



Sistema de reconocimiento de microterremotos en tiempo real del volcán Cotopaxi aplicando aprendizaje supervisado.

Altamirano Rodríguez, Brayan Santiago

Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Trabajo de titulación, previo a la obtención del título de Ingeniero en Electrónica y
Telecomunicaciones

Ing. Lara Cueva, Román Alcides PhD.

03 de marzo del 2021

URKUND

Document Information

Analyzed document Tesis_AltamiranoSantiago.pdf (D97138221)
Submitted 3/3/2021 9:54:00 PM
Submitted by
Submitter email bsaltamirano@espe.edu.ec
Similarity 2% 
Analysis address ralara.espe@analysis.arkund.com

Sources included in the report

- | | | |
|----------|---|--|
| W | URL: http://repositorio.espe.edu.ec/bitstream/21000/15514/1/T-ESPE-040582.pdf
Fetched: 11/29/2020 8:16:07 AM |  10 |
| W | URL: https://www.diegopeluffo.com/DegreeTheses/2019_DegreeThesis_volcano_segmentation.pdf
Fetched: 12/8/2020 6:00:21 AM |  1 |
| W | URL: https://bibdigital.epn.edu.ec/bitstream/15000/20723/1/CD%2010239.pdf
Fetched: 7/19/2020 3:58:46 AM |  3 |



**DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y
TELECOMUNICACIONES**

**CARRERA DE INGENIERÍA EN ELECTRÓNICA Y
TELECOMUNICACIONES**

CERTIFICACIÓN

Certifico que el trabajo de titulación, "**Sistema de reconocimiento de microterremotos en tiempo real del volcán Cotopaxi aplicando aprendizaje supervisado.**" fue realizado por el señor **Altamirano Rodríguez, Brayan Santiago**, el cual ha sido revisado y analizado en su totalidad por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 5 de marzo del 2021

Firma:



Firmado electrónicamente por:
**ROMÁN
ALCIDES LARA**

Dr. Lara Cueva, Román Alcides

C. C 1713988218



**DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y
TELECOMUNICACIONES**

**CARRERA DE INGENIERÍA EN ELECTRÓNICA Y
TELECOMUNICACIONES**

RESPONSABILIDAD DE AUTORÍA

Yo **Altamirano Rodríguez, Brayan Santiago**, con cédula de ciudadanía n°1723705701, declaro que el contenido, ideas y criterios del trabajo de titulación: **"Sistema de reconocimiento de microterremotos en tiempo real del volcán Cotopaxi aplicando aprendizaje supervisado."** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 5 de marzo del 2021

Altamirano Rodríguez Brayan Santiago

C.C.: 1723705701



**DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y
TELECOMUNICACIONES**

**CARRERA DE INGENIERÍA EN ELECTRÓNICA Y
TELECOMUNICACIONES**

AUTORIZACIÓN DE PUBLICACIÓN

Yo Altamirano Rodríguez, Brayan Santiago, con cédula de ciudadanía n°1723705701, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **"Sistema de reconocimiento de microterremotos en tiempo real del volcán Cotopaxi aplicando aprendizaje supervisado."** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 5 de marzo del 2021

Altamirano Rodríguez, Brayan Santiago

C.C.: 1723705701

DEDICATORIA

Dedico este proyecto y todos los logros a mi Dios, por darme salud, vida y fuerzas para culminar una de las etapas más importantes de mi vida, a mi familia, a mi madre María, mi padre Santiago y mi hermana Priscila por ser mi apoyo incondicional y mi inspiración para seguir adelante en culminar mi carrera universitaria esto es para ustedes, finalmente esto es para ti mi querida abuelita María Ernestina que me apoyaste hasta el último día que estuviste a mi lado.

Brayan Santiago Altamirano Rodríguez

AGRADECIMIENTO

Quiero dar gracias a Dios por cada día de mi vida, por la fuerza, la paciencia y la sabiduría que me ha brindado durante mi trayectoria universitaria, por la oportunidad de culminar con éxito una nueva etapa de mi vida, porque ha sido mi guía y mi pilar que me ha sostenido para alcanzar mis metas, y una de ellas el obtener mi título universitario.

A mi familia, mis padres y a mi hermana por su gran amor y su apoyo constante, por permanecer a mi lado en todos los momentos buenos y difíciles que se han presentado en mi vida, sus consejos y enseñanzas me han ayudado a salir adelante y ser cada día una mejor persona.

A mi novia que ha permanecido a mi lado y me ha brindado su apoyo en cada momento, por su confianza en mí y su comprensión que me ha llenado de fuerza en los momentos difíciles.

Un agradecimiento especial al Dr. Román Lara por su colaboración como director del proyecto, por su paciencia y su ayuda en el desarrollo de cada idea tomada para la culminación del proyecto y la revisión del mismo.

Brayan Santiago Altamirano Rodríguez

ÍNDICE DE CONTENIDOS

Reporte de verificaciones de similitud	2
Certificado del director del proyecto	3
Responsabilidad de Autoría	4
Autorización de Publicación	5
Dedicatoria.....	6
Agradecimiento.....	7
Índice de contenidos.....	8
Índice de tablas.....	11
Índice de Figuras	14
Resumen	19
Abstract	20
Capítulo I	21
Introducción	21
<i>Antecedentes</i>	21
<i>Justificación e Importancia</i>	23
<i>Alcance del Proyecto</i>	24
<i>Objetivos</i>	25
<i>Trabajos Relacionados</i>	26
<i>Organización</i>	28

Capítulo II	30
Marco Teórico.....	30
<i>Vulcanología</i>	30
<i>El Volcán Cotopaxi</i>	35
Conceptos Básicos.....	45
Capítulo III	66
Materiales y Métodos	66
Detector	68
Clasificación	78
Sistema de reconocimiento	84
Capítulo IV	85
Resultados	85
Detección	85
Clasificación	110
Sistema de Reconocimiento en tiempo real.....	141
Evaluación del sistema de reconocimiento de microterremotos en tiempo real con los expertos del IGEPN.....	142
Comparación con trabajos relacionados	146
Capítulo V	150
Conclusiones y Recomendaciones	150
Conclusiones	151

Trabajos Futuros	153
Referencias	154

ÍNDICE DE TABLAS

Tabla 1 <i>Registro de tipos de volcanes, su lava predominante, estilo y características eruptivas..</i>	31
Tabla 2 <i>Volcanes En estado eruptivo desde el año 1532.....</i>	34
Tabla 3 <i>Registro de los cinco grandes periodos eruptivos del volcán Cotopaxi.....</i>	36
Tabla 4 <i>Características de cada segmento de la señal xi.....</i>	49
Tabla 5 <i>Tipos de aprendizaje de Máquina.....</i>	53
Tabla 6 <i>Matriz de confusión.....</i>	63
Tabla 7 <i>Métricas de rendimiento para balance de datos 84 Características.....</i>	86
Tabla 8 <i>Métricas de rendimiento para balance de datos al remover los componentes DC 84 Características.....</i>	87
Tabla 9 <i>Métricas de rendimiento para balance de datos 257 Características PSD.....</i>	88
Tabla 10 <i>Métricas de rendimiento para balance de datos al remover componentes DC 257 Características PSD.....</i>	90
Tabla 11 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 30 Características.....</i>	94
Tabla 12 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 100 Características PSD.....</i>	96
Tabla 13 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 7 Características.....</i>	98
Tabla 14 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 10 Características PSD.....</i>	100
Tabla 15 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 21 Características.....</i>	103

Tabla 16 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 100 Características PSD.</i>	105
Tabla 17 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con las 13 mejores Características.</i>	106
Tabla 18 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 100 mejores Características PSD.</i>	107
Tabla 19 <i>Métricas de rendimiento sistema de votación con 5 y 3 modelos sistema de detección de microterremotos.</i>	109
Tabla 20 <i>Métricas de rendimiento sistema de votación para la clasificación de 82 y 563 microterremotos con 84 características.</i>	112
Tabla 21 <i>Métricas de rendimiento sistema de votación para la clasificación de 82 y 563 microterremotos con 257 características PSD.</i>	113
Tabla 22 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 39 Características.</i>	116
Tabla 23 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 50 Características PSD.</i>	119
Tabla 24 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 3 Características.</i>	122
Tabla 25 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 6 Características PSD.</i>	125
Tabla 26 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 6 Características.</i>	128
Tabla 27 <i>Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 65 Características PSD.</i>	133

Tabla 28 <i>Métricas de rendimiento sistema de votación con 39 Características</i>	135
Tabla 29 <i>Métricas de rendimiento sistema de votación con 50 Características PSD</i>	136
Tabla 30 <i>Métricas de rendimiento sistema de votación con 5 y 3 modelos sistema de clasificación de microterremotos</i>	139
Tabla 31 <i>Métricas de rendimiento detección y clasificación de microterremotos 21 de diciembre</i>	144
Tabla 32 <i>Métricas de rendimiento detección y clasificación de microterremotos 22 de diciembre</i>	146
Tabla 33 <i>Características de interés (Lara Cueva, 2015)</i>	147

ÍNDICE FIGURAS

Figura 1 <i>Volcán Cerro Azul de la Isla Isabela en Galápagos</i>	32
Figura 2 <i>El Volcán Tungurahua</i>	32
Figura 3 <i>El Volcán Sangay</i>	33
Figura 4 <i>Volcán Cotopaxi (2015)</i>	35
Figura 5 <i>Red de Monitorización del Volcán Cotopaxi</i>	37
Figura 6 <i>Ondas de un evento microsísmico</i>	38
Figura 7 <i>Forma de propagación de las ondas P</i>	39
Figura 8 <i>Forma de propagación de las ondas S</i>	39
Figura 9 <i>Forma de propagación de las ondas L</i>	40
Figura 10 <i>Ejemplo de evento LP y su espectro en frecuencia</i>	41
Figura 11 <i>Ejemplo evento VT y su espectro en frecuencia</i>	42
Figura 12 <i>Ejemplo evento TRE de frecuencia intermedia y su espectro</i>	44
Figura 13 <i>Ejemplo evento HYB y su espectro en frecuencia</i>	45
Figura 14 <i>Ejemplo de clasificación con algoritmo kNN</i>	55
Figura 15 <i>Estructura de algoritmo Árboles de Decisión</i>	56
Figura 16 <i>Ejemplo de regiones que se dividen una base de datos</i>	57
Figura 17 <i>Ejemplo de Hiperplano SVM</i>	59
Figura 18 <i>Ejemplo de Hiperplano con método de Kernel</i>	60
Figura 19 <i>Diagrama de bloques general del sistema de reconocimiento</i>	66
Figura 20 <i>Diagrama de Bloques de la Investigación</i>	67
Figura 21 <i>Base de datos 1</i>	69
Figura 22 <i>Comparación de señal VT original y normalizada</i>	70
Figura 23 <i>Señal segmentada en ventanas de 15 segundos</i>	71

Figura 24 <i>Primer segmento de 15 segundos señal de tipo VT.</i>	72
Figura 25 <i>Etiquetamiento de segmentos señal del tipo VT</i>	73
Figura 26 <i>Obtención de 84 Características segmentos de 15 segundos.</i>	74
Figura 27 <i>Obtención PSD con una resolución de 257 puntos para cada segmento de 15 segundos.</i>	75
Figura 28 <i>Proceso de entrenamiento detección de microterremotos.</i>	76
Figura 29 <i>Base de datos 2</i>	79
Figura 30 <i>Comparación entre la señal VT original a 50 Hz y normalizada a 100 Hz.</i>	80
Figura 31 <i>Obtención de las 84 características de las señales normalizadas</i>	81
Figura 32 <i>Obtención PSD con una resolución de 257 puntos para cada señal normalizada</i>	82
Figura 33 <i>Proceso de entrenamiento clasificación de microterremotos</i>	82
Figura 34 <i>Evolución del BER al balancear y retirar los componentes DC a las proporciones de segmentos con señal y sin señal.</i>	91
Figura 35 <i>Métricas de rendimiento para la selección de características usando método de filtro.</i>	93
Figura 36 <i>Análisis del BER para la selección de características usando método de filtro.</i>	93
Figura 37 (a) <i>Métricas de rendimiento para la selección de características PSD usando método de filtro, (b) Frecuencias de interés selección de características método de filtro.</i>	95
Figura 38 <i>Análisis del BER para la selección de características PSD usando método de filtro.</i>	96
Figura 39 <i>Representación de árbol de decisión considerando 84 características.</i>	97
Figura 40 (a) <i>Representación de árbol de decisión considerando 257 características PSD,(b) Frecuencias de interés selección de características método embebido.</i>	99
Figura 41 <i>Métricas de rendimiento para la selección de características PSD usando método de envoltura.</i>	101

Figura 42 <i>Análisis del BER para la selección de características PSD usando método de envoltura.</i>	102
Figura 43 (a) <i>Métricas de rendimiento para la selección de características PSD usando método de envoltura,</i> (b) <i>Frecuencias de interés selección de características método de envoltura.</i>	104
Figura 44 <i>Análisis del BER para la selección de características PSD usando método de envoltura.</i>	105
Figura 45 <i>Modelos finales sistema de detección de microterremotos</i>	108
Figura 46 <i>Sistema de votación con 5 modelos sistema de detección de microterremotos</i>	108
Figura 47 <i>Sistema de votación con 3 modelos sistema de detección de microterremotos</i>	109
Figura 48 <i>Matriz de confusión clasificación 82 microterremotos con 84 características.</i>	111
Figura 49 <i>Matriz de confusión clasificación 563 microterremotos con 84 características.</i>	111
Figura 50 <i>Matriz de confusión clasificación 82 microterremotos con 257 características PSD.</i> ..	112
Figura 51 <i>Matriz de confusión clasificación 563 microterremotos con 257 características PSD.</i> 113	
Figura 52 <i>Métricas de rendimiento para la selección de características del sistema de clasificación usando el método de filtro</i>	115
Figura 53 <i>Análisis del BER para la selección de características del sistema de clasificación usando el método de correlación</i>	115
Figura 54 <i>Matriz de confusión clasificación 82 microterremotos con 39 características.</i>	116
Figura 55 <i>Matriz de confusión clasificación 563 microterremotos con 39 características.</i>	117
Figura 56 (a) <i>Métricas de rendimiento para la selección de características PSD sistema de clasificación usando método de correlación,</i> (b) <i>Frecuencias de interés selección de características método de filtro.</i>	118
Figura 57 <i>Análisis del BER para la selección de características PSD sistema de clasificación usando método de correlación</i>	119

Figura 58 <i>Matriz de confusión clasificación 82 microterremotos con 50 características PSD.....</i>	120
Figura 59 <i>Matriz de confusión clasificación 563 microterremotos con 50 características PSD... </i>	120
Figura 60 <i>Representación de árbol de decisión considerando 84 características sistema de clasificación.</i>	121
Figura 61 <i>Matriz de confusión clasificación 82 microterremotos con 3 características.....</i>	122
Figura 62 <i>Matriz de confusión clasificación 563 microterremotos con 3 características.....</i>	123
Figura 63 <i>(a) Representación de árbol de decisión considerando 257 características PSD sistema de clasificación, (b) Frecuencias de interés selección de características método embebido.....</i>	124
Figura 64 <i>Matriz de confusión clasificación 82 microterremotos con 6 características PSD.....</i>	125
Figura 65 <i>Matriz de confusión clasificación 563 microterremotos con 6 características PSD.....</i>	126
Figura 66 <i>Métricas de rendimiento para la selección de características PSD sistema de clasificación usando el método de envoltura</i>	127
Figura 67 <i>Análisis del BER para la selección de características PSD sistema de clasificación usando el método de envoltura</i>	127
Figura 68 <i>Matriz de confusión clasificación 82 microterremotos con 6 características.....</i>	129
Figura 69 <i>Matriz de confusión clasificación 563 microterremotos con 6 características.....</i>	129
Figura 70 <i>(a) Métricas de rendimiento para la selección de características PSD sistema de clasificación usando el método de envoltura, (b) Frecuencias de interés selección de características método de envoltura.</i>	131
Figura 71 <i>Análisis del BER para la selección de características PSD sistema de clasificación usando el método de envoltura</i>	132
Figura 72 <i>Matriz de confusión clasificación 82 microterremotos con 65 características PSD.....</i>	133
Figura 73 <i>Matriz de confusión clasificación 563 microterremotos con 65 características PSD... </i>	134
Figura 74 <i>Matriz de confusión clasificación 82 microterremotos con 39 características.....</i>	135

Figura 75 <i>Matriz de confusión clasificación 563 microterremotos con 39 características.</i>	135
Figura 76 <i>Matriz de confusión clasificación 82 microterremotos con 50 características PSD.</i>	136
Figura 77 <i>Matriz de confusión clasificación 563 microterremotos con 50 características PSD.</i> ...	137
Figura 78 <i>Modelos finales sistema de clasificación de microterremotos</i>	137
Figura 79 <i>Sistema de votación con 5 modelos sistema de clasificación de microterremotos</i>	138
Figura 80 <i>Sistema de votación con 3 modelos sistema de clasificación de microterremotos</i>	138
Figura 81 <i>Matriz de confusión clasificación 82 microterremotos 5 modelos de clasificación.</i>	139
Figura 82 <i>Matriz de confusión clasificación 563 microterremotos 5 modelos de clasificación.</i> ...	140
Figura 83 <i>Matriz de confusión clasificación 82 microterremotos 3 modelos de clasificación.</i>	140
Figura 84 <i>Matriz de confusión clasificación 563 microterremotos 3 modelos de clasificación.</i> ...	140
Figura 85 <i>Interfaz gráfica sistema de reconocimiento de microterremotos del volcán Cotopaxi</i>	142
Figura 86 <i>Matriz de confusión detección todos los microterremotos 21 de diciembre 2020.</i>	143
Figura 87 <i>Matriz de confusión detección 8 eventos recopilados 21 de diciembre 2020.</i>	144
Figura 88 <i>Matriz de confusión Clasificación de microterremotos 21 de diciembre 2020.</i>	144
Figura 89 <i>Matriz de confusión detección de microterremotos 22 de diciembre 2020.</i>	145
Figura 90 <i>Matriz de confusión detección 60 eventos recopilados 21 de diciembre 2020.</i>	145
Figura 91 <i>Matriz de confusión Clasificación de microterremotos 22 de diciembre 2020.</i>	145

RESUMEN

Vivimos en un mundo continuamente cambiante, donde se genera una variedad de fenómenos físicos naturales. Muchos de estos fenómenos físicos han sido monitorizados constantemente por el peligro latente que generan. Los volcanes son un claro ejemplo de un fenómeno natural que puede traer consecuencias catastróficas si entra en etapa eruptiva. En este contexto, el Ecuador posee uno de los volcanes más activos del mundo, lo que genera la necesidad de monitorizar y recopilar información por expertos en geofísica, los cuales están interesados en entender el comportamiento de dicho fenómeno, por estas razones generan la necesidad de estudiar métodos que permitan identificar posibles erupciones mediante la monitorización y detección de la actividad microsísmica de un volcán, con el fin de salvaguardar vidas y pérdidas materiales. En este proyecto se desarrolló un sistema de reconocimiento automático de microterremotos en tiempo real del volcán Cotopaxi al aplicar técnicas de aprendizaje supervisado. El modelo de reconocimiento es obtenido aplicando modelos de aprendizaje supervisado k-Vecinos Cercanos (kNN), Máquina de Vectores Soporte (SVM) y Árboles de Decisión (DT) a características de tiempo, frecuencia y escala, aplicados a los datos proporcionados por el Instituto Geofísico de la Escuela Politécnica Nacional con la actividad sísmica presentada del volcán Cotopaxi en los años 2012, 2013 y 2014. El sistema de reconocimiento está constituido por una etapa de detección y clasificación, para lo cual se realizan procesos de segmentación de la señal en ventanas, etiquetamiento, extracción de características, selección de características y obtención de modelos unificados en un sistema de votación. Para las etapas de detección y clasificación de microterremotos presentaron resultados en términos de exactitud, precisión, sensibilidad, especificidad y tasa de error de balanceo (BER, del inglés *Balanced Error Rate*), para la detección el algoritmo kNN alcanzó porcentajes del 98.15 %, 95.20 % y 0.017 para la exactitud, precisión y BER, respectivamente, mientras que en la clasificación el algoritmo SVM alcanzó porcentajes del 94.49 %, 60.42 % y 0.16 para la exactitud, precisión y BER respectivamente.

