

**ESTUDIO COMPARATIVO DE ALGORITMOS PARA LA
SEGMENTACIÓN DE SEÑALES VOLCÁNICAS ORIENTADO A LA
CLASIFICACIÓN DE SISMOS**



YOINER JESÚS ERAZO BRAVO

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

**ESTUDIO COMPARATIVO DE ALGORITMOS PARA LA
SEGMENTACIÓN DE SEÑALES VOLCÁNICAS ORIENTADO A LA
CLASIFICACIÓN DE SISMOS**

YOINER JESÚS ERAZO BRAVO

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

ASESOR

**PhD. DIEGO HERNÁN PELUFFO ORDÓÑEZ
INGENIERO ELECTRÓNICO**

ASESORES EXTERNOS

**PAOLA ALEXANDRA CASTRO
ESTUDIANTE DE 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, 26/04/19

DEDICATORIA

“A la memoria de mi padre Martín Erazo y mi abuelito Alirio Erazo quienes han sido mi fuente de inspiración y el motor que me impulsa a ser una gran persona.

A mi abuelita, Blanca Solarte, no hay palabras suficientes, todo mi agradecimiento, por estar ahí en todo momento, por tus consejos, tus valores y motivación para ser una persona de bien, por el ejemplo, la responsabilidad, por el valor y la tenacidad para salir adelante. Por ser el pilar fundamental en todo lo que soy, por eso que lo es todo, pero más que nada por tu amor.

A mi madre, Luz Dary Bravo que siempre fue un apoyo en los momentos buenos y otros no tan buenos de mi vida.

Para ti Adriana, por ser parte de mi hoy y mi futuro, por nuestro hijo Martín, motivación para seguir triunfando juntos.

Finalmente, a toda mi familia que de una u otra manera me han motivado para que día tras día insista y pueda alcanzar esta meta.”

Yoiner Jesús Erazo Bravo

AGRADECIMIENTOS

“Agradezco a la Universidad de Nariño, me siento muy orgulloso de haber sido parte de esta alma mater; aprovechando la idoneidad, experiencia y competencia de tantos maestros que impulsaron nuestra formación.

De manera muy especial al Observatorio Vulcanológico y Sismológico de Manizales (OVSM) por haber facilitado un valioso insumo, como lo son los registros de las señales que se emplearon en esta investigación.

A usted, mi asesor, profesor Diego Hernán Peluffo Ordoñez por su invaluable apoyo y valiosa asesoría para plasmar nuestras ideas en este proyecto de y culminar satisfactoriamente.

A todos mis compañeros por lo vivido y compartido, por esos increíbles y buenos momentos, por eso tan valioso que es nuestra amistad y compañerismo.”

¡Gracias a todos ustedes!

Yoiner Jesús Erazo Bravo

RESUMEN

En el tratamiento de señales, la segmentación es una fase previa indispensable en el caso de que posteriormente se requiera caracterizar, clasificar o agrupar señales, las cuales son fases convencionales para diversas aplicaciones. La segmentación se define como el proceso de dividir una señal en sus partes características principales, y este proceso se realiza, tradicionalmente, en conjunto con un proceso denominado filtrado de la señal. Sobre diferentes volcanes se han realizado estudios sismológicos, los cuales han permitido registrar y observar una variedad de señales volcánicas. El procesamiento y análisis apropiado de estas señales permite la extracción de conocimiento sobre un evento sísmico-volcánico. A menudo sucede que una observación (una señal) contiene varios impulsos (descargas) relacionados con un evento sísmico. Para un análisis más detallado varios impulsos en una señal, el análisis conjunto no es factible y, por tanto, deben aislarse y cada choque debe analizarse por separado. Para conseguir dicha separación, se realizó el proceso llamado segmentación.

Se han desarrollado diversos algoritmos relativamente efectivos para estaciones vulcanológicas, en particular. No obstante, la identificación de un algoritmo para segmentación de señales sísmico-volcánicas que implique una máxima efectividad con bajo costo computacional implementable en todas o gran mayoría de estaciones vulcanológica aún se considera un problema abierto.

En este trabajo se estudió en implemento dos algoritmos de segmentación (uno en el tiempo y otro de tipo espectral) con el fin de determinar la técnica con mayor eficiencia, precisión y menor costo computacional frente a este tipo de señales, de forma que pueda darse un paso hacia delante en el reto de lograr una generalización de un sistema automático orientado a la segmentación de señales sísmico-volcánicas a tiempo real.

ABSTRACT

When dealing with signal processing problems, segmentation is an indispensable preliminary stage in case of characterization, classification and/or clustering are subsequently required -which are conventional phases for several applications. Segmentation is defined as the process of dividing a signal into its main characteristic parts, and it is traditionally performed in conjunction with a process called signal filtering. Different seismological studies have been carried out on different volcanoes, which have allowed registering and observing a variety of volcanic signals. The appropriate processing and analysis of these signals allows the extraction of knowledge about an earthquake-volcanic event. It often happens that an observation (a signal) contains several impulses (discharges) related to a seismic event. For a more detailed analysis of several pulses in a signal, the joint analysis is not feasible and, therefore, each shock must be isolated and separately analyzed. To achieve this separation, the process called segmentation is performed.

Several relatively effective algorithms have been developed for specific volcanic stations. Nonetheless, the identification of an algorithm for segmentation of earthquake-volcanic signals that implies a maximum effectiveness with low computational cost that can be implemented in all or most volcanological stations is still considered an open issue.

In this work, we propose to study segmentation algorithms (one in time-domain and another of spectral type) in order to determine the techniques with greater efficiency, precision and lower computational cost compared to this type of signals, so that a step can be taken forward in the challenge of achieving a generalization of an automatic system aimed at the segmentation of earthquake-volcanic signals in real time.

CONTENIDO

Introducción	16
Planteamiento del problema	17
Justificación.....	18
Contribuciones de esta tesis.....	19
Organización del documento	19
Objetivo general	20
Objetivos específicos.....	20
Marco teórico	21
Sismografía	21
Sismisidad volcánica	23
Tipos de señales sísmico-volcánicas.....	23
Volcano tectónicas.....	23
Largo periodo	24
Tremor	25
Explosiones	26
Híbridos.....	27
Estado del arte sobre algoritmos de segmentación de señales sísmicas ..	28
Eétodos de segmentación de señales sísmicas.....	28
Métodos de detección de ondas p y/o s.....	32
Resumen de los métodos	34
Algoritmos implementados	36
Algoritmo 1: sta - Ita.....	37
Algoritmo 2: descomposición tiempo-frecuencia con modelado estadístico.....	39
Caracterización.....	41
Selección de características.....	42
Clasificación	42
Clasificación supervisada	42

1. DESARROLLO DE LA INVESTIGACIÓN	43
1.1 BASE DE DATOS	44
1.2 PRE-PROCESAMIENTO.....	45
1.3 SEGMENTACIÓN	45
1.3.1. Algoritmo 1	47
1.3.2. Algoritmo 2	49
1.4 CARACTERIZACIÓN	50
1.5 SELECCIÓN DE CARACTERÍSTICAS.....	50
2. MARCO EXPERIMENTAL.....	51
2.1. CLASIFICACIÓN.....	51
2.2. MEDIDAS DE DESEMPEÑO	51
2.3. PRUEBAS QUE SE APLICARON SOBRE LA BASE DE DATOS.....	52
3. RESULTADOS.....	55
3.1. INTERFAZ DE VISUALIZACIÓN DE RESULTADOS.....	72
4. CONCLUSIONES Y TRABAJO FUTURO	76
BIBLIOGRAFÍA	78
ANEXOS.....	81

LISTA DE FIGURAS

Figura 1. Registro sismográfico general.....	22
Figura 2. Ejemplo de un registro volcano tectónico con su contenido espectral. ...	24
Figura 3. Ejemplo de un registro largo periodo con su contenido espectral.....	25
Figura 4. Ejemplo de un registro tremor con su contenido espectral.	26
Figura 5. Ejemplo de un registro tipo explosión con sus tres componentes.....	27
Figura 6. Ejemplo de un registro hibrido con su contenido espectral.	28
Figura 7 . Funciones y variables para el disparo de STA-LTA.	38
Figura 8. Diagrama de bloques funcionamiento del algoritmo 2.	41
Figura 9. Metodología propuesta.	43
Figura 10. Señal tipo VT con una envolvente de baja frecuencia.	45
Figura 11. Ejemplos de registros individuales por clase de la base de datos usada.	46
Figura 12. 10 señales concatenadas por clase VT, LP, TR, IC, SM, HB. Para señales tipo EX 4 señales.....	47
Figura 13. Segmentación del algoritmo 1, sin detección de tiempo final.....	56
Figura 14. segmentación realizada por algoritmo 1 de forma correcta.	56
Figura 15. Segmentación realizada por el algoritmo 1, detecciones falsas.....	57
Figura 16. Detección errónea (b) y detección exitosa (b) para una señal de 4 choques.	58
Figura 17. Segmentación y función de probabilidad para una señal.....	59
Figura 18. Segmentación realizada por el algoritmo 2 con filtro de suavizado aplicado a la función de probabilidad.....	60
Figura 19. Espectrograma, matriz binarizada y segmentación de la señal con $P=20$ y $K=0.8$	61
Figura 20. Cambios en la función de probabilidad según el valor de percentil. ...	61
Figura 21. Filtros: $AE= 50$ (A), $AE= 1000$ (B) y $EH>90\%$ (C), $EH>50\%$ (D), aplicados a la matriz binarizada.....	62
Figura 22. Detección errónea (a) y detección exitosa (b), señal TR con 2 choques.	63
Figura 23. Detección del algoritmo 1 con señales concatenadas y ruido excesivo.	64
Figura 24. Detección exitosa para 10 señales concatenadas tipo LP.....	65
Figura 25. Segmentación y función de probabilidad con señales TR concatenadas con $K=0.2$, $P=60$ y $Framelen=7$	66
Figura 26. Segmentación y función de probabilidad con señales TR concatenadas con $K=0.1$, $P=40$ y $Framelen=51$	67
Figura 27. Error en la clasificación con 14 características y base original con 3 clases.....	68

Figura 28. Error en la clasificación con 10 características con Relieff y la base de datos original.....	68
Figura 29. Error de clasificación con BestFirst y la base de datos original.	69
Figura 30. Error en la clasificación con 14 características con nueva base de datos 1	70
Figura 31. Error en la clasificación con BestFirst y la nueva base de datos 1	70
Figura 32. Error en la clasificación 14 características y nueva base de datos 2 ..	71
Figura 33. Error en la clasificación con BestFirst y nueva base de datos 2	72
Figura 34. Primera etapa de la interfaz, carga base de datos y selección de la clase.	73
Figura 35. Segunda etapa de la interfaz, se elige algoritmos y parámetros.....	73
Figura 36. Tercera etapa de la interfaz, número de señales y puesta en marcha.	74
Figura 37. Cuarta etapa de la interfaz, creación de bases de datos.	74
Figura 38. Quinta etapa de la interfaz, evaluación de las nuevas bases de datos (caracterización y clasificación).	75
Figura 39. Visión general de la interfaz de segmentación de señales volcánicas.	75

LISTA DE TABLAS

Tabla 1. Resumen de los algoritmos para la segmentación de señales sísmicas...	34
Tabla 2. Resumen sobre algunos algoritmos para la detección ondas s y p.....	36
Tabla 3. Número de registro por clase en la base de datos.	44
Tabla 4. Valores iniciales para las variables del algoritmo ST / LTA.	48
Tabla 5. Valores para las variables del algoritmo STA/ LTA por cada clase determinados en esta investigación.....	48
Tabla 6. Valor inicial de las variables del algoritmo 2.	49
Tabla 7. Valores usados para la generación de la nueva base de datos 2.	49
Tabla 8. Características representativas para señales volcánicas.	50
Tabla 9. Tiempos de ejecución del algoritmo 1 con señales individuales.....	55
Tabla 10. Valores utilizados en la segmentación de una señal tipo VT.....	55
Tabla 11. Valores utilizados en la segmentación de una señal tipo VT.....	57
Tabla 12. Parámetros usados a una señal tipo VT con múltiples choques.	57
Tabla 13. Parámetros usados con una señal tipo VT con múltiples choques para una identificación exitosa.	58
Tabla 14. Tiempos de ejecución promedio del algoritmo 2 con señales individuales.	59
Tabla 15. Tiempo y porcentaje de detección registrados por el algoritmo 1.....	63
Tabla 16. Parámetros para la detección de señales concatenadas tipo LP.	65
Tabla 17. Tiempo y porcentaje de detección del algoritmo 2 según intervalo.	66

ANEXOS

ANEXOS 1. Códigos(scripts) resultantes del estudio sobre señales volcánicas. ..	81
ANEXOS 2. Dos Artículos para enviar a: 6th IEEE Latin American Conference on Computational Intelligence, LA-CCI 2019.	89
ANEXOS 3. Página web	91

GLOSARIO

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

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

Clase: Agrupación de objetos que tiene características comunes.

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

Inteligencia artificial: Es un área multidisciplinaria que combina diferentes áreas como la computación y la lógica, cuyo objetivo es dar a la maquina la capacidad de resolver problemas o realizar tareas por sí mismos, utilizando algoritmos y paradigmas de comportamiento humano.

Ondas p: También llamadas primarias, son un tipo de onda sísmica, son las primeras en ser registradas por las estaciones sísmicas o volcánicas, su desplazamiento es longitudinal o compresional, lo cual significa que el suelo es alternadamente comprimido y dilatado en la dirección de la propagación.

Ondas s: También llamadas secundarias, son un tipo de onda sísmica, su desplazamiento es transversal a la dirección de propagación. Su velocidad es menor que la de las ondas primarias. Debido a ello, estas aparecen en el terreno algo después que las primarias.

Pre-procesamiento: 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.

Sismógrafo: es un dispositivo que se emplea para la medición de las ondas sísmicas que provocan los terremotos. Gracias a esta máquina que se utiliza en el ámbito de la sismología, es posible registrar la amplitud de las ondas y su propagación.

INTRODUCCIÓN

Durante las últimas tres décadas el problema de la segmentación de señales ha ganado considerable atención en el contexto de la investigación. Este proceso tiene una relación estrecha con la detección y el diagnóstico de cambio en la señal, al parecer es el tema central en varias áreas de aplicación [1]. La segmentación se define como el proceso de dividir una señal en sus partes características principales [2], este proceso se realiza en conjunto con un filtrado de la señal. Dependiendo de la aplicación, la segmentación se puede llevar a cabo de varias maneras, es decir, con el uso de varios criterios y en varios dominios [2]. Este proceso involucra tanto detectar si un cambio ha ocurrido o no, o si se han producido varios cambios, como identificar los tiempos de dichos cambios [3].

Una gran parte de la población del planeta se concentra, alrededor o sobre zonas volcánicas muy activas [4]. Esto implica un peligro contante e impredecible de una posible erupción. Por ejemplo, en Colombia, ocurrió una catástrofe el 13 noviembre de 1985, que arrasó la población de Armero (Tolima) y cobro la vida de 23000 personas [5]. Como respuesta a este evento se creó el Observatorio Vulcanológico y Sismológico de Manizales, para el monitoreo del Volcán Nevado del Ruiz (VNR), volcán responsable de este desastre natural [5]. Frente a situaciones como esta la comunidad científica dedicada al estudio de las actividades sísmico-volcánicas está en constante búsqueda de automatizar los procesos de reconocimiento de estas señales, uno de los estudios más actuales es la propuesta planteada en [6]. La idea en general es desarrollar un sistema de detección automática, segmentación, reconocimiento (clasificación) y finalmente visualización de eventos sísmicos a tiempo real que permita de cierto modo predecir o dar una alerta temprana ante un evento volcánico o de tipo tectónico como un terremoto.

Sin embargo, para lograr este objetivo es necesario el procesamiento de señales sísmicas, el primer paso en el procesamiento es la segmentación de estas señales, que es el corazón de este documento. La solución exitosa del problema de identificación de eventos sísmicos contenida a partir de datos sísmicos registrados continuamente constituye el ingrediente básico para lograr con éxito el objetivo final de selección, es decir, la estimación de los tiempos de llegada de las ondas sísmicas a las estaciones de registro. El enfoque común seguido, en las llamadas técnicas fuera de línea para resolver este problema es detectar primero la presencia de los eventos existentes y extraer segmentos del registro que contienen un evento cada uno y luego aplicar un método de selección a cada uno para estimar el tiempo de llegada correspondiente, la efectividad de este enfoque depende en gran medida de la capacidad del método de segmentación para obtener una división adecuada de la señal, que a su vez se ve muy afectada por factores incontrollables, como la magnitud y la duración de los eventos, así como su separación en el tiempo [7].

Se ha propuesto diversos algoritmos para la segmentación de señales sísmicas con enfoques en el tiempo, tiempo-frecuencia por mencionar algunos. Sin embargo, no existe un método universal que pueda aplicarse a todos los fenómenos [8]. El propósito de este trabajo fue estudiar dos algoritmos (uno en el tiempo propuesto en [9], y otro de tipo espectral propuesto en [10]) aplicados a la segmentación de señales registradas por la actividad del volcán Nevado del Ruiz en el año 2015. Se llevó estos algoritmos a un mismo lenguaje de programación y posteriormente fueron evaluados y comparados, determinó la técnica con mayor eficiencia, precisión y menor costo computacional frente a este tipo de señales, de forma que se consiguió dar un paso hacia delante en el reto de lograr una generalización de un sistema automático orientado a la segmentación de señales sísmico-volcánicas a tiempo real.

Modalidad: Trabajo de investigación

Línea de investigación: Procesamiento de señales sísmico-volcánicas.

PLANTEAMIENTO DEL PROBLEMA

De acuerdo con su comportamiento, estadísticamente, las señales pueden clasificarse como estacionarias o no estacionarias. En muchas aplicaciones, a menudo, es necesario dividir la señal no estacionaria en varias piezas, cada una de las cuales tiene aproximadamente las mismas propiedades estadísticas, como amplitud, frecuencia, entre otras. Este procedimiento se conoce como segmentación de la señal [4]. La solución al problema de la segmentación de señales sísmicas constituye una tarea muy interesante y desafiante. La principal dificultad para resolver este problema se le atribuye al hecho de que, en general, se desconocen tanto las propiedades estadísticas del ruido sísmico como también las características de los eventos registrados [3].

En la mayoría de los volcanes, se presenta actividad sísmica a diario, donde un bajo porcentaje de ellos, en la actualidad tienen un monitoreo adecuado. Algunos procesos se hacen de forma manual. Por otra parte, las señales registradas varían de volcán a volcán e inclusive dentro de un mismo sistema volcánico. Debido a esto se ha desarrollado sistemas de detección y clasificación de señales sísmicas para estaciones vulcanológicas en particular, sin embargo, no existe aún herramienta computacional para la detección y clasificación que, de forma genérica, flexible, adaptada, y versátil, funcione para cualquier estación vulcanológica con resultados adecuados.

A nivel técnico, la etapa final de las herramientas computacionales de asistencia en estaciones vulcanológicas es la clasificación, la cual generalmente requiere contar con una gran variedad de bases de datos de diferentes estaciones vulcanológicas, como también con procedimientos óptimos en las etapas previas de segmentación y caracterización de las señales sísmicas. En el caso de la segmentación, se han desarrollado diversos algoritmos relativamente efectivos; no obstante, la identificación de un algoritmo para segmentación de señales sísmico-volcánicas que implique un buen compromiso entre efectividad, coste computacional y factibilidad de implementación y uso versátil en diversas estaciones vulcanológicas, se considera aún un problema abierto.

JUSTIFICACIÓN

El tratamiento adecuado de señales sísmico-volcánicas conlleva a extraer conocimiento sobre el comportamiento de un volcán que en unión con conocimientos geofísicos y vulcanológicos es posible predecir el surgimiento de un evento, el cual puede afectar a poblaciones cercanas de un determinado volcán, gracias a esto es posible tomar algunas medidas de precaución.

Para los científicos en el área de vulcanología, reconocer el tipo de señales que se generaron por cierto evento volcánico es de gran ayuda para determinar la fuente que las generó, además de entender el sistema volcánico en estudio. La idea en general es implementar un sistema de detección y clasificación de eventos volcánicos a tiempo real. Para lograr resultados óptimos en la clasificación es necesario tener una gran cantidad y variedad de bases de datos vulcanológicas esta tarea hace uso de la segmentación por tanto identificar el algoritmo de segmentación más efectivo y con menor costo computacional conlleva a la reducción de los tiempos al realizar esta tarea. Por otra parte, al conocer el intervalo de tiempo (P-S) se puede estimar la distancia a la que se produjo el evento. De manera que, si esta detección se hace de forma más eficiente, el error en la estimación de la distancia será menor. Al implementar, analizar y evaluar algoritmos de segmentación se abre un escenario para proponer o modificar un algoritmo de segmentación de señales sísmo-volcánicas que sea óptimo con respecto a los existentes.

Actualmente existen sistemas de detección y clasificación automáticas para estaciones vulcanológicas específicas. Este proyecto busca aportar de tal manera que pueda darse un paso hacia delante en el reto de lograr una generalización de un sistema automático orientado a la segmentación de señales sísmico-volcánicas a tiempo real.

CONTRIBUCIONES DE ESTA TESIS

Se realizó la implementación de dos algoritmos de segmentación en el entorno del software MATLAB®, y se realizó diferentes pruebas principalmente con cuatro clases de señales volcánicas tales como: Hielos (IC), Tremor (TR), Largo Periodo (LP) y volcano-tectónicos (VT), de forma que se evalúan diversas maneras de preprocesar, segmentar, caracterizar, seleccionar características y clasificar señales volcánicas.

La metodología propuesta sirve de base para trazar caminos hacia donde deben enfocarse futuros estudios sobre segmentación de señales volcánicas aplicada a identificación de sismos. Además de facilitar la comparación de nuevos algoritmos de segmentación que se implementen.

En esta tesis se desarrolla una interfaz que permite visualizar la morfología y duración de las señales volcánicas. Además, de brindar la posibilidad de realizar la inspección visual uno a uno de los registros sísmico-volcánicos. Como también realizar segmentación con cada algoritmo por separado o en conjunto, variando sus parámetros, visualización de los procesos de forma gráfica, creación de nuevas bases de datos segmentadas, caracterización y clasificación.

En el campo de investigación sobre segmentación de señales volcánicas, este trabajo representa un aporte importante que puede contribuir para el desarrollo de nuevas investigaciones y el entendimiento de la naturaleza de estas señales, abre escenarios para proponer nuevos algoritmos de segmentación como también en la detección de diferentes patrones que se asocian a la actividad sísmico-volcánica.

ORGANIZACIÓN DEL DOCUMENTO

Este trabajo está dividido en **7 secciones** principales nombradas de la siguiente manera: Introducción, descripción del problema, objetivos, marco teórico, metodología, resultados y conclusiones.

En la **Sección 1**, se presenta el planteamiento del problema, la justificación de este trabajo y las contribuciones científicas de esta investigación.

En la **Sección 2**, se presenta el objetivo general y los objetivos específicos, que fueron planteados al inicio de este trabajo de grado.

En la **Sección 3**, se presenta una revisión bibliográfica donde el lector podrá encontrar la definición sobre segmentación aplicada a señales volcánicas, clases de señales volcánicas, algoritmos implementados. Además, se explica las

implicaciones de trabajar con estas señales y la importancia de esta etapa en el diseño de un sistema automático de detección y clasificación de este tipo de señales.

En la **Sección 4**, se describe la metodología desarrollada para lograr la comparación de los algoritmos implementados para la segmentación de señales volcánicas con 13 experimentos. El marco experimental se discute en la **Sección 5**.

Finalmente, después de discutir los resultados en la **Sección 6**, la **Sección 7** presenta las conclusiones que se obtuvieron a partir de este trabajo y trabajos futuros que pueden mejorar la metodología de comparación propuesta.

