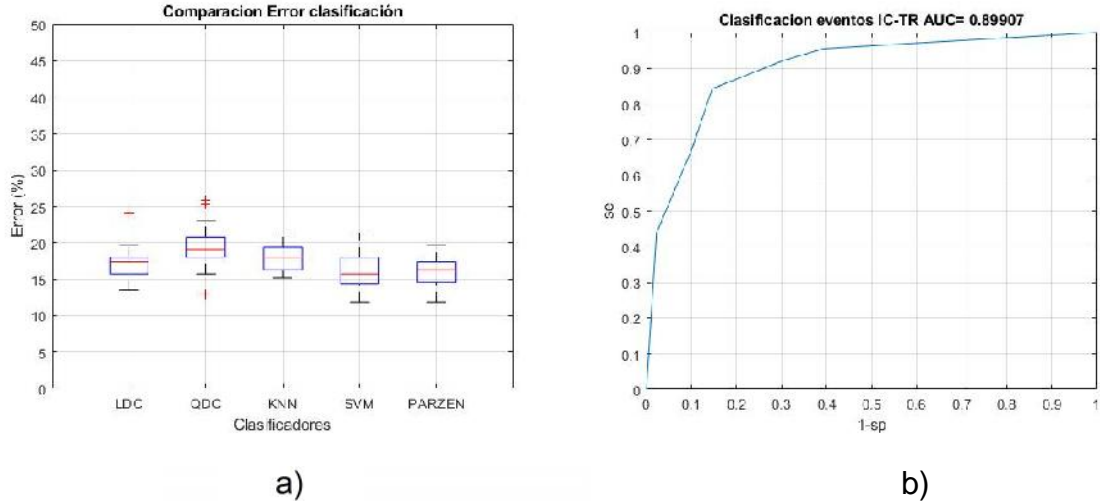
Tabla 10. Medidas de desempeño *Se*, *Sp* y *Exactitud* de clasificación de eventos IC-TR.

SENSIBILIDAD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,81	0,79	0,89	0,80	0,85
Frecuencia	0,58	0,66	0,71	0,60	0,58
Tiempo-frecuencia	0,74	0,71	0,87	0,65	0,87
MFCC	0,64	0,69	0,69	0,74	0,70
Matriz completa	0,83	0,82	0,83	0,78	0,78
ESPECIFICIDAD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,84	0,87	0,91	0,82	0,84
Frecuencia	0,62	0,60	0,58	0,63	0,62
Tiempo-frecuencia	0,70	0,74	0,65	0,79	0,65
MFCC	0,75	0,79	0,80	0,80	0,83
Matriz completa	0,83	0,88	0,88	0,88	0,87
EXACTITUD	CLASIFICADORES				
CARACTERÍSTICAS	C1	C2	C3	C4	C5
Tiempo	0,83	0,83	0,83	0,83	0,85
Frecuencia	0,60	0,63	0,65	0,61	0,60
Tiempo-frecuencia	0,72	0,72	0,76	0,72	0,76
MFCC	0,70	0,74	0,74	0,77	0,76
Matriz completa	0,83	0,85	0,85	0,83	0,82

En el experimento 6 se encontró la Sensibilidad entre altos valores para la clasificación utilizando el conjunto de características de tiempo, y la matriz completa de características, que a su vez fueron valores altos en la especificidad en la misma clasificación. De acuerdo a esto se podría clasificar utilizando los atributos de tiempo o bien la matriz completa, como se observó en la exactitud.

Se muestra en la figura 15 a), el diagrama de cajas y bigotes del error para cada clasificador en las 20 iteraciones. Se observó que la mejor clasificación se alcanza utilizando el conjunto de características de tiempo con el clasificador PARZEN. En b) se muestra la curva ROC para el clasificador PARZEN con un AUC = 0,89907.

Figura 15. a) Diagrama cajas y bigotes clasificación de eventos IC-TR utilizando las características de tiempo, b) Curva ROC para el clasificador PARZEN.