PALABRAS CLAVES:

- **ACTIVIDAD MICROSÍSMICA**
- **MONITORIZACIÓN VOLCÁNICA**
- **EXTRACCIÓN DE CARACTERÍSTICAS**
- **APRENDIZAJE SUPERVISADO**
- **SELECCIÓN DE CARACTERÍSTICAS**

ABSTRACT

We live in a continuously changing world, where a variety of natural physical phenomena are generated. Many of these physical phenomena have been constantly monitored for the latent danger they generate. Volcanoes are a clear example of a natural phenomenon that can have catastrophic consequences if it erupts. In this context, Ecuador has one of the most active volcanoes in the world, which generates the need to monitor and collect information by experts in geophysics, who are interested in understanding the behavior of this phenomenon, for these reasons generate the need to study methods to identify possible eruptions by monitoring and detecting the microseismic activity of a volcano, in order to safeguard lives and material losses. In this project, an automatic real-time micro-earthquake recognition system for the Cotopaxi volcano was developed by applying supervised learning techniques. The recognition model is obtained by applying supervised learning models k-Nearest Neighbors (kNN), Support Vector Machine (SVM) and Decision Trees (DT) to time, frequency and scale features, applied to the data provided by the Instituto Geofísico de la Escuela Politécnica Nacional with the seismic activity presented Cotopaxi volcano in the years 2012, 2013 and 2014. The recognition system is constituted by a detection and classification stage, for which signal segmentation processes are performed in windows, labeling, feature extraction, feature selection and obtaining unified models in a voting system. For the stages of detection and classification of micro-earthquakes presented results in terms of accuracy, precision, sensitivity, specificity and Balanced Error Rate (BER), for detection the kNN algorithm reached percentages of 98.15 %, 95.20 % and 0.017 for accuracy, precision and BER, respectively, while in classification the SVM algorithm reached percentages of 94.49 %, 60.42 % and 0.16 for accuracy, precision and BER respectively.

KEYWORDS:

- **MICROSEISMIC ACTIVITY**
- **VOLCANIC MONITORING**
- **FEATURE EXTRACTION**
- **SUPERVISED LEARNING**
- **FEATURE SELECTION**

Capítulo I

Introducción

Antecedentes

El anillo de Fuego del Pacífico es una zona geográfica con una intensa actividad sísmica y volcánica que comprende un área de 40000 km, en este territorio están ubicados 452 volcanes y concentra el 75% de los volcanes más peligrosos del mundo los cuales han generado el 90% de las actividades sísmicas registradas (Leire, 2014). Los volcanes Cotopaxi, Ojo del Salado y Cerro Azul ubicados en esta zona geográfica son considerados los más activos y peligrosos del mundo (Molina, 2018).

El Ecuador es considerado uno de los países que posee los volcanes más activos del mundo, esto se debe a varios factores como su ubicación geográfica al estar situado en el Cinturón de Fuego del Pacífico donde existe una alta frecuencia en la actividad volcánica y sísmica. Los volcanes Cotopaxi, Tungurahua y Sangay son considerados los volcanes actualmente activos y más peligrosos del Ecuador (IGEPN, 2020).

En este contexto, el volcán Cotopaxi es un estratovolcán nevado con una frecuente actividad sísmica y volcánica, el cual se encuentra ubicado sobre la Cordillera Oriental a una distancia de 35 km al Noreste de Latacunga y de 45 km al Sureste de Quito, con una altura de 5897 m sobre el nivel del mar. Los registros de la actividad eruptiva del volcán Cotopaxi se han recopilado desde inicios de la conquista española, se encuentran documentados cinco grandes episodios eruptivos en los siguientes años: 1532-1534, 1742-1744, 1766-1768, 1853-1854 y 1877-1880. Las consecuencias que se han registrado con cada periodo eruptivo han sido pérdidas socio-económicas en el Ecuador. El verdadero peligro que radica en cada episodio eruptivo es dar lugar a formaciones de lahares que pueden afectar actualmente a un estimado

de 300.000 personas de poblaciones aledañas al volcán Cotopaxi, como Latacunga y una parte del Valle de los Chillos (IGEPN, 2017).

La frecuente actividad sísmica que presenta el volcán Cotopaxi ha llevado a la necesidad de mantenerlo permanentemente monitorizado. El Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN) se encarga de realizar la monitorización de las actividades sísmicas y volcánicas en el Ecuador. La monitorización de volcán Cotopaxi se realiza a través de una red de 59 estaciones compuestas por detectores de lahares, deformaciones por movimiento de magma, medidores de emisión de gases, detectores de sismos, pluviómetros, cámaras y termocupla (IGEPN, 2019). Los datos recopilados por las estaciones son almacenados y posteriormente analizados para determinar si el volcán Cotopaxi entra en un proceso eruptivo que puede causar daño al equilibrio del ecosistema del Ecuador.

Algunas de las consecuencias que se pueden presentar si el volcán Cotopaxi entra en un proceso eruptivo son: expulsión de gases volcánicos como SO_2 y CO_2 , que en altas cantidades pueden ser mortales para los seres vivos; flujos piroclásticos formados por gases, fragmentos de rocas y ceniza que son expulsados por el cono del volcán; lluvia de ceniza que se puede extender por grandes áreas afectando a las poblaciones, vegetación y fauna aledaña; lahares que son desplazamientos de lodo y escombros que llegan a alcanzar una velocidad de hasta 100 km/h y finalmente presenta microterremotos los cuales son de muy baja magnitud que no son percibidos por las poblaciones aledañas (Hall, Andrade, Mothes, Troncoso, & ...Yepes, 2005).

El IGEPN realiza una monitorización permanente las 24 horas del día y 365 días del año para la recopilación de las distintas alteraciones que presenta el volcán Cotopaxi. Disponen de un Sistema Informático para el Procesamiento y Análisis de Señales Sísmicas de los volcanes del Ecuador (SIPASS), el cual fue implementado en el 2011 y permite identificar el tipo de

microterremoto que se produce mediante el cálculo de la transformada rápida de Fourier (FFT, del inglés *Fast Fourier Transform*) para obtener el espectro de la señal y verificar el rango de frecuencia en la se encuentra, posteriormente los expertos del IGEPN lo etiquetan de forma visual y almacenan la información recopilada en una base de datos (Viracucha & De la Batida, 2014). El proceso de etiquetación de los microterremotos al ser de forma visual requiere una gran cantidad de tiempo, lo cual puede causar represión de la información si existe una elevada actividad sísmica en el volcán Cotopaxi.

La frecuencia de los distintos tipos de microterremotos que presenta un volcán puede dar aviso si el mismo se encuentra en una etapa eruptiva, de ahí la importancia para realizar el reconocimiento de los microterremotos en el menor tiempo posible así se puede generar una alerta temprana que ayude a salvaguardar a las poblaciones aledañas y prevenir tragedias frente a una posible erupción (IGEPN, 2015).

En la sección 1.5 se presenta de forma detallada los trabajos relacionados con el reconocimiento de microterremotos con distintos métodos de análisis.

Justificación e Importancia

Los volcanes son uno de los peligros naturales permanentemente latentes que amenazan con el equilibrio de un ecosistema, los cuales son monitorizados para detectar características relacionadas con deformación del suelo, flujo de gas, sismicidad y otros factores. La sismología es una herramienta importante y efectiva para monitorizar volcanes, la sismicidad es el método más rápido y usado para detectar cambios en los volcanes. Los analistas pueden interpretar la firma sismológica para identificar los diferentes tipos de microterremotos, los que se presentan más frecuentemente son eventos de Largo Período (LP), terremotos Volcano Tectónicos (VT), Temblores (TRE, del inglés *Tremor*) e híbridos (HYB, del inglés *Hybrids*) (Lara,

Benítez, Carrera, Ruiz, & Rojo Álvarez, 2016). La frecuencia de estos eventos puede dar aviso del incremento en la actividad sísmica de un volcán.

El volcán Cotopaxi posee una monitorización constante por el IGEPN (24 h / 7 días a la semana) (Ortiz Herazo, 2013). La actividad sísmica del volcán Cotopaxi, genera más de 100 MB de datos por día, lo que significa un promedio de 17 y 21 eventos por día en periodos de actividad normal (Lara, Benítez, Carrera, Ruiz, & Rojo Álvarez, 2016). Existen 16 estaciones sismológicas desplegadas en el volcán Cotopaxi, por lo tanto, los científicos expertos deben analizar visualmente para etiquetar y clasificar las señales volcánicas. Cuando el volcán entra en estado eruptivo se pueden presentar un promedio de 130 eventos por día, como la registrada en el año 2015; por lo tanto, la evaluación de los datos puede convertirse en un proceso extremadamente lento, que puede también causar represión de la información y evitar dar una alerta oportuna frente a una erupción.

Alcance del Proyecto

Esta investigación busca crear un sistema de reconcomiendo automático de microterremotos en tiempo real, que sea una herramienta de apoyo para los expertos en el tema de sismología en el Ecuador. Para la obtención de este sistema se abarcan temas referentes a bases de datos, procesamiento digital de señales (PDS), extracción de características (tiempo, frecuencia y escala), selección de características, aprendizaje supervisado (kNN, DT y SVM) e implementación de un sistema en tiempo real.

Para la etapa de entrenamiento se dispone de dos bases de datos de microterremotos del volcán Cotopaxi etiquetadas por el IGEPN. Las bases de datos recopiladas en los años 2010 y 2012 se usarán para obtener modelos de entrenamiento aplicando técnicas de aprendizaje supervisado mediante el software matemático MATLAB® para la detección y clasificación de

microterremotos respectivamente. Posteriormente se realizan pruebas para comparar las etiquetas del IGEPN con las obtenidas en los modelos de entrenamiento evaluado mediante medidas de desempeño, que son: exactitud, precisión, especificidad, sensibilidad y tasa de error de balanceo (BER, del inglés *Balanced Error Rate*).

Finalmente se incorporan todos los modelos de detección y clasificación en una interfaz gráfica en el software matemático MATLAB® para facilitar el uso del sistema de reconocimiento de microterremotos por los expertos de IGEPN. La presente investigación busca generar una herramienta de apoyo para los expertos del IGEPN en la detección y clasificación de eventos sísmicos del volcán Cotopaxi en tiempo real, facilitando el reconocimiento de los distintos tipos de microterremotos que pueden dar aviso si el volcán Cotopaxi entra en una etapa eruptiva.

Objetivos

General

Implementar un sistema que permita reconocer microterremotos en tiempo real del volcán Cotopaxi aplicando aprendizaje supervisado.

Específicos

- Adquirir información mediante un estudio previo sobre el estado del arte de reconocimiento de microterremotos.
- Implementar un modelo de reconocimiento de microterremotos en tiempo, frecuencia y escala.
- Implementar los algoritmos de aprendizaje supervisado para clasificación de microterremotos en tiempo real.

- Evaluar el desempeño del sistema en función de la precisión, sensibilidad y especificidad del sistema de reconocimiento de microterremotos con características de tiempo, frecuencia y escala.
- Evaluar el sistema de reconocimiento de microterremotos con los expertos del Instituto Geofísico de la Escuela Politécnica Nacional.

Trabajos Relacionados

Los centros de investigación de vulcanología en el mundo utilizan métodos sísmicos para la monitorización y pronóstico de erupciones (Roman & Cashman, 2018) (Zobin, 2017). Estos métodos son utilizados como una herramienta para la detección y clasificación correcta de los eventos sísmicos que se generan en un volcán (Havskov & Ottemoller, 2010) (Bean, y otros, 2014) (Cusano, Palo, & West, 2015). Los enfoques seguidos para el reconocimiento automático de microterremotos plantean distintos tipos de metodologías. Para la detección se han seguido enfoques en el procesamiento de una señal en dominios de tiempo, frecuencia y escala (Gutiérrez, Ramirez, Ibañez, & Bentez, 2011) (Lara-Cueva R. , Benítez, Paillacho, Villalva, & Rojo-Álvarez, 2017) . Por otro lado, para la clasificación se calculan ciertas características en el dominio de tiempo, frecuencia o escala relacionadas a la señal e implementar algoritmos de aprendizaje automático (Jaramillo, 2015) (Cortés, Benitez, Garca, & Álvarez, 2015). Solo para la detección de eventos microsísmicos se han aplicado distintos métodos que son: algoritmo promedio a corto plazo / promedio a largo plazo (STA/LTA, del inglés *short-term average/long-term average*) como los presentados (Vaezi & Van der Baan, 2015); detección basada en actividad de voz (Lara-Cueva, Moreno, Larco, & Benítez, 2016); detección aplicando técnicas espectrales (Aguinaga, 2018); detección mediante algoritmos de descomposición variacional (VMD, del inglés *Variational Mode Decomposition*) (Dragomiretskiy & Zosso, 2014). Solo para la clasificación de eventos microsísmicos se han aplicado distintos métodos que son : clasificación

basados en aprendizaje automático aplicando redes neuronales (ANN, del inglés, artificial neural network) (Orozco, Garcia, Duin, & Castelanos, 2006); clasificación mediante modelos de Markov (HMM del inglés, hidden Markov models) (Benitez, 2007); clasificación mediante un enfoque semi-supervisado (Brusil, Grijalva, Lara-Cueva, & Ruiz, 2019); clasificación mediante un descriptor basado en imágenes (Pérez, Venegas, Benítez, Lara-Cueva, & Ruiz, 2020). Estas son algunas de las técnicas que han sido empleadas por los centros de investigación para el reconocimiento y estudio de la actividad sísmica que se presenta en un volcán.

Este trabajo de investigación está enfocado en reconocimiento de microterremotos aplicando técnicas de aprendizaje supervisado como los observados; en el artículo de (Lara, Paillacho, & Villalva, 2016) habla sobre la detección de eventos del volcán Cotopaxi empleando una clasificación supervisada, su metodología radica en un algoritmo empleado para realizar una detección de manera automática de las señales extraídas siendo estas sismo-volcánicas y no volcánicas, estas señales contienen características tales que permiten realizar su extracción por medio de la transformada de Fourier y Wavelet para interpretar rápidamente una actividad sísmica del volcán Cotopaxi y anunciar una posible erupción, este método empleado para la detección presenta una precisión alrededor de 98%. Otro método usado para la detección de los eventos sísmicos del volcán Cotopaxi en tiempo real es el algoritmo de VAD (Rodríguez, 2018) (Rosero, 2019) este es un detector basado en la actividad de voz y en puntos extremos que al adaptarlo al sistema la precisión de obtener resultados válidos en tiempo real es de un 96% y de 100% para cada investigación respectivamente.

La detección ESV (Eventos Sismo Volcánicos) del volcán Cotopaxi se realiza también mediante un proceso para caracterizar las señales sísmicas presentado por (Jaramillo, 2015), este otro método utiliza estimadores espectrales para analizar la frecuencia de dicha señal y conocer el tipo de sismo, su aplicación presenta una precisión de un 79.8% y la sensibilidad para

detectar el evento es de un 93.8%. Existe un detector y clasificador automático de los eventos sísmicos del volcán Cotopaxi en tiempo real propuesto por (Lara-Cueva R. , Benítez, Carrera, Ruíz, & Rojo, 2016), para la segmentación de las señales se utilizó ventanas de tiempo deslizantes no superpuestas que ayuda para la detección de los eventos sísmicos que se puedan presenciar de forma automática. El mejor resultado se obtuvo utilizando el clasificador Máquina de Vectores Soporte (SVM, del inglés *Support Vector Machines*), en este análisis se obtuvo una precisión de hasta el 99% en la etapa de detección, y del 97% en la precisión y sensibilidad en la etapa de clasificación.

Organización

El presente trabajo de investigación está constituido de cinco capítulos que fueron organizados de manera que facilite su lectura. El primer capítulo describe una breve introducción de la investigación, justificación e importancia, también se detalla el alcance del proyecto, objetivos y los trabajos previamente realizados que están relacionados con el tema de investigación.

En el segundo capítulo se detalla el marco teórico de la investigación, abarcando los temas principales que deben ser conocidos para entender el proceso que se va a implementar. Se profundiza las técnicas implementadas, dando el trasfondo necesario para la comprensión del siguiente capítulo.

En el tercer capítulo se explica la metodología utilizada para el desarrollo propuesto en la investigación y cumplimiento de los objetivos propuestos. Se describe de forma detallada cada proceso realizado y técnicas implementadas, para la obtención de los modelos de detección y clasificación de microterremotos.

En el cuarto capítulo se describen los resultados obtenidos del sistema de detección y clasificación con las bases de datos disponibles. Los sistemas son evaluados mediante las métricas de rendimiento.

Finalmente, en el quinto capítulo se realiza una discusión de los resultados obtenidos en este trabajo de investigación, se detallan las conclusiones y recomendaciones obtenidas y se proponen trabajos futuros que complementen este trabajo de investigación.

Capítulo II

Marco Teórico

Vulcanología

La vulcanología es una ciencia que estudia los fenómenos eruptivos, como es el transporte y la erupción del magma; a nivel mundial existen alrededor de 1555 volcanes registrado por el GVP-SI (Programa Global de Vulcanismo-Institución Smithsonian) (Tele13, 2015) de los cuales más de 1300 han entrado en un proceso eruptivo en los últimos 10000 años, durante este periodo los volcanes han provocado desastres que han afectado a más de 300000 personas por lo que se tuvo la necesidad de instalar observatorios vulcanológicos que registran constantemente la actividad volcánica para alertar a la población. Anualmente unos 50 volcanes entran en un estado eruptivo, y se estima que el 10% de la población vive cerca o incluso sobre volcanes considerados potencialmente peligrosos, esto se debe a que esta actividad volcánica proporciona terrenos fértiles que dan sustento a la civilización humana, por el contrario, al igual que es un beneficio para la parte geológica trae peligros a la vida misma cuando comienza a presenciar actividad en el volcán llevando a una posible erupción (Tilling & Punongbayan, 1993).

Cada uno de los volcanes presenta su propio estilo y características eruptivas que se mencionan en la Tabla 1, la Organización Mundial de Observatorios Vulcanológicos desarrollan múltiples mapas ubicando las zonas de mayor peligrosidad de un volcán, los países con mayor índice de volcanes son los que presenten su ubicación geográfica sobre el Cinturón de Fuego del Pacífico o el Cinturón Alpino ya que se consideran zonas volcánicas.

Tabla 1

Registro de tipos de volcanes, su lava predominante, estilo y características eruptivas

Tipo de Volcán	Lava Predominante		Estilo Eruptivo	Características Eruptivas
	<i>Composición</i>	<i>Viscosidad Relativa</i>		
Escudo-Volcán	Basáltica	Fluidal	Generalmente no explosivo o débilmente explosivo.	Fuentes de lava, flujos de lava (extensos), lagos y pozas de lava.
	Andesítica	Menos fluidal	Generalmente explosivo, pero a veces no explosivo.	Fluidos de lava (medianos), productos balísticos explosivos, caídas de tefra, flujos y oleadas de piroclásticos.
Estrato-Volcán	Dacítica a Riolítica	Viscosa a muy viscosa	Típicamente muy explosivo, pero puede ser no explosivo, especialmente después de una explosión grande.	Productos balísticos explosivos, caídas de tefra, flujos y oleadas de piroclásticos, flujos de lava cortos y domos de lava.

Nota. Recuperado de: (Tilling & Punongbayan, 1993)

El volcán Cerro Azul ubicado en la Isla Isabela en Galápagos es considerado un escudo volcán por los flujos de lava basáltica que se genera en el interior del volcán y una suave pendiente, en la Figura 1 se observa el volcán Cerro Azul.

Figura 1

Volcán Cerro Azul de la Isla Isabela en Galápagos



Nota. Recuperado de IGEPN.

El volcán Tungurahua es considerado un estrato volcán compuesto, llamado así por generar fluidos de lava andesítica razón por la que su cono tiene una forma compuesta como se observa en la Figura 2.

Figura 2

El Volcán Tungurahua



Nota. Recuperado de IGEPN.

En la Figura 3 se presenta al volcán Sangay el cual es considerado un estrato volcán por su frecuente actividad volcánica y la viscosidad de su lava que impide que recorra extensas áreas formando domos de lava a su alrededor (IGEPN, 2017).

Figura 3

El Volcán Sangay



Nota. Recuperado de IGEPN.