OBJETIVO GENERAL

Desarrollar un estudio comparativo de algoritmos de segmentación de señales volcánicas con el fin de determinar aquellos que presenten eficiencia, precisión y bajo costo computacional.

OBJETIVOS ESPECÍFICOS

- Realizar el estado del arte sobre algoritmos de segmentación de señales sísmico-volcánicas con el fin de identificar los que actualmente presentan mejor desempeño.
- Implementar los algoritmos de segmentación de señales volcánicas para evaluar su desempeño en términos de precisión y coste computacional.
- Desarrollar una metodología de comparación de los algoritmos de segmentación que permita identificar el óptimo en términos de costo computacional y efectividad para la segmentación de señales volcánicas.

MARCO TEÓRICO

El problema de la detección y el diagnóstico del cambio ha ganado considerable atención durante las últimas tres décadas en un contexto de investigación y parece ser el tema central en diversas áreas de aplicación [3].

Específicamente, en el área de sismología se ha presentado diferentes propuestas de segmentación de señales sísmicas y señales volcánicas, algunas basadas en el tiempo, frecuencia o ambas. La principal dificultad para resolver este problema se atribuye al hecho de que, en general, se desconocen tanto las propiedades estadísticas del ruido sísmico como las características de los eventos registrados. Para entrar en materia lo primero será definir sismografía, actividad volcánica y describir los tipos de señales que se han identificado en los registros de actividad sísmico-volcánica.

SISMOGRAFÍA

Es la ciencia dedicada al estudio de los terremotos y las ondas sísmicas que se mueven a través y alrededor de la Tierra. Las ondas sísmicas son las ondas de energía causada por la ruptura repentina de la roca dentro de la tierra o una explosión. Esta energía viaja a través de la tierra y se registra en los sismógrafos [11]. Existen diferentes tipos de ondas sísmicas las cuales se comportan de forma distinta, entre las principales están las ondas de compresión y superficiales. Las ondas superficiales son allegas que se detectan antes que las ondas superficiales, estas tienen un mayor ancho de banda que las de superficie. A Continuación, se define ondas de compresión conocidas.

Ondas P

También conocidas como ondas primarias, son el primer tipo de ondas de compresión. Es la más rápida de las ondas sísmicas y en consecuencia la primera en ser detectada, registrada por una estación sísmica o volcánica. La característica principal de las ondas P es que el movimiento que realizan es a través de las rocas y los fluidos, comprimen y descomprimen las partículas de la misma forma como lo hacen las ondas sonoras con el aire, con este fenómeno se define la dirección propagación de la onda [11].

Ondas S

La onda S u onda secundaria, es el segundo tipo de onda de las que se desplazan a través de la tierra es que es la segunda. La velocidad de una onda S es menor

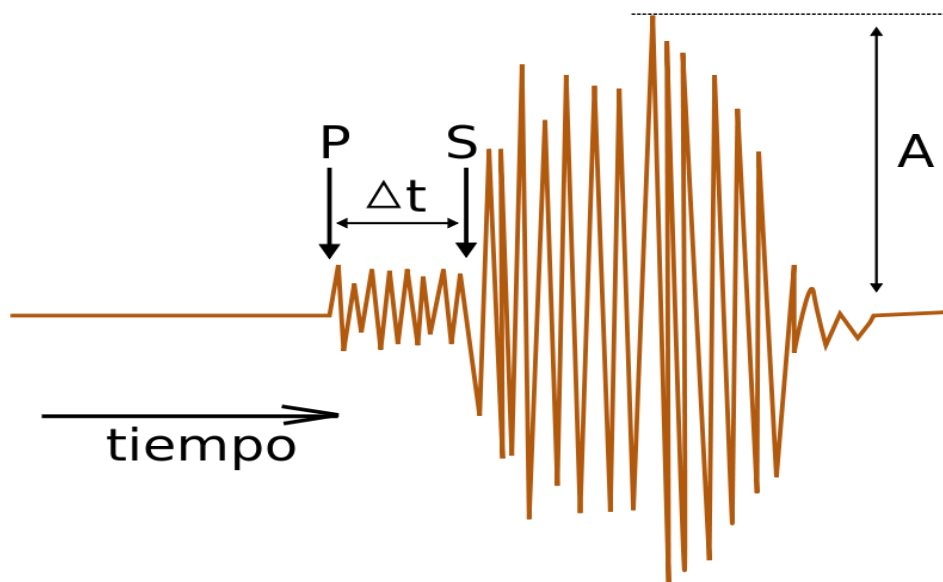
que la de una onda P. En consecuencia, el movimiento a través de cualquier medio líquido no es posible, únicamente a través de roca sólida. Gracias a esta característica los sismólogos llegaron a la conclusión de que el núcleo de la Tierra es líquido. Las ondas S logran que las partículas de roca se muevan hacia arriba y hacia abajo o de lado a lado de forma perpendicular a la dirección en que la onda viaja [11].

Ondas Superficiales

También llamadas ondas Love y Rayleigh, por los primeros investigadores que las descubrieron y analizaron. Tienen un comportamiento peculiar dado que viajan sólo a través de la corteza de la tierra y su frecuencia es menor que las ondas S y P. Debido a esto es posible distinguirlas fácilmente en un registro sismográfico, a pesar de llegar después que las ondas de S y P. La destrucción y los daños causados por terremotos están directamente asociada con este tipo de ondas [11].

Antes de analizar diferentes tipos de ondas que se va a tratar en este documento, en la Figura 1, se observa un sismograma generalizado con las respectivas ondas p, s y ondas superficiales.

Figura 1. Registro sismográfico general.



Fuente: [12]

SISMISIDAD VOLCÁNICA

Desde el punto de vista sísmico, realizar estudios sobre una región volcánica es fundamental ya que permite determinar y conocer aspectos con respecto al sistema volcánico en cuestión. Por mencionar algunos, la dinámica y los mecanismos de transporte de fluidos, así como los efectos o consecuencias que se derivan de dicha dinámica. Mediante el monitoreo de estas regiones se ha logrado determinar que en este tipo de ambientes las señales registradas son diversas entre sí, y algunas de ellas con diferencias muy grandes con respecto a las que se pueden esperar en ambientes puramente tectónicos. Con base en esto, resulta evidente que la comprensión, desde un punto de vista sísmico de una región volcánica, es una tarea complicada, pero no por ello se debe hacer un estudio superficial, sino por lo contrario aprovechar los datos al máximo en conjunto con los instrumentos de los que se tenga a nuestra disposición [13].

Para realizar estos estudios es necesario el uso de conceptos y técnicas que en un gran número de casos no son exclusivas del área de sismología Volcánica, sino que es determinante utilizar las técnicas, conceptos que usualmente son de la sismología en general aplicada sobre zonas volcánicas.

Para cualquier tipo de señal volcánica o tectónica, los instrumentos y técnicas de procesamiento y análisis tienen varios puntos en común, sin embargo, las diferencias que ha encontrado radican en el tipo de señal y principalmente en la fuente que las generan. De modo que, no se debe hablar de terremotos volcánicos o tectónicos, en su lugar de una forma más simple, únicamente de terremotos, y por llamarlo de alguna manera su “apellido” estaría relacionado directamente con su fuente de origen [13].

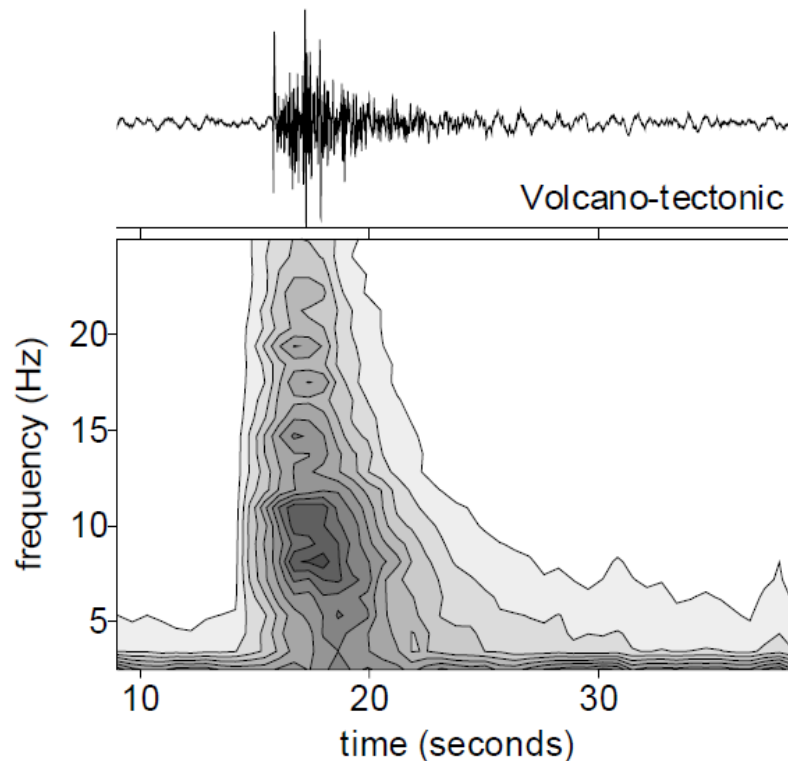
TIPOS DE SEÑALES SÍSMICO-VOLCÁNICAS

Volcano tectónicas

Son señales que ocurrieron en regiones o ambientes volcánicos, reciben el nombre de terremotos volcano-tectónico. Estos registros se caracterizan por ser señales con una duración cambiante, desde los pocos segundos para con pequeños terremotos, hasta pocos minutos para los terremotos de mayor magnitud.

El inicio de esta señal habitualmente es más o menos impulsivo. Por otra parte, en la mayoría es posible identificar la llegada de la onda S, en especial si dicho registro fue captado en sus tres componentes. En la Figura 2 se puede apreciar un ejemplo de terremoto volcano-tectónico que fue registrado en la Isla Decepción (Antártida), donde se aprecia la señal en uno de sus ejes, en conjunto con su contenido frecuencial [13].

Figura 2. Ejemplo de un registro volcano tectónico con su contenido espectral.



Fuente: [13].

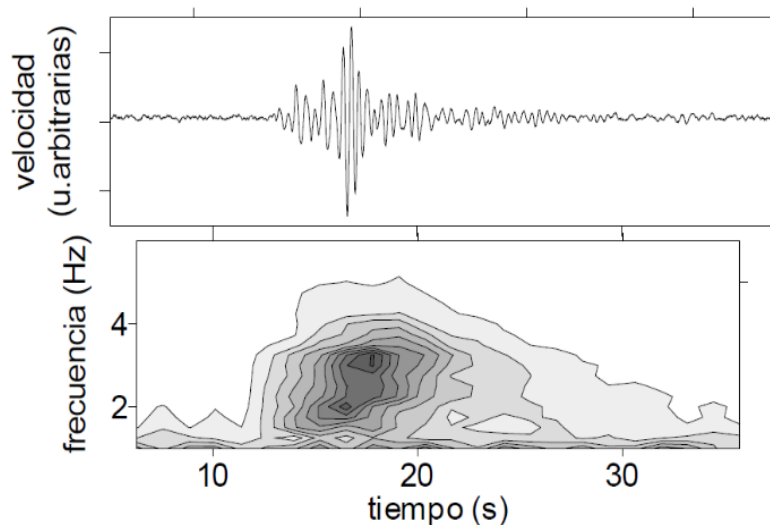
Largo periodo

Estas señales se suelen conocer como eventos de baja frecuencia o terremotos de tipo B. Estas señales son típicas de ambientes volcánicos. Entre sus características se destaca, su duración que puede ser entre los pocos segundos hasta un poco más de un minuto, su contenido espectral es de limitado con algunas bandas de frecuencia relativamente estrechas entre 0.5 Hz hasta menos de 5 Hz.

Habitualmente es posible encontrar eventos largo periodo (LP), que aparecen precedidos por llegadas con menor energía, con frecuencias superiores a los 5 Hz.

Por lo general, el contenido espectral de esta señal logra variar de volcán a volcán e inclusive en el mismo sistema volcánico [13]. En la figura 3, se aprecia un ejemplo de un registro largo periodo (LP) con su espectrograma.

Figura 3. Ejemplo de un registro largo periodo con su contenido espectral.



Fuente: [13].

Tremor

El tremor volcánico es una señal sísmica que tiene como característica principal el hecho de que mantiene la amplitud constante durante un largo período de tiempo, que puede variar entre algunos minutos hasta horas, con contenido espectral centrado en bandas de frecuencia relativamente reducidas. En la figura 4, se aprecia un ejemplo de una señal tipo tremor (TR) con su contenido espectral. Según los numerosos estudios sobre el tremor volcánico se ha establecido una serie de características comunes, por ejemplo, que frecuencias centrales de los picos dominantes son extremadamente estables y usualmente no se ven influenciadas por las características de erupción. Por otra parte, no es posible identificar la presencia de llegadas de las ondas primeras o secundarias [13]. Según el contenido espectral del tremor, se divide en tres subgrupos:

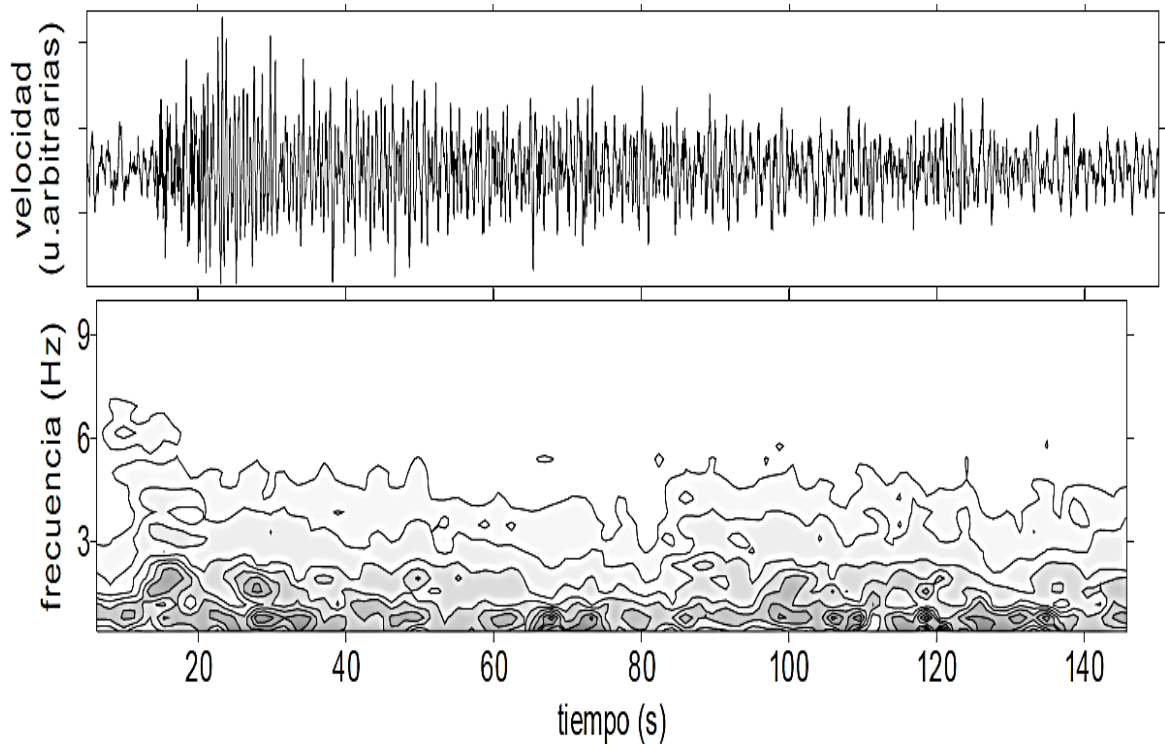
T_H: Altas frecuencias mayor a 6 Hz, suelen después de explosiones, se destaca por que dentro de su contenido espectral hay frecuencias superiores a 25 Hz.

T_I: Frecuencias intermedias entre 1 a 6 Hz. En determinados volcanes el contenido espectral de este tipo de tremor se encuentra relacionado con el tipo de actividad.

Con fases tranquilas bajas frecuencias, aproximadamente 1 Hz, pero en fases pre-eruptivas la frecuencia aumenta hasta los 3 Hz.

T_L: Bajas frecuencias, rondan los 0.5 Hz o un poco menos, para lograr estos registros, es necesario contar con sensores de banda estrecha.

Figura 4. Ejemplo de un registro tremor con su contenido espectral.



Fuente: [13].

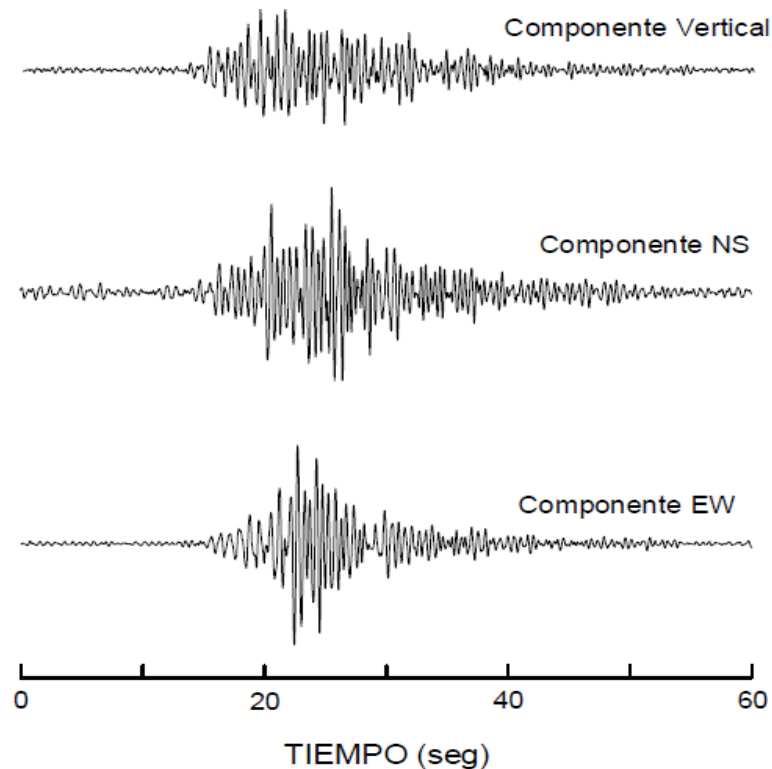
Explosiones

Estas señales aparecen cuando está en marcha un proceso eruptivo, junto con las señales tipo tremor, están entre las señales más caracterizan de estos procesos. Es posible que las explosiones aparezcan superpuestas a una señal de tremor; si este es el caso, son fácilmente identificables por su aumento inesperado en la amplitud y un incremento en las frecuencias de esta señal. Como regla general, esta señal tiene como mínimo dos llegadas diferentes y claras.

La primera está relacionada con la propagación en forma de ondas internas o superficiales de la explosión. La segunda llegada es lo que comúnmente se

denomina ondas de aire, ondas de choque y ondas sonoras, con una velocidad de propagación igual a 340 m/s. En la figura 5 se observa un ejemplo de esta señal registrada sus tres componentes. Cuando no ha sido posible distinguirlos durante su ocurrencia, se vuelve posible debido a esta velocidad de propagación tan lenta, lo cual facilita el poder identificar este tipo de eventos sobre los sismogramas [13].

Figura 5. Ejemplo de un registro tipo explosión con sus tres componentes.



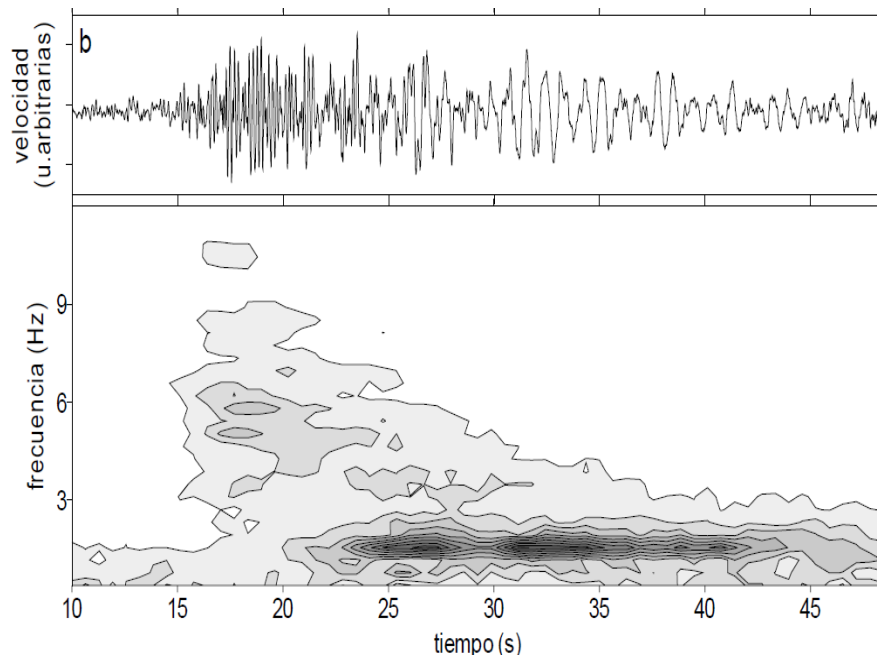
Fuente: [13].

Híbridos

Comúnmente este tipo de eventos tiene un inicio caracterizado por señales a altas frecuencias, con una amplia banda espectral logrando alcanzar hasta un poco más de 10 Hz. Otra característica de estas señales es que es posible identificar ondas llegadas de las ondas primarias y secundarias. Debido a esto, normalmente se asocia con un terremoto de pequeña magnitud. Luego de la primera llegada aparece una señal con gran similitud en cuanto a la morfología de la onda, duración y contenido espectral, a los eventos de largo período. No obstante, de ninguna manera se debe confundir los eventos tipo híbridos con eventos largo periodo.

Por norma general, en la llegada en alta frecuencia de los eventos tipo híbridos presenta ondas P y S de manera clara. Además, cuando estas señales con altas frecuencias son terremotos, cuando son registradas por una red sísmica convencional, presentan cambios en la dirección del primer impulso. Además, se deberá observar las distintas polaridades para las llegadas de las ondas primarias, en función de las estaciones sísmicas [13], En la figura 6 se aprecia un registro tipo híbrido junto con su contenido en frecuencia.

Figura 6. Ejemplo de un registro híbrido con su contenido espectral.



Fuente: [13].

ESTADO DEL ARTE SOBRE ALGORITMOS DE SEGMENTACIÓN DE SEÑALES SÍSMICAS

Métodos de segmentación de señales sísmicas

Para el desarrollo del presente estado del arte se realizó una investigación profunda y exhaustiva sobre algoritmos que se han propuesto para solucionar el tema de segmentación automática de señales sísmicas. Sobre este tema la literatura no es muy rica, pese a ello de los artículos encontrados se eligió 10 donde se prestan

métodos distintos y relevantes en relación con el tema, estas propuestas se dan a conocer a partir el año 1980. A continuación, se menciona estos artículos de manera resumida con la información más sobresaliente y en orden cronológico.

En [14] se menciona que el algoritmo de segmentación y detección de eventos sísmicos que usa su sistema está basado en un algoritmo estándar en el dominio del tiempo conocido por sus siglas, STA/LTA (Short Term Average to Long Term Average; Lee y Stewart, 1981). Este algoritmo tiene como desventaja el detectar como eventos sísmicos los ruidos producidos por el ambiente. Se busco más información al respecto, pero no se fue posible encontrar datos relevantes, de modo que es posible afirmar que fue el primer algoritmo de detección y segmentación de señales sísmica, tanto en [9] como en [15] se describe la teoría de su funcionamiento.