- **Experimento 7:** En la Figura 16 se muestra los valores medios para las medidas *Se*, *Sp* y *Exactitud*, en los cuatro grupos de características, tiempo, frecuencia, tiempo-frecuencia, MFCC obtenidos en este experimento. Se utilizó como método de clasificación K-NN. Para el número de muestras de la clase de interés correctamente clasificadas, el grupo que mejor resultado obtuvo fue el de las características temporales con un promedio de 0,76. También se calculó el porcentaje de clasificación que indica una relación entre las muestras correctamente clasificadas y el número total de muestras que se introducen al clasificador, obteniendo el mejor resultado con el grupo de las características de tiempo, aunque no es una buena clasificación. En la tabla 11 se muestra la matriz de confusión y las medidas de desempeño por cada clase, igualmente para cada grupo de características, con el clasificador K-NN.

Figura 16. Grafica promedio medidas de desempeño para cada grupo de características, clasificación multiclase con KNN.

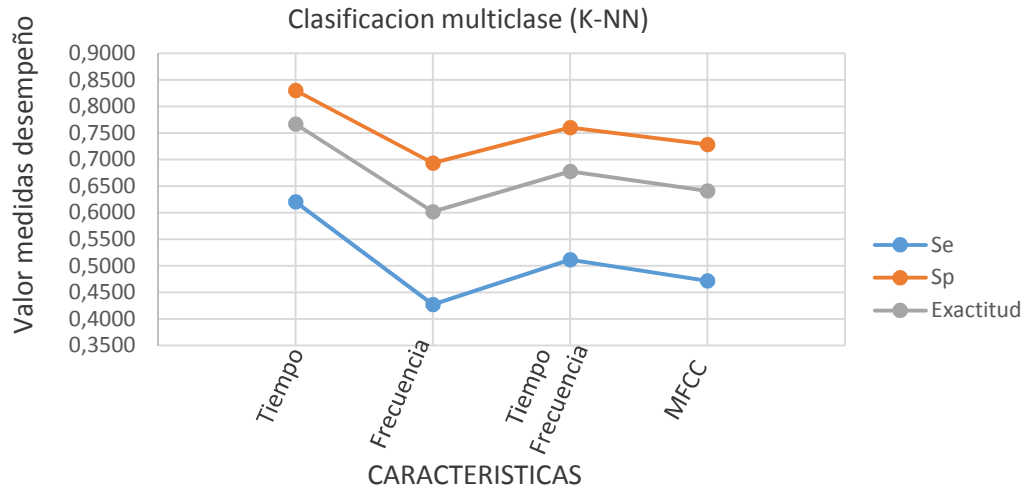


Tabla 11. Matriz de confusión y medidas de desempeño para clasificación multiclase con K- NN con diferentes grupos de características.

		PREDICCIÓN DE CLASES				MEDIDAS DE DESEMPEÑO		
		IC	LP	TR	VT	SENSIBILIDAD	ESPECIFICIDAD	EXACTITUD
i. TIEMPO	IC	70	6	10	3	0,7865	0,7947	0,7921
	LP	20	34	16	19	0,3820	0,8578	0,7199
	TR	5	17	62	5	0,6966	0,8071	0,7727
	VT	14	8	12	55	0,6180	0,8601	0,7837
ii. FRECUENCIA	IC	32	22	14	21	0,3596	0,6593	0,5609
	LP	24	34	19	12	0,3820	0,6484	0,5609
	TR	19	33	29	8	0,3258	0,7688	0,6104
	VT	19	9	4	57	0,6404	0,6985	0,6756
iii. TIEMPO-FRECUENCIA	IC	46	16	18	9	0,5169	0,7640	0,6816
	LP	16	40	17	16	0,4494	0,7282	0,6408
	TR	10	23	52	4	0,5843	0,7222	0,6766
	VT	16	14	15	44	0,4944	0,8263	0,7109
iv. MFCC	IC	40	18	16	15	0,4494	0,7485	0,6462
	LP	16	39	18	16	0,4382	0,7247	0,6292
	TR	10	16	54	9	0,6067	0,6706	0,6486
	VT	17	15	22	35	0,3933	0,7688	0,6412

El conjunto test está conformado por el 20 % del total de las señales trabajadas, que fueron 89 señales por cada clase para realizar el test del clasificador. La matriz de confusión permite analizar las medidas de desempeño por cada clase, en este experimento se observó que para los eventos IC de las representaciones planteadas la que mejor se ajustó son las características que se calcularon en el dominio temporal, con una sensibilidad de 0.78, esta representación muestra un resultado similar para los eventos TR con Sensibilidad de 0,69. De acuerdo al análisis realizado la representación que más se ajustó a los eventos sísmico volcánicos trabajados es en el espacio temporal, sin embargo no alcanzó una buena clasificación.