El Ecuador es uno de los países ubicado en el Cinturón de Fuego del Pacífico que se divide en cuatro regiones: el Oriente o Amazonía Ecuatoriana, Sierra o Región Interandina, Costa o Región Litoral y la Región Insular conformado por las Islas Galápagos; este país abarca a 68 volcanes de estos únicamente 33 están considerados potencialmente activos (Corominas & Martí, 2011), desde el año 1532 se ha presenciado erupciones de 7 volcanes continentales y 7 volcanes ubicados en las Islas Galápagos cuyos nombres se enlista en la Tabla 2 (IGEPN, 2019).

Tabla 2

Volcanes En estado eruptivo desde el año 1532

Volcanes	
Continetales	Islas Galápagos
Cayambe	Marchena
Reventador	Cerro Azul
Guagua Pichincha	Fernandina
Cotopaxi	Santo Tomás
Tungurahua	Alcedo
Sangay	Darwin

La monitorización de los volcanes en el Ecuador lo realiza el IGEPN al emplear una monitorización para geoquímica, deformación, cambios físicos y vigilancia sísmica y volcánica; para el análisis del comportamiento de un volcán concentran información de los microsismos vulcanológicos, estos eventos son un alto número de sismos que se generan en el interior del volcán y por medio de sus respectivos análisis se clasifican al obtener datos que pueden ayudar evidenciar una posible erupción volcánica y realizar planes de contingencia para salvaguardar la vida de las personas de poblaciones aledañas. El IGEPN ha prestado mayor vigilancia a los volcanes que presentan un gran riesgo a los pobladores, flora y fauna de la zona, entre ellos se encuentra el volcán Cotopaxi considerado de alto riesgo y activo, su última erupción fue en el año 1877 cuyos glaciares se derritieron creando lahares que llegaron hasta el Océano Pacífico destruyendo todo a su paso (Brusil C. , 2020).

El Volcán Cotopaxi

Sismicidad

En el territorio del Ecuador se encuentra uno de los volcanes más peligrosos del mundo; el volcán Cotopaxi, considerado así por su intensa actividad volcánica y sísmica, se encuentra ubicado a 35 km al noreste de Latacunga y a 45 km al sureste de Quito sobre la Cordillera Oriental con una altura de 5897 m sobre el nivel del mar, se encuentra cubierto por una densa capa glaciara como se puede observar en la Figura 4. Su historial eruptivo demanda una gran vigilancia ya que amenaza a una parte muy significativa de población por la formación de enormes lahares y la caída de ceniza que afectaría a gran parte de la sierra y costa ecuatoriana (IGEPN, 2020).

Figura 4

Volcán Cotopaxi (2015)



Nota. Recuperado de IGEPN.

Este volcán es una gran amenaza por lo que es el más vigilado de todo el Ecuador, se han registrado cinco episodios eruptivos que se encuentra redactado en la Tabla 3. Estos episodios han presentado una alta actividad sísmica y varios fenómenos volcánicos peligrosos para todo ser vivo que se encuentre en zonas cercanas al volcán, el IGEPN ha monitorizado el volcán Cotopaxi desde el año 1976. En este mismo periodo se colocó la primera estación permanente para la vigilancia de los eventos sísmicos en el Cotopaxi, se instalaron 4 estaciones para la monitorización mediante un sismómetro y sensores inalámbricos. Esta red de monitorización ha registrado varios eventos sísmicos durante todo este tiempo.

Tabla 3

Registro de los cinco grandes periodos eruptivos del volcán Cotopaxi

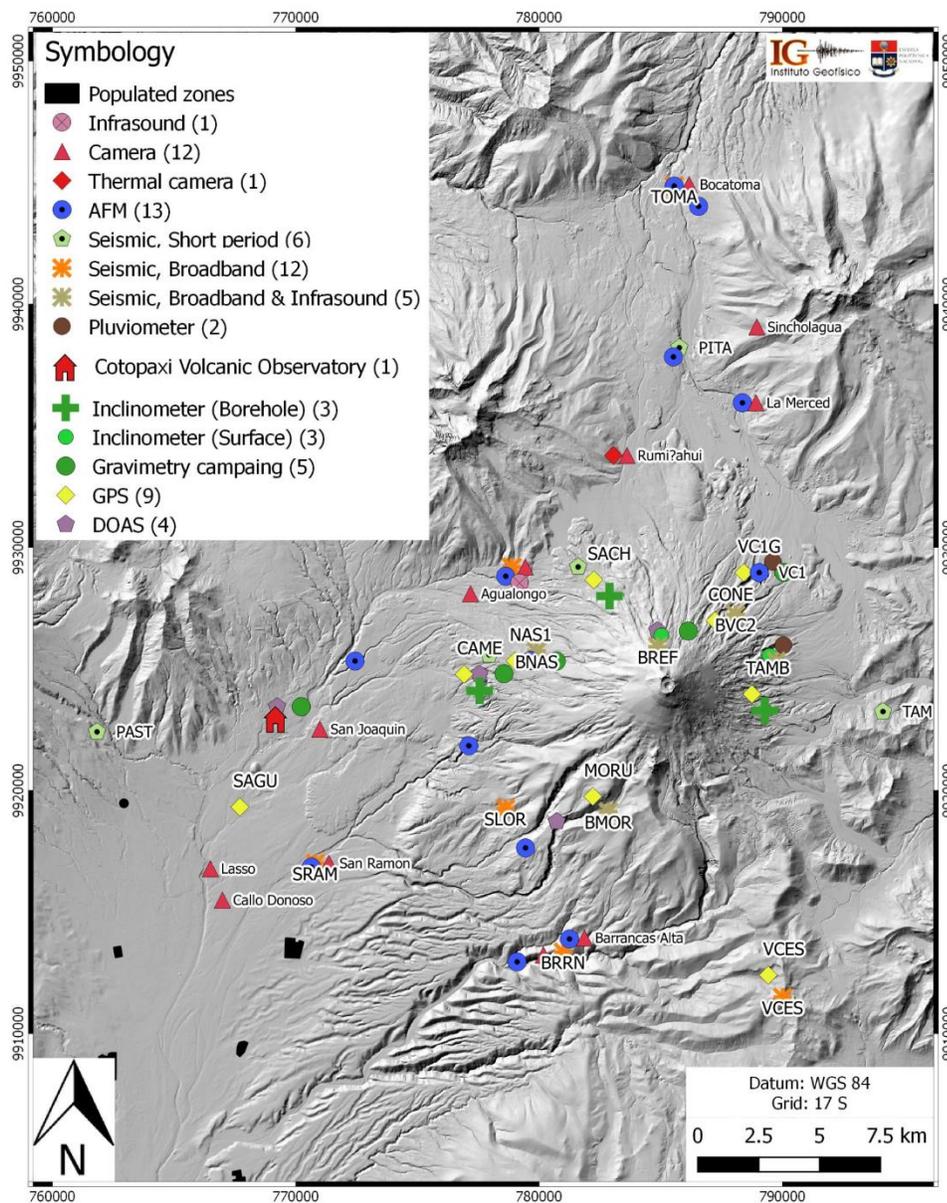
Periodos Eruptivos del Volcán Cotopaxi	
Inicio	Fin
1532	1534
1742	1744
1766	1768
1853	1854
1877	1880

Monitorización

El volcán Cotopaxi dispone de un sistema de monitorización constante (24 h / 7 días a la semana) al año (Ortiz Herazo, 2013). Esta red de monitorización que se puede observar en la Figura 5 ha registrado varios eventos sísmicos durante todo este tiempo.

Figura 5

Red de Monitorización del Volcán Cotopaxi



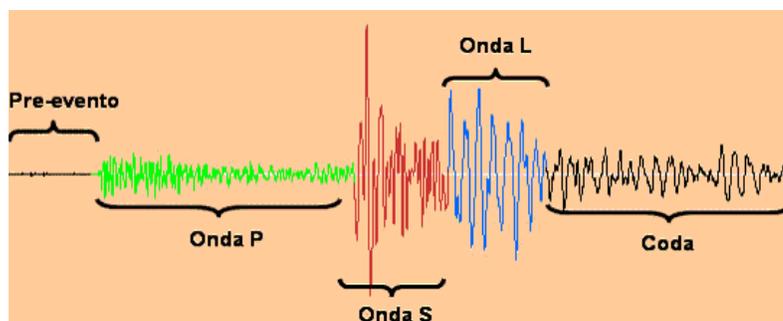
Nota. Recuperado de: (IGEPN, 2020)

Las ondas microsísmicas captadas por la red de sensores del IGEPN están compuestas por varias subondas, las cuales ocurren simultáneamente al existir actividad sísmica, estas subondas se dividen en dos grupos que son: las ondas de cuerpo que viajan desde el interior del

planeta y ondas superficiales que viajan por la superficie terrestre. En la Figura 6 se pueden observar las subondas que ocurren al producirse un microterremoto.

Figura 6

Ondas de un evento microsísmico



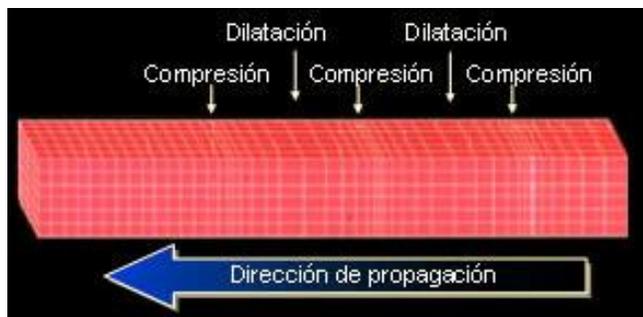
Nota. Recuperado de: (Universidad de Costa Rica, 2011)

Los componentes que conforman un evento microsísmico son: pre-evento; onda P; Onda S; Onda L; Coda. Estos componentes proporcionan información importante con respecto al evento microsísmico. El pre-evento está constituido por el ruido ambiente (agitaciones del suelo por factores naturales ejemplo: viento, lluvia, etc.).

Las ondas P son las ondas primarias que se generan después del pre-evento, son las primeras en captar el sensor, tiene una velocidad promedio de propagación del doble de las ondas S, su propagación se puede realizar por medio de cualquier tipo de material por ello son más rápidas y mientras se traslade va comprimiendo y dilatando el suelo, su forma de propagación se muestra en la Figura 7.

Figura 7

Forma de propagación de las ondas P

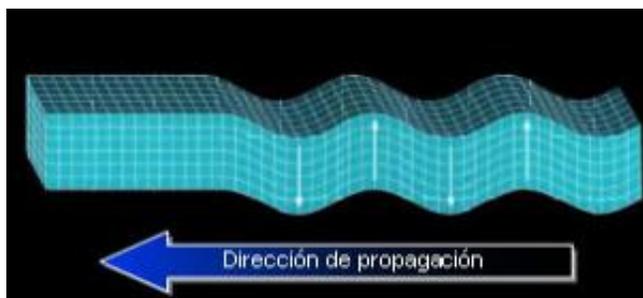


Nota. Recuperado de: (Universidad de Costa Rica, 2011)

Las ondas S o secundarias son ondas transversales caracterizadas por tener la mayor amplitud en el evento, estas ondas tienen una menor velocidad con respecto a las ondas P ya que no pueden propagarse por medios líquidos y su amplitud hace que corte el medio por el cual se propaga, su forma de propagación se observa en la Figura 8.

Figura 8

Forma de propagación de las ondas S



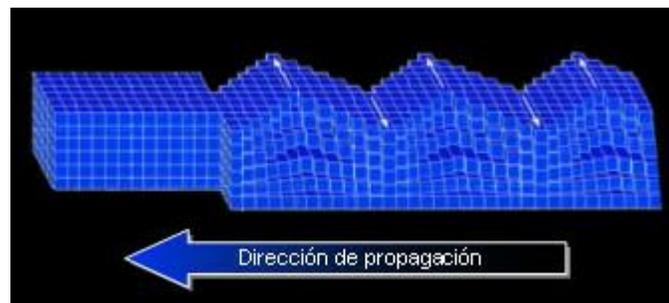
Nota. Recuperado de: (Universidad de Costa Rica, 2011)

Las ondas L son ondas caracterizadas por tener una amplitud similar a las ondas S con una variación, que su periodo es mucho más largo. La velocidad es menor que las ondas S y se la

considera la más catastrófica ya que se trasladan por capas muy cercanas a la superficie de la tierra, la forma de propagación de estas ondas se observa en la Figura 9 en la cual se muestra su oscilación de forma horizontal.

Figura 9

Forma de propagación de las ondas L



Nota. Recuperado de: (Universidad de Costa Rica, 2011)

Al final posee una Coda, que es un segmento donde el sismógrafo decae hasta obtener una forma similar al pre-evento (Universidad de Costa Rica, 2011).

Tipos de eventos sísmicos

La monitorización realizada por el IGEPN recopila distintos microterremotos cuya ocurrencia es común en una posible erupción volcánica, los cuales son:

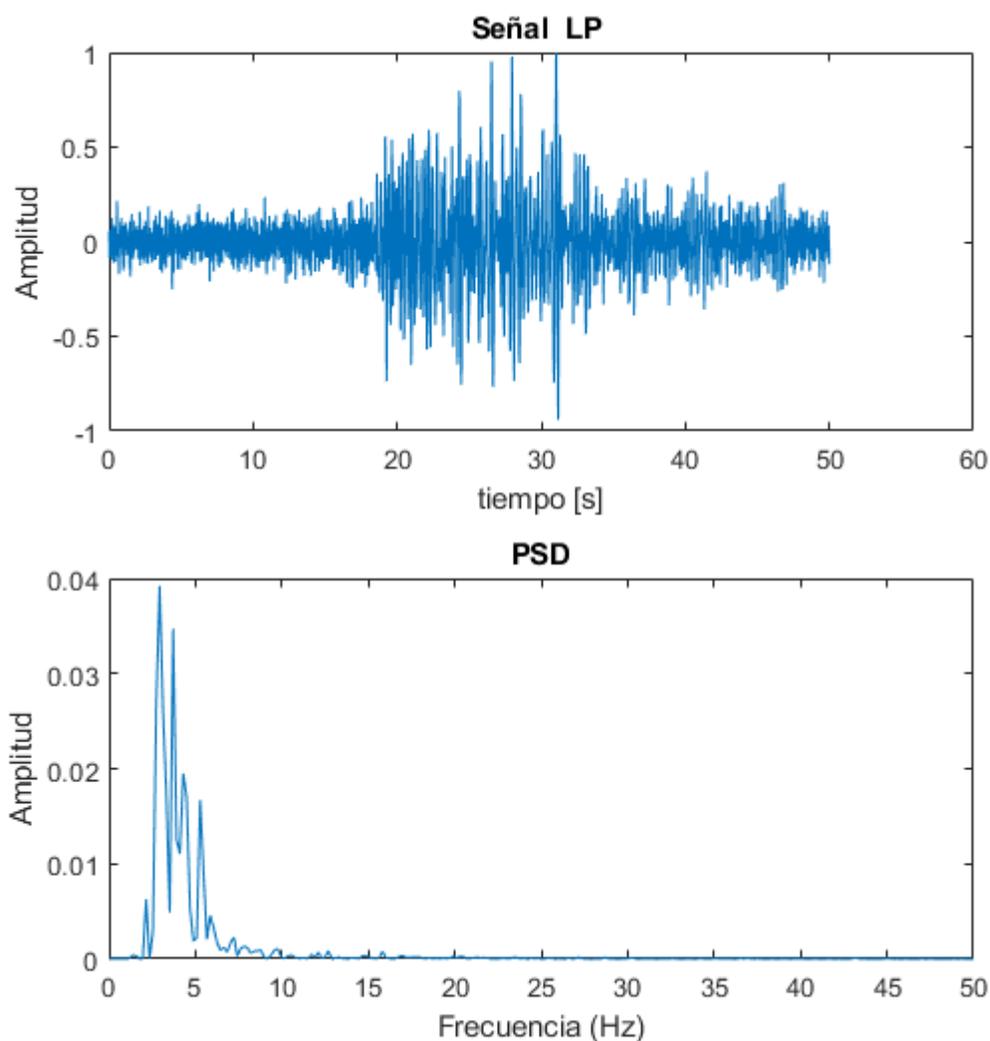
a. Eventos de Largo Período.

Los eventos de Largo Período (LP) son sismos caracterizados por ser de baja frecuencia, originados por movimientos de duración corta producidos en la parte interna del volcán. La frecuencia en el cual se produce este tipo de microterremoto está en un rango de 0.5 Hz a 5 Hz (Ibáñez & Carmona, 1997).

La forma de onda de una señal del tipo LP se caracteriza por tener un comienzo emergente en la cual no se puede identificar la onda P y S, el tiempo de duración de este tipo de microterremoto no sobrepasa el minuto. La frecuencia de eventos LP suele presentarse cuando el volcán está entrando a una etapa eruptiva (Chouet, 1996). En la Figura 10 se puede observar un ejemplo de un microterremoto del tipo LP y su espectro de frecuencia.

Figura 10

Ejemplo de evento LP y su espectro en frecuencia



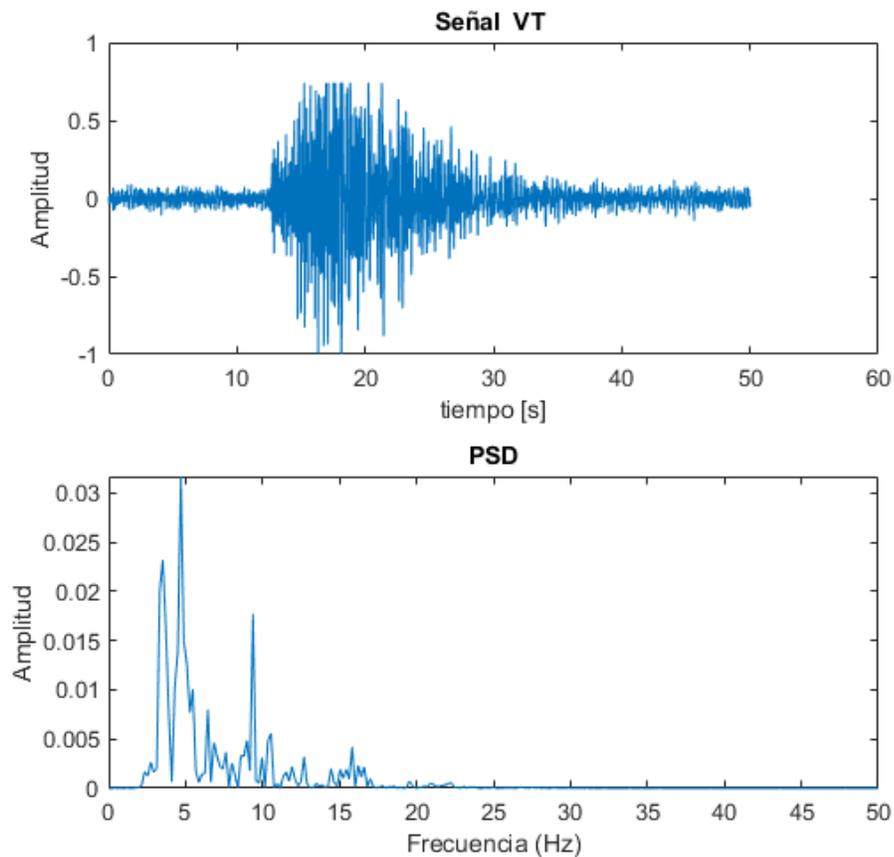
b. Eventos Volcano Tectónicos.

Los microterremotos Volcano Tectónicos (VT) se caracterizan por ser eventos de alta frecuencia, las ondas generadas en los eventos del tipo VT superan la frecuencia de 10 Hz. En este tipo de señales se puede distinguir las ondas S y P, seguidas por la onda del tipo L y su coda (Barquero, Lesage, Metaxaian, Creusot, & Fernandez).

Este tipo de eventos está asociado con aperturas de grietas, rupturas de fallas locales, rompimiento de rocas entre otras. En la Figura 11 se puede observar un ejemplo de un microterremoto del tipo VT y su espectro de frecuencia.