En 1984 se encontró dos publicaciones la primera en el mes de abril [16], en donde los autores proponen solucionar el problema mediante afinidad, que se define como una técnica de agrupación jerárquica que se puede utilizar para segmentar formas de onda como parte de un algoritmo sintáctico (o estructural) de reconocimiento de patrones. El algoritmo de afinidad comienza con la forma de onda misma o con una segmentación inicial primitiva de la misma. Construye una descripción jerárquica de la forma de onda combinando segmentos adyacentes similares. Una ventaja de la afinidad sobre otros métodos de segmentación es que no requiere ningún umbral preestablecido. Además, en lugar de producir solo una única segmentación de la forma de onda, son posibles muchas segmentaciones según cómo el analizador estructural utilice la descripción jerárquica. Las versiones del algoritmo de afinidad se aplican a señales sísmicas locales utilizando información de amplitud, pendiente y cruce por cero.

La segunda se publicó en el mes de mayo, en [17] se presenta un algoritmo secuencial efectivo que emplea un modelo autorregresivo de los datos y una prueba de razón de verosimilitud generalizada para detectar los cambios estadísticos significativos en la forma de onda. El algoritmo no requiere conocimiento previo de los datos. El modelado de bajo orden requerido reduce en gran medida los requisitos computacionales. Mediante ejemplos de formas de onda de explosión nuclear y terremoto natural, se muestra que el algoritmo localiza el tiempo de inicio de las ondas P con buena precisión. El algoritmo puede calcular secuencialmente y es bastante adecuado para el procesamiento en línea o en tiempo real. Sin embargo, se necesita más trabajo para emplear el algoritmo en la descomposición del campo de ondas sísmicas.

Para el año 1998 en el mes de agosto en [9] se realizó un estudio detallado sobre la comprensión y configuración de parámetros del algoritmo de activación STA-LTA. Aquí se habla sobre un software que incluye un nuevo algoritmo: promedio a corto plazo / promedio a largo plazo (STA / LTA) disparador. Este algoritmo aumenta significativamente la sensibilidad del disparador del instrumento y lo hace más robusto a los falsos disparadores en muchas aplicaciones sismológicas de movimiento particularmente débil. La captura exitosa de eventos sísmicos depende de la configuración adecuada de los nuevos parámetros de activación. La configuración de los parámetros STA / LTA es más exigente que la configuración de un disparador de umbral de amplitud simple que tradicionalmente se construye en grabadores sísmicos de movimiento fuerte. Para ayudar con esta tarea, esta nota de aplicación explica el funcionamiento del disparador STA / LTA y brinda instrucciones generales sobre cómo seleccionar sus parámetros. Este documento consta de 33 páginas que proporcionan una gran cantidad de información gran detalle como por ejemplo como escoger el tamaño de las ventanas de corto y largo alcance. Además, trata el tema de filtros, explicación de los softwares mencionados entre otros datos de gran importancia para comprender este algoritmo tan relevante en la sismología. En septiembre de 1999 en [15] se menciona una explicación sobre el mismo algoritmo con una nueva propuesta de sintonización de parámetros con esto dos artículos se aclaran todas las ideas relacionadas con el algoritmo STA / LTA.

En el mes de septiembre del siguiente año en [14] se diseñó un sistema automático específicamente para la detección de eventos sísmicos. Con el fin de reducir la detección de eventos falsos propone usar redes neuronales para el reconocimiento de patrones sísmicos en el dominio de la frecuencia como un paso siguiente al algoritmo de filtrado en el dominio de tiempo del tipo STA / LTA, que por lo general es usado en una variedad sistemas. Como resultado se obtuvo un algoritmo muy eficiente, de tal forma que consume tan pocos recursos de computacionales que es posible operarlo a tiempo real desde cualquier ordenador personal. como ventaja adicional se logró la detección de los eventos sísmicos independientemente del nivel de ruido que presente la señal sísmica, sino únicamente los patrones de la base de conocimientos. Este sistema lleva operando en la red sismológica de los Andes Venezolanos sin perder información, aproximadamente desde el año 1996. Se menciona que las primeras versiones del sistema presentaban el problema de que el 70% de los eventos grabados eran falsos, debido a que el algoritmo de detección de eventos sísmicos que utilizaba un sistema estaba basado en un algoritmo estándar en el dominio del tiempo. Como para resaltar se concluye que con la combinación del algoritmo STA/LTA en el dominio del tiempo y una red neuronal para el reconocimiento de patrones en el dominio de la frecuencia, se desarrolló un

detector de eventos que cuentas con las siguientes ventajas: Una disminución del 95% en grabaciones de eventos falsos, además de la detección de microsismos.

En el año 2012 en el artículo [7] los autores plantean que la principal dificultad para resolver este problema de la segmentación de estas señales se atribuye al hecho de que, en general, se desconocen tanto las propiedades estadísticas del ruido sísmico como las características de los eventos registrados. En este documento, al explotar la naturaleza particular de las señales que se están tratando, y al usar algunas propiedades interesantes que obedecen a una estadística de prueba basada en diferencias, así como sus ingredientes, proponen un enfoque que resulta en un método de detección automática robusto y eficiente. A partir de una serie de experimentos que realizan en datos sísmicos tanto sintéticos como reales, confirman la efectividad de la técnica propuesta.

Para el año siguiente en el mes de octubre en [3] menciona que, a partir de la experiencia en el campo, en su artículo introduce un nuevo enfoque para la detección de eventos en señales sísmicas utilizando el estimador de probabilidad máxima a posteriori (MAP), operando en la entropía R'enyi de tiempo y frecuencia a corto plazo. El problema se transfiere desde el espacio de medidas originales al espacio de la entropía R'enyi de frecuencia de tiempo, donde se introducen y utilizan algunas medidas de concentración de distribución tiempo-frecuencia y algoritmo MAP para la segmentación. Después en octubre de 2014 en [1] tienen como objetivo dar a conocer un método de segmentación, basado en el estimador de probabilidad máxima a posteriori (MAP), con aplicación en el procesamiento de señales sísmicas; también se discuten algunas interpretaciones y conexiones con otros enfoques en la detección y segmentación de cambios, así como los aspectos computacionales en este campo. Se presentan los resultados experimentales obtenidos por simulaciones de Monte Carlo para la segmentación de señales usando diferentes modelos de señales, incluyendo modelos con cambios en la media, en los parámetros del modelo FIR, AR y ARX, así como comparaciones con otros métodos, y la efectividad del enfoque propuesto está probado. Finalmente, se discute una aplicación de segmentación en el análisis de los registros de Terremotos durante el terremoto Kocaeli, Turquía, agosto de 1999, estación Arcelik (ARC). Los resultados de segmentación aparentemente son óptimos se comparan con el análisis de frecuencia y tiempo para la distribución de interferencia reducida (RID). Los resultados del análisis confirman la eficiencia del enfoque de segmentación utilizado, los instantes de cambio resultan de que el MAP aparece claro en los contenidos de energía y frecuencia de la distribución tiempo-frecuencia.

Sobre una mina de carbón subterránea se realizaron los siguientes estudios en relación con el tema de segmentación. El primer estudio se realizó en agosto del

año 2015, aquí se propone el algoritmo que es el corazón de este documento sobre el cual se realizaron las pruebas de comparación junto con el algoritmo STA / LTA. En el artículo [10] los autores plantean un procedimiento automático de segmentación de señales sísmicas. Este procedimiento está motivado por el análisis de señales de vibración sísmica real adquiridas en la mina. El algoritmo utiliza una descomposición de tiempo y frecuencia de la señal y modelar cada sub-señal en cada intervalo de frecuencia usando métodos estadísticos junto con medidas basadas en cuantiles empíricos estimados para cada sub-señal y, finalmente, agrega matriz 2D a vector de probabilidad 1D que indica la ubicación donde las características estadísticas han cambiado de un régimen a otro. El procedimiento propuesto puede aplicarse para mejorar el aislamiento en el dominio del tiempo de evento único para el caso, cuando la duración de la adquisición de la señal es más larga que la duración del evento o para aislar un evento único de la secuencia de eventos.

El segundo estudio se realizó en el siguiente año en el mes de marzo, en [8], se propone dos nuevos métodos de segmentación que toman en consideración las propiedades estocásticas de las señales analizadas en el dominio del tiempo. Las adquiridas en la misma mina subterránea de [10]. En tales señales se observa eventos sísmicos que aparecen después de la actividad minera, como voladuras, relajación provocada de la roca, y algunos eventos inesperados, como el estallido de la roca natural. Los procedimientos de segmentación propuestos permiten la extracción de tales partes de las señales analizadas que se relacionan con los mencionados eventos. Por último, en el tercer estudio realizado sobre estas señales [2], se basa en la observación de cambios en la estructura del espectro de señales basados exclusivamente en estadísticas de señales de segundo orden. El punto de partida del método propuesto son los coeficientes de reflexión variables en el tiempo del filtro de innovación adaptativa.

El algoritmo de Schur, que realiza la estimación de los coeficientes de reflexión, se conoce como un algoritmo rápido, preciso y estable capaz de rastrear rápidamente los cambios en las estadísticas de la señal. Desafortunadamente, la segmentación de señales sísmicas es muy difícil debido a. El artículo presenta la estructura del algoritmo, define sus propiedades, especifica los valores de los parámetros y presenta los resultados de la efectividad del proceso de detección.

Métodos de detección de ondas p y/o s

Dentro de la segmentación de las señales sísmicas existe un gran interés por detectar con exactitud el tiempo de inicio y fin de las ondas p y s, porque así se puede saber a qué distancia y en que coordenadas dio inicio un determinado sismo.

Los artículos que se encontraron se mencionan, ya que están relacionados con el tema principal de esta tesis que es la segmentación, Sin embargo, no se realizó un estudio lo suficientemente profundo.

Para el año 2006 en [18] plantean y prueban un algoritmo de ventanas múltiples para la detección y selección automáticas de las fases P impulsivas de eventos sísmicos en entornos de baja relación señal / ruido (SNR). Este método emplea tanto la amplitud absoluta instantánea como la promediada de las trazas en varias ventanas de tiempo antes y después de cada punto de tiempo (muestra) como funciones características. Cuando el valor absoluto instantáneo de una función característica supera un umbral dinámico ajustado automáticamente, las proporciones basadas en los promedios de las ventanas a lo largo del tiempo proporcionan parámetros para diferenciar un evento esperado de un ruido no deseado. El examen del algoritmo mediante el uso de datos sintéticos y reales muestra que la precisión de selección de los primeros arribos impulsivos puede ser inferior a 1-2 muestras, incluso cuando la relación señal / tono es inferior a 3.

En el año 2007 en [19] se implementan un algoritmo capaz de identificar parámetros de las ondas sísmicas, tales como: ondas primarias, secundarias, amplitud, período y duración. Para conseguir esto, se hace uso del procesamiento de señal digital y el lenguaje de programación en el software Matlab®, para realizar la lectura de forma automática de los parámetros de una señal sísmica. Se hace uso de algunas herramientas matemáticas, como la transformada de Fourier discreta (FFT) para estudiar los componentes en frecuencia de la señal y la transformada de Hilbert con el fin de determinar envolvente de la señal. Además, se diseñó e implementó una interfaz gráfica de usuario GUI mediante la herramienta GUIDE de Matlab®. La lectura es posible hacerla con funciones de lectura automática o de forma manual, con señales sísmicas con tres componentes. De este modo, que facilita la rutina de leer las señales sísmicas dentro de un gran conjunto. Los resultados alcanzados del algoritmo automático concuerdan con los que realiza un operador manual para señales con una excelente relación señal / ruido.

En el mes de octubre del año 2015 en [20] se publica un trabajo en donde se presentan nuevas estrategias de identificación automática de ondas P y S en sismogramas. Utilizando una base de datos con los tiempos de llegada identificados previamente por un lector humano, se procedió a comparar la identificación utilizando técnicas automáticas, basadas en la Transformada de Fourier en Tiempo Reducido (espectrogramas), Análisis Fractal y el Algoritmo Basic Matching Pursuit. Las dos primeras técnicas se usaron para identificación de la onda P, mientras que la segunda para la identificación de la onda S. Para la validación de los resultados,

se comparó con el identificador STA/LTA utilizado por Rietbrock et al. (2012) para la localización de hipocentros de réplicas del terremoto del Maule 2010.

Resumen de los métodos

Para una mayor comprensión y visualización de la información suministrada anteriormente, a continuación, se presenta un resumen en la **Tabla 1**.

Tabla 1. Resumen de los algoritmos para la segmentación de señales sísmicas.

ALGORITMOS DE SEGMENTACIÓN DE SEÑALES SÍSMICAS			
Título del artículo	Año	País	Método(s) usado
Hierarchical segmentation of seismic waveforms using affinity	1984	U.S.A.	Afinidad, agrupación jerárquica, amplitud, pendiente, cruce por cero, no requiere umbral
On a segmentation algorithm for seismic signal analysis	1984	U.S.A.	Modelo autorregresivo, prueba de razón de verosimilitud generalizada, no requiere conocimiento previo de los datos, modelado bajo orden
Understanding & Setting STA/LTA Trigger Algorithm Parameters for the K2	1998	U.S.A.	Conceptos básicos, parámetros de activación de, selección de la duración de las ventanas de promedio de tiempo corto y largo, nivel de umbral de activación y dettrigger de STA / LTA, selección de filtros de disparo, etc.
Understanding and parameter setting of STA/LTA trigger algorithm	1999	U.S.A.	Explicación del funcionamiento del algoritmo, otro método de sintonización de parámetros
Detector de eventos sísmicos en tiempo real utilizando redes neuronales	2000	Venezuela	Combinación de un algoritmo clásico STA/LTA en el dominio del tiempo con una red neuronal para el reconocimiento de patrones en el dominio de frecuencia, ventajas del resultado: reducción de casi 95% de grabaciones de

			eventos falsos, detección de microsismos, posibilidad de implantar la localización automática de hipocentros
A new automatic method for seismic signals segmentation	2012	Grecia	Usa algunas propiedades interesantes que obedecen a una estadística de prueba basada en diferencias, así como sus ingredientes, proponen un enfoque que resulta en un método de detección automática robusto y eficiente.
Signal Segmentation in Time-Frequency Plane using Rényi Entropy - Application in Seismic Signal Processing	2013	Francia	Estimador de probabilidad máxima a posteriori (MAP), entropía Rényi de tiempo, frecuencia a corto plazo, espacio de la entropía Rényi de frecuencia de tiempo, medidas de concentración de distribución tiempo-frecuencia, algoritmo MAP
Signal segmentation using changing regression models with application in seismic engineering	2014	Rumania	Estimador de probabilidad máxima a posteriori (MAP), detección de cambios, aspectos computacionales en este campo, resultados experimentales obtenidos por simulaciones de Monte Carlo, modelos FIR, AR y ARX, comparaciones con otros métodos.
Seismic signal segmentation procedure using time-frequency decomposition and statistical modelling	2015	Polonia	Descomposición de tiempo y frecuencia, métodos estadísticos, medidas basadas en cuantiles empíricos, agrega una matriz 2D y un vector de probabilidad 1D que indica el cambio de un régimen a otro.
Stochastic Modelling as a Tool for Seismic Signals Segmentation	2016	Polonia	Se propone dos nuevos métodos de segmentación que toman en consideración las propiedades estocásticas de las señales analizadas en el dominio del tiempo
Automatic segmentation of seismic signal with support of innovative filtering	2017	Polonia	Estructura del espectro de señales basados exclusivamente en estadísticas de señales de

			segundo orden, coeficientes de reflexión variables en el tiempo del filtro de innovación adaptativa, algoritmo de Schur, para la estimación de coeficientes de reflexión, es un algoritmo rápido, preciso y estable
--	--	--	---

De igual forma en la **Tabla 2** se presenta un resumen de los algoritmos detectores de las ondas p-s.

Tabla 2. Resumen sobre algunos algoritmos para la detección ondas s y p.

ALGORITMOS PARA DE DETECCIÓN DE LAS ONDAS P-S			
Título del artículo	Año	País	Método(s) usado
A multi-window algorithm for real-time automatic detection and picking of P-phases of microseismic events	2006	-	Vventanas múltiples en eventos sísmicos en entornos de baja relación señal / ruido (SNR), amplitud absoluta instantánea, trazas en varias ventanas de tiempo antes y después de cada punto de tiempo
Procesamiento digital de señales sísmicas con Matlab	2007	Perú	Identificación fases P y S, amplitud, periodo y duración, transformada discreta de Fourier, transformada de Hilbert para obtener la envolvente
Nuevas Estrategias de Identificación Automática de Tiempos de Llegada de Ondas P y S	2015	Chile	Tiempos identificados por un humano, Transformada de Fourier (espectrogramas), análisis Fractal y el Algoritmo Basic Matching Pursuit.

ALGORITMOS IMPLEMENTADOS

Después de analizar una serie de algoritmos encontrados en el estado del arte y al ser este estudio un primer acercamiento que se hace en la región sobre el tema de segmentación de señales sísmico-volcánicas, la selección de estos métodos fue debido a su versatilidad, facilidad en la implementación, además de que existe buena documentación al respecto. A continuación, se muestran los detalles de su teoría y funcionamiento.

Algoritmo 1: STA - LTA

El algoritmo denominado STA/LTA es uno de los más usados en la detección de eventos sísmicos, que se puede traducir como razón promedio de corto plazo versus largo plazo, donde STA es la amplitud media en un intervalo temporal corto y LTA es la amplitud media en un intervalo temporal largo, en comparación con un umbral de activación y desactivación.

En [6] utilizan ecuaciones basados en [9] y [15] donde se define los parámetros iniciales del algoritmo, N_s es el número de muestras de la ventana corta STA y N_L , el número de muestras de la ventana larga LTA. Se denota como F_s la frecuencia de muestreo de la señal. Los parámetros t_{STA} , t_{LTA} son los tiempos en segundos que se definen para cada ventana.

$$N_s = F_s * t_{STA} \quad (1)$$

$$N_L = F_s * t_{LTA} \quad (2)$$

En [6] consideran x_i cada valor del registro en función del tiempo, así se define el cálculo de STA/LTA como:

$$STA = \frac{1}{N_s} \sum_{j=i-N_s}^i x_j^2 \quad (3)$$

$$LTA = \frac{1}{N_L} \sum_{j=i-N_L}^i x_j^2 \quad (4)$$

Si,

$$j \leq 0, x_j = \frac{x_1 + x_2}{2} \quad (5)$$

De esta forma el cociente STA/LTA está dado por:

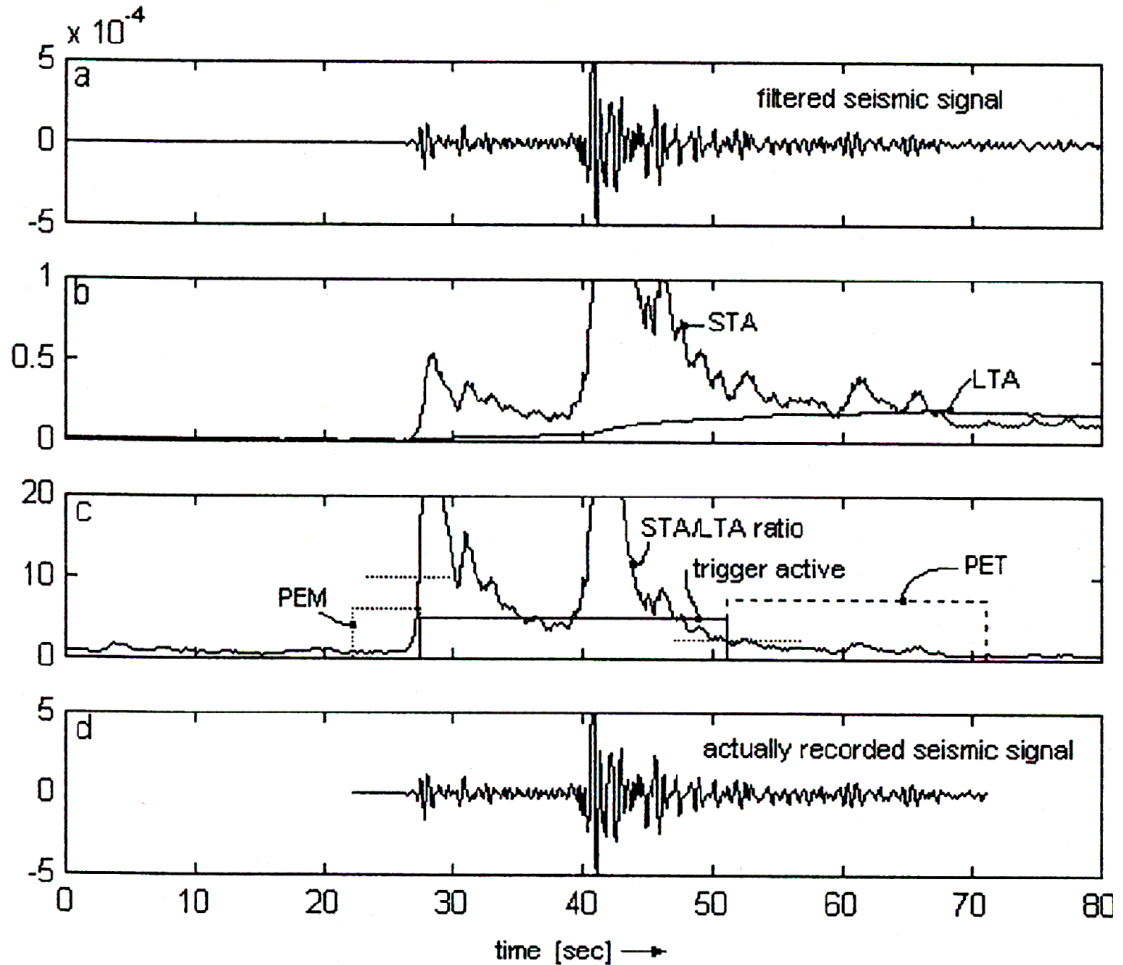
$$Radio_i = \frac{STA}{LTA} \quad (6)$$

Niveles de Umbral

La relación STA/LTA se compara continuamente con el nivel de umbral de activación (*trigger*); si la relación excede este umbral indica la presencia de un evento sísmico y se comienzan a guardar los datos hasta que la relación STA/LTA

cae por debajo de otro parámetro denominado umbral de desactivación (*detrigger*), el cual indica el final del evento [6].

Figura 7 . Funciones y variables para el disparo de STA-LTA.



Fuente: [9].

El disparador STA / LTA es más beneficioso en sitios sísmicamente tranquilos donde el ruido sísmico natural es el tipo dominante de ruido sísmico. También es efectivo en caso de cambios de ruido sísmico artificial "continuo". Tales cambios, por ejemplo, ocurren debido a la variación día / noche de la actividad humana cercana o en áreas urbanas. El algoritmo STA / LTA es menos efectivo en presencia de ruido sísmico irregular, de gran amplitud hecho por el hombre, que a menudo es de tipo ráfaga y/o pico [15]. En este documento se utiliza este algoritmo con el fin de encontrar un punto de inicio y fin en el tiempo de la señal para lograr obtener solamente la información más relevante.

Algoritmo 2: Descomposición tiempo-frecuencia con modelado estadístico

Este algoritmo se propone en [10], las ecuaciones que a continuación se presentan son tomadas de este artículo y se hace una breve explicación de las etapas necesarias.

Como se menciona en el título del artículo la primera fase se basa en la representación en tiempo-frecuencia de la señal, para ello se aplica el espectrograma, que se basa en la transformada de Fourier a corto plazo definida por la siguiente ecuación:

$$S(t, f) = \int_{-\infty}^{\infty} X(\tau)W(\tau - t)e^{-j2\pi f\tau} d\tau \quad (7)$$

Donde $X(\tau)$ es la señal por analizar y $W(t)$ es una ventana apropiada. En este caso la señal sísmico-volcánica X_1, X_2, \dots, X_n , es una señal discreta, así la transformada de Fourier discreta está dada por:

$$S(t, f) = \sum_{k=1}^N X_k W_{k-t} e^{-j2\pi f k} \quad (8)$$

Por medio de la función $S(t, f)$, la señal se puede analizar en el dominio de tiempo-frecuencia (como un conjunto de sub-señales correspondientes a las frecuencias apropiadas), así como en el dominio de la frecuencia. El cuadrado del valor absoluto de la matriz $S = \{S(t, f)\}$ se llama espectrograma. En la segunda fase del procedimiento, para cada frecuencia f analizamos las series de tiempo $S(:, f)$ por separado. Estas series de tiempo se llaman sub-señales. Para cada sub-señal se calcula cuantiles de orden dado p . Desde el punto de vista teórico, el cuantil de orden p de una variable aleatoria X es un número $x(p)$, donde se cumple que:

$$P(X \in (-\infty, x(p)]) \geq p, \quad P(X \in [x(p), \infty)) \geq 1 - p \quad (9)$$

La distribución teórica de sub-señales es desconocida, por lo tanto, se calcula los cuantiles empíricos, es decir, los cuantiles calculados sobre la base de las sub-señales correspondientes a las frecuencias dadas. El cuantil empírico de orden p es una observación tomada de los datos ordenados ascendentemente en el número de posición $N * p/100$, donde N es la longitud de una sub-señal dada. Debido a que para cada frecuencia f obtenemos diferentes cuantiles de orden p , por lo tanto, en el análisis posterior de los cuantiles apropiados que se denotan como $x(p, f)$.