A continuación, se realizó la clasificación de los eventos sismo-volcánicos utilizando una matriz concatenada con los grupos de características desarrollados en este trabajo. Se muestra la matriz de confusión. Tabla 12

Tabla 12. Matriz de confusión y medidas de desempeño para clasificación multiclase con K- NN con matriz de características concatenando los grupos anteriores.

	PREDICCIÓN DE CLASES				MEDIDAS DE DESEMPEÑO		
	IC	LP	TR	VT	SENSIBILIDAD	ESPECIFICIDAD	EXACTITUD
IC	50	22	7	10	0,5618	0,8066	0,7259
LP	10	43	32	4	0,4831	0,7183	0,6490
TR	10	19	56	4	0,6292	0,7487	0,7101
VT	15	19	8	47	0,5281	0,8922	0,7656

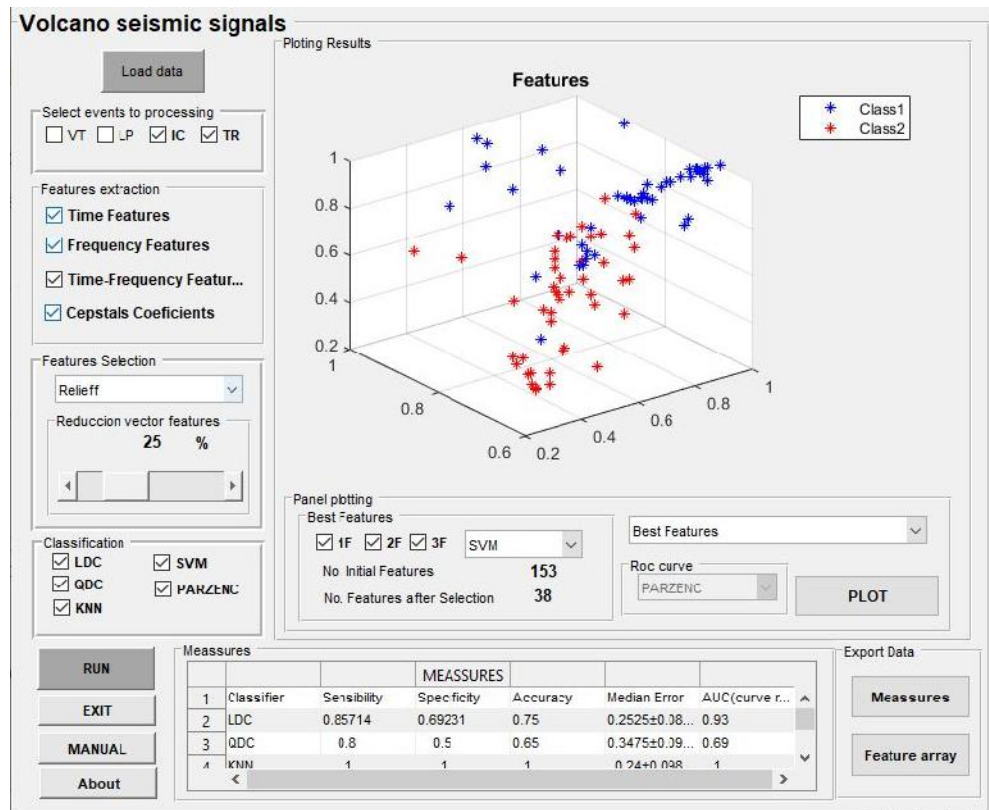
Realizando una clasificación con la matriz de características combinando los grupos antes trabajados, se obtuvo una mejor clasificación sin embargo los niveles de *Se* son bajos.

3.1. INTERFAZ DE VISUALIZACIÓN DE RESULTADOS

Se construyó una interfaz con la herramienta GUIDE de Matlab, mostrada en la figura 17, con el fin de dar a conocer el trabajo y permitir interactividad con el usuario, exponiendo métodos de caracterización y clasificación, además de las medidas de desempeño, con el propósito de que el usuario pueda hacer un estudio

exploratorio y comparativo tanto de las señales sismo-volcánicas como de los métodos de caracterización y clasificación.