Figura 11

Ejemplo evento VT y su espectro en frecuencia



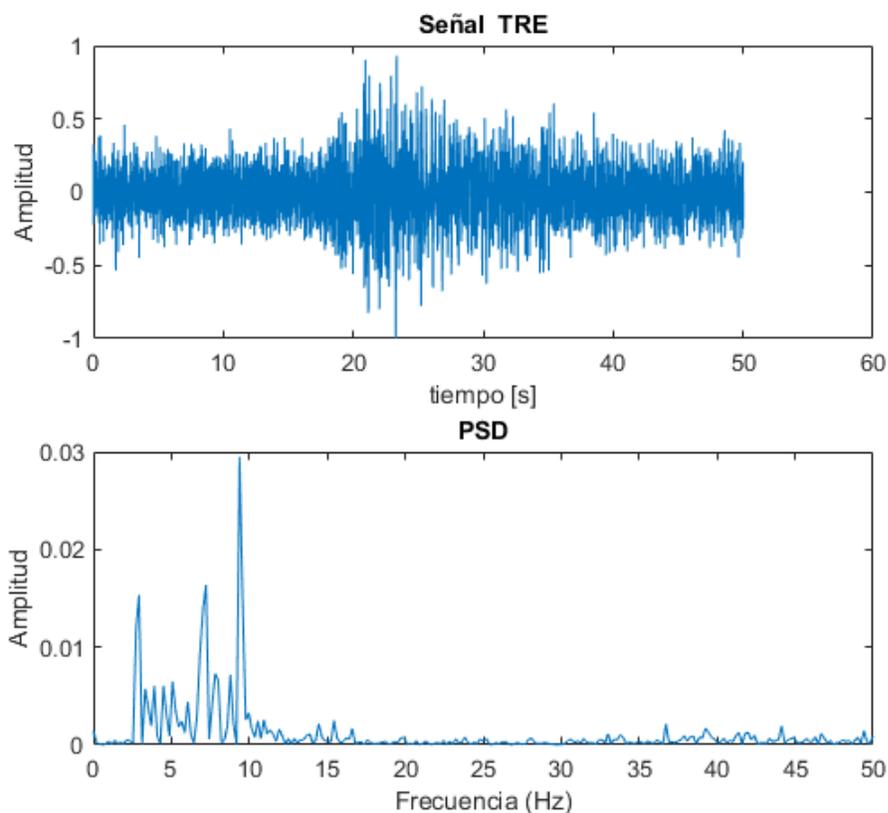
c. Tremores volcánicos.

Los eventos TRE son vibraciones que generalmente son de baja amplitud y baja frecuencia que se mantienen constantes, estos eventos sísmicos tienen una duración de algunos segundos que se pueden prolongar varios días (Mora & Alvarado, 2001). Estos eventos sísmicos generalmente aparecen cuando existe un movimiento de fluidos volcánicos. En la Figura 12 se puede observar un ejemplo de un microterremoto del tipo TRE y su espectro de frecuencia.

Los eventos TRE se subdividen en tres grupos debido a su espectro. Estos eventos de altas frecuencias, que pueden alcanzar una frecuencia alrededor de 25 Hz, se producen durante una explosión y se caracterizan por tener una duración prolongada. Las vibraciones de los eventos TRE de frecuencias intermedias duran minutos y las de frecuencias bajas tienen una duración de varios segundos (Aguinaga, 2018).

Figura 12

Ejemplo evento TRE de frecuencia intermedia y su espectro



d. Eventos Híbridos.

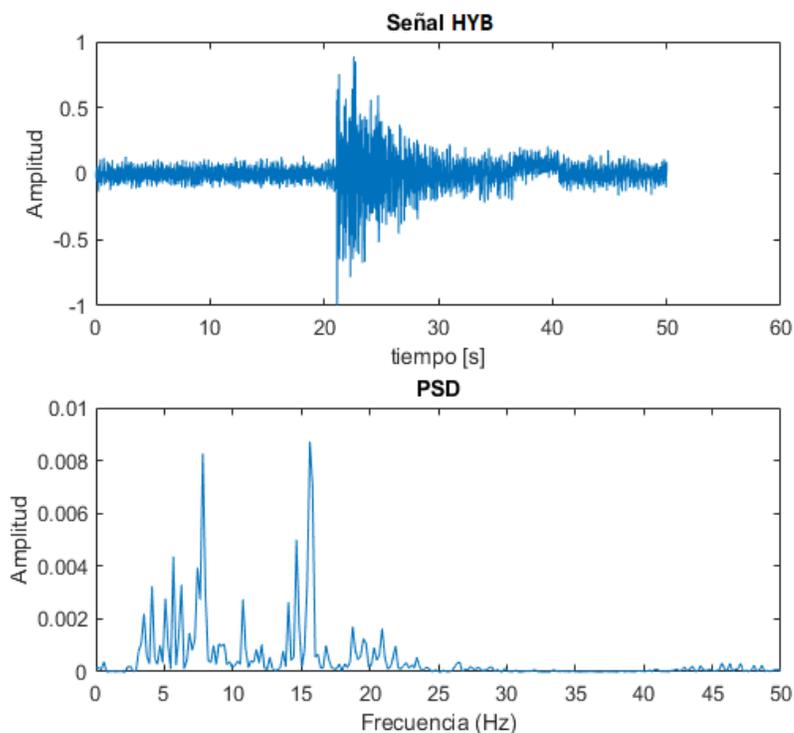
Los tipos de eventos Híbridos (HYB) se producen cuando existe movimiento de fluido o fracturas internas, este tipo de evento tiene la característica de estar compuesto por eventos LP y VT. Posee un espectro muy similar al de los eventos del tipo VT, es decir, frecuencias superiores de 10 Hz y presenta frecuencias similares a los eventos del tipo LP, es decir, por debajo de 5 Hz. La formación de un evento HYB comienza de un evento del tipo VT seguido de un evento LP.

Se puede identificar las partes de los eventos VT y LP por medio de su espectro en frecuencias, con esto se puede notar claramente las ondas P y S de cada tipo de evento (Ibáñez

& Carmona, 1997). En la Figura 13 se puede observar un ejemplo de un microterremoto del tipo Híbrido y su espectro de frecuencia.

Figura 13

Ejemplo evento HYB y su espectro en frecuencia



Conceptos Básicos

En esta sección se presentan los conceptos básicos y técnicas implementadas para el desarrollo del presente trabajo. Se especifican los métodos utilizados para el pre-procesamiento y procesamiento de cada evento. Se explica los cálculos de características en Tiempo, Frecuencia y Escala. Se explican las técnicas de aprendizaje supervisado a implementar para obtener los modelos para la detección y la clasificación. Se optimizan estos modelos al realizar una selección de características (del inglés, *Feature Selection*) al aplicar métodos de Filtros (del inglés, *Filters*),

métodos Embebidos (del inglés, *Embedded*) y métodos de envoltura (*del inglés, Wrapper*), para determinar las características más importantes.

Pre-procesamiento y Procesamiento.

Los microterremotos son señales analógicas que se consideran estocásticas y estacionarias, para poder trabajar con estas señales se necesitan realizar ciertos pasos previos antes de su análisis como: Filtrado, normalización, segmentación y etiquetamiento. Se detalla cada uno de los pasos a seguir.

a. Filtrado.

Las señales sísmicas que se generan cuando existe un evento volcánico están en un rango de 0.5 Hz a 50 Hz, independiente del tipo de evento (Ibáñez & Carmona, 1997). Este rango de frecuencia es de interés para la presente información, si existen frecuencias fuera de este rango se consideran como alteraciones o ruido presente en la señal. Es necesario realizar un filtrado a cada señal para eliminar el ruido presente en la señal.

En este trabajo no se realiza un filtrado previo, las bases de datos entregados por el IGEPN ya poseen este parámetro previamente calculado.

b. Normalización.

Los eventos recopilados en cada base de datos provienen de distintas estaciones sismológicas ubicadas en puntos estratégicos en el volcán Cotopaxi. Estas estaciones que almacenan los datos no trabajan con la misma escala en amplitud y frecuencia, es necesario llevarlo a una misma escala y frecuencia. En este trabajo se realiza una normalización de amplitud.

Se normalizan los datos retirando la media y dividiendo para el máximo absoluto. Siendo r_i la señal sísmica a ser analizada como se observa en la ecuación (1) y en la ecuación (2).

$$\hat{r}_i = \frac{r_i - \mu_i}{\sigma_i}, \quad (1)$$

donde \hat{r}_i es la señal de microterremoto normalizada, μ_i y σ_i son media y desviación estándar respectivamente.

$$x_i = \frac{\hat{r}_i}{\max|\hat{r}_i|}, \quad (2)$$

donde x_i es la señal estandarizada en un rango de [-1,1].

c. Segmentación.

La base de datos entregada por el IGEPN dispone de eventos sísmicos con una duración de 20 minutos, es necesario realizar segmentaciones en ventanas de tiempo. En este trabajo se realiza una segmentación en ventanas de 15 segundos como las realizadas en el trabajo de investigación (Lara, Benítez, Carrera, Ruiz, & Rojo Álvarez, 2016), donde se obtuvieron los mejores resultados en la detección de microterremotos con distintas ventanas de tiempo.

Al realizar una segmentación de 15 segundos se obtienen 80 segmentos de cada microterremoto. Estos segmentos son almacenados en un vector s_i como se observa en la ecuación (3), y finalmente se agrega cada señal segmentada a una matriz como se observa en la ecuación (4).

$$S_i = \left[s_{i,1}^T, s_{i,2}^T, \dots, \dots, s_{i,80}^T \right]^T, \quad (3)$$

donde $s_{i,j}$ es un segmento de señal de 15 [s] de la señal normalizada x_i .

$$S = \left[s_1^T, s_2^T, s_3^T, \dots, \dots, s_M^T \right]^T, \quad (4)$$

donde S_M es un vector de cada señal segmentada y M es el número de señales que se encuentran en la base de datos .

d. Etiquetamiento.

Posteriormente de realizar la segmentación y almacenamiento, se proceden a etiquetar los segmentos que contienen las señales sísmicas al colocar un +1 y los segmentos que no contienen señal con un -1. Esta información se encuentra incluida en la base de datos entregada por IGEPN, este proceso se realiza para cada señal y se almacena en un vector como se observa en la ecuación (5) y ecuación (6).

$$Y_i = \left[y_{i,1}^T, y_{i,2}^T, \dots \dots \dots, y_{i,80}^T \right]^T, \quad (5)$$

donde $y_{i,j}$ es el etiquetamiento de cada segmento de señal $s_{i,j}$.

$$Y = \left[Y_1^T, Y_2^T, Y_3^T, \dots \dots, Y_M^T \right]^T, \quad (6)$$

donde Y_M es un vector de cada señal etiquetada y M es el número de señales que se encuentran en la base de datos .

Extracción de características.

Es necesario realizar un proceso para extraer las características de las señales sísmicas que son filtradas para el análisis de los microterremotos, esta caracterización permite transformar tales muestras a vectores de características que trabajan por medio de algoritmos que se representan en modelos estadísticos, esta transformación permite clasificar cada una de las señales sísmicas registradas y obtener sus características específicas (Pérez A. , 2018).

Las características extraídas de cada segmento de la señal x_i se las considera como cada uno de los eventos por separado de esta forma permite realizar la clasificación; tales características se basan en tres dominios que son: tiempo, frecuencia y escala. Se consideran los tres dominios para el cálculo de características de la señal por medio de descriptores

estadísticos, entre otros (Pérez, y otros, 2019); cada una tiene sus componentes que lo distingue y su enfoque específico a una determinada función, por tal motivo al utilizar características en tiempo, frecuencia y escala permite acelerar el proceso, detallar un estudio más relevante, tener un mejor análisis de las características, y un buen entendimiento que mejora los resultados.

En la Tabla 4 se anexa a detalle todas las 84 características que se calcula de cada segmento de la señal x_i en los tres dominios analizados: tiempo, frecuencia y escala, al aplicar las fórmulas relacionadas con el trabajo de investigación de (Pérez, y otros, 2019).

Tabla 4

Características de cada segmento de la señal x_i

Dominio del Tiempo		Dominio de Frecuencia		Dominio de Escala	
<i>ID</i>	<i>Características</i>	<i>ID</i>	<i>Características</i>	<i>ID</i>	<i>Características</i>
f1	Media	f29	Potencia RMS	f57	Porcentaje de energía para D1
f2	Desviación Estándar	f30	Densidad de picos en RMS.	f58	Porcentaje de energía para D2
f3	Varianza	f31	2do Pico con valor más alto	f59	Porcentaje de energía para D3
f4	Entropía	f32	Frec. de 2do pico más alto	f60	Porcentaje de energía para D4
f5	Kurasis	f33	3er Pico con valor más alto	f61	Porcentaje de energía para D5
f6	Entropía Multiescala (MSE).	f34	Frec. de 3er pico más alto	f62	Porcentaje de energía para D6
f7	Tiempo hasta máximo pico	Dominio de Escala		f63	A6 RMS en tiempo-dominio
f8	Valor RMS	f35	A6 Max. pico en frec-dominio	f64	D1 RMS en tiempo-dominio
f9	Valor pico a pico.	f36	D1 Max. pico en frec-dominio	f65	D2 RMS en tiempo-dominio
f10	Pico a RMS porción	f37	D2 Max. pico en frec-dominio	f66	D3 RMS en tiempo-dominio
f11	Energía	f38	D3 Max. pico en frec-dominio	f67	D4 RMS en tiempo-dominio

f12	Ratio del cero	f39	D4 Max. pico en frec-dominio	f68	D5 RMS en tiempo-dominio
f13	Density de picos above RMS	f40	D5 Max. pico en frec-dominio	f69	D6 RMS en tiempo-dominio
	Dominio de Frecuencia	f41	D6 Max. pico en frec-dominio	f70	A6 Pico a pico en tiempo-dominio
f14	Frecuencia de máximo pico.	f42	A6 Frec. de max. pico	f71	D1 Pico a pico en tiempo-dominio
f15	Bandawidth de 90% energía	f43	D2 Frec. de max. pico	f72	D2 Pico a pico en tiempo-dominio
f16	Entropía	f44	D3 Frec. de max. pico	f73	D3 Pico a pico en tiempo-dominio
f17	Media	f45	D4 Frec. de max. pico	f74	D4 Pico a pico en tiempo-dominio
f18	Desviación Estándar	f46	D5 Frec. de max. pico	f75	D5 Pico a pico en tiempo-dominio
f19	Varianza	f47	D6 Frec. de max. pico	f76	D6 Pico a pico en tiempo-dominio
f20	Energía	f48	A6 Media en frec.-dominio	f77	A6 Pico a RMS porción en tiempo-dominio
f21	Kurtosis	f49	D1 Media en frec.-dominio	f78	D1 Pico a RMS porción en tiempo-dominio
f22	Entropía Multiescala	f50	D2 Media en frec.-dominio	f79	D2 Pico a RMS porción en tiempo-dominio
f23	Máximo pico en 10-20 Hz banda	f51	D3 Media en frec.-dominio	F80	D3 Pico a RMS porción en tiempo-dominio
f24	Frec. de max. Pico en 10-20 Hz banda	f52	D4 Media en frec.-dominio	f81	D4 Pico a RMS porción en tiempo-dominio
f25	Máximo pico en 20-30 Hz banda	f53	D5 Media en frec.-dominio	f82	D5 Pico a RMS porción en tiempo-dominio
f26	Frec. de max. Pico en 20-30 Hz banda	f54	D6 Media en frec.-dominio	f83	D6 Pico a RMS porción en tiempo-dominio
f27	Valor RMS	f55	Media energía de componentes	f84	Media energía de wavelet coeficientes
f28	Pico a RMS porción	f56	Porcentaje de energía para A6		

Nota. Recuperado de: (Vásconez, 2020)

Las 84 características de cada segmento son almacenadas en un vector $c_{i,j}$ como se observa en la ecuación (7) y posteriormente se almacena en una matriz como se observa en la ecuación (8).

$$C_i = \left[c_{i,1}^T, c_{i,j}^T, \dots, c_{i,80}^T \right]^T, \quad (7)$$

donde $c_{i,j}$ representa el vector que contiene las 84 características de cada segmento s_i .

$$C = \left[C_1^T, C_2^T, C_3^T, \dots, C_M^T \right]^T, \quad (8)$$

donde C_M es un vector donde se almacena las características de cada señal y M es el número de señales que se encuentran en la base de datos .

Densidad Espectral

Los microterremotos son señales que se consideran parte de un proceso estocástico estacionario, para poder realizar un análisis a este tipo de señales se debe aplicar distintas técnicas, en este caso se desea obtener un análisis espectral. Para poder realizar este análisis se implementa el uso de densidad espectral de potencia (PSD, del inglés *Power Spectral Density*) con el método de Welch. Se implementa el método de Welch puesto que, en el trabajo de (Lara-Cueva, y otros, 2020) se realiza un sistema de reconocimiento de microsismos del volcán Cotopaxi al aplicar métodos de densidad espectral como: Yule-Walker; Burg; y Welch, donde se obtuvo que el método de Welch alcanza una exactitud del 90% y un BER de 0.05. Este método tiene las mejores métricas de rendimiento en comparación del método de Yule-Walker y Burg.

PSD con método de Welch.

El método de Welch consiste en dividir a la señal en segmentos en el dominio del tiempo, donde a cada segmento se calcula el periodograma en ventanas y, finalmente, se

promedian los periodogramas para obtener la estimación de PSD. Este proceso de la creación de ventanas tiene como objetivo minimizar el sesgo y la varianza en el cálculo de la PSD (Alesio, 2015). En cada ventana se pueden determinar los puntos de resolución. En esta investigación se coloca una ventana de Kaiser con una resolución de 512 puntos, lo cual entrega un análisis de PSD con 257 características en frecuencia. Esta ventana de 512 puntos se usa puesto que es una resolución media para la extracción de características en frecuencia, al aumentar esta ventana en función 2^n se obtiene una mejor resolución con mayor número de características, pero aumenta el costo computacional al momento de obtener los modelos de aprendizaje supervisado, por este motivo se usa una ventana con resolución media de 512 puntos.

Se realiza el cálculo de la PSD con el método de Welch a cada segmento $S_{i,j}$, y se almacena en un vector como se observa en la ecuación (9) y (10).

$$w_i = \left[w_{i,1}^T, w_{i,2}^T, \dots, w_{i,80}^T \right]^T, \quad (9)$$

donde $w_{i,j}$ son las 257 características obtenidas mediante la PSD de cada segmento $S_{i,j}$.

$$W = \left[w_1^T, w_2^T, w_3^T, \dots, w_M^T \right]^T, \quad (10)$$

donde w_M es la PSD de cada señal extraída y M es el número de señales que se encuentran en la base de datos.

Aprendizaje de Máquina.

Es una rama de la inteligencia artificial, también conocida como aprendizaje automático, que tiene la finalidad de que las máquinas puedan aprender sin la necesidad de ser programadas. El aprendizaje de una máquina está ligado al comportamiento de la información,

que permite detectar patrones entre los datos para determinar el comportamiento de la máquina (Otero, 2008).

El aprendizaje de máquina se subdivide en distintos modelos de aprendizaje como se muestra en la Tabla 5.

Tabla 5

Tipos de aprendizaje de Máquina

Tipos de aprendizaje de máquina	
<i>Tipo de aprendizaje</i>	<i>Características</i>
Supervisado	Tiene una relación directa la entrada del sistema con la salida. Para realizar este tipo de entrenamiento se necesita ejemplos o datos etiquetados, para que la máquina sea capaz de etiquetar nuevos datos.
No supervisado	Solo se conocen los datos de entrada, no se conoce la etiqueta de estos datos, la máquina debe reconocer patrones y lograr dar una etiqueta a estos datos.
Semi-supervisado	Este tipo de aprendizaje es una mezcla de aprendizaje supervisado y no supervisado.
Por esfuerzo	Realiza pruebas de ensayo-error, donde los resultados retroalimenta a la entrada de la máquina para poder aprender el comportamiento de los datos.

Nota. Recuperado de: (Otero, 2008).

Este trabajo se enfoca en el uso de técnicas de aprendizaje supervisado, puesto que disponemos el etiquetamiento de los microterremotos en las bases de datos dispuestas por el IGEPN.

Las técnicas de aprendizaje supervisado que se implementan para la obtención de los modelos de detección y clasificación en esta investigación son: k-Vecinos Cercanos (kNN, del

inglés *k Nearest Neighbors*), Máquina de Vectores Soporte (SVM, del inglés *Support Vector Machines*) y Árboles de Decisión (DT, del inglés *Decision Tree*). Estas técnicas de aprendizaje supervisado son usadas por los métodos de selección de características que se implementan. Se aplica tres métodos de selección de características que son: método de filtro que está ligado a la dependencia lineal entre los datos similar a los algoritmos kNN; método de envoltura que depende de los coeficientes de entrenamiento del modelo de aprendizaje supervisado ligados a los coeficientes de SVM y método embebido que para la selección de características depende de su propio modelo de entrenamiento ligado a los niveles de poda del método de DT. A continuación, se detalla cada una de estos algoritmos.