Para la fase 3 del procedimiento de segmentación, se calcula la denominada matriz binaria, que contiene valores de cero y uno, e indica si la observación dada de una

señal secundaria correspondiente a la frecuencia f excede el cuantil $x(p, f)$. Dicha matriz binaria se define como:

$$S_{bin}(t, f) = \begin{cases} 1, & |S(t, f) > x(p, f), \\ 0, & |S(t, f) \leq x(p, f) \end{cases} \quad (10)$$

A continuación, en la fase tres se procede a calcular el vector de probabilidades para cada punto de tiempo t , que es igual a la media de la matriz binaria en cada punto de tiempo. Por lo tanto, se define de la siguiente manera:

$$P(t) = \frac{1}{\#F} \sum_{f \in F} S_{bin}(t, f). \quad (11)$$

Luego de calcular el vector de probabilidades, se determina el umbral de porcentaje K . Sobre la base de este umbral, se calcula el límite del vector normalizado $\{P(t)\}$. Este límite es como se muestra en la siguiente ecuación:

$$l = [k * \max(p(t))]. \quad (12)$$

Sobre la base del límite calculado, se analiza un vector de decisión definido de la siguiente manera:

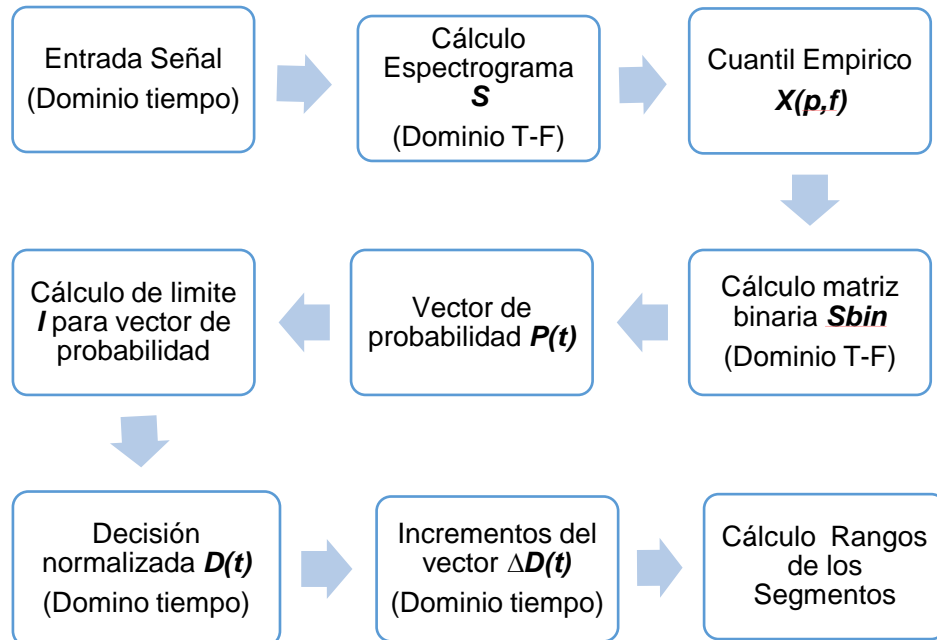
$$D(t) = \begin{cases} 1, & p(t) > l, \\ 0, & \text{otro caso} \end{cases} \quad (13)$$

Para la fase final se debe analizar los incrementos del vector de decisión $D(t)$, así:

$$\Delta D(T) = D(t) - D(t - 1). \quad (13)$$

Los inicios de los segmentos apropiados están en los puntos de tiempo donde el vector $\Delta D(t)$ tiene un valor negativo, mientras que los finales de los segmentos, donde $\Delta D(t)$ es positivo. Para una mejor comprensión del metro descrito en la Figura 8 se presenta el esquema del método.

Figura 8. Diagrama de bloques funcionamiento del algoritmo 2.



Fuente: [10]

CARACTERIZACIÓN

Se refiere a la acción de extraer características o atributos a un conjunto de señales, esto implica en primera instancia el reducir la cantidad de recursos necesarios para describir un gran conjunto de datos (reducción de dimensión). En segunda parte, se define como el proceso de formar un conjunto de características (de dimensión n) partiendo de los datos de entrada (de dimensión $m > n$).

Debido que el tema central de este trabajo no está basado en la caracterización, la búsqueda sobre este tema no fue exhausta, sin embargo, en la literatura se encontró un artículo reciente [21], en donde se hace una gran recolección de características de forma de onda, espectrales, pseudo-espectrograma y atributos de polaridad.

Las características de polaridad se descartaron, ya es necesario tener las señales en tres ejes es decir tres señales, y como se mencionó el mayor énfasis se hace sobre la segmentación y no sobre caracterización. Dichas características son aplicadas sobre dos clases de señales volcánicas. De las características utilizadas en [21] se decide emplear 14, debido a que se encontró buenos resultados con la base de datos usada.

SELECCIÓN DE CARACTERÍSTICAS

Las señales sísmico-volcánicas son datos altamente dimensionales que pueden contener muchas características irrelevantes. La selección del subconjunto de características es una técnica para reducir el espacio de atributos de un conjunto de características. En otras palabras, se está identificando un subconjunto de características mediante la eliminación de características irrelevantes o redundantes. En esta investigación para la etapa de selección de características se utilizó tres métodos de selección el primero, llamado mejor primero (BestFirst).

El algoritmo ReliefF pondera cada característica de acuerdo con su relevancia para la clase. El peso de cada una de las características se actualiza iterativamente. 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.

CLASIFICACIÓN

Clasificación supervisada

La clasificación supervisada es aquella en la cual a partir de un conjunto de datos conocidos por la maquina (conjunto de entrenamiento), se clasifica un segundo conjunto de datos desconocido. Los algoritmos de clasificación tienen ya un conocimiento del conjunto estudiado, por lo que le es necesario utilizar datos previamente etiquetados con el fin de agrupar en el conjunto o clase que corresponda. Entre algunos ejemplos de clasificadores supervisados están:

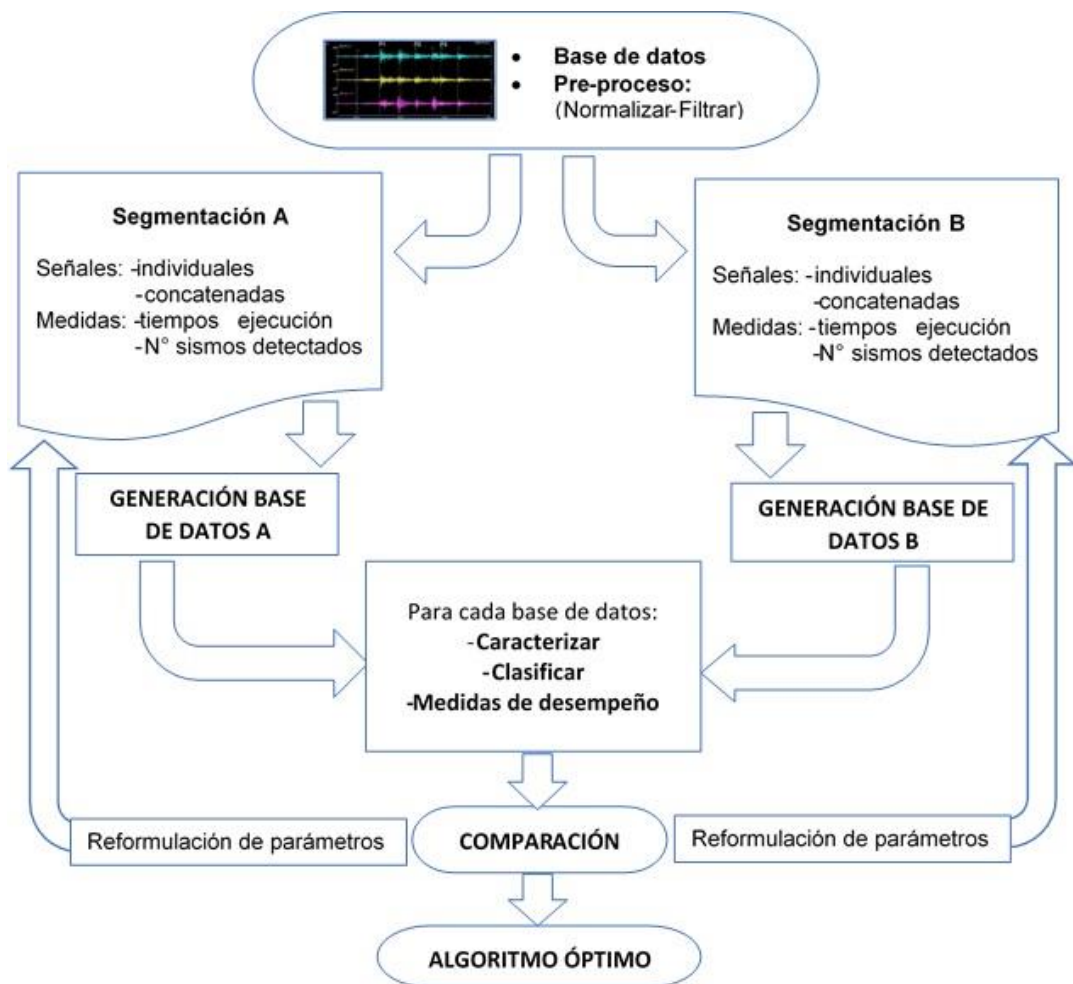
- **K vecinos más cercanos (k-NN Nearest Neighbour):** K-NN se basa en la idea en que los nuevos ejemplos serán clasificados a la clase a la cual pertenezca la mayor cantidad de vecinos más cercanos del conjunto de entrenamiento más cercano a él [22].
- **Máquina de vector soporte (Support Vector Machine – SVM):** Este método puede clasificar datos linealmente separables, es resistente al sobreajuste porque buscan una frontera de decisión específica y es eficiente en el caso no lineal porque no crea explícitamente el espacio transformado y su transformación no lineal es implícita. Puede procesar gran número de entradas [23].

- **Bosques aleatorios (Random-Forest):** Este método de clasificación es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos [24].

1. DESARROLLO DE LA INVESTIGACIÓN

La metodología propuesta con el fin de lograr la comparación de los algoritmos de segmentación implementados aplicados a señales volcánicas se resume en la Figura 9 y consta de las siguientes etapas: preprocesamiento, segmentación, extracción de características y clasificación.

Figura 9. Metodología propuesta.



Fuente: Esta investigación

1.1 BASE DE DATOS

Se trabajó sobre una la base de datos facilitada por el observatorio sismológico de Manizales (OVSM), aquí se registró la actividad del volcán Nevado del Ruiz durante el año 2015. Los registros se hacen con las diferentes componentes o ejes del sismómetro, cada eje se identifica como Z (vertical), E (East), y N (North); y según corresponde al nombre de la estación que registró el evento. Para este caso los eventos fueron grabados por 11 estaciones vulcanológicas que según la base de datos tienen las siglas: OLL, TOL, LIS, BIS, ELA, ALF, RC3, CL2, GLA, REB, OC2, en los 3 ejes, de este modo la base de datos original contiene 7081 señales.

Debido a esto hay una gran variedad de señales que no son exactamente iguales, sino replicadas en ejes con diferentes direcciones, por ende, se debía escoger una estación y un eje. En este estudio se eligió la estación Olleta (OLL) y el eje Z, esto fue sugerido por el observatorio sismológico, por lo tanto, la base de datos se redujo a 1871 registros individuales con 7 clases. Estos datos fueron adquiridos con una frecuencia de muestreo de 100 Hz. Los registros están etiquetados de la siguiente manera: Volcano tectónico (VT), Tremor (TR), Largo periodo (LP), Desprendimiento de hielos (IC), Híbridos (HB), Sismos múltiples (SM), Explosiones (EX). A continuación, en la **tabla 3** se muestran la cantidad de señales por cada clase.

Tabla 3. Número de registro por clase en la base de datos.

Cantidad de registros	Clase
450	VT
450	TR
450	LP
450	IC
56	HB
11	SM
4	EX

Como se observa en la tabla anterior la base de datos no está balanceada debido a que las clases no tienen igual número de registros, por este motivo y para efectos de clasificación se decidió trabajar con las 4 clases que tienen igual número de registros tales como: VT, TR, LP, IC. La mayoría de las pruebas se realizaron con 3 clases VT, TR, LP, ya que con estas se consiguió mejores resultados de clasificación. Sin embargo, para las pruebas de segmentación con señales individuales se trabajó con las 7 clases.

1.2 PRE-PROCESAMIENTO

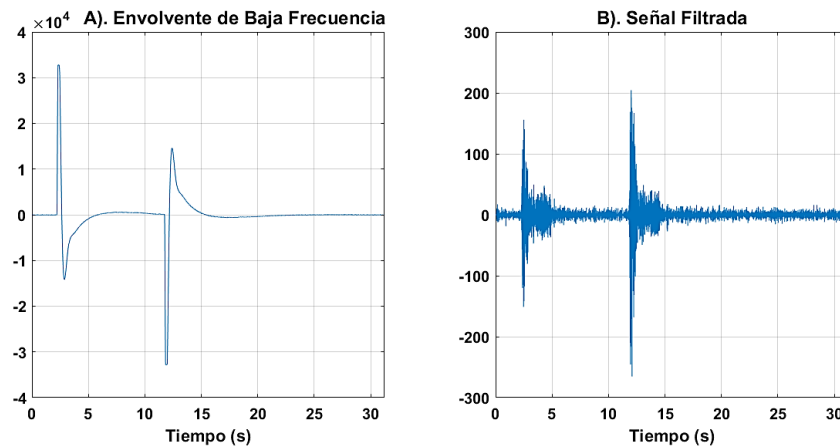
Antes de someter las señales volcánicas a un procesamiento de señales se normalizan los registros en el rango de $[-1,1]$, logrando quitar niveles de offset para esto se hizo uso la ecuación 14.

$$S_v = \frac{S_v - \overline{S_v}}{\max |S_v|} \quad (14)$$

Donde S_v representa la señal volcánica, $\max |S_v|$ el máximo del valor absoluto de la señal y $\overline{S_v}$ es la media de la señal.

Con lo anterior se asegura que las señales no sobrepasen un cierto rango que esta entre $[-1,1]$, lo que facilita el respectivo procesamiento a seguir para las muestras. Además, seguido a este proceso se aplica un filtro butterworth pasa altas con una frecuencia de corte de 0,5 Hz y sexto orden, con el fin de eliminar frecuencias bajas, que en algunos casos era la envolvente de las señales como se muestra en la Figura 10.

Figura 10. Señal tipo VT con una envolvente de baja frecuencia.



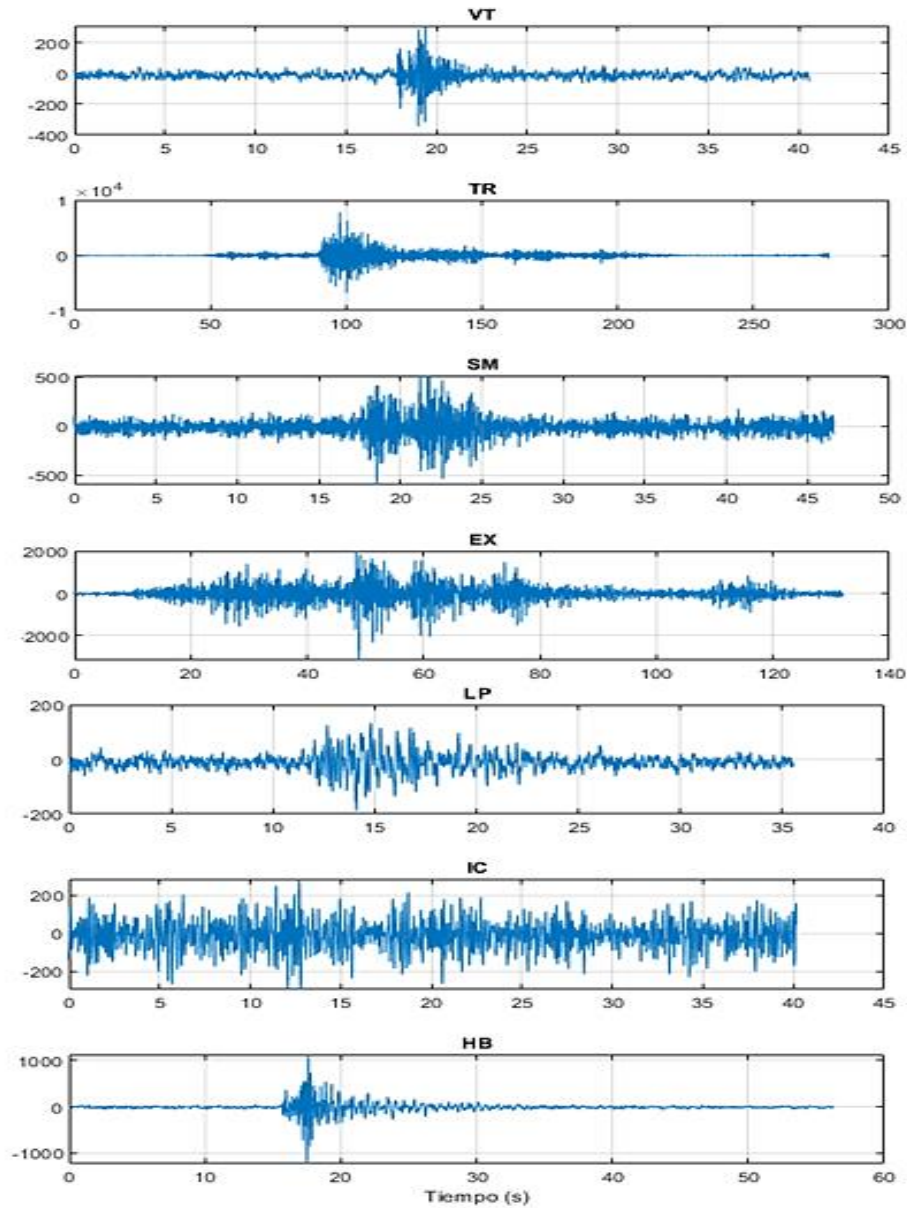
Fuente: Señal del volcán Nevado del Ruiz registrada por el OVSM en el año 2015.

1.3 SEGMENTACIÓN

Para esta etapa se propone evaluar los algoritmos con dos formas de disposición de las señales, la primera con las señales individuales como viene dada la base de datos. En la Figura 11 se observa un ejemplo de registro tomado una señal de la base de datos de forma aleatoria por cada clase de forma individual.

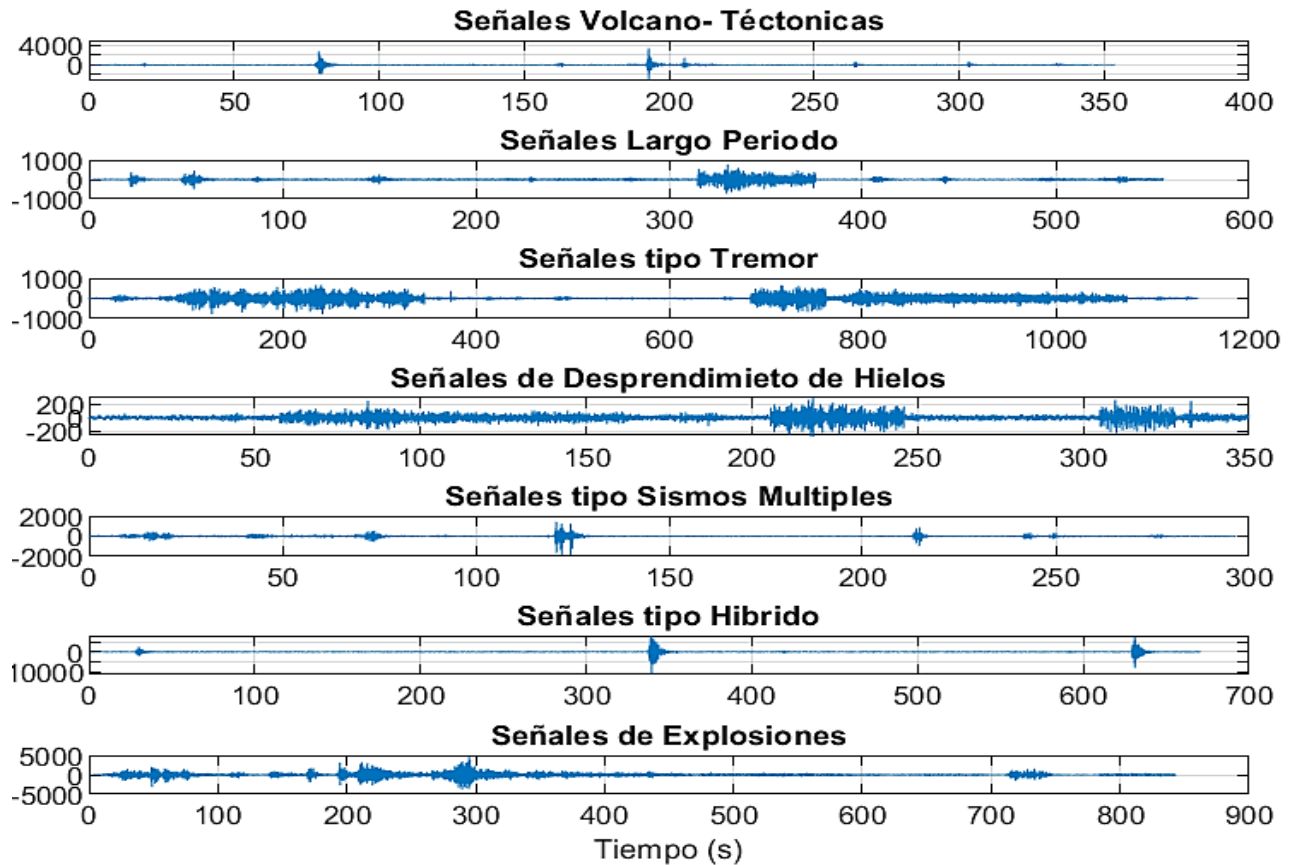
La otra forma que se realizó las pruebas a los algoritmos de segmentación fue concatenando las señales, es decir, poner las señales una a continuación de otra simulando un registro real de los datos, cabe aclarar que para este estudio se realizó una concatenación entre la misma clase. Un ejemplo de señales cocatenadas se observa en la Figura 12.

Figura 11. Ejemplos de registros individuales por clase de la base de datos usada.



Fuente: Señales del volcán Nevado del Ruiz registradas por el OVSM en el año 2015.

Figura 12. 10 señales concatenadas por clase VT, LP, TR, IC, SM, HB. Para señales tipo EX 4 señales.



Fuente: Señales del volcán Nevado del Ruiz registradas por el OVSM en el año 2015.

1.3.1. Algoritmo 1

Definición de parámetros

La configuración de los parámetros de activación de STA / LTA siempre es una compensación entre varias consideraciones sismológicas e instrumentales. El objetivo de buscar ajustes de parámetros óptimos es la mayor sensibilidad posible de la estación sísmica para un tipo dado de señal sísmica en este trabajo la sintonización se realizó de forma manual.