Figura 17. Interfaz gráfica de clasificación de Señales sismo volcánicas.



4. CONCLUSIONES Y TRABAJO FUTURO

Las técnicas de machine learning, ofrecen soluciones automatizadas para diferentes problemáticas de la sociedad. En el entorno de la vigilancia volcánica aportan en gran magnitud a la automatización de reconocimiento de señales sismo-volcánicas. Se diseñó una metodología de clasificación supervisada de eventos sísmicos y se comprobó la complejidad de esta tarea dada la naturaleza de las señales y su variabilidad que no permite que se generalice un método para lograr su clasificación automatizada.

Se realizó la comparación de diferentes técnicas de extracción de características, que de acuerdo a los resultados la que más se ajusta a una clasificación con buen desempeño fue la matriz de características obtenida a partir de la caracterización de la señal en su estado natural escala de tiempo, se obtuvo una exactitud del 88% clasificando las clases VT y TR, con PARZEN. No obstante, la concatenación de todos los grupos de características obtenidos en una matriz, también fue buena, con una exactitud del 88% clasificando las clases de eventos VT y TR con SVM.

La etapa de selección de características fue fundamental en el modelo de clasificación de señales sismo-volcánicas, ya que permitió optimizar el proceso de clasificación y además disminuyó el coste computacional por procesar estas señales. Entre las características seleccionadas de la matriz completa, se encontró 11 características de tiempo, 8 de frecuencia, 36 de tiempo-frecuencia y 22 MFCC.

Los mejores resultados de esta investigación fueron dirigidos a la clasificación biclase de señales sismo-volcánicas, donde se obtuvo buenos porcentajes de clasificación, específicamente citamos la clasificación de los eventos VT y TR donde se alcanzó un porcentaje de clasificación cerca del 88%, y un AUC de 0,95, con el clasificador SVM, los eventos fueron caracterizados satisfactoriamente por la representación planteada de las características de Tiempo, y matriz completa. Esta clasificación se justifica ya que los eventos VT se caracterizan por su particular morfología impulsiva, con respecto a los eventos TR, que muestran una forma de onda casi constante. Además, la duración de los TR es mucho mayor de los VT.

De acuerdo a los experimentos realizados, el clasificador que mejor se comportó y logró rendimiento sobresaliente en la mayoría de experimento fue el clasificador por densidad de PARZEN. Además, cabe resaltar los clasificadores SVM y LDC que también lograron un buen porcentaje de clasificación de evento sismo-volcánicos.

TRABAJO FUTURO

Teniendo en cuenta los resultados obtenidos en la clasificación multiclase, donde se obtiene el mayor porcentaje de clasificación usando las características de tiempo, se podría orientar el estudio hacia este tipo de características con el fin de proponer una matriz más nutrida y realizar una clasificación multiclase con alto porcentaje de acierto. Además, optimizarla al unir características espectrales que logren discriminar mejor las clases.

Siguiendo la metodología planteada se puede explorar otras técnicas de extracción de características como los coeficientes de LPC (linear prediction coding), transformada wavelet, técnicas aplicadas a otro tipo de señales, además se podría trabajar con bases de datos de diferentes volcanes para discutir la generalización de una metodología de clasificación automática.

Trabajar en conjunto con un segmentador de señales sismo-volcánicas, y recrear un sistema de monitoreo volcánico completo, con el propósito de optimizar su funcionamiento y obtener las señales con punto inicial y final bien marcado, entre otras características, para que se pueda proceder a realizar un buen preprocesamiento y caracterización.