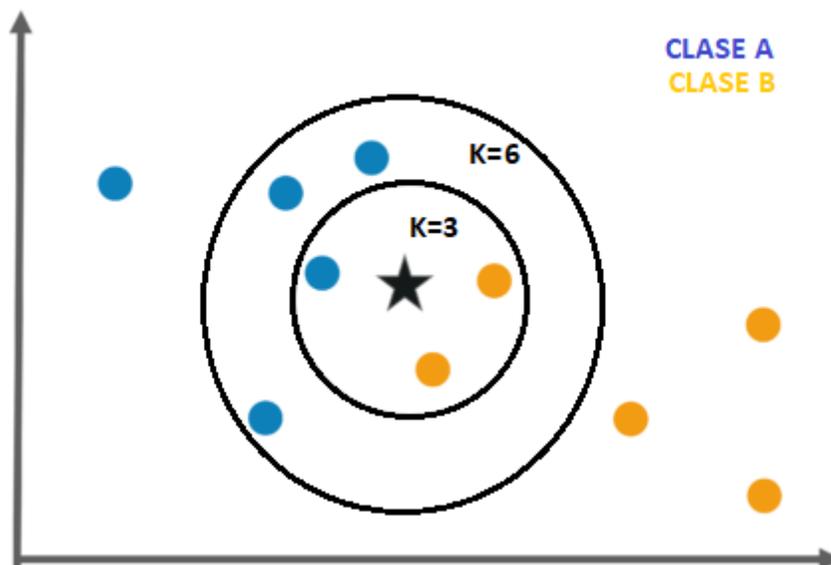
k Nearest Neighbors (kNN).

El algoritmo kNN es un algoritmo de clasificación, caracterizado por su simplicidad, alta precisión e insensible a los valores atípicos. Este algoritmo es usado más frecuentemente en extracción de datos, reconocimiento de patrones y detección de intrusos (Gonzales, AprendelA, 2018). El funcionamiento de este algoritmo consiste en colocar en un espacio las características de entrenamiento y los datos a ser evaluados. Una vez colado estos elementos, el algoritmo clasifica a que característica pertenece los datos ingresados, dependiendo de una distancia denominada **k**.

En la Figura 14 se explica el funcionamiento del algoritmo kNN, para realizar la clasificación de a qué clase pertenece un nuevo objeto caracterizado por una estrella, siendo las opciones a clasificarse en la clase A y la clase B.

Figura 14

Ejemplo de clasificación con algoritmo kNN



Nota. Recuperado de: (edureka!, 2020).

Para determinar a qué clase pertenece el objeto, primero se asigna un valor entero positivo a k , posteriormente se calcula la distancia Euclidiana como se observa en la ecuación (11) entre el dato a ser clasificado y los k elementos más cercanos. Finalmente, el algoritmo determina los k elementos con la distancia euclidiana más corta y realiza un sistema de votación entre ellos para determinar a qué clase pertenece el evento.

$$Distancia_{Euclidiana} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \quad (11)$$

Árboles de Decisión.

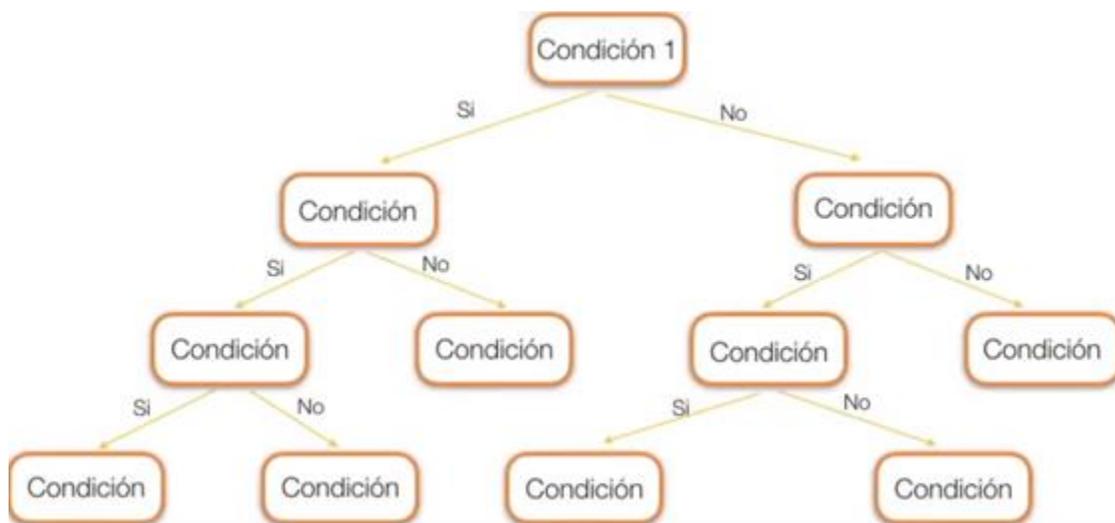
DT es un algoritmo de clasificación de aprendizaje supervisado, caracterizado por dividir a los datos de entrenamiento en dos o más conjuntos homogéneos basados en los valores más significativos de los datos de entrada. Este algoritmo es uno de los más usados por sus diversas

ventajas que son: fácil entendimiento, estabilidad, menor alteración con valores atípicos de la data, no tiene una restricción con el tipo de datos entre otras (Gonzales, AprendalA, 2018).

El algoritmo DT posee una estructura muy similar a un árbol invertido compuesto por nodos y ramas. Los nodos son los espacios de decisión y las ramas son segmentos que conectan a los nodos como se puede observar en la Figura 15. Los nodos están organizados desde los más significativos y cada nodo posee dos condiciones si cumple o no el condicional.

Figura 15

Estructura de algoritmo Árboles de Decisión

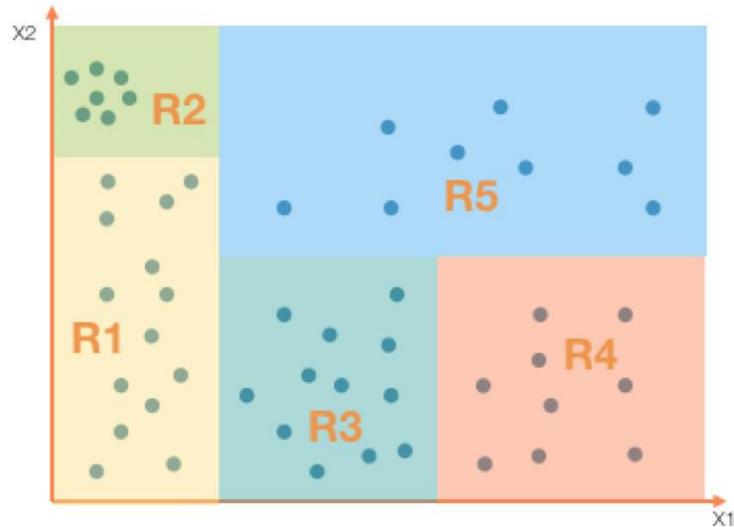


Nota. Recuperado de: (Gonzales, AprendalA, 2018).

Para la construcción de un árbol de decisión se debe considerar algunos aspectos. Dividir los datos de entrenamiento en regiones denominados espacios predictores, donde se encuentran los valores posibles que pueden tomar los datos de entrenamiento. Los datos x_1, x_2, \dots, x_n , son los datos de entrenamiento o de entrada, estos datos se dividen en j regiones distintas R_1, R_2, \dots, R_j . Esto se puede observar en la Figura 16.

Figura 16

Ejemplo de regiones que se dividen una base de datos



Nota. Recuperado de: (Gonzales, AprendelA, 2019).

El algoritmo Árboles de decisión se puede representar de manera matemática, como se observa en la ecuación (12).

$$f(x) = \sum_{j=1}^J w_j \phi(x; v_j), \quad (12)$$

donde w_j es la respuesta media de la región particular (R_j), v_j representa como se divide cada variable (Gonzales, AprendelA, 2019).

Cada una de las regiones que se construyen con los resultados obtenidos con el modelo de clasificación, tienen el objetivo de minimizar la tasa de error de clasificación (Acuña, 2000). El cálculo de la tasa de error de clasificación es usado para calcular el crecimiento del árbol de decisión como se observa en la siguiente ecuación (13).

$$E_{\text{clasificación}} = 1 - (P_{Rk}), \quad (13)$$

donde P_{Rk} es la cantidad de datos en la región de entrenamiento de la clase k.

Para complementar el cálculo del crecimiento del árbol se usa el índice de Gini, que es la medida de la varianza total en cada clase k y se considera una medida de pureza del nodo, como se observa en la ecuación (14). Donde se calcula si P_{Rk} esta cerca de cero o uno, para determinar si el nodo contiene la observación o condición. Esto permite que el árbol incremente sus niveles de poda con respecto a su modelo de clasificación.

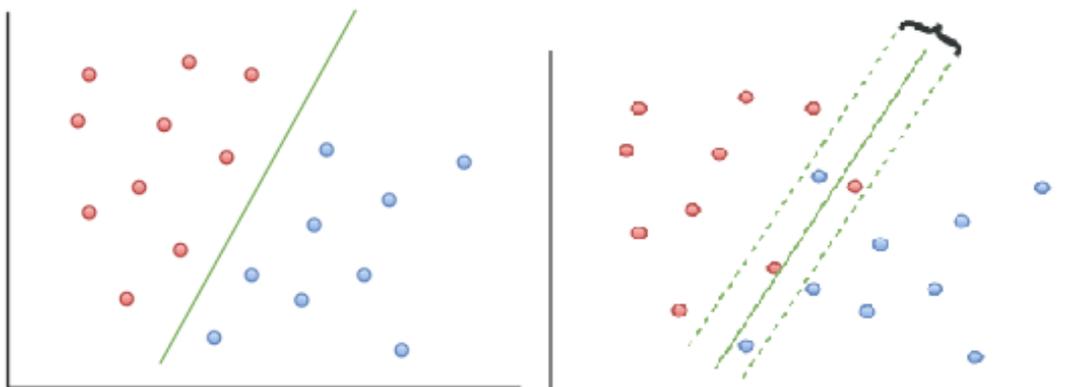
$$G = \sum_{k=1}^{Rk} P_{Rk}(1-P_{Rk}) , \quad (14)$$

Máquina de Vector de Soporte (SVM)

SVM es un algoritmo de clasificación de aprendizaje supervisado, basado en teoría estadística. Este algoritmo se caracteriza por generar un hiperplano de separación donde se colocan los datos de entrenamiento como se puede observar en la Figura 17. La formulación matemática del hiperplano se puede observar en la ecuación (15) (Mathworks, 2020). Este hiperplano solo puede dividirse en dos partes, esto quiere decir que solo puede clasificar en dos opciones.

$$F(x) = x' \beta + b = 0, \quad (15)$$

donde β y b son números reales positivos.

Figura 17*Ejemplo de Hiperplano SVM*

Nota. Recuperado de: (Kailash, 2017).

Para mejorar el funcionamiento del algoritmo SVM, se han implementado algunos métodos que permiten que el hiperplano se acople a cualquier tipo de conjunto de datos como por ejemplo el método de Kernel. Este método está enfocado en transformar la entrada dimensional baja a un alta, es decir, separa los datos de entrenamiento al agregar más dimensiones para evitar los problemas de separación entre los datos. Las ventajas de implementar este método en SVM permiten mejorar la precisión, el rendimiento y la flexibilidad.

El método de Kernel tiene algunas variantes, pero en esta investigación se usa el método de Kernel Polinomial. Este método se caracteriza por ser más generalizado que el lineal y distingue las separaciones de entrada curva o lineal. La representación matemática de esta ecuación se puede ver en la ecuación (16).

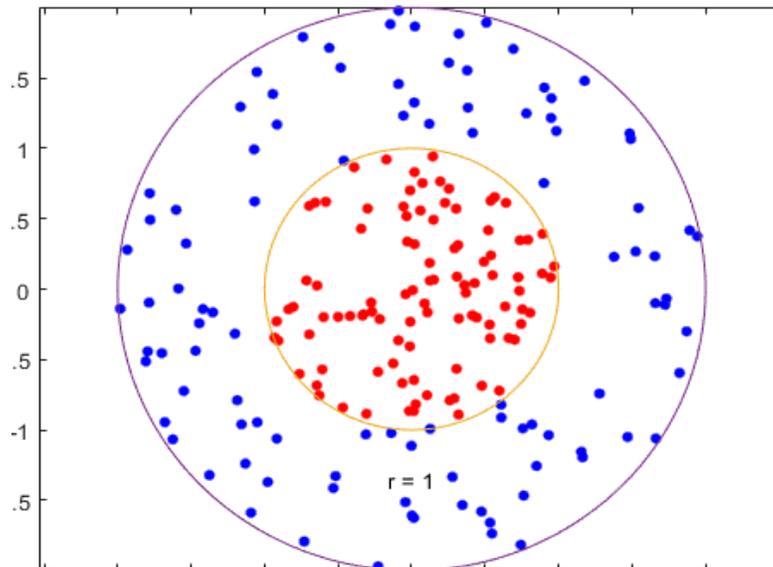
$$K(X, X_i) = 1 + \sum (X \times X_i)^d, \quad (16)$$

donde d es el grado del polinomio.

Al agregar este método, se puede obtener un hiperplano con mejor adaptación a los datos de entrenamiento como se observa en la Figura 18.

Figura 18

Ejemplo de Hiperplano con método de Kernel



Nota. Recuperado de: (Mathworks, 2020).

Selección de Características.

La selección de características es una técnica implementada para reducir la cantidad de datos de un conjunto original, con la finalidad de seleccionar los atributos con la mayor relevancia. Esta técnica es fundamental aplicar en algoritmos de aprendizaje supervisado como los vistos en la sección 2.3.4, esto permite que el aprendizaje sea más rápido con una mejor respuesta en comparación de trabajar con el conjunto de datos originales (Sposova, 2017).

Algunas de las ventajas de aplicar esta técnica son:

- Mejor velocidad de respuesta.

- Menor costo computacional.
- Reduce el sobreajuste.
- Modelos de entrenamiento más simples.
- Mejores resultados de los algoritmos de aprendizaje supervisado.

En este trabajo se aplican tres técnicas para la selección de características (*Feature Selection*) al aplicar métodos de Filtros (*Filters*), método de Embebidos (*Embedded*) y método de envoltura (*Wrapper*). A continuación, se detallan cada uno de los métodos aplicados.

Filtros.

La selección de características mediante métodos de filtros son independientes del modelo de entrenamiento de aprendizaje supervisado que se usa, los filtros son un preprocesamiento de datos según los porcentajes estadísticos que tienen los datos de entrenamiento, este porcentaje estadístico puede ser medido por la correlación que se usa para encontrar una combinación lineal entre los datos de entrenamiento y sus etiquetas (aprendeIA, 2019). Para el cálculo de la correlación se consideran los datos de entrada x_1, x_2, \dots, x_j es decir, las características de entrenamiento con sus respectivas etiquetas. Para cada dato se calcula una correlación, que permite dar un orden a las características que tengan la mejor correlación. El cálculo de la correlación se puede observar en la siguiente ecuación (17).

$$r_{XY} = \frac{\sum (X - \bar{X}) \times (Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \times \sum (Y - \bar{Y})^2}}, \quad (17)$$

donde X son todas las características, Y son las etiquetas de cada característica.

Envoltura (Wrapper).

Este es un método de eliminación de características que está relacionado directamente con el algoritmo de clasificación que se emplea en el entrenamiento, que busca las características más adecuadas para la mejora de su rendimiento (aprendeIA, 2019). Este método usa los coeficientes obtenidos en el entrenamiento para realizar un proceso de eliminación de características con menor relevancia. Mediante la eliminación de características recursivas (RFE, del inglés *Recursive Feature Elimination*), se seleccionan las mejores características para obtener el modelo de aprendizaje supervisado. El método de envoltura es principalmente usado para reducir el número de características al aplicar el algoritmo SVM (Huang, Hung, Lee, Li, & Jiang, 2014).

La formulación matemática que explica el método RFE se puede observar en la ecuación (18), que es muy similar al método de la correlación.

$$RFE = \left(\sum_k \alpha_k \times Y_k \times X_k \right)^2, \quad (18)$$

donde X_k son todas las características, Y_k son las etiquetas de cada característica y α_k son los coeficientes del modelo de entrenamiento SVM (Huang, Hung, Lee, Li, & Jiang, 2014). Con esta formulación se permite obtener el orden de las características más importantes.

Embebidos

Este método de selección de características es una combinación entre los métodos de filtros y de envoltura que se implementa mediante algoritmos de aprendizaje supervisado que tiene sus propios métodos de selección de características incorporados y se relacionan con el clasificador, un claro ejemplo es el algoritmo de árboles de decisión. Para la selección de características mediante este método se utiliza el índice de Gini visto en la ecuación (14) que

determina los niveles de poda con las mejores características obtenidas al realizar el proceso de entrenamiento.

Métricas de desempeño

Para calcular el desempeño de los sistemas de aprendizaje supervisado se usa de apoyo la matriz de confusión, que contiene la información de la predicción realizada por los modelos de clasificación. En esta matriz se coloca el número de predicciones como se puede observar en la Tabla 6. En esta matriz se calculan los verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y los falsos negativos (FN).

Tabla 6

Matriz de confusión

		Predicción	
		Positivos	Negativos
Observaciones	Positivos	Verdadero Positivo (VP)	Falso Negativo (FN)
	Negativo	Falso Positivo (FP)	Verdadero Negativo (VN)

Los valores que se calculan en esta matriz permiten obtener métricas de rendimiento como la exactitud (A, del inglés *Accuracy*), precisión (P, del inglés *Precision*), sensibilidad o recuperación (R, del inglés *Sensitivity o Recall*), especificidad (S, del inglés *Specificity*) y la tasa de error de balanceo (BER, del inglés *Balanced Error Rate*), (Lara, Benítez, Carrera, Ruiz, & Rojo Álvarez, 2016).

a. Exactitud.

Es la relación entre las predicciones correctas con el total de predicciones. La formulación matemática para esta métrica se define en la ecuación (19).

$$A(\%) = \frac{VP + VN}{VP + VN + FP + FN} \quad (19)$$

b. Precisión.

Es la relación entre las predicciones verdaderas positivas con el total de las predicciones falsos positivas. La formulación matemática para esta métrica se define en la ecuación (20).

$$P(\%) = \frac{VP}{VP + FP} \quad (20)$$

c. Sensibilidad

Es la relación entre las predicciones verdaderas positivas con las predicciones falsas negativas. La formulación matemática para esta métrica se define en la ecuación (21).

$$R(\%) = \frac{VP}{VP + FN} \quad (21)$$

d. Especificidad

Es la relación entre las predicciones verdaderas negativas con las predicciones falsas positivas. La formulación matemática para esta métrica se define en la ecuación (22).

$$S(\%) = \frac{VN}{VN + FP} \quad (22)$$

e. Tasa de error de balanceo

Es la relación presente entre la Sensibilidad y la Especificidad. La formulación matemática para esta métrica se define en la ecuación (23).

$$BER = 1 - \frac{R + S}{200} \quad (23)$$

Capítulo III

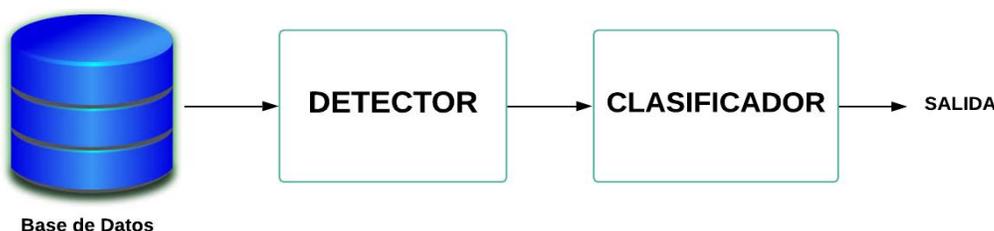
Materiales y Métodos

En este trabajo de investigación se utiliza el software Matemático MATLAB® versión R2020a el cual está instalado en un ordenador I5 con procesador de 2.4 GHz. Este programa permite manipular las bases de datos para obtener los modelos de entrenamiento del sistema de reconocimiento de microterremotos, adicionalmente permite incorporar todos los modelos obtenidos en una interfaz gráfica que es la presentación final del sistema de reconocimiento.