Inicialmente para la segmentación con el algoritmo STA-LTA los parámetros se determinaron basado en [15], [6] donde se propone usar los siguientes valores:

Tabla 4. Valores iniciales para las variables del algoritmo ST / LTA.

Parámetro	Valor
Sta	1
Lta	8
Activación	2
Desactivación	1.6
Salto	0
Duración mínima	1.7
Tiempo antes	1
Tiempo después	1

Se tomó como base estos valores para realizar los primeros procedimientos de segmentación, sin embargo, el valor de los parámetros se cambió según lo indicaban los resultados de los experimentos realizados. Como se menciona en los artículos relacionados con el algoritmo STA / LTA la definición de los parámetros depende de las señales específicas a tratar.

Según la práctica se definió una tabla de valores con los parámetros óptimos en la detección aplicado a la base de datos según el tipo de señal. Bajo estos parámetros se realizó la segmentación de una nueva base de datos usando tres clases VT, LP y TR.

Tabla 5. Valores para las variables del algoritmo STA/ LTA por cada clase determinados en esta investigación.

	VT	LP	TR	IC	SM	HB	EX
Sta	1	1	1	1	1	1	1
Lta	7	5	4	8	8	8	6
Activación	1	1.2	1.5	2	2.2	2.5	1.8
Desactivación	1	1.1	1.6	1.6	1.3	1.8	2
Salto	0	0	0	0	0	0	0
Duración mínima	2	2	2	2	2	2	2
Tiempo antes	1	1	1	1	1	1	1
Tiempo después	1	1	1	1	1	1	1

1.3.2. Algoritmo 2

Definición de parámetros

Además de las etapas propuestas por [10] se aumenta dos etapas extras propuestas en este trabajo con el fin de mejorar la detección. La primera se llama Área a Eliminar (AE), una vez obtenida la matriz binarizada, se elimina las áreas pequeñas que por lo general son ruido y causan falsas detecciones. Por otra parte, una vez se encuentra la función de probabilidad, se aplica un suavizado a la señal, esto beneficia al proceso aumentando la precisión de detección, el filtro tiene dos variables, el orden y la longitud del cuadrado (framelen). En las pruebas, el filtro se fijó de orden 3, el valor de “framelen” debe ser impar y mayor que el valor del orden.

Para el proceso de segmentación de señales individuales y señales concatenadas, inicialmente se usó los parámetros sugeridos en [10], en la tabla 6 se pueden observar.

Tabla 6. Valor inicial de las variables del algoritmo 2.

Percentil(p)	K
60	80%

Sin embargo, fue necesario realizar cambios de empíricamente según lo requerían las señales de la base de datos con el fin de mejorar la detección. En la tabla 7 se puede ver los valores usados para estos parámetros durante los experimentos.

Tabla 7. Valores usados para la generación de la nueva base de datos 2.

Percentil (p)	90
Área Eliminar (AE)	500
Framelen (FI)	21
K	30%

1.4 CARACTERIZACIÓN

Se extrajeron 2 conjuntos de características para el estudio comparativo de la siguiente manera:

- i) 8 características en el tiempo.
- ii) 6 características en frecuencia.

Tabla 8. Características representativas para señales volcánicas.

Característica N°	Atributo
C1	Desviación estándar
C2	Media
C3	Mediana
C4	Valor máximo
C5	Curtosis
C6	Simetría (Sesgo)
C7	Frecuencia media de los 5 picos más altos
C8	Descomposición wavelet
C9	Duración de la señal
C10	Impulsividad de comienzo de señal
C11	Frecuencia Domínate
C12	Rango de frecuencia definido por el contorno del espectro
C13	Frecuencia del centro de gravedad (centroide) del contorno espectral
C14	Frecuencia de máxima de energía en el espectrograma

1.5 SELECCIÓN DE CARACTERÍSTICAS

Para evaluar las señales segmentadas se necesita realizar un proceso de clasificación, pero antes de esto es necesario identificar el menor número de características para extraer, de modo que se logra disminuir la cantidad de datos que entran al sistema, además conseguir un error aceptable en la clasificación, para ello se propuso comparar la siguiente selección de características:

- i) Tomando el total de 14 características
- ii) Tomando las 10 mejores características del conjunto original.
- iii) Selección de características realizada por el algoritmo mejor primero (BestFirst).

2. MARCO EXPERIMENTAL

En esta sección se realiza la descripción de los experimentos propuestos llevados a cabo, con el propósito de cumplir 3 objetivos, el primero es identificar un grupo de características que representen este tipo de señales. En segundo lugar, identificar el algoritmo de clasificación que tenga mejor desempeño en la identificación de señales volcánicas. Y por último evaluar el desempeño de los dos algoritmos de segmentación de señales volcánicas implementados, en cuanto a detección de señales sísmicas, tiempo empleado en la segmentación de cada clase, tanto individual como en grupos de señales concatenadas. Finalmente, evaluar las bases de datos segmentadas con los algoritmos 1 y 2 respectivamente con los mejores parámetros de caracterización y clasificación determinados en los **experimentos 5 a 7**.

2.1. CLASIFICACIÓN

Después de realizar la caracterización y selección de características propuesta se prueba con 4 clasificadores en cascada que son: parzen, knn, nusv y random-forest, con 10 iteraciones. Esta clasificación se aplica sobre la base de datos original y con las bases de datos segmentadas con 3 clases VT, LP, TR.

2.2. MEDIDAS DE DESEMPEÑO

Para calificar el resultado de la segmentación se utilizaron las siguientes medidas de desempeño: porcentaje de señales detectadas (PD) en señales concatenadas según la ecuación 15. Además, el tiempo empleado por cada algoritmo el realizar la segmentación tanto en señales concatenadas como en señales individuales.

$$PD = \frac{Nd}{Nsc} * 100\% \quad (15)$$

- Nd: Numero de señales detectadas.
- Nsc: Numero de señales concatenadas.

Para calificar el resultado de la clasificación se utilizaron las siguientes medidas de desempeño: Sensibilidad (Se), Especificidad (Sp) y porcentaje de clasificación (PC); estas medidas son calculadas con las ecuaciones 7, 8 y 9.

$$PC = \frac{Vn \times Vp}{Vn + Fp + Vp + Fn} \quad (16)$$

Los valores para calcular estas medidas son:

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

Para medir el costo computacional se registró el tiempo de ejecución de los diferentes procesos realizados con la base de datos principalmente en la etapa de segmentación.

2.3. PRUEBAS QUE SE APLICARON SOBRE LA BASE DE DATOS

En la metodología propuesta se indica realizar pruebas tanto con señales individuales como con señales concatenadas y tomar sus medidas de desempeño, para así obtener y evaluar nuevas bases de datos por medio de la caracterización y clasificación. Para ello se debe identificar qué características extraer y que clasificadores usar. Se ha dividido los experimentos en 3 etapas. Etapa1 (Segmentación), etapa 2 (extracción y selección características con clasificadores), etapa 3 (clasificación base de datos segmentadas).

ETAPA 1: SEGMENTACIÓN

La idea de estos experimentos es observar, analizar y cuantizar el desempeño de los algoritmos 1 y 2 de segmentación con todas las clases de forma individual y concatenada, mostrar ejemplos de detección exitosa, fallas en la detección, variación de parámetros, así como determinar el costo computacional en la ejecución de estas tareas, contabilizando el tiempo de ejecución.

Experimento 1: Se realiza la segmentación con el **algoritmo 1** sobre todas las clases de forma individual. Se muestra ejemplos de detección exitosa y fallida. Además, se realiza variación en los parámetros y analiza cambios.

Experimento 2: Se realiza la segmentación con el **algoritmo 2** sobre todas las clases de forma individual. Se muestra ejemplos de detección exitosa y fallida. Además, se realiza variación en los parámetros y analiza cambios.

Experimento 3: Se realiza la segmentación con el **algoritmo 1** sobre todas las clases concatenando grupos de 1 a 50, 100, 150, 200, 250, 350, 400 y 450 para el caso de las clases VT, LP, TR, IC. Para señales tipo HB grupo de 1 a 10, 20, 30, 40, 50. Para señales de SM y EX un solo grupo de 11 y 4 señales respectivamente.

Experimento 4: Se realiza la segmentación con el **algoritmo 2** de segmentación sobre todas las clases concatenando grupos de 1 a 50, 100, 150, 200, 250, 350, 400 y 450 para el caso de las clases VT, LP, TR, IC. Para señales tipo HB grupo de 1 a 10, 20, 30, 40, 50. Para señales de SM y EX un solo grupo de 11 y 4 señales respectivamente.

ETAPA 2: EXTRACCIÓN Y SELECCIÓN CARACTERÍSTICAS CON CLASIFICADORES

La idea de estos experimentos es observar, analizar y cuantizar la precisión de clasificación que se obtiene con el conjunto total de características, y con los selectores ReliefF y BestFirst, aplicados sobre la **base de datos original**, así como determinar el clasificador a usar para la evaluación de los algoritmos de segmentación.

Experimento 5: Se extrae el conjunto total de 14 características y se clasifica la **base de datos original** con 3 clases VT, LP y TR. Se usa los clasificadores parzen, knn, nusv y random-forest, en cascada. Se obtiene errores de clasificación, se muestran de forma gráfica.

Experimento 6: Se realiza selección de características con el método ReliefF tomado las 10 mejores características con la **base de datos original** con 3 clases VT, LP y TR. Se usa los clasificadores parzen, K-NN, nusv, random-forest, en cascada. Se obtiene errores de clasificación, se muestran de forma gráfica.

Experimento 7: Se realiza selección de características con el método mejor primero (BestFirst) con base de datos original con 3 clases VT, LP y TR. Se usa los clasificadores parzen, K-NN, nusv y random-forest, en cascada. Se obtiene errores de clasificación, se muestran de forma gráfica.

ETAPA 3: SEGMENTACIÓN DE BASES DE DATOS Y CLASIFICACIÓN

La idea de estos experimentos es aplicar la caracterización, selección de características y clasificación usadas sobre la **base de datos original** en las **nuevas bases de datos** generadas por los algoritmos 1 y 2. Para así determinar

cuál algoritmo de segmentación tuvo mejor desempeño en cuanto a precisión en la clasificación.

Experimento 8: Se realiza la segmentación con el **algoritmo 1** a las 450 señales concatenadas de cada clase VT, LP, TR. Esto con los parámetros obtenidos de los experimentos anteriores, se etiqueta y se genera una nueva **Base de Datos 1**.

Experimento 9: Se realiza la segmentación con el **algoritmo 2** a las 450 señales concatenadas de cada clase VT, LP, TR. Esto con los parámetros obtenidos de los experimentos anteriores, se etiqueta y se genera una nueva **Base de Datos 2**.

Experimento 10: Se extrae el conjunto total de 14 características, se clasifica la **Base de Datos 1**. Se usa los clasificadores parzen, K-NN, nusv, random-forest, en cascada. Se obtiene errores de clasificación, se muestran de forma gráfica.

Experimento 11: Se realiza selección de características con el método BestFirst. Esto se hace sobre la **Base de Datos 1**. Se usa los clasificadores parzen, K-NN, nusv, random-forest, en cascada. Se obtiene errores de clasificación, se muestran de forma gráfica.

Experimento 12: Se extrae el conjunto total de 14 características, se clasifica la **Base de Datos 2**. Se usa los clasificadores parzen, knn, nusv, random-forest, en cascada. Se obtiene errores de clasificación, se muestran de forma gráfica.

Experimento 13: Se realiza selección de características con el método ReliefF, tomado las 10 mejores características y el método BestFirst. Esto se hace sobre la **Base de Datos 2**. Se usa los clasificadores parzen, knn, nusv, random-forest, en cascada. Se obtiene errores de clasificación, se muestran de forma gráfica.

Finalmente, se diseña e implementa una interfaz en GUI del software MatLab®, la cual resume el resultado de los scripts obtenidos tanto en segmentación, caracterización, selección de características, clasificación usados en este trabajo. El usuario puede realizar variaciones en los parámetros de cada algoritmo de segmentación. Además de usarlos de forma individual y conjunta. Con señales individuales o concatenadas por clase. Generar nuevas señales, grupos de señales o bases de datos segmentadas, para posteriormente ser evaluadas. A estas señales se les extraen 5 características: valor medio, curtosis, duración la señal, frecuencia del contorno del espectro y frecuencia del máximo de energía en el espectrograma. Para clasificar se implementó el algoritmo Random-Forest, con la posibilidad de visualizar el error de clasificación por medio de diagrama de cajas y bigotes. Cuenta una visualización de las señales que permite observar en una figura con 45 señales. También guarda medidas de desempeño del tiempo de ejecución y porcentaje de detección.

3. RESULTADOS

Por medio de tablas y figuras se exponen los resultados obtenidos de las pruebas mencionadas en la sección anterior.

Experimento 1: En la tabla 9 se observa que el menor tiempo de ejecución fue con señales de la clase Sismos Múltiples (SM) con un tiempo de ejecución de 0,017 segundos.

Tabla 9. Tiempos de ejecución del algoritmo 1 con señales individuales.

Tiempo(s)	Señal
0,020	VT
0,022	LP
0,050	TR
0,018	IC
0,017	SM
0,020	HB
0,063	EX

La señal roja que se observa en la figura 13 y en las siguientes figuras indica los puntos de inicio y fin de la segmentación indicado por el algoritmo en cuestión en este caso STA \ LTA. Para mostrar los prendimientos llevados a cabo se tomó aleatoriamente una señal de la clase VT; cabe resaltar que este proceso se llevó a cabo con todas señales de las 7 clases.

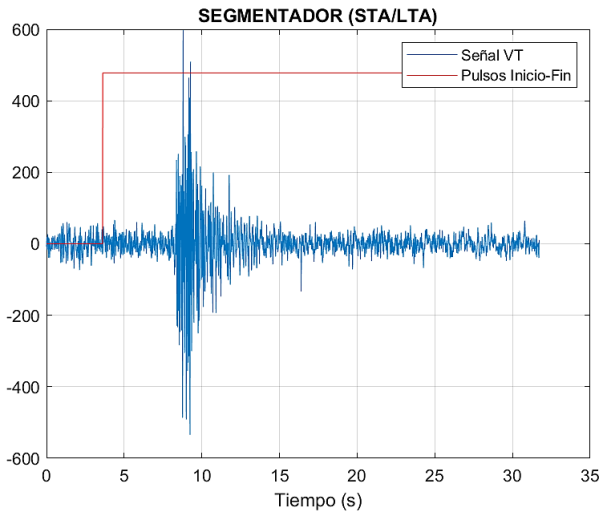
Se identificó que una de las variables más críticas de este algoritmo es el tiempo de desactivación (dettrigger) inicialmente se realizó pruebas con los valores indicados en la tabla 10, se tuvo en cuenta las variables, donde los resultados de las pruebas presentaron mayor relevancia en el funcionamiento del algoritmo.

Tabla 10. Valores utilizados en la segmentación de una señal tipo VT.

Parámetro	Sta	Lta	Activación	Desactivación
Valor	0.8	4	1	0.5

En la figura 13 se observa que con estos valores el algoritmo únicamente identifica el tiempo de inicio, pero no encontró el tiempo final, para este caso el algoritmo entrega como tiempo final la duración máxima de la señal.

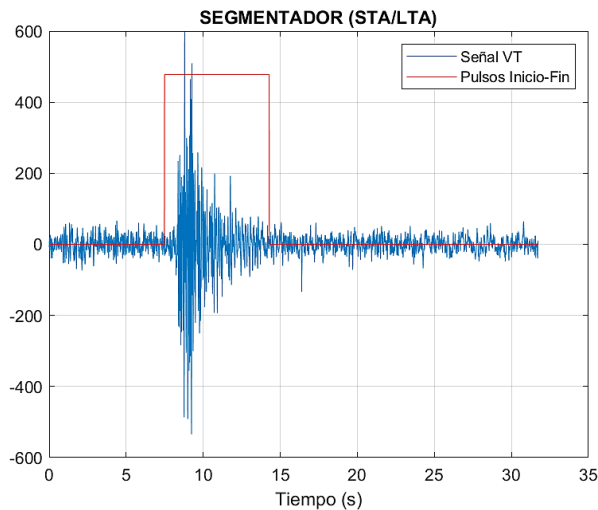
Figura 13. Segmentación del algoritmo 1, sin detección de tiempo final.



Fuente: Señal del volcán Nevado del Ruiz registrada por el OVSM en el año 2015.

Para dar solución a este inconveniente presentado la mejor opción fue aumentar el valor de la desactivación a un valor igual a 1.2, la nueva detección se puede visualizar en la figura 14.

Figura 14. segmentación realizada por algoritmo 1 de forma correcta.



Fuente: Señal del volcán Nevado del Ruiz registrada por el OVSM en el año 2015.

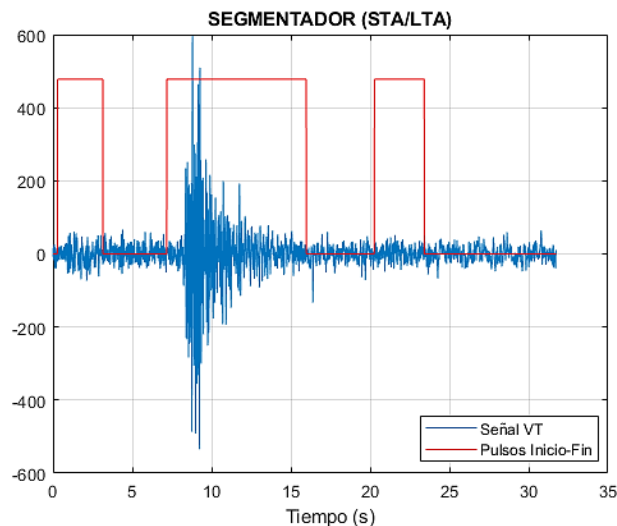
Se realizaron otras variaciones a los parámetros como disminuir el tiempo de activación y la ventana Lta, en la tabla 11 se muestran los valores concretos. El

resultado de estos cambios se puede observar en la figura 15, esto causo que el algoritmo realice falsas detecciones.

Tabla 11. Valores utilizados en la segmentación de una señal tipo VT.

Parámetro	Sta	Lta	Activación	Desactivación
Valor	0.8	1	0.5	1

Figura 15. Segmentación realizada por el algoritmo 1, detecciones falsas.



Fuente: Señal del volcán Nevado del Ruiz registrada por el OVSM en el año 2015.

Dentro de la base de datos de registros individuales se encontró señales con múltiples choques etiquetadas como único registro. Para este caso en particular se aplicó el proceso de segmentación con los valores indicados en la tabla 12. En la figura 16 se puede observar tanto el tipo de señal encontrada como la segmentación realizada por el algoritmo 1.

Tabla 12. Parámetros usados a una señal tipo VT con múltiples choques.

Parámetro	Sta	Lta	Activación	Desactivación
Valor	0.2	1	0.1	3

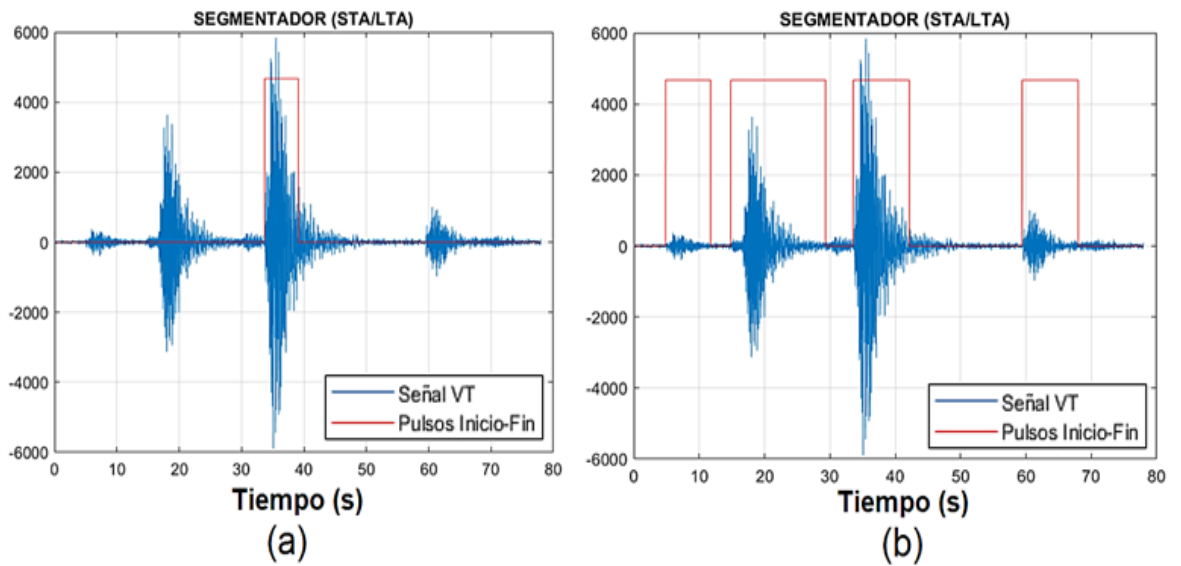
En la figura 16 (a) se observa que con los parámetros indicados en la tabla 12 el algoritmo detecta una señal, pero en realidad existen 4 señales. Para solucionar este error se modificó los parámetros los cuales se registraron en la tabla 13. El

algoritmo realizo una detección correcta con estas modificaciones esto se observa con detalle en la figura 16 (b).

Tabla 13. Parámetros usados con una señal tipo VT con múltiples choques para una identificación exitosa.

Parámetro	Sta	Lta	Activación	Desactivación
Valor	0.2	3	0.1	1.2

Figura 16. Detección errónea (a) y detección exitosa (b) para una señal de 4 choques.



Fuente: Señales del volcán Nevado del Ruiz registradas por el OVSM en el año 2015.

Como conclusión del experimento 1 se puede decir que se mostraron las pruebas más relevantes realizadas a los registros de la base de datos de manera individual. Al observar el comportamiento del resultado en la segmentación con respecto a las modificaciones en las variables se logró comprender que variaciones se deben hacer con cada clase de señales para mejorar el desempeño del **algoritmo 1** en la detección, así como buscar un equilibrio en los parámetros.

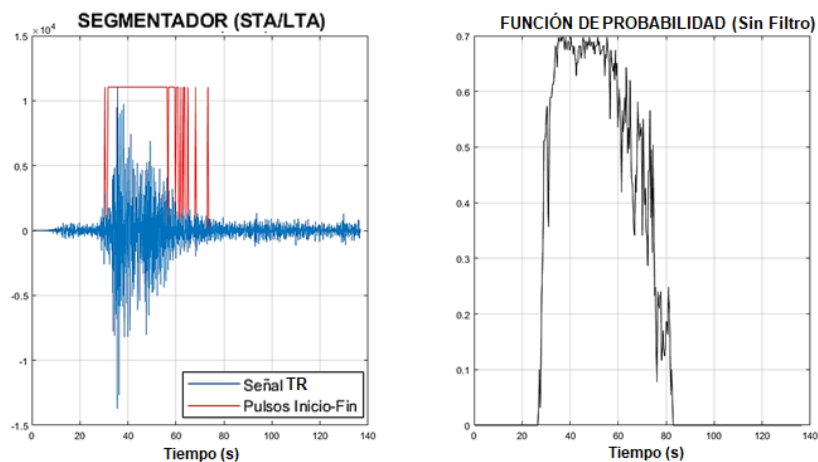
Experimento 2: En la tabla 14 se observa que el menor tiempo de ejecución fue con señales de la clase volcano tectónica (VT), con un tiempo de ejecución de 0,12 segundos.

Tabla 14. Tiempos de ejecución promedio del algoritmo 2 con señales individuales.