REFERENCIAS

- [1] C. Hibert et. al., «Automatic identification of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise volcano using a Random Forest algorithm», *Journal of Volcanology and Geothermal Research*, vol. 340, pp. 130-142, 2017.
- [2] M. Malfante, M. Dalla, y J.P. Metaxion, «Machine Learning for volcano-seismic Signals: Challenges and Perspectives», *IEEE Signal Processing Magazine*, vol. 35, pp. 20–30, 2018.
- [3] G. Thompson, «Seismic Monitoring of volcanoes», *Encyclopedia of Earthquake Engineering*, pp.1-25, 2015.
- [4] M. Ibs-von Seht, «Detection and identification of seismic signals recorded at Krakatau volcano (Indonesia) using artificial neural networks», *Journal of Volcanology and Geothermal Research*, vol. 176, pp. 448-456, 2008.
- [5] M. Curilem, J. Vergara, C. San Martin, G. Fuentealba, C. Cardona, F. Huenupan, M. Chacon, M. Salman Khan, W. Hussein, N. Becerra, «Pattern recognition applied to seismic signal of de Llaima volcano (Chile), An analysis of the events features», *Journal of Volcanology and Geothermal Research*, vol. 282, pp. 137-147, 2014.
- [6] S. Falsaperla, S. Graziani, G. Nunnari, S. Spampinato, «Automatic classification of volcanic Earthquakes by using Multi-Layered neural networks», *Natural Hazards*, vol. 13, pp. 205-228, 1996.
- [7] M. C. Benítez, J. Ramírez, J. C. Segura, J. M. Ibáñez, J. Almendros, A.i Garcías y G. Cortés, « Continuous HMM-Based Seismic-Event Classification at Deception Island, Antarctica», *IEEE Trans. Inf. Geo. R. Sens.*, vol. 45, n.º1, pp 138-146, 2007.
- [8] Instituto geográfico Nacional, «Técnicas de vigilancia volcánica», [En línea]. Disponible en: http://www.ign.es/resources/docs/IGNCnig/TecnicasDeVigilancia_v8.pdf.
- [9] S. R. McNutt, «Seismic 11 of Volcanoes: A review of the state-of-the-art and case histories"», pp. 1095-1129, 1996.
- [10] R.S.J. Spark, J. Bigg, J. Neuberg, «Monitoring Volcanoes», *Science*, vol. 335, pp. 1310-1311,2012.
- [11] «Erupción Volcánica», DESASTRENATURALPEDIA, [En línea]. Disponible en: <http://www.desastrenaturalpedia.com>

- [12] T. Minakami, «Seismology of Volcanes in Japan», Devp. in Solid Earth Geophys., vol. 6, pp.1-27, 1974.
- [13] P. L. Ward, T. Matumoto, «A summary of volcanic and seismic activity in Katmai National Monument, Alaska», vol. 31, pp. 107-129, 1967.
- [14] D. Shimozuru, « Observation of Volcanic Eruption by an Infrared Radiation Meter», Nat. Int. Jour. of sci., vol. 234, pp. 457-459, 1971.
- [15] B. Chouet, « Long-Period Volcano Seismicity: Its Source and Use in Eruption Forecasting», Nat. Int. Jour. of sci., vol. 380, pp. 309-316, 1996.
- [16] R. Tilling, « Volcanic hazards and their mitigation: Progress and problems», Rev. Geophys., vol. 27, pp. 237-269, 1989.
- [17] S. C. Loughlin, S. S. Sparks, K. Brown, S. F. Jenkins, C. Brown, «Global Volcanic Hazards and Risk», Chapter 15, pp. 299-304, 2015.
- [18] R. Tilling, «Apuntes para un Curso Breve sobre Los Peligros Volcánicos», Chapter 15, pp. 299-304, 1993.
- [19] J. Havskov, L. Ottemoller, «Routine Data Processing in Earthquake Seismology», Springer Dordrecht Heidelberg, New York, p. 7, 2010.
- [20] R. V. Allen, «Automatic earthquake recognition and timing from single traces», Bull. of the Seis. Soc. of America, vol. 68, pp.1521-1522., 1978», Bull. of the Seis. Soc. of America, vol. 72, n.º6, pp.225-242., 1982.
- [21] N. Langet, A. Maggi, A. Michelini, F. Brenguier, « Continuous Kurtosis-Based Migration for Seismic Event Detection and Location, with Application to Piton de la Fournaise Volcano, La Réunion», Bull. of the Seis. Soc. of America, vol. 72, n.º1 pp.229-246., 2014.
- [22] R. Lara-cueva, A. Moreno, J. Larco, D. Benitez, « Real-Time Seismic Event Detection Using Voice Activity Detection Techniques», IEEE Geoscience & Remote Sensing Society, vol. 01 pp. 5533 - 5542., 2016.
- [23] J. Gomberg, « The failure of earthquake failure models», JGR Solid Earth, vol. 106, pp. 16253-16263, 2001.
- [24] J. A. Power, S. C. Moran, S. R. McNutt, S. D. Stihler, J. J. Sanchez, « Seismic Response of the Katmai Volcanoes to the 6 December 1999 Magnitude 7.0 Karluk Lake Earthquake, Alaska », Bull. of the Seis. Soc. of America, vol. 91 n.º.1, pp. 57 - 63, 2001.