La metodología aplicada en este trabajo es experimental, donde el sistema consta de dos procesos generales como se observa en la Figura 19. En la Figura 20 se muestra un mejor detalle de cada una de las etapas que constituyen esta metodología.

Figura 19

Diagrama de bloques general del sistema de reconocimiento

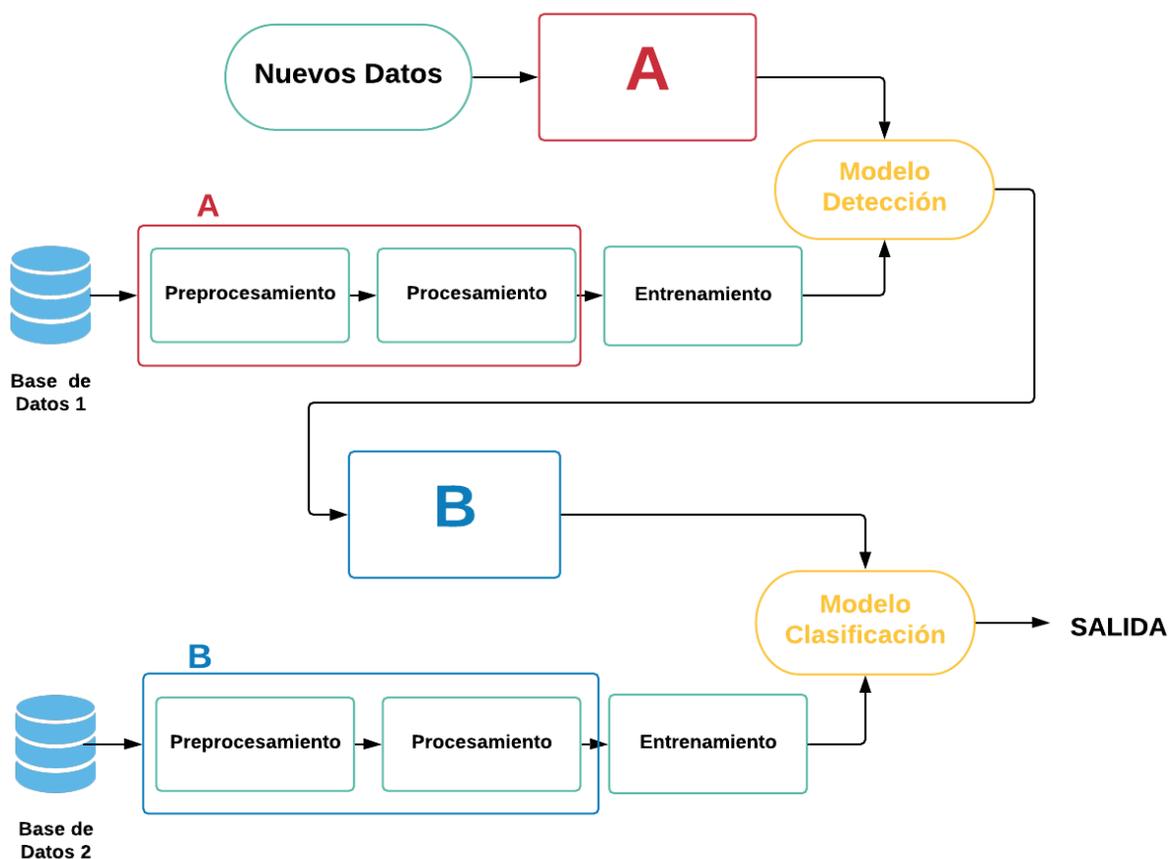


El sistema de reconocimiento consta de dos procesos generales, uno para detección y otro de clasificación. Cada uno de estos procesos son secuencias de etapas de preprocesamiento, procesamiento, extracción de características, selección de características, aplicación de algoritmos de aprendizaje supervisado y obtención de modelos tanto para detección y clasificación respectivamente. En la etapa del procesamiento existe algunas variaciones en el caso de aplicarse para detección y clasificación, se presenta una breve descripción de cada etapa y más adelante en este capítulo se detalla la diferencia en la etapa de

procesamiento (Figura 20). A continuación, se presenta una breve descripción de las etapas que constituyen el sistema de reconocimiento de microterremotos.

Figura 20

Diagrama de Bloques de la Investigación



En la primera etapa se realiza el preprocesamiento de los microterremotos almacenados en la base de datos, donde se realiza un proceso de normalización de la señal para facilitar el análisis de la misma. Como segunda etapa se encuentra el procesamiento, donde se realiza la segmentación, etiquetamiento, extracción de características y estimación espectral con el método de Welch. Posteriormente, en la tercera etapa se realiza el entrenamiento al efectuar un proceso repetitivo, a las características extraídas se aplican algoritmos de aprendizaje

supervisado k-NN, SVM y DT para obtener un modelo de detección. Se realiza la selección de las mejores características al usar métodos de filtros, embebidos y envoltura, donde nuevamente se emplean algoritmos de aprendizaje supervisado para obtener un nuevo modelo de detección. Finalmente se obtienen las métricas de rendimiento en cada proceso repetitivo realizado para comparar los resultados que se presentan en el capítulo IV de esta investigación.

Detector

Base de Datos.

La base de datos usada en este trabajo de investigación fue proporcionada por el IGEPN, el cual es el encargado de la monitorización de la actividad sísmica y volcánica en el Ecuador. Esta base de datos fue recopilada en el año 2012, donde se almacenó cada evento producido. Estos eventos fueron detectados por varias estaciones simultáneamente, lo que permitió verificar que se constataba de microterremotos, estas estaciones se encuentran ubicadas en las faldas del volcán Cotopaxi, entre las cuales se encuentran 5 estaciones de corto periodo que trabajan en un rango de frecuencia de 1 a 50 Hz y 12 estaciones de banda ancha que trabajan a una frecuencia de 0,1 a 50 Hz (IGEPN, 2017). En este trabajo de investigación se utiliza la base de datos recogida de la estación VC1 (Volcán Cotopaxi 1), la cual trabaja a una frecuencia de muestreo de digitalización de 100 Hz.

La base de datos consta de 350 señales de microterremotos etiquetados por los expertos del IGEPN, en las cuales poseen 282 eventos del tipo LP, 50 eventos del tipo VT, 15 eventos tipo TRE y 3 eventos HYB, estas señales están ordenadas por mes y tipo de microterremoto como se muestra en la Figura 21. A continuación, se realiza una descripción de la información que constituye cada base de datos entregada por IGEPN.

En la Figura 21 se muestra la estructura de la información proporcionada por la base de datos utilizada para la presente investigación. Se utiliza el software matemático MATLAB® R2020a para visualizar la tabla de datos. Esta base está constituida por 10 columnas; la primera columna hace referencia al mes en el cual fue recopilado el microterremoto; en la segunda columna denominada “Signal” posee los microterremotos recopilados en cada mes respectivamente, cada señal tiene un total de 120000 muestras y una duración de 20 minutos; en la tercera y cuarta columna denominadas “InicioEvento” y “FinEvento”, se halla el punto de inicio y fin de cada microterremoto que se encuentra en la columna “Signal”; en la quinta columna denominada “SoloEvento” se encuentra la extracción de solo el evento sísmico, que tiene una duración variable dependiendo del tipo de microterremoto; en la sexta columna denominada “Etiqueta” se encuentra la identificación del tipo de microterremoto, donde los expertos del IGEPN colocaron las siguientes etiquetas para cada microterremoto: 0 para microterremotos del tipo LP, 1 para microterremotos del tipo VT, 2 para microterremotos del tipo TRE y 3 para microterremotos del tipo HYB; finalmente en la séptima, octava, novena y décima columna se encuentran los microterremotos divididos por su tipo.

Figura 21

Base de datos 1

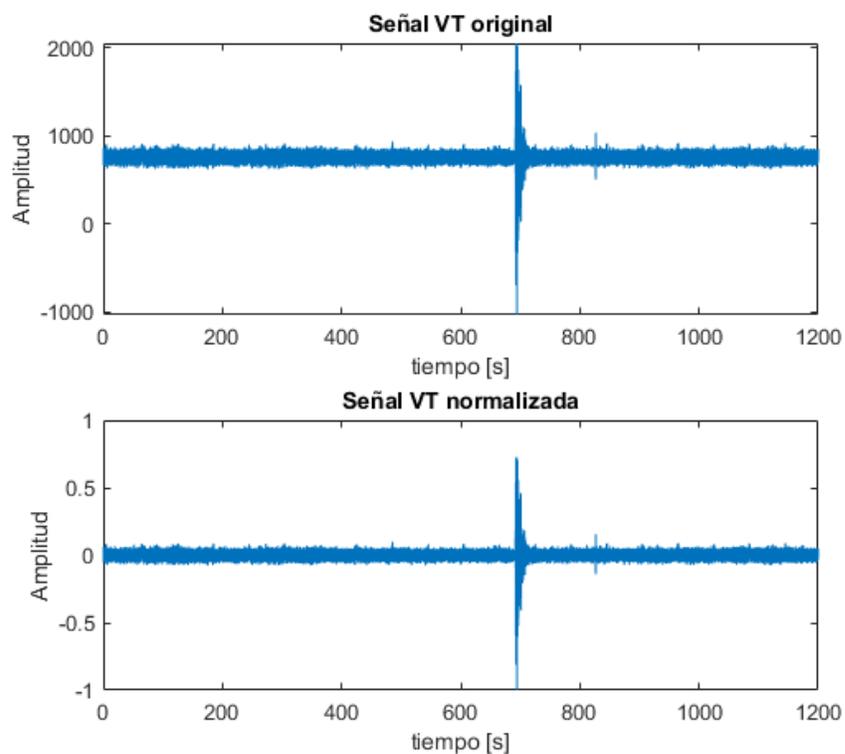
	1 Mes	2 Signal	3 InicioEvento	4 FinEvento	5 SoloEvento	6 Etiqueta	7 LP	8 VT	9 TR	10 HYB
1	'ENERO'	1x71 cell	1x71 cell	1x71 cell	1x71 cell	1x71 cell	1x53 cell	1x15 cell	1x1 cell	1x2 cell
2	'FEBRERO'	1x27 cell	1x27 cell	1x27 cell	1x27 cell	1x27 cell	1x23 cell	1x3 cell	1x1 cell	[]
3	'MARZO'	1x25 cell	1x25 cell	1x25 cell	1x25 cell	1x25 cell	1x19 cell	1x5 cell	1x1 cell	[]
4	'ABRIL'	1x21 cell	1x21 cell	1x21 cell	1x21 cell	1x21 cell	1x17 cell	1x1 cell	1x3 cell	[]
5	'MAYO'	1x32 cell	1x32 cell	1x32 cell	1x32 cell	1x32 cell	1x24 cell	1x7 cell	1x1 cell	[]
6	'JUNIO'	1x39 cell	1x39 cell	1x39 cell	1x39 cell	1x39 cell	1x28 cell	1x7 cell	1x3 cell	1x1 cell
7	'JULIO'	1x63 cell	1x63 cell	1x63 cell	1x63 cell	1x63 cell	1x59 cell	1x2 cell	1x2 cell	[]
8	'AGOSTO'	1x32 cell	1x32 cell	1x32 cell	1x32 cell	1x32 cell	1x31 cell	1x1 cell	[]	[]
9	'SEPTIEMBRE'	1x28 cell	1x28 cell	1x28 cell	1x28 cell	1x28 cell	1x24 cell	1x4 cell	[]	[]
10	'OCTUBRE'	1x6 cell	1x6 cell	1x6 cell	1x6 cell	1x6 cell	1x4 cell	[]	1x2 cell	[]
11	'NOVIEMBRE'	1x2 cell	1x2 cell	1x2 cell	1x2 cell	1x2 cell	[]	1x2 cell	[]	[]
12	'DICIEMBRE'	1x4 cell	1x4 cell	1x4 cell	1x4 cell	1x4 cell	[]	1x3 cell	1x1 cell	[]

Preprocesamiento

Las señales entregadas por el IGEPN son señales ya filtradas en un rango de 1 a 50 Hz, previamente han sido eliminados los picos de baja amplitud que se encuentran en frecuencias de 0.1 a 0.3 Hz, debido principalmente a movimientos de placas oceánicas. Las señales recopiladas por la estación VC1 son señales que tienen distintas amplitudes que dependen de la intensidad de cada microterremoto, no tiene una escala definida por lo cual a todas las señales de la base de datos se deben normalizar. Para realizar la normalización se debe remover la media y llevar a cada señal a una escala de $[-1 \text{ a } 1]$ centrada en cero, para esto se aplica las ecuaciones (1) y (2). En la Figura 22 se puede observar la comparación de una señal original con otra normalizada.

Figura 22

Comparación de señal VT original y normalizada



Procesamiento

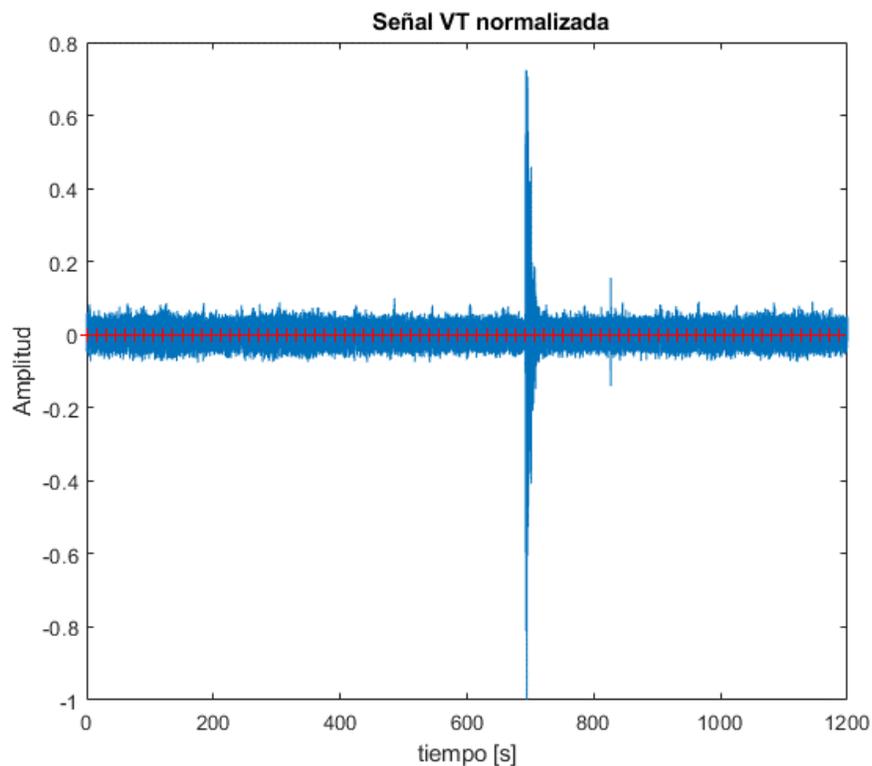
En esta etapa se realiza el proceso de la segmentación, etiquetamiento y obtención de características de las señales normalizadas. A continuación, se explica cada uno de los procesos que se realiza a cada señal.

a) Segmentación

Para aplicar este proceso se toma la señal normalizada con una duración de 1200 segundos o 20 minutos y se divide en segmentos de 15 segundos como se observa en la Figura 23. Al realizar esto se obtiene de cada señal 80 segmentos que se almacenan en un vector como se observa en la ecuación (3), como existe un total de 350 señales se obtiene un total de 28000 segmentos que son almacenados en una matriz como se observa en la ecuación (4).

Figura 23

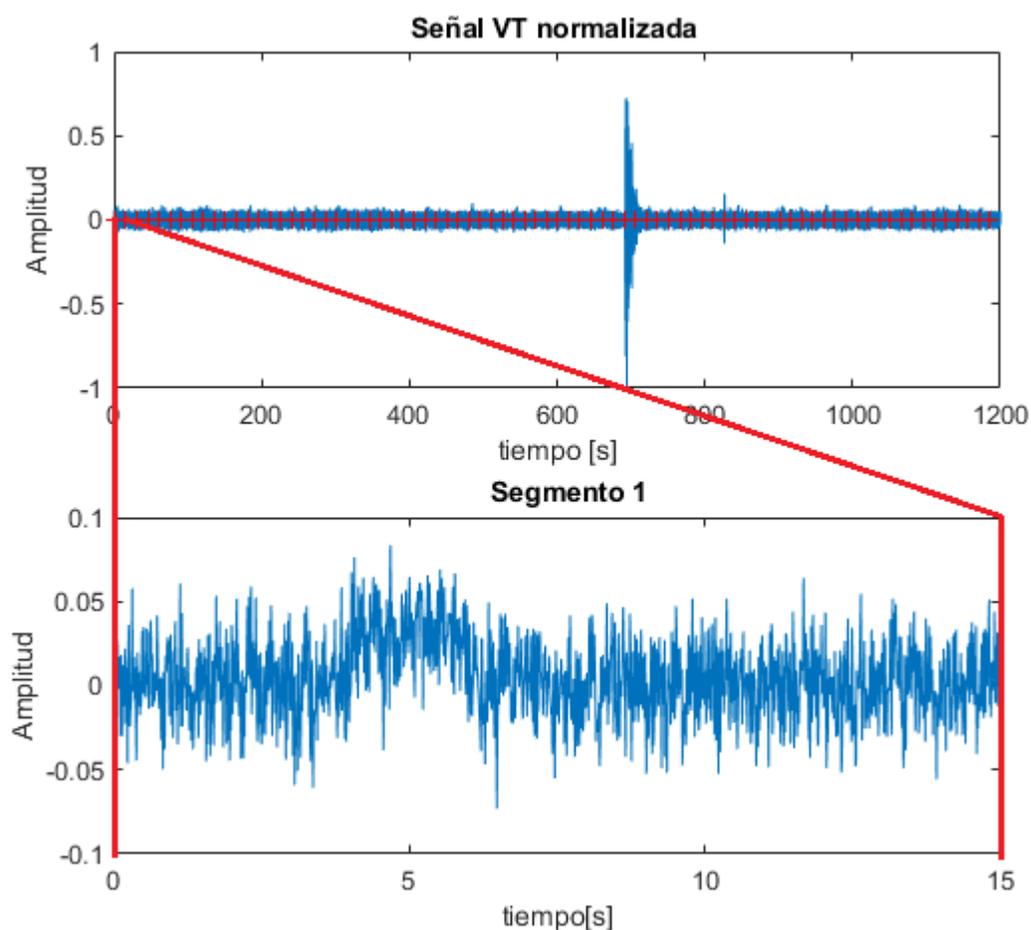
Señal segmentada en ventanas de 15 segundos.



Cada señal al tener un total de 120000 muestras al segmentarlas en ventanas de 15 segundos se toma un total de 1500 muestras, esto se debe a que cada señal esta digitalizada con una frecuencia de muestreo de 100 Hz. En la Figura 24 se muestra el primer segmento de la señal de la Figura 23. En este trabajo se realiza una ventana de segmentación (w), con $w=15$ s como las realizadas en el trabajo de investigación (Lara, Benítez, Carrera, Ruiz, & Rojo Álvarez, 2016), donde se obtuvieron los mejores resultados en la detección de microterremotos con distintas ventanas de tiempo.

Figura 24

Primer segmento de 15 segundos señal de tipo VT.