Tiempo(s)	Señal
0,12	VT
0,23	LP
0,55	TR
0,23	IC
0,3	SM
0,30	HB
0,43	EX

Los siguientes resultados son producto de procedimientos realizados sobre señales de la clase TR, se tomaron aleatoriamente. Sin embargo, este proceso se llevó a cabo con todas las señales de las 7 clases. Inicialmente en pruebas se empleó los valores: $P=60$ y $K=0.8$. Estos valores se sugerían en [10], como resultado, el algoritmo realizó múltiples detecciones, esto por causa de que la función de probabilidad contiene ruido pronunciado lo conlleva a dichas detecciones falsas.

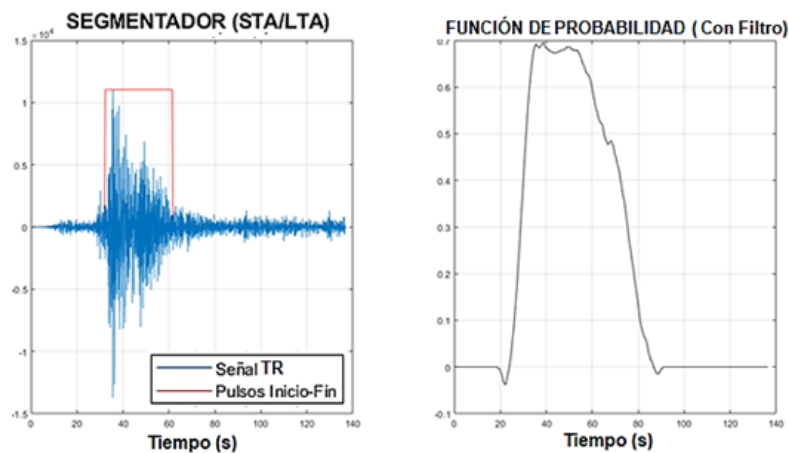
Figura 17. Segmentación y función de probabilidad para una señal.



Fuente: Esta investigación con señal del volcán Nevado del Ruiz registrada por el OVSM en el año 2015.

Se realizó una variedad de pruebas con las etapas del algoritmo 2 descritas en la sección 2, sin embargo, el resultado era el mismo, múltiples detecciones, como solución a esta anomalía se decide adicionar una etapa al proceso propuesto originalmente, esta se trata de un filtro de suavizado a la función de probabilidad, en la figura 18 se observa que el resultado fue exitoso, la detección se realiza de forma correcta.

Figura 18. Segmentación realizada por el algoritmo 2 con filtro de suavizado aplicado a la función de probabilidad.



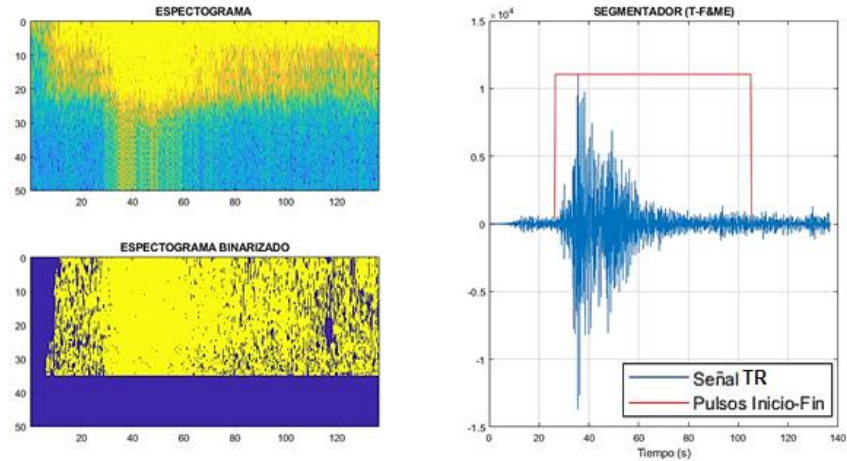
Fuente: Esta investigación con señal del volcán Nevado del Ruiz registrada por el OVSM en el año 2015.

La primera etapa de este algoritmo de segmentación es determinar el valor absoluto del espectrograma de la señal, para este trabajo se calculó con una ventana Haming de 64 muestras y un traslape por defecto del 50%. Esto da como resultado una matriz con diversos valores proporcionales a la energía de la señal. En segunda instancia esta binarizar dicha matriz, esto se hace con un vector de umbrales de comparación que recorre todas las filas de la matriz dada por el espectrograma.

La función de probabilidad depende de la información que se extraiga de la matriz binarizada. Uno de los parámetros relevantes del algoritmo de segmentación es el percentil, para la segmentación mostrada en la en figura 18 se tomó el valor de $P=60$, se disminuye este valor a $P=20$, conservando el valor de K utilizado anteriormente y con la etapa de filtro adicional. Como resultado se ve un aumento en la amplitud ventana de la detección, tanto el espectrograma,

la matriz binarizada y la segmentación resultante con esta modificación se observa en la figura 19.

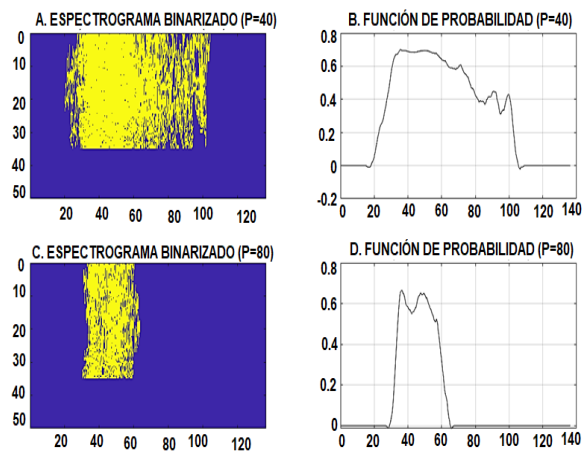
Figura 19. Espectrograma, matriz binarizada y segmentación de la señal con $P=20$ y $K=0.8$.



Fuente: Esta investigación con señal del volcán Nevado del Ruiz registrada por el OVSM en el año 2015.

En la figura 19 se muestra los cambios que sufre la función de probabilidad al aumentar el valor del percentil, la relación entre ambas características es inversamente proporcional.

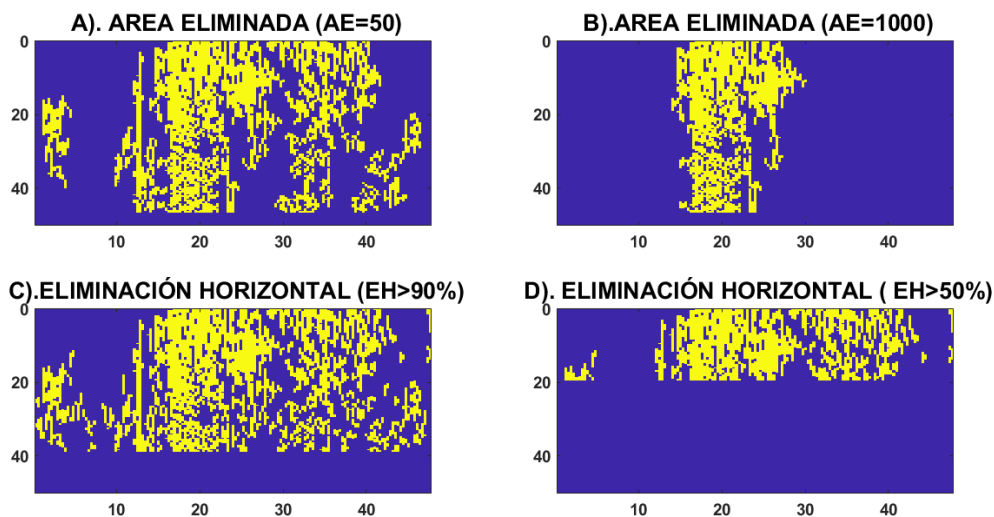
Figura 20. Cambios en la función de probabilidad según el valor de percentil.



Fuente: Esta investigación

En medio de las pruebas realizadas a este algoritmo, se encontró que en señales con ruido abundante la matriz binarizada termina contaminada con áreas sin información útil, debido a esto, se adicionó dos etapas más al proceso, llamadas eliminación de área (AE), elimina las áreas de pixeles de la aimagen binarizada con un determinado valor de AE, si el área de pixeles es menor a este valor, se eliminan. Este proceso se puede detallar en la figura 20 (A) y (B), donde se muestra el resultado de eliminar áreas menores a 50 y áreas menores 1000. Por otra parte se adiciono la etapa que la llamó eliminación horizontal (EH), esta se hace con el fin de mejorar la función de probilidad. Se convierte en cero todos los valores superiores a un numero de fila determinado, esto se observa en la figura 20 (C) y (D). En donde para una señal se convierten en cero todas las filas superiores al 90% del total de filas y para el otro caso superiores al 50%.

Figura 21. Filtros: AE= 50 (A), AE= 1000 (B) y EH>90% (C), EH>50% (D), aplicados a la matriz binarizada.

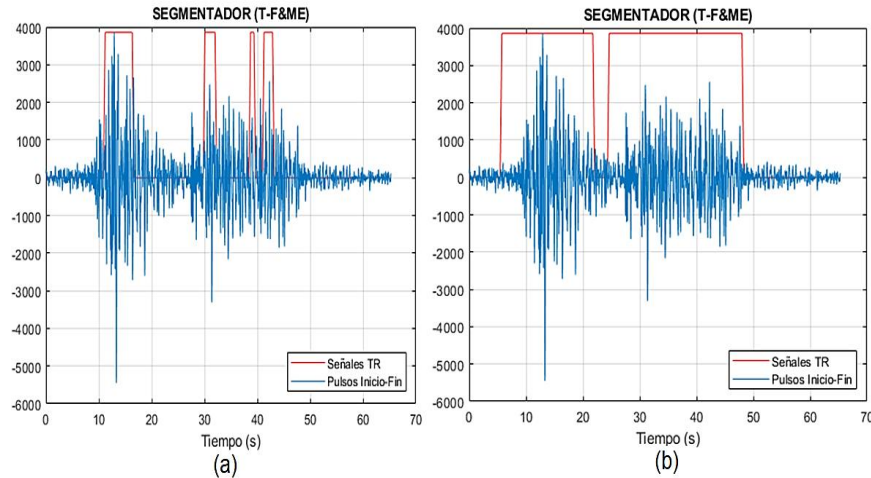


Fuente: Esta investigación

Al igual que en la clase VT en las señales TR también se encontraron registros con múltiples choques etiquetados como una sola señal, así que de la misma forma se realizó la segmentación, los parámetros con valores de $P=80$, $K=0,4$ y un $\text{framelen} = 7$ para el filtro. Como resultado se tiene una decisión fallida que se muestra en la figura 22 (a). Posteriormente se hacen las pruebas y modificaciones correspondientes dando como resultado nuevos parámetros:

$P=55$, $K=0,1$ y $\text{frameleng}=7$. Obteniendo una detección exitosa como se observa en la figura 22 (b).

Figura 22. Detección errónea (a) y detección exitosa (b), señal TR con 2 choques.



Fuente: Señal del volcán Nevado del Ruiz registrada por el OVSM en el año 2015.

Experimento 3: La prueba de porcentaje y tiempo de ejecución para cada intervalo propuesto de señales concatenadas se observa en la tabla 15, este procedimiento se realizó sobre las 4 clases VT, LP, TR y LP, porque cuentan con igual número de registros como se muestra en la tabla 3. Entre las observaciones más relevantes que se pueden hacer al respecto están: el mayor éxito en la relación detección / tiempo, es para el intervalo de 1 a 50 señales con la clase VT.

Por otra parte, a medida que aumenta el rango, el porcentaje de detección disminuye considerablemente. Las señales que presentaron mayor costo computacional son las de la clase TR. Así mismo, la clase que arrojó el porcentaje de detección más bajo es IC, este es uno de los motivos para descartar esta clase de las pruebas de clasificación.

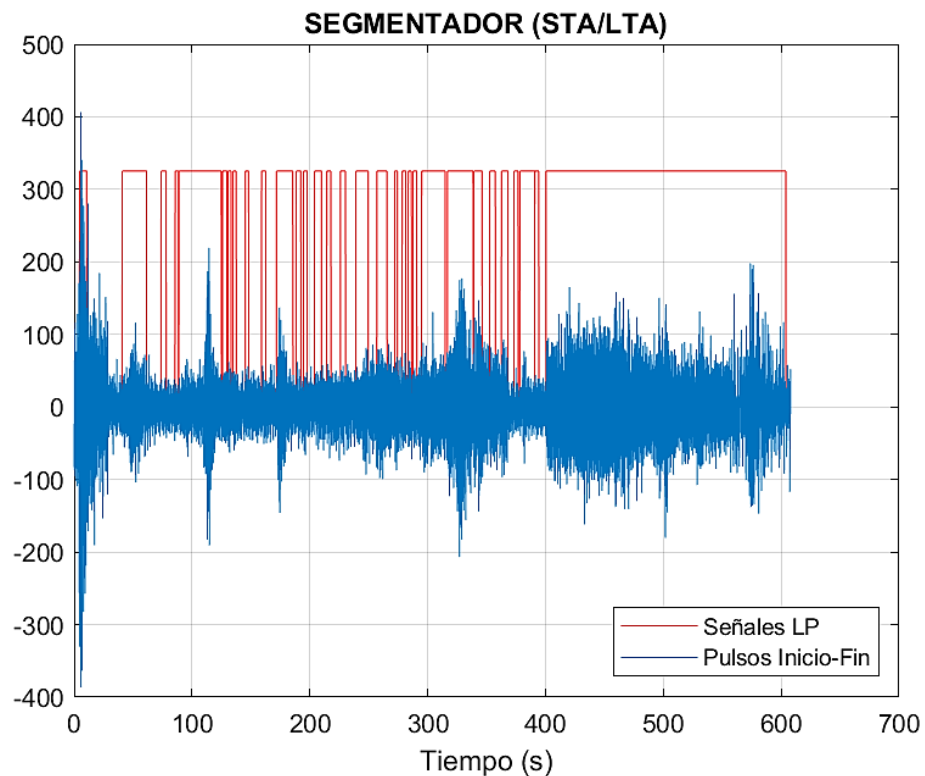
Tabla 15. Tiempo y porcentaje de detección registrados por el algoritmo 1.

Intervalos	VT		LP		TR		IC	
	PD	t(s)	PD	t(s)	PD	t(s)	PD	t(s)
1 a 50	70	0.6	46.00	0.74	18	1.68	18	0.54
1 a 100	62	1.2	44.00	1.55	20	3.32	18	1.21
1 a 150	64	1.8	44.00	2.25	17	4.87	17	1.82

1 a 200	59	2.6	43.00	2.99	25	8.08	19	2.40
1 a 250	54	4	41.00	4.50	30	14.20	17	3.58
1 a 350	50	4.6	40.00	5.15	28	15.53	16	4.11
1 a 400	47	5.1	37.50	5.88	26	16.31	15	4.62
1 a 450	43	5.7	38.00	6.66	25	17.68	14	5.18

Dentro del marco experimental llevado a cabo, con señales concatenadas se tomó una muestra aleatoria de la clase LP, se encontró que el algoritmo 1 funciona de forma incorrecta ante señales con ruido abundante, esto se pueden observar en la figura 23, se realizó numerosas variaciones en los parámetros del algoritmo, pero no se encontró mejoría frente a este tipo de señales.

Figura 23. Detección del algoritmo 1 con señales concatenadas y ruido excesivo.



Fuente: Señales del volcán Nevado del Ruiz registradas por el OVSM en el año 2015.

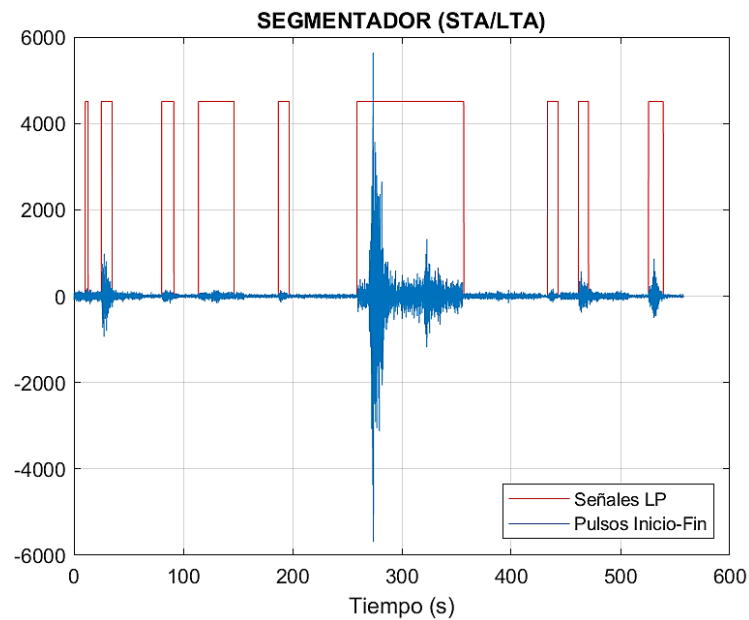
Se tomó una nueva muestra aleatoria de 10 señales de la clase VT y se concatenan. A esta señal se le aplica el algoritmo 1. En la figura 24, se observa que el espacio

entre impulsos es considerable permitiendo así, que el algoritmo realice una detección adecuada, los parámetros usados se muestran en la tabla 16.

Tabla 16. Parámetros para la detección de señales concatenadas tipo LP.

Parámetro	Sta	Lta	Activación	Desactivación
Valor	0.5	5	1	1.2

Figura 24. Detección exitosa para 10 señales concatenadas tipo LP.



Fuente: Señales del volcán Nevado del Ruiz registradas por el OVSM en el año 2015.

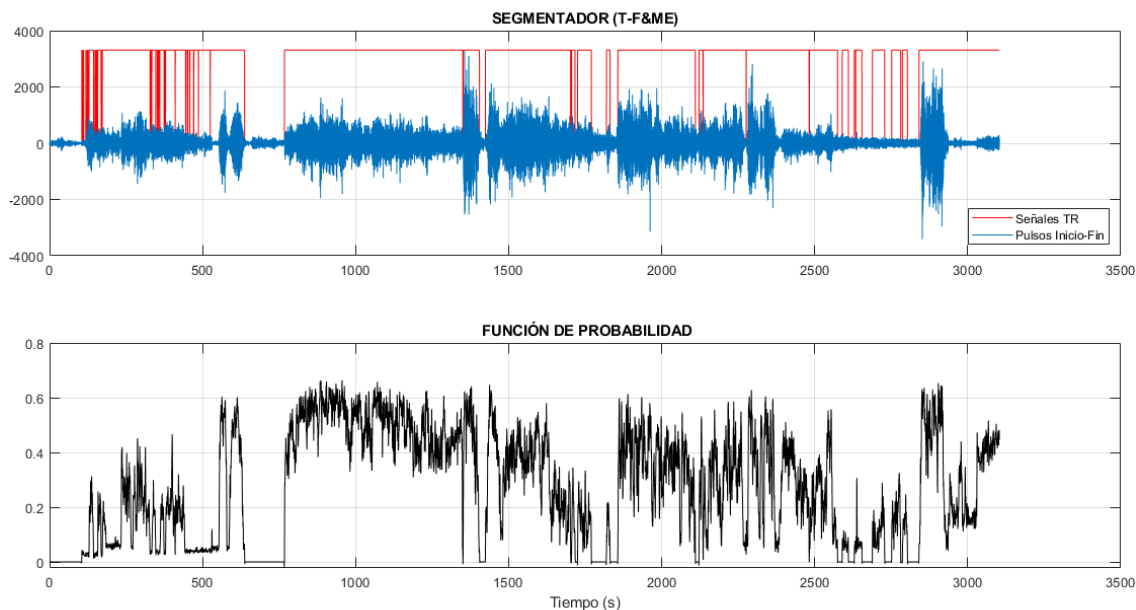
Experimento 4: La prueba de porcentaje de detección y tiempo de ejecución para cada intervalo propuesto de señales concatenadas y segmentadas por el algoritmo 2 se observa en la tabla 17. Las consideraciones son igual a las que se tuvo con la segmentación de señales concatenadas con el algoritmo 1. Entre las observaciones más relevantes que se pueden hacer al respecto están: el mayor éxito en la relación detección / tiempo es para el intervalo de 1 a 50 señales con la clase VT con un aumento del 15% en el porcentaje detección con respecto al algoritmo 1. En general, la detección es mejor con todas las clases. Por otra parte, se conserva la relación, al aumentar el rango el porcentaje de detección disminuye, pero en este caso la disminución es menor. Las señales que presentaron mayor costo computacional son las de la case TR y la clase IC es la que arrojó el porcentaje de detección más bajo, mostrando el mismo comportamiento que el prestado con el algoritmo 1.

Tabla 17. Tiempo y porcentaje de detección del algoritmo 2 según intervalo.

Intervalos	VT		LP		TR		IC	
	PD	t	PD	t	PD	t	PD	t
1 a 50	85	0.8	66	1.1	50	2.68	35	0.54
1 a 100	70	5.24	65	2.85	50	6.32	33	1.21
1 a 150	70	5.86	51	4.25	48	8.87	31	2.2
1 a 200	65	6.60	50	4.99	45	15.08	27	3.40
1 a 250	64	8.05	50	7.50	42	20.5	25	5.8
1 a 350	62	8.60	45	8.15	38	28	23	6.11
1 a 400	60	10.15	43	11.88	35	33.2	21	8.62
1 a 450	57	12.70	45	12.66	33	40.2	20	10.18

De la concatenación de 10 señales de clase TR, y usando los valores de $K=0.2$, $P=60$, $\text{framelen}=7$, en la figura 25 se puede observar que el resultado son detecciones falsas. Además, en la función de probabilidad a medida que aumenta el número de señales concatenadas, aumenta el ruido en gran escala.

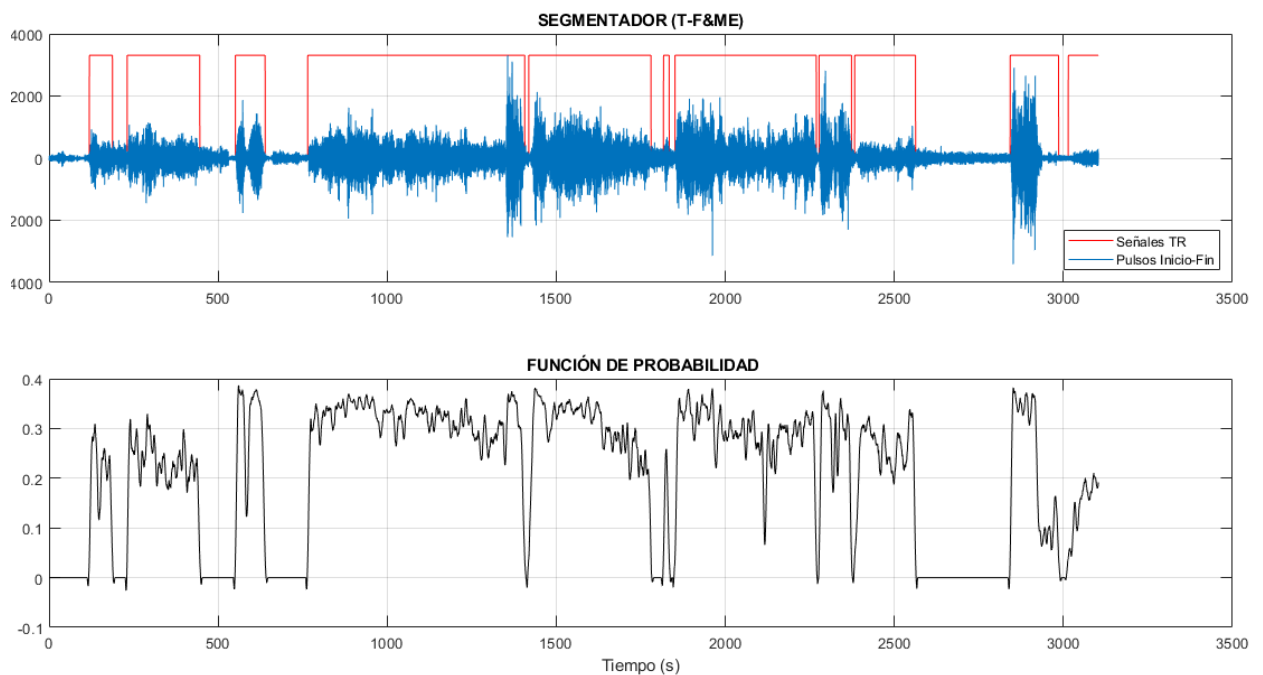
Figura 25. Segmentación y función de probabilidad con señales TR concatenadas con $K=0.2$, $P=60$ y $\text{Framelen}=7$.