- [25] S. R. McNutt, « Volcanic Seismology», Annu. Rev. Earth Planet. Sci, vol. 33, n°. 1, pp. 461-491, 2005.
- [26] B. Chouet, R. Y. Koyanagi, K. Aki, « Origin of volcanic tremor in Hawaii, Part II: Theory and discussion », Volcanism in Hawaii, vol. 1, Chapter: 45, pp.1259-1280, 1987.
- [27] C. Weave, S. Malone, « Volcanic earthquakes or glacial noises», Geophysical Research Letters, vol. 3 n°.3, 1976.
- [28] A. Oppenheim, A. Willsky, S H. Nicanawab, « Señales y sistemas», Geophysical Research Letters, 2ª ed, 1976. Naucalpan de Juarez, Prentice Hall Hispanoamericana S.A, pp. 941, 1998.
- [29] F. I. Echeverría, «Evaluación de parámetros dinámicos en estructuras en movimiento y registros sísmicos utilizando la transformada wavelet». Trabajo de grado, pp. 135, 2010.
- [30] «Transformada de Fourier de Tiempo Reducido», [En línea]. Disponible en: https://es.wikipedia.org/wiki/Transformada_de_Fourier_de_Tiempo_Reducido.
- [31] G. Cortés, «Reconocimiento de Señales sismo-volcánicas mediante canales específicos basado en Modelos Ocultos de Markov». Tesis Doctoral, pp. 319,2015.
- [32] I. Alvares, G. Cortés, «Improving feature extraction in the automatic classification of seismic events. Application to colima and arenal volcanoes», IEEE International Geoscience and Remote Sensing Symposium, 2009.
- [33] P. A. Castro, M. Orozco, A. Adami, M. Bicego, J. Londoño, German Castellanos, « A Comparison between Time-Frequency and Cepstral Feature Representations for the Classification of Seismic-Volcanic Signals». Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 19th Iberoamerican Congress, pp.440-447.2014.
- [34] «MFCC», [En línea]. Disponible en: <https://es.wikipedia.org/wiki/MFCC>.
- [35] A. Moreno, E. Armengol, J. Béjar, L. Belanche, U. Cortés, R. Gavaldà, J. Gimeno, B. López, M. Martín, M. Sánchez, « Aprendizaje automático». Barcelona, Edición UPC, pp. 321, 1994.

- [36] C. Bishop, «Pattern Recognition and Machine Learning». Nueva York, 6 Springer Science+ Business Media, LLC, pp. 729, 2006.
- [37] J. Ruiz, J. Carrasco, J. Martínez, «I Reconocimiento de patrones: conceptos y metodología», Reporte Técnico Reconocimiento de patrones, CENATAV, 2009.
- [38] A. Blum, P. Langley, « Selection of relevant features and examples in machine learning», *Artificial Intelligence*, vol. 97, pp. 245-271, 1997.
- [39] S. Ruiz, «El algoritmo K-NN y su importancia en el modelado de datos», *AD TECH & ANALYTICS*, 2017, [En línea]. Disponible en: <https://www.analiticaweb.es/algoritmo-knn-modelado-datos/>.
- [40] M. Peixeiro, «Classification Linear Discriminant Analysis: An explanation of Bayes' theorem and linear discriminant analysis», *Towards Data Science*, 2018, [En línea]. Disponible en: <https://towardsdatascience.com/classification-part-2-linear-discriminant-analysis-ea60c45b9ee5>
- [41] C. A. González, «SVM: Máquinas de Vectores Soporte», p. 52.
- [42] L. Breiman, «Random Forests», *Mach. Learn.*, vol. 45, pp. 5-32, oct. 2001.
- [43] J. M. M. Diazaraque, «Introducción a las Redes Neuronales Aplicadas», p. 95.
- [44] H. He, E. García, « Learning from Imbalanced Data », *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263-1284, 2009.
- [45] M. Zmeskal, A. Plesinge, « Simulation of standard seismograms, in PREPROC-Software for Automated Preprocessing of Digital Seismic Data », vol. 2, 1995, Ed. E. Bergman, NEIC Golden/GI Prag, 1995.
- [46] P. Leseage, « Interactive Matlab software for the analysis of seismic volcanic signals», *Computers & Geosciences*, vol. 35, n. ° 10, pp.2137-2144, 2009.
- [47] M. Bicego, C. Acosta, M. Orozco, « Classification of Seismic Volcanic Signals Using Hidden-Markov-Model-Based», *Ieee Transactions on Geoscience and Remote Sensing*, vol. 51, n°. 6, pp. 3400-3408, 2013.
- [48] R. Avesani, A. Azzoni, M. Bicego, M. Orozco, « Automatic Classification of Volcanic Earthquakes in HMM-Induced Vector Spaces», *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp 640-647, 2012.
- [49] Subba Rao, T. (2011). Classification, parameter estimation and state estimation an engineering approach using MATLAB. *Journal of Time Series Analysis*, 32(2), 194-194.]