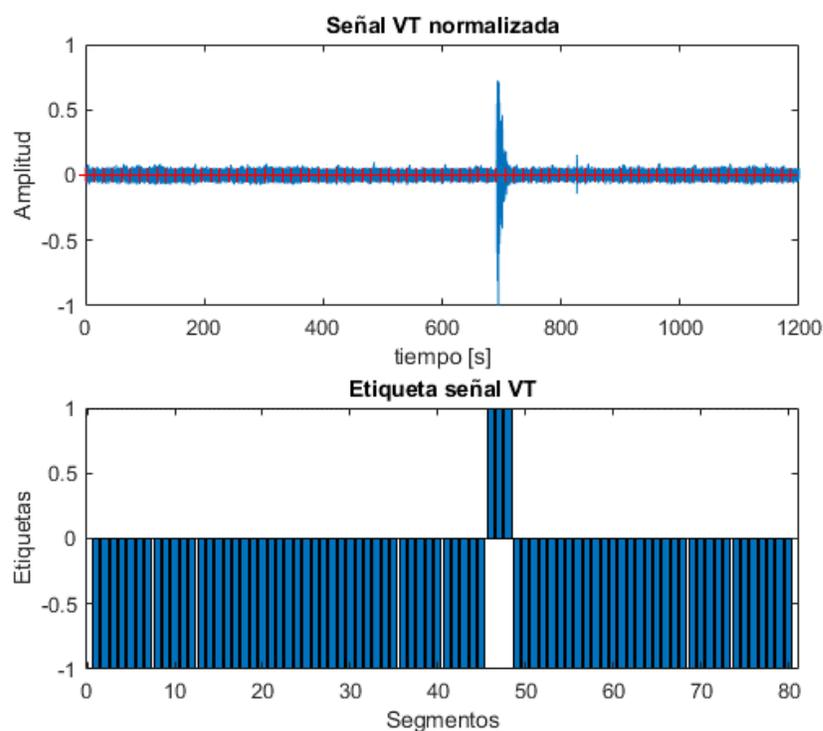


b) Etiquetamiento

El etiquetamiento se realiza a cada ventana de segmentación (w), con $w=15$ s, donde se coloca una etiqueta de un +1 cuando el segmento contenga la señal y un -1 en el caso de no contener el evento sísmico como se observa en la Figura 25. Se obtiene un total de 80 etiquetas de cada señal que se almacena en un vector como se observa en la ecuación (5), este proceso se realiza para cada señal que se almacena en una matriz como se observa en la ecuación (6). La información para determinar los segmentos que contienen el evento sísmico se encuentra en la base de datos en la columna denominada "InicioEvento" y "FinEvento". Se considera que un segmento contiene la señal cuando los puntos de inicio y fin se encuentran dentro de un determinado segmento.

Figura 25

Etiquetamiento de segmentos señal del tipo VT

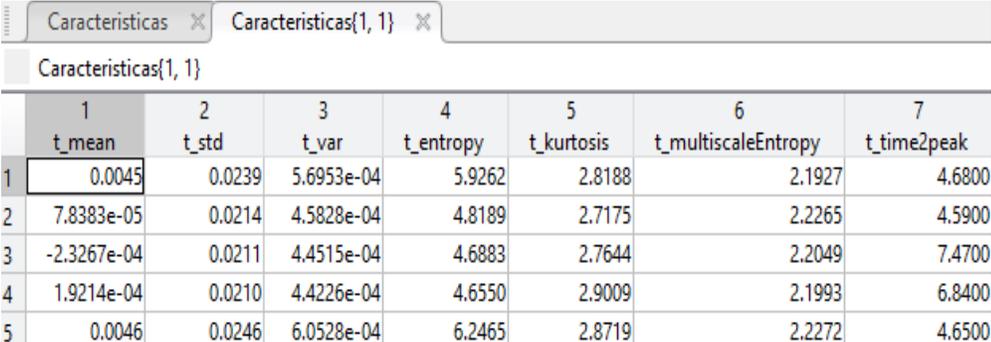


c) Extracción de características

En esta etapa se trabaja con cada ventana de $w=15$ s y se obtienen características en tiempo, frecuencia y escala. Se calculan un total de 84 características las cuales se especifican en la Tabla 4. Las características están divididas en: 13 en tiempo, 21 en frecuencia y 50 en escala. Estas características más adelante son usadas para el entrenamiento, en la Figura 26 se puede observar algunas características obtenidas. Estas características son almacenadas en un vector $c_{i,j}$ como se observa en la ecuación (7), y finalmente se agregan las características de la señal segmentada a una matriz como se observa en la ecuación (8).

Figura 26

Obtención de 84 Características segmentos de 15 segundos.



	1	2	3	4	5	6	7
	t_mean	t_std	t_var	t_entropy	t_kurtosis	t_multiscaleEntropy	t_time2peak
1	0.0045	0.0239	5.6953e-04	5.9262	2.8188	2.1927	4.6800
2	7.8383e-05	0.0214	4.5828e-04	4.8189	2.7175	2.2265	4.5900
3	-2.3267e-04	0.0211	4.4515e-04	4.6883	2.7644	2.2049	7.4700
4	1.9214e-04	0.0210	4.4226e-04	4.6550	2.9009	2.1993	6.8400
5	0.0046	0.0246	6.0528e-04	6.2465	2.8719	2.2272	4.6500

d) Estimación espectral

En esta etapa se trabaja con cada ventana $w=15$ s y se calcula la PSD con el método de Welch, se almacena en un vector como se observa en la ecuación (9). La PSD se calcula con una ventana de 512 puntos y entrega una resolución de 257 puntos, lo cual es una resolución media para el análisis espectral de la señal, estos puntos de resolución son considerados como características en frecuencia. Estas características más adelante son usadas para el entrenamiento, en la Figura 27 se pueden observar algunas características obtenidas. La PSD de la señal segmentada se almacena en una matriz como se observa en la ecuación (10).

Figura 27

Obtención PSD con una resolución de 257 puntos para cada segmento de 15 segundos.

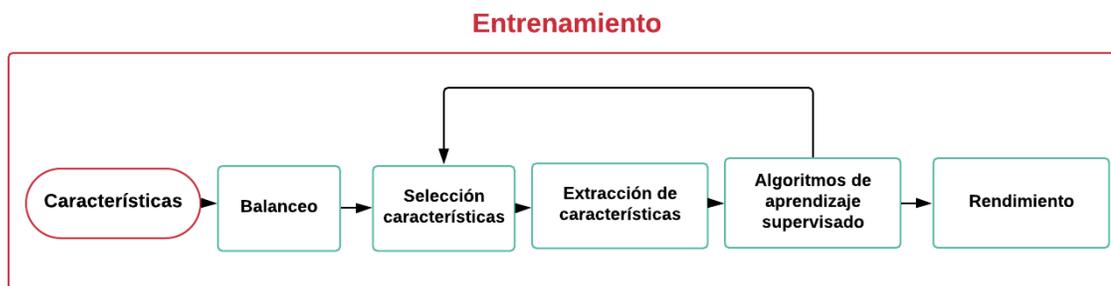
W{1, 1}	
1	
1	1x257 double
2	1x257 double
3	1x257 double
4	1x257 double
5	1x257 double
6	1x257 double
7	1x257 double
8	1x257 double
9	1x257 double

Entrenamiento

Esta etapa es la más importante del sistema de detección, es un proceso repetitivo como se muestra en la Figura 28, donde entran las características y se obtienen los modelos para la detección de microterremotos. Primero se realiza un balance entre los segmentos que contiene y no contiene el evento sísmico, posteriormente se realiza la selección de características que se ingresan a los algoritmos de aprendizaje supervisado para la obtención de modelos de clasificación, y finalmente se calculan las métricas de rendimiento para la comparación de cada modificación en esta etapa. A continuación, se realiza una descripción de los procesos en la etapa de entrenamiento.

Figura 28

Proceso de entrenamiento detección de microterremotos.



Las características que ingresan al sistema de entrenamiento son las calculadas en la sección 3 de este documento.

a) Balance

Para realizar el entrenamiento los algoritmos de aprendizaje supervisado necesitan partir de un equilibrio de datos, es decir, los datos de entrenamiento deben estar constituidos de segmentos que contengan el evento sísmico y el mismo número de segmentos que no contengan el evento (Roman V. , 2019).

La base de datos usada para obtener los modelos de detección contiene un total de 350 eventos sísmicos, de los cuales se toma el 50% de todos los tipos de microterremotos almacenados en esta base de datos, es decir, se toma 175 señales para el entrenamiento y la otra mitad de datos para las pruebas. Cada señal tiene 80 segmentos que están etiquetados con un +1 cuando contiene el evento sísmico y con -1 cuando no contiene el evento.

El balance de los datos es una parte fundamental en esta investigación, como se puede observar en la Figura 25, existe una mayor cantidad de segmentos que no contienen el evento sísmico. Cada señal en promedio solo tiene un 4% de segmentos que contienen el evento sísmico, es decir, entre [2 a 4] segmentos de los totales que depende del microterremoto que se

analiza. Bajo este contexto se propone aumentar los segmentos que no contienen el evento sísmico, puesto que existe mayor información sobre tales segmentos. Esto permite que exista una mayor tendencia a calcular los segmentos que no contienen la señal.

Para realizar el balance de datos primero se contabilizan los segmentos que contienen la señal sísmica de los datos usados para el entrenamiento, se selecciona el mismo número de segmentos que no contengan el evento. Posteriormente se incrementan la cantidad de segmentos que no contienen señal para observar cual es la mejor proporción de datos a utilizarse. Se contabilizan un total de 361 segmentos que contienen el evento sísmico de las señales de entrenamiento, con este valor equilibran los datos con la misma cantidad de segmentos sin señal y posteriormente se aumentan en proporciones del mismo número. Se realizan pruebas con cada proporción, donde se aplican los algoritmos de aprendizaje supervisado para determinar la mejor proporción mediante las métricas de rendimiento. Adicionalmente se realiza este proceso con los datos originales y posteriormente se remueve los componentes DC, es decir $\mu = 0$ y $\sigma = 1$ a las características del mismo tipo como se observa en la ecuación (1).

En la Figura 35 se pueden observar los 3 modelos de aprendizaje supervisado obtenidos y ver como varía el BER a medida que aumentan las proporciones de segmentos que no contienen el evento sísmico. Los resultados de este proceso se pueden observar en el capítulo 4 de la presente investigación.

b) Selección de Características y Aprendizaje Supervisado

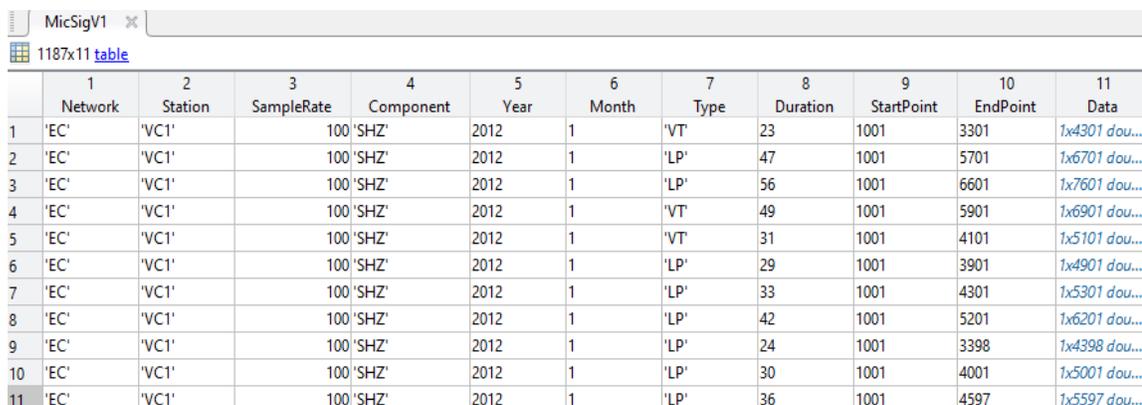
En este proceso se seleccionan las características con mayor relevancia para obtener los modelos de aprendizaje supervisado, para esto se siguen los siguientes pasos: primero, se obtienen los algoritmos de aprendizaje supervisado con todas las características, es decir, las 84

características en tres dominios distintos y las 257 características obtenidas de la PSD; segundo, se aplican los métodos de selección de características mencionados en la sección 2.3.5, donde se determina con cada método las mejores características; posteriormente, se obtienen nuevos algoritmos de aprendizaje supervisado con las características con mayor relevancia; finalmente se determinan las mejores características con los tres métodos distintos observando las métricas de rendimiento que se obtuvieron en cada prueba. A continuación, se observan la selección de características con cada método. En la sección IV de este trabajo de investigación se detalla la selección de características para el sistema de detección.

Clasificación

Base de Datos

La base de datos usada para obtener los modelos de clasificación de microterremotos fue proporcionada por el IGEPN, estos datos son recopilados a lo largo del año 2012, enero 2013 y enero 2014. Los eventos son detectados por las estaciones ubicadas en el volcán Cotopaxi que se muestra en la Figura 5. Los eventos son recopilados por las estaciones (VC1, BREF y BVC2) que trabajan en un rango de frecuencias de respuesta de 0.1 a 50 Hz. Cada estación digitaliza la señal con una tasa de muestreo, las estaciones BREF y BVC2 digitalizan a una frecuencia de muestreo de 50 Hz y la estación VC1 digitaliza a una frecuencia de muestreo de 100 Hz. Esta base de datos consta con un total de 1187 microterremotos con distintas variaciones de tiempo, esta base está compuesta de 1044 eventos del tipo LP, 101 eventos VT, 27 eventos Regionales, 8 eventos del tipo HYB y 7 eventos de rompimiento de glaciares (ICQ, del inglés *Icequake*). En la Figura 29 se observa la información que proporciona la base de datos.

Figura 29*Base de datos 2*


	1	2	3	4	5	6	7	8	9	10	11
	Network	Station	SampleRate	Component	Year	Month	Type	Duration	StartPoint	EndPoint	Data
1	'EC'	'VC1'	100	'SHZ'	2012	1	'VT'	23	1001	3301	1x4301 dou...
2	'EC'	'VC1'	100	'SHZ'	2012	1	'LP'	47	1001	5701	1x6701 dou...
3	'EC'	'VC1'	100	'SHZ'	2012	1	'LP'	56	1001	6601	1x7601 dou...
4	'EC'	'VC1'	100	'SHZ'	2012	1	'VT'	49	1001	5901	1x6901 dou...
5	'EC'	'VC1'	100	'SHZ'	2012	1	'VT'	31	1001	4101	1x5101 dou...
6	'EC'	'VC1'	100	'SHZ'	2012	1	'LP'	29	1001	3901	1x4901 dou...
7	'EC'	'VC1'	100	'SHZ'	2012	1	'LP'	33	1001	4301	1x5301 dou...
8	'EC'	'VC1'	100	'SHZ'	2012	1	'LP'	42	1001	5201	1x6201 dou...
9	'EC'	'VC1'	100	'SHZ'	2012	1	'LP'	24	1001	3398	1x4398 dou...
10	'EC'	'VC1'	100	'SHZ'	2012	1	'LP'	30	1001	4001	1x5001 dou...
11	'EC'	'VC1'	100	'SHZ'	2012	1	'LP'	36	1001	4597	1x5597 dou...

La base de datos está constituida por 11 columnas las cuales contienen toda la información de cada evento sísmico; la primera columna hace referencia a la red de monitorización donde se recopilaron los datos; la segunda columna proporciona la información sobre la estación que recopiló el microterremoto; la tercera columna informa la frecuencia de muestreo de cada estación; la quinta y sexta columna contienen el año y mes en que se produjo el microterremoto; la séptima columna es el etiquetamiento realizado por los expertos del IGEPN del tipo de microterremoto registrado; la octava columna es la duración en segundos de cada evento sísmico; la novena y décima columna es la información del punto de inicio y fin del evento sísmico de cada microterremoto que se muestra en la onceava columna, donde se encuentra la señal de cada evento.

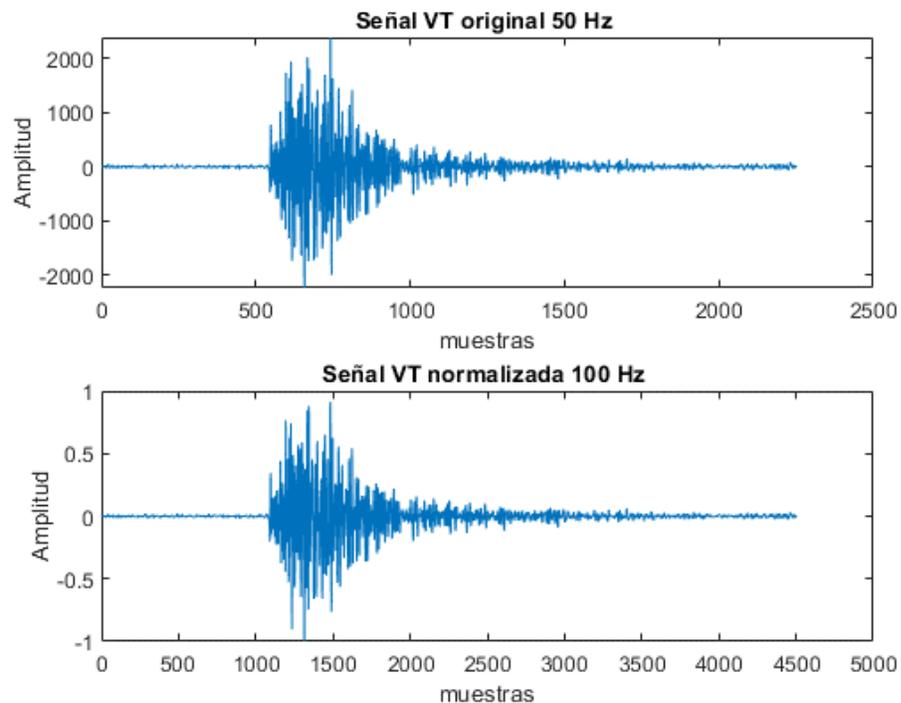
Preprocesamiento

Las señales entregadas por el IGEPN son señales ya filtradas en un rango de 1 a 50 Hz. Las señales al provenir de distintas estaciones de monitorización tienen distintas frecuencias de muestreo, se deben remuestrear a todas a una frecuencia de muestreo de 100 Hz. En la base de datos existen señales muestreadas a una frecuencia de 50 Hz y para llevarlas a 100 Hz se debe

remuestrear los datos. Para poder remuestrear (*resample*) la señal aplica una función en el software MATLAB. Posteriormente a las señales muestreadas a una frecuencia se realiza la normalización, para ello se debe remover la media y llevar a cada señal a una escala de [-1 a 1] centrada en cero, para esto se aplican las ecuaciones (1) y (2). En la Figura 30 se puede observar la comparación de una señal original con otra normalizada y remuestreada a 100 Hz.

Figura 30

Comparación entre la señal VT original a 50 Hz y normalizada a 100 Hz.



Procesamiento

La etapa de procesamiento en el sistema de clasificación tiene variaciones con el sistema de detección. La principal variación es que en la etapa de clasificación se obtiene el etiquetamiento y la extracción de características de las señales normalizadas, donde solo se consideran el tiempo de inicio y fin de la base de datos. A continuación, se explica cada uno de los procesos que se realiza a cada señal.

a) Etiquetamiento

Para el etiquetamiento de las señales solo se consideran los 1044 eventos del tipo LP y los 101 eventos del tipo VT. Solo estos dos tipos de eventos se van a usar para obtener los modelos de entrenamiento para la clasificación de microterremotos; para los eventos del tipo LP se coloca una etiqueta de -1 y para los eventos del tipo VT se coloca una etiqueta de +1.

b) Extracción de características

En esta etapa se toma la señal normalizada desde los puntos de inicio y fin que corresponde a cada microterremoto, y se obtienen características en tiempo, frecuencia y escala. Se calculan un total de 84 características las cuales se especifican en la Tabla 4. Este proceso se realiza a los 1187 eventos almacenados en la base de datos de la Figura 29. En la Figura 31 se puede observar algunas características obtenidas.

Figura 31

Obtención de las 84 características de las señales normalizadas

	1	2	3	4	5	6	7	8	9
	t_mean	t_std	t_var	t_entropy	t_kurtosis	t_multiscaleEntropy	t_time2peak	t_rms	t_peak2peak
1	-8.0519e-05	0.1593	0.0254	125.5754	8.3045	1.6596	1.4900	0.1592	1.7357
2	-3.9743e-05	0.1978	0.0391	393.8532	5.2904	2.0046	10.0600	0.1978	1.8378
3	-1.5126e-05	0.1909	0.0365	396.6835	6.6663	1.7445	14.8700	0.1909	1.9341
4	-1.4593e-05	0.3631	0.1318	620.9504	3.5809	1.6144	11.9300	0.3630	1.8073
5	-4.0935e-05	0.2178	0.0474	235.1077	6.6473	1.5676	3.7600	0.2178	1.7310

c) Estimación espectral

En esta etapa se toma cada señal y se calcula la PSD con el método de Welch, se almacenan en un vector como se observa en la ecuación (9). La PSD se calcula con una ventana de 512 puntos y entrega una resolución de 257 puntos, estos puntos de resolución son considerados como características en frecuencia. Estas características más adelante son usadas para el entrenamiento, en la Figura 32 se puede observar algunas características obtenidas. La PSD de la señal segmentada se almacena en una matriz como se observa en la ecuación (10).