Fuente: Esta investigación con señales del volcán Nevado del Ruiz registradas por el OVSM en el año 2015.

Después de realizar las pruebas y cambios respectivos, se obtiene como resultado la función de probabilidad y los pulsos de la señal de detección mostrados en la figura 26. Donde se observa que el algoritmo funcionó de manera correcta, fue necesario aumentar el valor de framelen, es decir que se mejoró el filtrado de la función de probabilidad, así como tomar mayor parte de la matriz binarizada esto mediante la disminución en el valor del percentil.

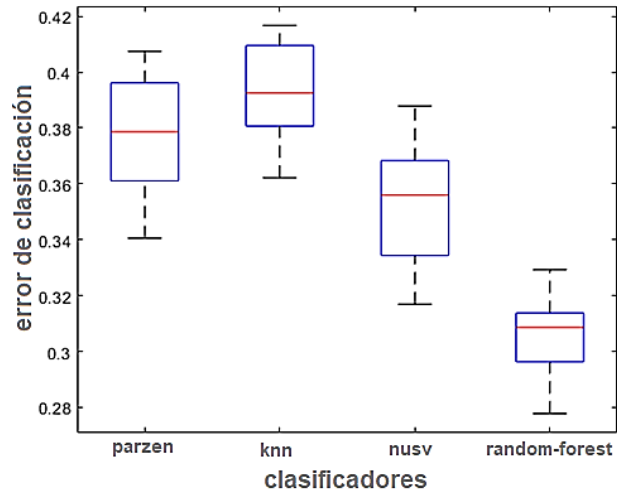
Figura 26. Segmentación y función de probabilidad con señales TR concatenadas con $K=0.1$, $P=40$ y Framelen=51.



Fuente: Esta investigación con señales del volcán Nevado del Ruiz registradas por el OVSM en el año 2015.

Experimento 5: En la figura 27 se observa que el menor error fue para el clasificador 4, correspondiente a Random-Forest, alcanzando un promedio aproximado del 69% en la precisión de clasificación con 3 clases. El segundo lugar en mejor clasificación fue para el clasificador 3, que corresponde a soporte vectorial(nsv), con una precisión en la clasificación del 65% aproximadamente. Con un tiempo de ejecución para 10 iteraciones de 156 segundos.

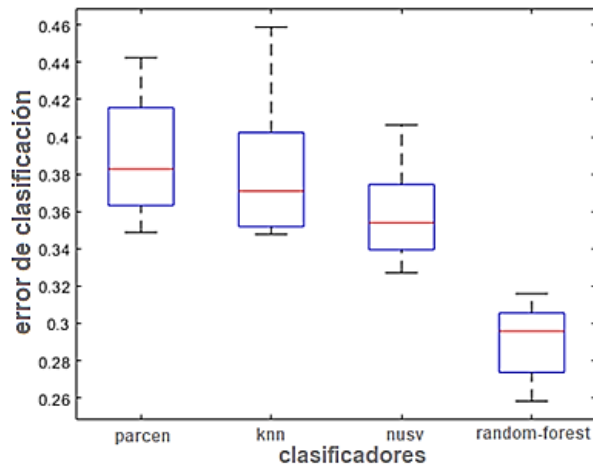
Figura 27. Error en la clasificación con 14 características y base original con 3 clases.



Fuente: Esta investigación

Experimento 6: En la figura 28 se observa que el menor error fue para el clasificador 4 correspondiente a Random-Forest, alcanzando un promedio aproximado del 71% en la precisión de clasificación. Con un tiempo de ejecución para 10 iteraciones de 145 segundos.

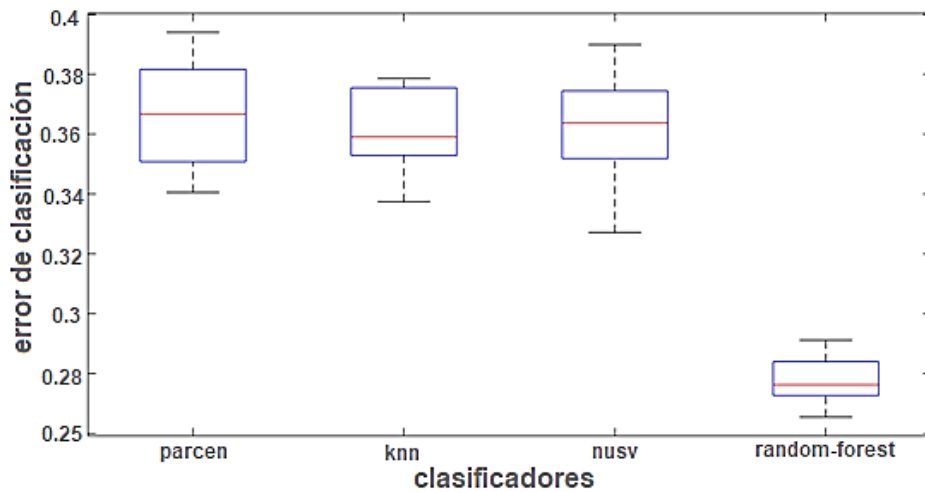
Figura 28. Error en la clasificación con 10 características con Relieff y la base de datos original.



Fuente: Esta investigación

Experimento 7: En la figura 29 se observa que el menor error fue para el clasificador 4 correspondiente a Random-Forest, alcanzando un promedio aproximado del 73% en la precisión de clasificación. Con un tiempo de ejecución para 10 iteraciones de 140 segundos.

Figura 29. Error de clasificación con BestFirst y la base de datos original.



Fuente: Esta investigación

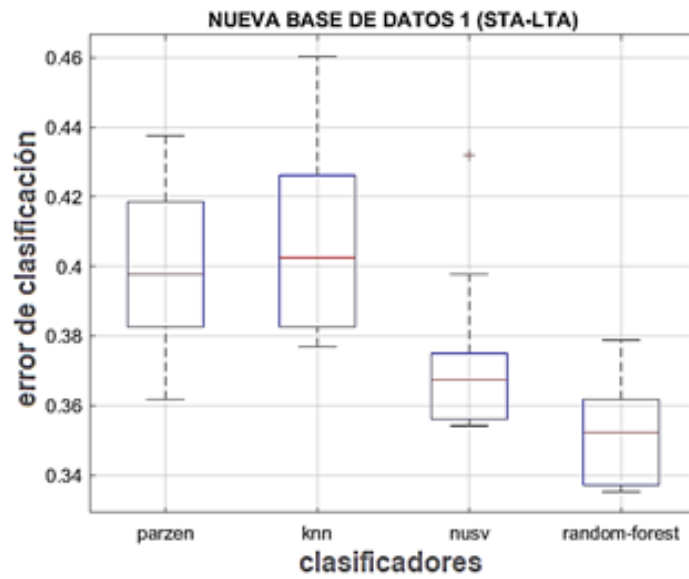
Experimento 8: Con el algoritmo 1 (STA / LTA) de segmentación, se obtuvo una detección exitosa para la clase VT de 350 señales, para LP de 320 señales y 310 para TR, con el fin de que la base de datos generada este balanceada, se tomó los grupos iguales de 310 señales para las 3 clases generando así, la nueva **base de datos 1**, con 930 registros debidamente etiquetados.

Experimento 9: Con el algoritmo 2 (T-F&ME) de segmentación, se obtuvo una detección exitosa para la clase VT de 376 señales, para LP de 332 señales y 317 para TR, con el fin de que la base de datos generada este balanceada, se tomó los grupos iguales de 310 señales. Igual que el algoritmo 1, para las 3 clases generando así, la nueva **base de datos 2** con 930 registros debidamente etiquetados.

Experimento 10: En la figura 30 se observa que el menor error fue para el clasificador Random-Forest alcanzando un promedio aproximado del 63% en la

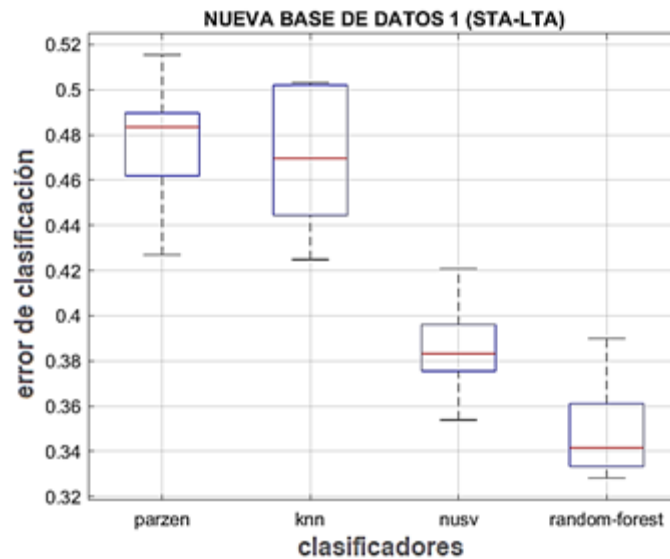
precisión de clasificación. Con un tiempo de ejecución para 10 iteraciones de 109 segundos.

Figura 30. Error en la clasificación con 14 características con nueva **base de datos 1**.



Fuente: Esta investigación

Figura 31. Error en la clasificación con BestFirst y la nueva **base de datos 1**.

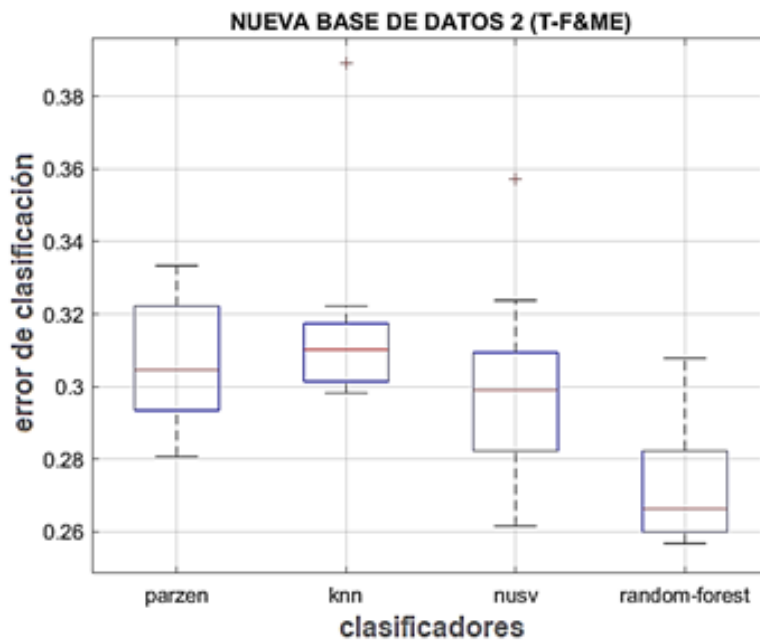


Fuente: Esta investigación

Experimento 11: En la figura 31 se observa que el menor error fue para el clasificador Random-Forest alcanzando un promedio aproximado del 66% en la precisión de clasificación. Con un tiempo de ejecución para 10 iteraciones de 109 segundos.

Experimento 12: En la figura 32 se observa que el menor error fue para el clasificador Random-Forest alcanzando un promedio aproximado del 73,5% en la de clasificación. Con un tiempo de ejecución para 10 iteraciones de 127 segundos.

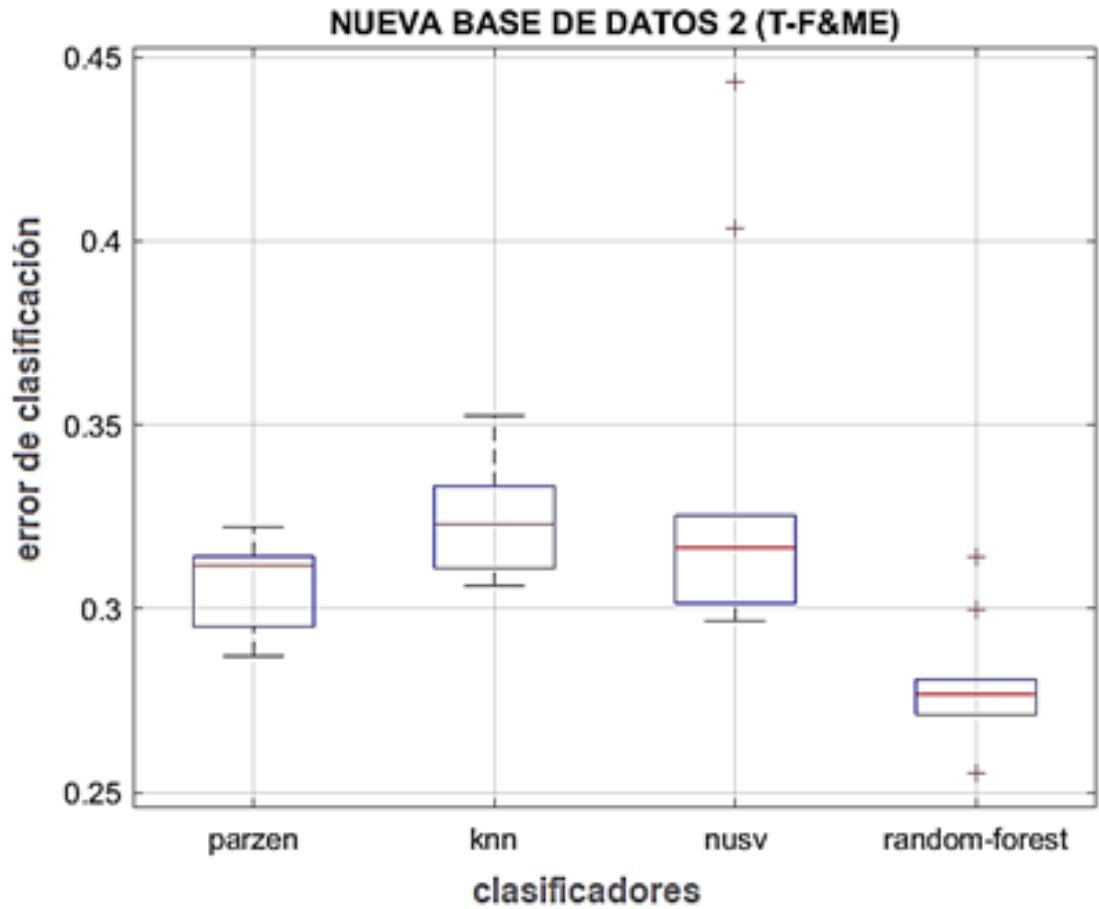
Figura 32. Error en la clasificación 14 características y nueva **base de datos 2**.



Fuente: Esta investigación

Experimento 13: En la figura 33 se observa que el menor error fue para el clasificador Random-Forest, que alcanzó un promedio aproximado del 73% en la precisión de clasificación. Con un tiempo de ejecución para 10 iteraciones de 118 segundos aproximadamente.

Figura 33. Error en la clasificación con BestFirst y nueva **base de datos 2**.



Fuente: Esta investigación

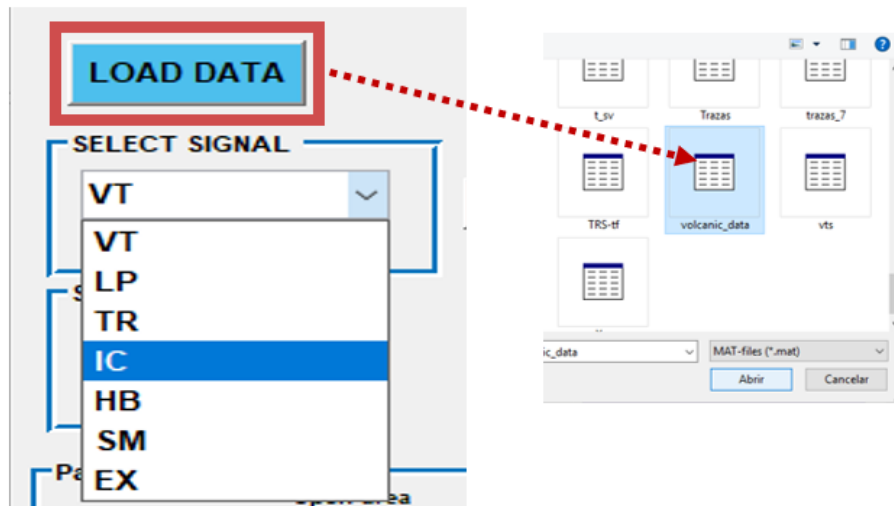
3.1. INTERFAZ DE VISUALIZACIÓN DE RESULTADOS

A continuación, se muestra la forma como trabaja la interfaz desarrollada, se toma pantallas de sus principales funcionalidades.

La metodología sugerida para su uso se divide en 5 etapas principales, se muestran a continuación:

- CARGA DE DATOS Y SELECCIÓN DE LA SEÑAL A TRABAJAR.

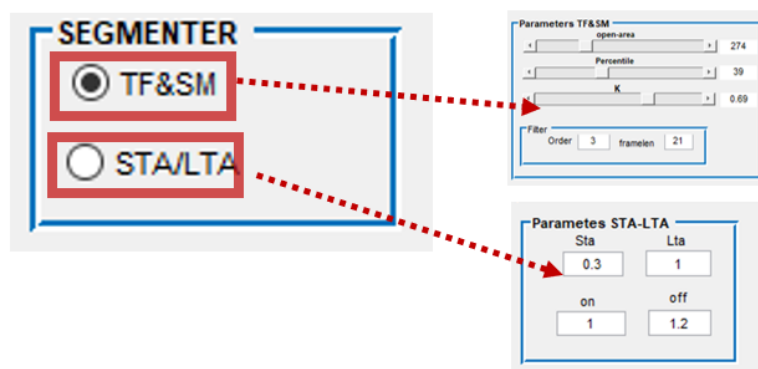
Figura 34. Primera etapa de la interfaz carga base de datos y selección de la clase.



Fuente: Esta investigación

- ELECCIÓN DE ALGORITMOS DE SEGMENTACIÓN Y CONFIGURACIÓN DE SUS PARÁMETROS.

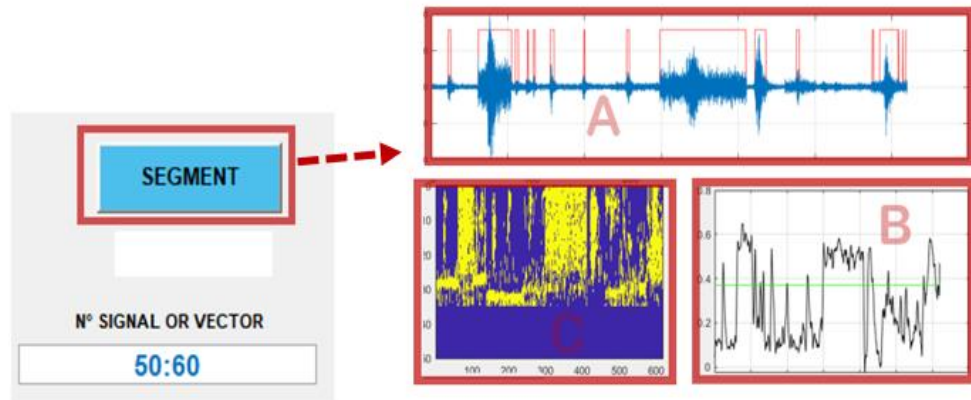
Figura 35. Segunda etapa de la interfaz, se elige algoritmos y parámetros.



Fuente: Esta investigación

- SE ELIGE NÚMERO DE SEÑAL, RANGO Ó VECTOR DE SEÑALES, SE PONE EN MARCHA Y VISUALIZA.

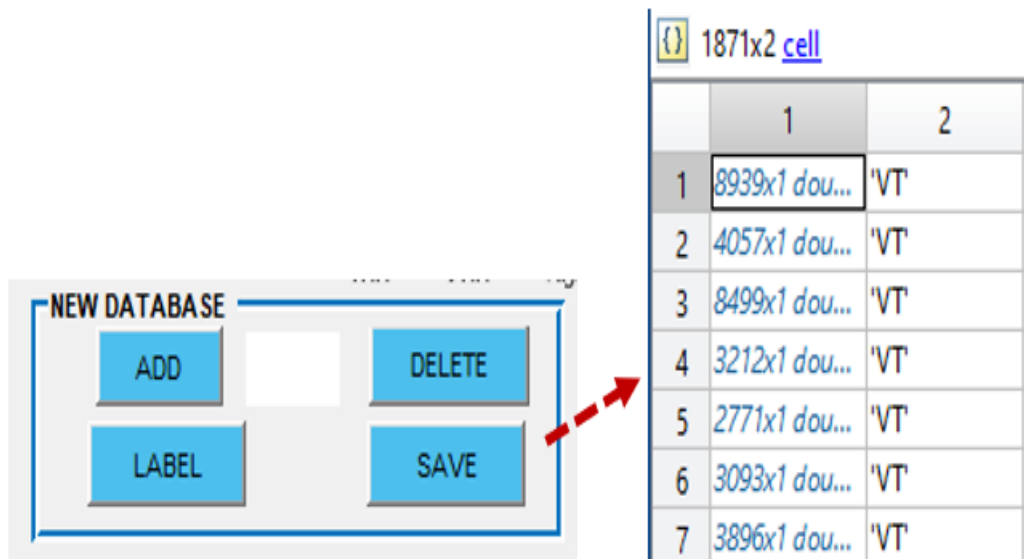
Figura 36. Tercera etapa de la interfaz, número de señales y puesta en marcha.



Fuente: Esta investigación

- SE ADICIONA, ETIQUETA Y GUARDA SEÑALES SEGMENTANDAS, CREANDO BASES DE DATOS NUEVAS.

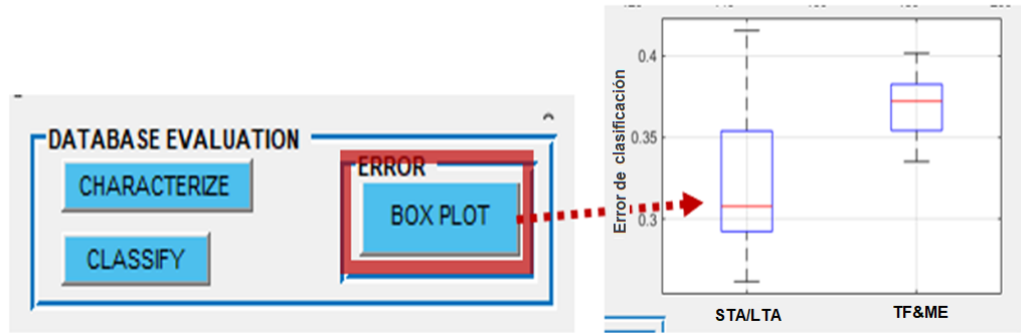
Figura 37. Cuarta etapa de la interfaz, creación de bases de datos.



Fuente: Esta investigación

- SE CARACTERIZA Y CLASIFICA LAS NUEVAS BASES DE DATOS.