- [50] McDaid, A. F., Greene, D., & Hurley, N. (2011). Normalized mutual information to evaluate overlapping community finding algorithms. *arXiv preprint arXiv:1110.2515*.
- [51] C. Hibert, A. Mangeney, G. Grandjean. Provost, « Automated identification, location, and volume estimation of rockfalls at Piton de la Fournaise volcano», *Journal of Volcanology and Geothermal Research*, vol. 119, pp. 1082-1105, 2014.
- [52] G. Curilem, J. Vergara, G. Fuentealba, G. Acuña, Max Chacón, « Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms », *Journal of Volcanology and Geothermal Research*, vol. 180, pp 1-8, 2009.
- [53] R. Kumar, A. Indrayan, « Receiver Operating Characteristic (ROC) Curve for Medical Researchers», From the Department of Biostatistics and Medical Informatics, University College of Medical Sciences, Vol. 48, 2011.
- [54] «curva ROC». [En línea] Disponible en https://es.wikipedia.org/wiki/Curva_ROC
- [55] M. Robnik, I. Kononenko, « Theoretical and Empirical Analysis of ReliefF and RReliefF», *Machine learning journal*, Vol. 53, pp. 23-69, 2003.
- [56] S. Roa, H. Loiza, «Evaluación de técnicas para el análisis de relevancia basadas en filtros sobre imágenes radiológicas», *Revista Ingeniería Biomédica*, Vol. 5, pp. 26-34, 2011.
- [57] T. Hastie, R. Tibshirani, J. Friedman, «The elements for statistical learning», Springer Science, pp. 106-112, 2009.

ANEXOS

ANEXO 1. Página web

Dentro del desarrollo de este proyecto, uno de los productos esperados era el diseño una página web en la que se pueda subir información relacionada a la investigación, como interfaz, algoritmos, ejecutables, datos y otros productos adicionales como los artículos y videos. La página web fue creada en Google sites y se puede acceder mediante el siguiente vínculo:

https://sites.google.com/site/degreethesisdiegopeluffo/volcanic_signal_car

Characterization of volcanic signals for the supervised classification of seismic events

Edison Gerardo Rozero Narvaez y Angela Patricia Cabrera Riobamba Universidad de Nariño, San Juan de Pasto 2019.

The monitoring of a volcanic system, through seismic events generated from the activity of a volcano, commonly, requires detection and recording of signals, carried out by seismic stations that usually have excellent instrumentation and telemetry tools that allow detecting earthquakes vulcanological and in turn achieve a correct transmission to the observatory, where the signals are processed, characterized and classified each event separately. An active volcano can present hundreds of seismic events a day, often in swarms, groups in time and space of similar earthquakes, where each event represents a different volcanic activity. The data obtained from the volcanic seismicity are of large volume, but its processing is not so agile, often these steps are carried out by human operators, who helped by their experience look for changes in the speed, wave form, duration, dominant frequencies among others, to classify the seismic events, and to make decisions about forecasts and early warning of risk with respect to the volcanic activity that describes each type of earthquake, according to this, the analysis of the information obtained from the volcano is very important. For this is necessary implement computerized automatic methods to independently or assist process the seismic signals of a volcano, with the purpose of reducing the work of scientists in observatories and minimizing the error in reading data.