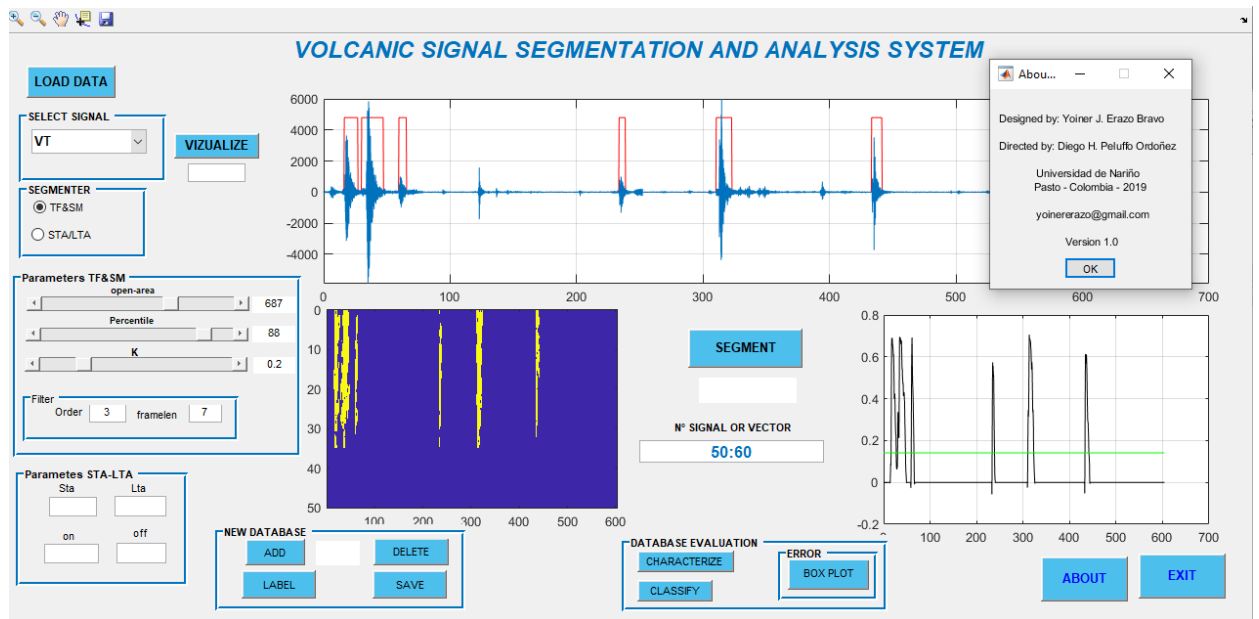
Figura 38. Quinta etapa de la interfaz, evaluación de las nuevas bases de datos (caracterización y clasificación).



Fuente: Esta investigación

Para finalizar la sección de interfaz de resultados, en la Figura 34 se observa una visión general de la primera versión de la interfaz de usuario para la segmentación y análisis de señales sísmico - volcánicas.

Figura 39. Visión general de la interfaz de segmentación de señales volcánicas.



Fuente: Esta investigación

4. CONCLUSIONES Y TRABAJO FUTURO

En términos generales, se implementó un algoritmo diseñado para una mina de carbón, y posteriormente se evaluó con señales sísmico - volcánicas. Además, se realizó una comparación con uno de los algoritmos más usados en la detección de sismos el STA / LTA. Se comprobó la complejidad que implica trabajar con estas señales dada su naturaleza. Además, se evidenció la importancia la segmentación en el desarrollo de sistemas de detección e identificación de señales sísmico - volcánicas.

A nivel más técnico y específico:

- El algoritmo con menor costo computacional en la tarea de segmentación fue el STA / LTA. En comparación el algoritmo T-F&ME, el tiempo promedio de este fue 10 veces mayor que empleado por el algoritmo STA / LTA.
- En cuanto al porcentaje de detección, el algoritmo T-F&ME es el que presentó porcentajes más altos, especialmente con señales de la clase VT. Por otro lado, los dos algoritmos coinciden en que las señales Tremor son las que requirieron mayor tiempo para su procesamiento y las señales de la clase IC presentaron menor porcentaje de identificación en los dos casos.
- Entre los tres casos de selección de características, 10 más altas de ReliefF, BestFirst y con el conjunto total de 14 características, el subconjunto de características dado por el algoritmo BestFirst fue el que mostró mayor porcentaje en la clasificación.
- En cuanto a los métodos utilizados en la clasificación, el método Random-Forest fue el que mejor porcentaje de clasificación presentó tanto con la base de datos original como las bases de datos nuevas 1 y 2. En esta investigación se encontró que el mejor porcentaje de clasificación obtenido fue de 75% con tres clases VT, LP, TR.
- En cuanto a la precisión en la clasificación el algoritmo T-F&ME tuvo un mejor desempeño alcanzando un 73% frente al STA /LTA que alcanzó un 65% como máximo.
- El algoritmo T-F&ME es más robusto al ruido frente al STA / LTA, este último con señales con ruido excesivo perdió su funcionalidad por completo.

Como trabajo futuro, se propone implementar nuevos algoritmos de segmentación encontrados en la literatura y realizar las pruebas con señales sísmicas reales ya que la mayoría de estos métodos funcionan de manera teórica. Además, se propone mejorar el algoritmo STA / LTA con umbrales de activación y desactivación variables, ya que este presenta un costo computacional bajo. Por otro lado, si se realiza un estudio a fondo de los métodos de segmentación de señales sísmico – volcánicas, es posible proponer un nuevo método que cumpla con los requerimientos que requieren los sistemas vulcanológicos a tiempo real.

Adicionalmente, se propone realizar una búsqueda de bases de datos confiales, balanceadas o al menos con un número de registro mayor a 100 por clase para realizar las pruebas de clasificación, en este caso se realizó la mayoría de los procesos de segmentación, caracterización y clasificación con 3 clases, se sugiere ampliar el estudio con las 7 clases VT, LP, TR, IC, SM, HB y EX.

BIBLIOGRAFÍA

- [1] T. D. Popescu, «Signal segmentation using changing regression models with application in seismic engineering,» *Digital Signal Processing*, vol. 24, pp. 14-26, 2014.
- [2] R. M. R. & Z. R. Hossa, «Automatic segmentation of seismic signal with support of innovative filtering,» *International Journal of Rock Mechanics Mining Sciences*, vol. 91, pp. 29-39, 2017.
- [3] T. D. Popescu y A. Dorel, «Signal Segmentation in Time-Frequency Plane using Rényi Entropy - Application in Seismic Signal Processing,» *Conference on Control and Fault-Tolerant Systems (SysTol)*, p. 6, 2013.
- [4] L. A. G. Espinoza, «Sistema de Detección y Clasificación de Señales Sísmico-Volcánicas utilizando Modelos Ocultos de Markov (HMMS): Aplicación a Volcanes Activos de Nicaragua e Italia,» *Editorial de la Universidad de Granada*, 2017.
- [5] J. V. ., Elsa y A. M. ., José, «Análisis de la actividad volcánica tipo VT y LP del volcán Nevadodel Ruiz entre 1985-2012,» *Ventana Informatica*, vol. 33, 2015.
- [6] A. S. Pérez Quisaguano, «Detección automática de eventos sísmicos en el volcán cotopaxi mediante técnicas de aprendizaje de máquinas,» *Universidad de las Fuerzas Armadas (ESPE)*, pp. 9-11, 2018.
- [7] E.-V. Pikoulis y E. Z. Psarakis, «A New Automatic Method For Seismic Signals Segmentation,» *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3973-3976, 2012.
- [8] K. Daniel, W. Agnieszka, O. Jakub, Z. RadosBaw y M. Maciej, «Stochastic Modelling as a Tool for Seismic Signals Segmentation,» *Shock and Vibration*, vol. 2016, pp. 1-13, 2016.
- [9] A. Trnkoczy, «Understanding & Setting STA/LTA Trigger Algorithm Parameters for the K2,» *Application Note, 41*, pp. 16-20, 1998.
- [10] Z. Radoslaw, M. Maciej, Ż. Grzegorz y W. Agnieszka, «Sismic signal segmentation procedure using time-frequency decomposition and statistical modelling,» *Journal of Vibroengineering*, vol. 17, nº 6, pp. 3111-3121, 2015.

- [11] Análisis de señales geológicas Señales sísmicas, Proyecto Especial 1er Cuat 2011.
- [12] «Francisco Gonzales,» *Wikipedia*, 28-03-2019.
- [13] J. Ibáñez y E. Carmona, «Sismicidad volcánica,» *Instituto Andaluz de Geofísica*, pp. 1-14.
- [14] C. G. Barráez, «Detector de eventos sísmicos en tiempo real utilizando redes neuronales,» *Interciencia*, vol. 25, nº 6, pp. 293-298, 2000.
- [15] A. Trnkoczy, «Topic Understanding and parameter setting of STA/LTA trigger algorithm,» *New Manual of Seismological Observatory Practice*, pp. 1-10, 1999.
- [16] J. E. Gabyab y K. R. Anderson, «Hierarchical segmentation of seismic waveforms using affinity,» *Geoexploration*, vol. 23, pp. 1-16, 1984.
- [17] C. CHEN, «On a segmentation algorithm for seismic signal analysis,» *Geoexploration*, vol. 23, nº 1, pp. 35-40, 1984.
- [18] C. Zuolin y R. . Robert, «A multi-window algorithm for real-time automatic detection and picking of P-phases of microseismic events,» *CREWES Research Report*, vol. 18, pp. 1-9, 2006.
- [19] O. J. . César, «Procesamiento digital de señales sísmicas con Matlab,» *Revista de Investigación de Física*, vol. 10, nº 2, pp. 45-50, 2007.
- [20] C. Rodrigo, C. Diana, J. Silva y D. Marcos, «Nuevas Estrategias de Identificación Automática de Tiempos de Llegada de Ondas P y S,» *XVI CONGRESO GEOLÓGICO CHILENO*, vol. 1, pp. 1-4, 2015.
- [21] H. Clément, P. Floriane, M. Jean-Philippe, M. Alessia, S. André y F. Valérie, «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.
- [22] «Caparrini, Clasificación Supervisada y No Supervisada - Fernando Sancho,» [En línea]. Available: <http://www.cs.us.es/~fsancho/?e=77>.
- [23] C. A. González, «SVM: Máquinas de Vectores Soporte,» pp. 52.

- [24] L. Breiman, «Random Forests,» *Mach. Learn*, vol. 45, nº 1, pp. 5-32, Oct. 2001.
- [25] H. Clément, P. Floriane, M. Jean-Philippe, M. Alessia, S. André y F. Valérie, «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.
- [26] H. Azami, S. Sanei y K. Mohammadi, «A Novel Signal Segmentation Method based on Standard Deviation and Variable Threshold,» *International Journal of Computer Applications*, vol. 34, nº 2, 2011.
- [27] M. Basseville y I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, Inc., 1993.
- [28] b. L.-M. D. N. Z. J.-H. X. ,. X.-N. W. H. L. J.-W. Cai-Ping Lu a, «Microseismic frequency-spectrum evolutionary rule of rockburs ttriggeredbyrooffall,» *International Journal of Rock Mechanics & Mining Sciences*, vol. 64, pp. 6-16, 2013.
- [29] R. A. Cueva Lara, V. Paillacho y M. Villalva, «Detección de eventos del Volcán Cotopaxi empleando clasificación supervisada,» *Revista Iberoamericana de las Ciencias Computacionales e Informática*, vol. 5, nº 9, pp. 123-143, 2016.

ANEXOS

ANEXOS 1. Códigos(scripts) resultantes del estudio que se realizó sobre señales volcánicas.

1. ALGORITMO 1 DE SEGMENTACIÓN: STA/LTA

```
clear;clc;close all;
load('vts');load('lps');load('trs');load('ics.mat');
load('trazas_7');

SV=LPS(1:10);

Sx=[];
L=length(SV);
for i=1:L
D=SV{i,1};
Sx=[Sx;D-mean(D)];
End

s=Sx;

tiempo=trazas{8,4};
componente='SHZ';
nombre = 'OLLZ';
startdate = datestr(719529+(tiempo/86400),'yyyy/mm/dd
HH:MM:SS');
tic
%Parametros de sintonización del STA/LTA

l_sta = 0.5; % STA window length (s)
l_lta = 5; % LTA window length (s)
t_on = 1; % STA/LTA trigger on threshold
t_off =1.2; % STA/LTA trigger off threshold
skip_int = 0; % Skip ahead after end of event
min_dur = 2; % Minimum event duration
pre_t = 0; % Added pre-event time
post_t = 0; % Added post-event time
Myedp = [l_sta l_lta t_on t_off skip_int min_dur pre_t
post_t];
```

```

w = waveform(nombre, componente, 100, startdate, s); % Nombre
estación, componente, tasa muestreo, tiempo inicial, traza
FUNCIÓN DEL ALGORITMO
events = sta_lta(w, 'edp', Myedp);
toc;
%%
fs=100; %Frecuencia de muestreo
ls=length(s);%%tamaño del vector
t=(0:ls-1)/fs;%Declaracion variable tiempo
Le = size(events,1);
Lt = length(s);
Dl = zeros(1,Lt);
a=0.8*max((s));
for i=1:Le
    Dl(events(i,1):events(i,2))=0.8*max(s);
end

figure(1)
plot(t,Dl,'r')
hold on
plot(t,s)
title('SEGMENTADOR (STA/LTA)');
xlabel('Tiempo (s)')
legend('Señales LP','Pulsos Inicio-
Fin','Location','southeast')
grid on
hold off

%SEGMENTACIÓN DE LA SEÑAL
Sx_seg={};
[f c]=size(events);
for m=1:f
    Sg=Sx(events(m,1):events(m,2));
    Sx_seg=[Sx_seg;Sg];
end

```

2. ALGORITMO 2 DE SEGMENTACIÓN: TF&ME

```

clear; close all; clc;
load('vts');load('lps');load('trs');load('ics');

SV=VTS(1:20); fs=100;

```

```

Sx=[];L=length(SV);
for i=1:L
D=SV{i,1};
Sx=[Sx;D-mean(D)];
end
t = 0:1/fs:(1/fs)*length(Sx);
t = t(1:end-1);

figure(1)
subplot(321)
plot(t,Sx)
title('Señal Sismica','FontSize',10,'FontWeight','bold')

%% Spectrograma de la señal Sx con una ventana hamming de 64
muestras y un traslape de .5 por defecto
[S,W,T]=spectrogram(Sx,64,[],[],100);
S=20*log10(abs(S));% saca la poencia en dB

% Grafica spectrograma
subplot(323);
image(T,W,S);
title('Espectrograma','FontSize',10,'FontWeight','bold')

%% Binarizacion spectrogram
[f c]=size(S);
Xpf=[]; % Vector cuantil
q=60; %numero cuantil%90
qt=round((c*q)/100); %cuantil
for i=1:f
    X=(S(i,:));
    AX=sort(X);
    Xpf=[Xpf;AX(qt)];
end

%% Binarizacion
BS=S;
for i=1:f
    for j=1:c
        if BS(i,j)>Xpf(i)
            BS(i,j)=1;
        else
            BS(i,j)=0;
        end
    end
end

```

```

        end
    end
end
for i=1:f
    for j=1:c
        if i>50
            BS(i,j)=0;
        end
    end
end
end

BS = bwareaopen(BS,500);

%% Grafica
subplot(325)
imagesc(T,W,BS)
title('Espectrograma
Binarizado','FontSize',10,'FontWeight','bold')

%% Funcion probabilidad de frecuencias
Pt=[];
for i=1:c
    P1=(BS(:,i));
    Sp1=sum(P1)/f;
    Pt=[Pt;Sp1];
end
Pt=sgolayfilt(Pt, 3, 15);
subplot(322)
plot(T,Pt,'k') %% Grafica funcion probabilidad
title('Función de
Probabilidad','FontSize',10,'FontWeight','bold')

%% Funcion pulsos inicio y fin de cada evento
D=Pt;
for n=1:10
    D(n)=0;
end
k=0.05;
L=(k*max(D));

for i=1:c
    if D(i)>L
        D(i)=max(Sx);
    else

```

```

        D(i)=0;
    end
end
subplot(324)
plot(T,D,'r');
title('Pulsos Inicio-Fin','FontSize',10,'FontWeight','bold')
hold on
plot(t,Sx);

%% calculo tiempos inicio y fin de la señal
in=[];fin=[];j=0;k=0;
fsD=length(D)/(length(Sx)/fs);%% Calculo frecuencia de
muestreo vector pulsos i f
for n=2:c
    if (D(n)-D(n-1))>0
        j=j+1;
        in(j)=floor(n/fsD);
    end
    if (D(n)-D(n-1))<0
        k=k+1;
        fin(k)=ceil(n/fsD);
    end
end

%% calculo muestras inicio y fin de la señal
mi=[];mf=[]; h=0; k=0;% muestra inicialy final-puntos de
corte

for i=1:length(fin)
    for j=1:length(t)
        if round(t(j),2) == in(i)
            h=h+1;
            mi(h)=j;
        end
        if round(t(j),2) == fin(i)
            k=k+1;
            mf(k)=j;
        end
    end
end

%% Segmentación de la señal

```

```

toc
Sx_seg={};
for i=1:length(mf)
    Sx_seg{i,1}=Sx(mi(i):mf(i));
end

```

3. CARACTERIZACIÓN

```

clear;clc;close all;
%Inicializar DIP_IMAGE TOOLBOX
dip_initialise;
%%
load('Soll.mat');
%%
SX=SV;
L=length(SX);fs=100;n=2;
MCAR1=[];
for i=1:L
    sn=normalizar(SX{i});
    matriz_car=CaracteristicasCurilem(sn,fs);
    MCAR1=[MCAR1;matriz_car];
end
MCAR2=[];
for i=1:L
    sn=normalizar(SX{i});
    IBON=CaracteristicasIbsVonSeht(sn,fs,0);
    MCAR2=[MCAR2;IBON];
end
MCAR=[MCAR1 MCAR2];
save('Matriz_de_caracteriticas.mat','MCAR');

```

4. SELECCIÓN DE CARACTERÍSTICAS

```

clear;clc;close all;
load('
Matriz_de_caracteriticas.mat');load('Etiquetas.mat');

%SELECCIÓN POR RELIEFF
[ranks,weights] = relieff(MCAR,Etiquetasn,10);
%%
ranks=C; pos=find(ranks); CS=pos(1:10);
[f c]=size(MCAR);
MCARS=[];

```

```

for i=1:length(CS)
    a=CS(i);
    MCARS=[MCARS MCAR(:,a)];
end
save('MCAR3C_Featself.mat','MCARS');
%%
%SELECCIÓN POR FEATSELF
A = prdataset(MCAR_1,Etiquetasn);
[W R]= featself(A,'NN',0);

```

5. CLASIFICACIÓN

```

clear;clc;close all;
load('Etiquetas');load('Matriz_de_caratecriticas');
Datos1 = MCAR;
Resultado =Etiquetas;
%%
tic
Datos1 = Datos1./max(abs(Datos1)),size(Datos1,1);
carmay =[Datos1 Resultado];
Con=0;
Time=0;
Errores=[];
for i=1:10
cont=0;Xbc=[];XBc=[];YBc1=[];XTest=[];X_CT=[];
Y_CT=[];porcentaje=0.2;porctest=0.1;flag1=1;

for k = 1:length(unique(carmay(:,end)))
    cc = find(carmay(:,end)==k);
    num = length(cc);
    Data(k).Carclas = carmay(cc(1:num),:);
    [Data(k).Row numattrs] = size(Data(k).Carclas);
    R = randperm(num);
    Data(k).rand=Data(k).Carclas(R(1:num),:);
    tamatest=porctest*(length(Data(k).Carclas(R(1:num),:)));
    tamatest=floor(tamatest);
    for i=1: tamatest
        Data(k).Xtest(i,:)=Data(k).rand(i,:);
    end
    Data(k).rand(1:tamatest,:)=[];
    Xbc=Data(k).rand;
    XBc=[XBc;Xbc];

```

```

Xtest=Data(k).Xtest;
XTest=[XTest;Xtest];
X_test=XTest;
X_test(:,end)=[];
Y_test=XTest(:,end);
X_BC=XBc;
X_BC(:,end)=[];
Y_BC= XBc(:,end);
Toma(k)=porcentaje*length(Y_BC);
Toma(k)=floor(Toma(k));
end

% Base de casos Entrenamiento

A = prdataset(X_BC,Y_BC);
[C,D] = gendat(A,Toma(k));

% Clasificadores en Cascada

T10= parzenc; W10= C*parzenc;
T11= knnc; W11= C*T11;
T12= nusvc; W12= C*T12;
T13= randomforestc; W13= C*T13;

% Errores de los clasificadores
[e10,c10] = testc(D*W10);
[e11,c11] = testc(D*W11);
[e12,c12] = testc(D*W12);
[e13,c13] = testc(D*W13);

errores =[e10 e11 e12 e13];
Errores=[Errores;errores];
end
toc
boxplot(Errores);
x1=Errores(:,1);x2=Errores(:,2);x3=Errores(:,3);x4=Errores(:,
4);
boxplot([x1,x2,x3,x4], 'Labels', {'parzen', 'knn', 'nusv', 'random
-forest'})
xlabel('Clasificadores'); ylabel('Error de clasificación');
title('NUEVA BASE DE DATOS 1 (STA-LTA)');grid on

```


ANEXOS 2. Dos Artículos para enviar a: 6th IEEE Latin American Conference on Computational Intelligence, LA-CCI 2019.

1.

A Comparison of algorithms for the seismic-volcanic signals segmentation : preliminary results

Y. Erazo ¹, A. Cabrera ¹, E. Rosero ¹, D. H. Peluffo-Ordóñez^{2,3}, and P. A. Castro-Cabrera⁴

¹ Universidad de Nariño, Colombia,

² Corporación Universitaria Autónoma de Nariño, Colombia

³ Yachay Tech, Ecuador.

³ Universidad Nacional de Colombia.

Abstract. In a seismic-volcanic event, it is often found that an observation (a signal) contains several impulses (discharges) related to a seismic event. Therefore, for a detailed analysis of several pulses of a signal requires a separate process. For this purpose, a stage called segmentation is carried out. In this paper, a comparative study of two segmentation algorithms is presented. The first is known by the acronym *sta / lta*, which is based on the quotient of the averages of short (*sta*) and long (*lta*) term with respect to a threshold. The second algorithm is based on the transformation of the signal to time frequency to obtain a statistical model and in this way determine a probability function. The measures of performance to take into account for comparison are: computational cost, classification, among others. From waveform, spectral and pseudo-spectral characteristics.

2.

Seismic-Volcanic Signals Glacial events and low frequency : Preliminary results

A. Cabrera ¹, Y. Erazo ¹, E. Rosero ¹, D. H. Peluffo-Ordóñez^{2,3}, and P. A. Castro-Cabrera⁴

¹ Universidad de Nariño, Colombia,

² Corporación Universitaria Autónoma de Nariño, Colombia

³ Yachay Tech, Ecuador.

⁴ Universidad Nacional de Colombia.

Abstract. The seismological behavior of the volcanoes presents numerous seismic events of different kinds, where each one provides great information about the volcanic building on its external and internal activity, such as the movement of flows and magma. A detailed analysis of the activity of the volcano from its seismic events requires detection and recording of the signals from the vulcanological stations, characterization and finally a quantitative analysis of the same in order to obtain accurate information of the volcanic system, carried out by expert personnel in vulcanological observatories, where in most cases a manual classification is performed, being complicated and easy to fail in tempopradas where the volcano is more active and the seismic events recorded are hundreds per day. At present, research has been developed that proposes the automatic classification of volcanic events, based on machine learning techniques, in order to determine the type of event recorded. However, there are no specific methods of characterization and classification. In the literature there are several characterization methods developed by different authors in many cases restricted to the volcano under study, for which the study of volcanic earthquakes and the evaluation of the associated risks are still open and timely issues. This article presents a study of two classes of volcanic seismic events which are at risk of confusion in the detection of warning systems, the events generated by fractures and ice sliding in volcanoes that have a glacial layer and the events of low frequency that due to its waveform and composition are difficult to locate and often found at shallow depths of the surface. It is proposed to optimize the characterization using different techniques and methods of machine learning through supervised techniques from a database of volcanic seismic events supplied.

ANEXOS 3. Página web

Dentro del desarrollo de este proyecto, uno de los productos esperados fue el diseño una página web en la que se pueda subir información relacionada a la 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-signals-segmentation>