Figura 32

Obtención PSD con una resolución de 257 puntos para cada señal normalizada

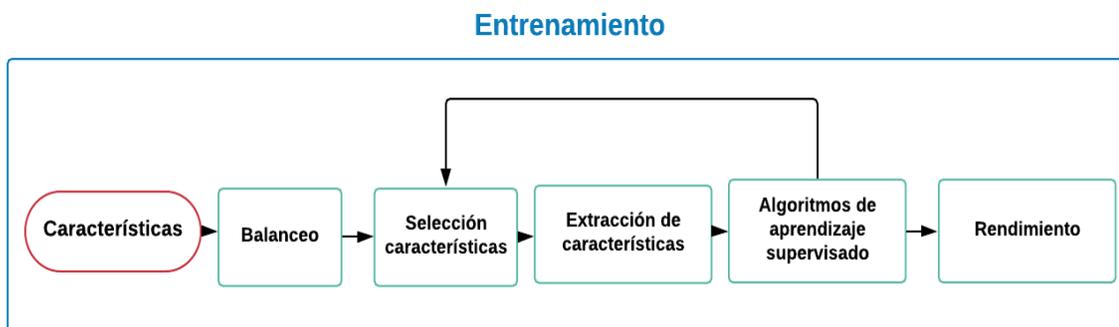
1187x1 cell	
	1
1	1x257 double
2	1x257 double
3	1x257 double
4	1x257 double
5	1x257 double
6	1x257 double

Entrenamiento

Esta etapa es un proceso repetitivo como se muestra en la Figura 33, donde entran las características y se obtienen los modelos para la clasificación de microterremotos. Primero se realiza un balanceo entre los datos a usarse para obtener los modelos de clasificación, posteriormente se realiza la selección de características que se ingresan a los algoritmos de aprendizaje supervisado para la obtención de los modelos de clasificación, y finalmente se calculan las métricas de rendimiento para la comparación de cada modificación en esta etapa. A continuación, se realiza una descripción de los procesos que se realizan en la etapa de entrenamiento.

Figura 33

Proceso de entrenamiento clasificación de microterremotos



Las características que ingresan al sistema de entrenamiento son las calculadas en la sección 3 de este documento.

c) Balance

La base de datos usada para obtener los modelos de detección contiene un total de 1187 eventos sísmicos, de los cuales son 1044 eventos del tipo LP y 101 eventos del tipo VT. Se seleccionan el 50% de los eventos del tipo LP para el entrenamiento, es decir, 522 eventos; de los 522 eventos del tipo LP usados para entrenamiento se divide estos datos en KFold, partición aleatoria no estratificada para la validación cruzada para datos, de donde, se eligen 101 eventos para nivelar a la cantidad de datos que tiene los eventos del tipo VT, de los 101 eventos del tipo LP y VT se eligen el 60% de los eventos sísmicos para entrenamiento y el 40% son para las pruebas, este porcentaje de eventos sísmicos se usan puesto que en el trabajo (Lara-Cueva, y otros, 2020) se obtuvieron las mejores métricas de rendimiento al usar ese porcentaje de datos para el entrenamiento. Para las pruebas se utilizan el resto de los eventos del tipo LP, es decir, los 522 eventos del tipo LP que no se usaron para el entrenamiento.

d) Selección de Características y Aprendizaje Supervisado

En este proceso se seleccionan las características con mayor relevancia para obtener los modelos de aprendizaje supervisado, para esto se siguen los siguientes pasos: Primero, se obtienen los algoritmos de aprendizaje supervisado con todas las características, es decir, las 84 características en tres dominios distintos y las 257 características obtenidas de la PSD; segundo, se aplican los métodos de selección de características mencionados en la sección 2.3.5, donde se determinan con cada método las mejores características; posteriormente, se realiza una comparación de cómo evoluciona el BER en los tres métodos de selección de características para determinar las que usaran en los modelos finales. En la sección IV de este trabajo de investigación se detalla la selección de características para el sistema de clasificación.

Sistema de reconocimiento

El sistema de reconocimiento consiste en unir el sistema de detección y clasificación en una sola interfaz gráfica, para esto se usa la herramienta appdesigner donde se incorporan los modelos obtenidos en los procesos de entrenamiento. La interfaz gráfica proporciona la siguiente información de cada señal ingresada al sistema de reconocimiento: porcentaje de análisis del sistema de reconocimiento, tiempo de procesamiento, gráfica del evento sísmico que va a ser analizado, segmentos donde se encuentra el evento sísmico, extracción del evento sísmico, PSD del evento sísmico, punto de inicio y fin de cada microterremoto, y la etiqueta del tipo de evento que está siendo analizado. Finalmente se determina el desempeño del sistema de reconocimiento que se presenta en el capítulo IV de esta investigación.

Capítulo IV

Resultados

En esta sección se detallan las medidas de rendimiento obtenidas en cada proceso planteado en la sección III. Las medidas de rendimiento que se obtienen son para la etapa de detección y clasificación del sistema de reconocimiento de microterremotos.

Detección

Para el entrenamiento y obtención de modelos se realizó el proceso de balance de datos, extracción de características, selección de características y obtención de modelos finales. A continuación, se detallan los resultados obtenidos en cada proceso realizado.

Balance

Se plantea la realización de un balance de datos, se efectuaron variaciones con respecto a las proporciones de segmentos sin señal para observar la variación de las medidas de rendimiento. Este proceso se realiza con las características originales y las características donde se remueve los componentes DC, es decir $\mu = 0$ y $\sigma = 1$.

Se obtienen las métricas de rendimiento de los algoritmos de aprendizaje supervisado para los modelos obtenidos con las 84 características y las 257 características PSD.

- **84 Características**

Como se observa en la Tabla 7, se obtienen las medidas de rendimiento con las características originales con un modelo kNN con $k=20$, el modelo DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, se puede notar que las mejores métricas de rendimiento se obtienen en una proporción de datos de 1 a 6, esta es la máxima proporción que se presenta, puesto que al aumentar esta proporción existe un sobre entrenamiento y los

modelos clasificación consideran como segmentos sin señal a todos los segmentos. Esta proporción maximiza las medidas de rendimiento en los algoritmos de clasificación empleados.

Tabla 7

Métricas de rendimiento para balance de datos 84 Características.

Proporción	A(%)	P(%)	R(%)	S(%)	BER
Algoritmos de detección con kNN					
1:1	89,77	63,59	98,49	87,90	0,070
1:2	92,24	72,65	98,52	90,75	0,060
1:3	92,85	74,45	97,85	91,60	0,055
1:4	93,48	76,67	97,87	92,36	0,048
1:5	93,84	78,09	97,88	92,83	0,046
1:6	94,38	80,00	97,88	93,25	0,043
Algoritmos de detección con DT					
1:1	71,76	22,45	98,31	69,37	0,160
1:2	77,44	33,51	97,61	75,81	0,140
1:3	79,35	39,42	96,50	76,65	0,134
1:4	80,49	39,87	97,63	77,93	0,122
1:5	81,32	42,80	95,52	79,00	0,127
1:6	82,88	47,29	94,24	80,81	0,120
Algoritmos de detección con SVM					
1:1	84,29	48,67	99,22	81,68	0,100
1:2	87,58	57,20	98,44	85,42	0,080
1:3	88,45	59,91	98,48	86,38	0,075
1:4	90,40	67,00	96,45	89,00	0,072
1:5	90,70	67,50	97,12	89,22	0,068
1:6	91,07	69,19	96,47	89,78	0,069

En la Tabla 8, se obtienen las métricas de rendimiento con las características al remover los componentes DC, es decir $\mu = 0$ y $\sigma = 1$, con un modelo kNN con $k=20$, el modelo DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, se puede notar que las mejores

métricas de rendimiento se obtienen en una proporción de datos de 1 a 6, esta es la máxima proporción que se presenta, puesto que al aumentar esta proporción existe un sobreentrenamiento y los modelos clasificación consideran como segmentos sin señal a todos los segmentos que evalúan. Esta proporción maximiza las medidas de rendimiento en los algoritmos de clasificación empleados.

Tabla 8

Métricas de rendimiento para balance de datos al remover los componentes DC 84

Características.

Proporción	A(%)	P(%)	R(%)	S(%)	BER
Algoritmos de detección con kNN					
1:1	91,84	70,58	97,77	90,45	0,059
1:2	93,16	75,00	97,77	92,04	0,050
1:3	93,75	77,27	97,84	92,71	0,047
1:4	93,85	77,71	97,84	92,84	0,046
1:5	93,30	75,95	97,88	92,14	0,049
1:6	94,00	78,16	97,84	93,00	0,042
Algoritmos de detección con DT					
1:1	86,90	57,91	95,86	84,97	0,096
1:2	90,47	69,65	93,96	89,60	0,082
1:3	89,17	65,87	93,29	88,15	0,093
1:4	89,85	68,65	92,00	89,34	0,093
1:5	90,00	70,51	92,17	89,40	0,092
1:6	93,62	82,51	92,07	94,10	0,069
Algoritmos de detección con SVM					
1:1	89,98	64,85	97,76	88,26	0,070
1:2	91,70	71,50	95,68	90,73	0,067
1:3	92,71	75,00	95,74	91,95	0,061
1:4	92,75	75,27	95,80	91,96	0,061
1:5	92,75	75,13	95,77	92,00	0,061
1:6	93,15	77,77	94,60	92,76	0,063

Al comparar los resultados, se usa la proporción de datos de 1 a 6 al remover los componentes DC, es decir $\mu = 0$ y $\sigma = 1$ para los procesos de selección de características y obtención de modelos finales. El proceso de estandarización y remoción de la media se lleva a cabo para mejorar los rangos de valores de características del mismo tipo con las que se obtienen los algoritmos de aprendizaje supervisado, donde se puede observar que las métricas de rendimiento son mejores al realizar este proceso.

- **257 Características PSD**

Como se observa en la Tabla 9, se obtienen las métricas de rendimiento con las características donde se aplica un modelo kNN con $k=30$, el modelo DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, se puede notar que las mejores métricas de rendimiento se obtienen en una proporción de datos de 1 a 6, esta es la máxima proporción que se presenta, puesto que al aumentar esta proporción existe un sobre entrenamiento y los modelos clasificación consideran como segmentos sin señal a todos los segmentos que evalúan.

Tabla 9

Métricas de rendimiento para balance de datos 257 Características PSD.

Proporción	A(%)	P(%)	R(%)	S(%)	BER
Algoritmos de detección con kNN					
1:1	77,29	35,01	96,52	74,55	0,145
1:2	95,12	85,02	96,60	94,67	0,044
1:3	96,95	91,50	96,55	97,09	0,032
1:4	98,47	97,36	96,73	99,08	0,021
1:5	98,64	98,03	96,77	99,31	0,020
1:6	98,98	99,34	97,79	99,75	0,017

Métricas de rendimiento para balance de datos 257 Características PSD.

Algoritmos de detección con DT					
1:1	69,10	17,97	92,00	67,30	0,203
1:2	71,43	22,23	95,83	69,18	0,170
1:3	79,72	41,98	94,30	77,16	0,143
1:4	76,08	35,48	89,15	74,00	0,184
1:5	84,28	55,10	91,01	82,72	0,132
1:6	69,10	17,97	92,00	67,30	0,204
Algoritmos de detección con SVM					
1:1	68,02	17,85	88,48	66,32	0,220
1:2	76,31	35,77	89,94	74,12	0,180
1:3	79,06	42,97	90,50	76,85	0,163
1:4	96,05	93,66	89,86	98,04	0,061
1:5	95,93	93,87	89,61	98,05	0,061
1:6	95,72	93,54	89,50	97,86	0,063

En la Tabla 10, se obtienen las medidas de rendimiento con las características al remover los componentes DC y obtener el un modelo kNN con $k=30$, el modelo DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, se puede notar que las mejores métricas de rendimiento se obtienen en una proporción de datos de 1 a 6. Esta es la máxima proporción que se presenta, puesto que al aumentar esta proporción existe un sobreentrenamiento y los modelos clasificación consideran como segmentos sin señal a todos los segmentos que evalúan.

Tabla 10

Métricas de rendimiento para balance de datos al remover componentes DC 257 Características

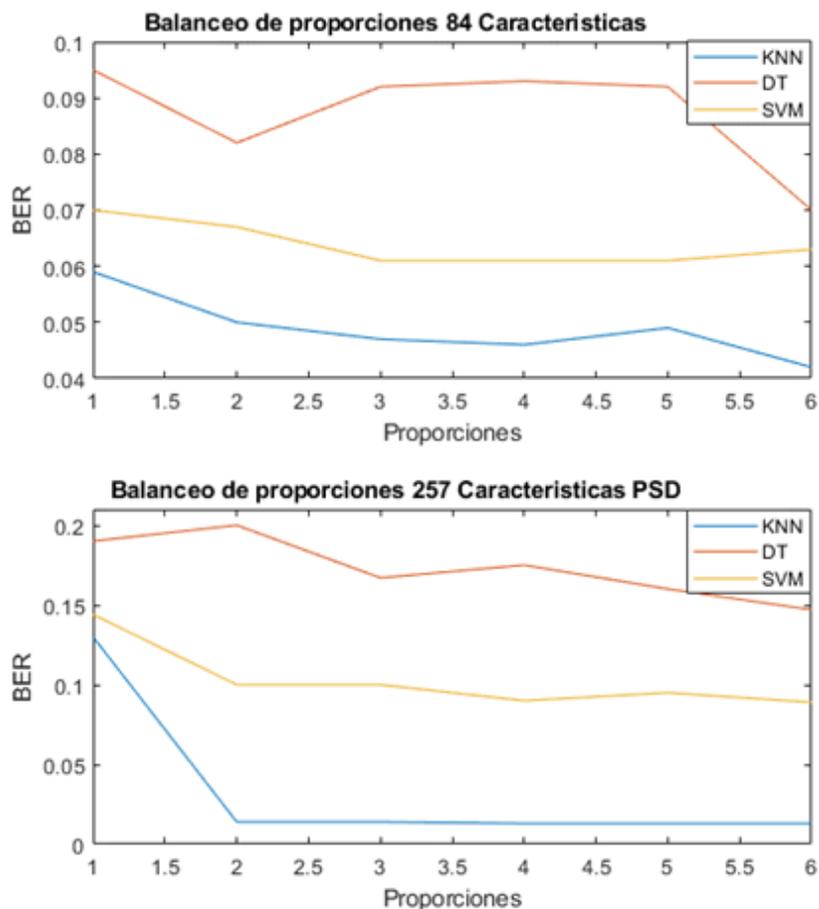
PSD.

Proporción	A(%)	P(%)	R(%)	S(%)	BER
Algoritmos de detección con kNN					
1:1	79,92	44,85	97,57	76,45	0,130
1:2	99,00	98,81	97,66	99,54	0,014
1:3	98,99	98,81	97,66	99,53	0,014
1:4	99,15	99,40	97,66	99,76	0,013
1:5	99,15	99,40	97,66	99,76	0,013
1:6	99,15	99,40	97,66	99,76	0,013
Algoritmos de detección con DT					
1:1	69,93	22,74	94,63	67,43	0,190
1:2	72,00	28,01	90,85	69,53	0,200
1:3	78,77	41,87	89,95	76,68	0,167
1:4	76,59	36,72	90,64	74,28	0,175
1:5	79,04	41,68	91,33	76,81	0,160
1:6	84,05	56,98	87,36	83,21	0,147
Algoritmos de detección con SVM					
1:1	85,11	61,68	86,56	84,71	0,144
1:2	91,51	82,91	85,93	93,55	0,100
1:3	91,64	83,33	85,93	93,73	0,100
1:4	93,10	88,77	86,01	95,82	0,090
1:5	92,71	87,76	85,50	95,46	0,095
1:6	93,23	89,24	86,01	96,01	0,089

Se puede notar tanto en la Tabla 8 y Tabla 10, la importancia de balancear y remover los componentes DC de los datos para obtener los modelos de entrenamiento, donde se observa que las métricas de rendimiento aumentan al realizar este proceso.

Figura 34

Evolución del BER al balancear y retirar los componentes DC a las proporciones de segmentos con señal y sin señal.



Selección de características

Se presentan las métricas de rendimiento de las pruebas realizadas a 14000 segmentos de las 175 señales no usadas para el entrenamiento. Los métodos usados para la selección de características son: filtro, embebido y envoltura. Este proceso se realizó con las 84 características y las 257 características PSD.

- Filtro

Este método permite dar un orden a las características que tiene una mejor correlación.

Este proceso se lo realiza tanto con las 84 características y 257 características PSD.

Para las 84 características se obtienen el siguiente orden: 'f33', 'f31', 'f74', 'f67', 'f39', 'f68', 'f75', 'f9', 'f2', 'f8', 'f4', 'f40', 'f55', 'f84', 'f3', 'f20', 'f11', 'f60', 'f29', 'f13', 'f15', 'f73', 'f66', 'f52', 'f38', 'f12', 'f57', 'f53', 'f61', 'f30', 'f51', 'f5', 'f58', 'f6', 'f69', 'f76', 'f48', 'f77', 'f54', 'f41', 'f70', 'f80', 'f28', 'f44', 'f34', 'f81', 'f32', 'f23', 'f83', 'f72', 'f10', 'f14', 'f50', 'f82', 'f24', 'f7', 'f63', 'f37', 'f65', 'f79', 'f35', 'f46', 'f47', 'f42', 'f43', 'f71', 'f59', 'f25', 'f78', 'f45', 'f26', 'f64', 'f56', 'f62', 'f36', 'f1', 'f16', 'f17', 'f18', 'f19', 'f21', 'f22', 'f27', 'f49'; con el orden ya establecido se aplica un modelo kNN con $k=20$, el modelo DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, donde se aumentan el número de características. Para determinar cuáles son las mejores características se evalúan las métricas de rendimiento del sistema de votación. Este proceso se puede observar en la Figura 35, donde se determinan que las mejores características son: 'f33', 'f31', 'f74', 'f67', 'f39', 'f68', 'f75', 'f9', 'f2', 'f8', 'f4', 'f40', 'f55', 'f84', 'f3', 'f20', 'f11', 'f60', 'f29', 'f13', 'f15', 'f73', 'f66', 'f52', 'f38', 'f12', 'f57', 'f53', 'f61', 'f30'; las cuales corresponden a las 30 primeras características obtenidas por el método de filtro de las cuales 8 son en tiempo, 6 en frecuencia y 16 en escala. En la Figura 36 se observa como disminuye el BER a medida que se aumenta las características.

Figura 35

Métricas de rendimiento para la selección de características usando método de filtro.

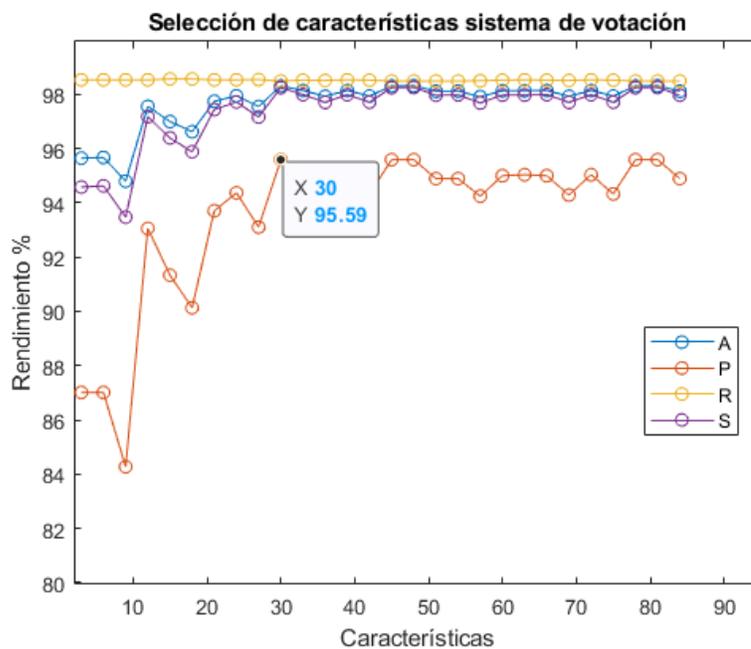
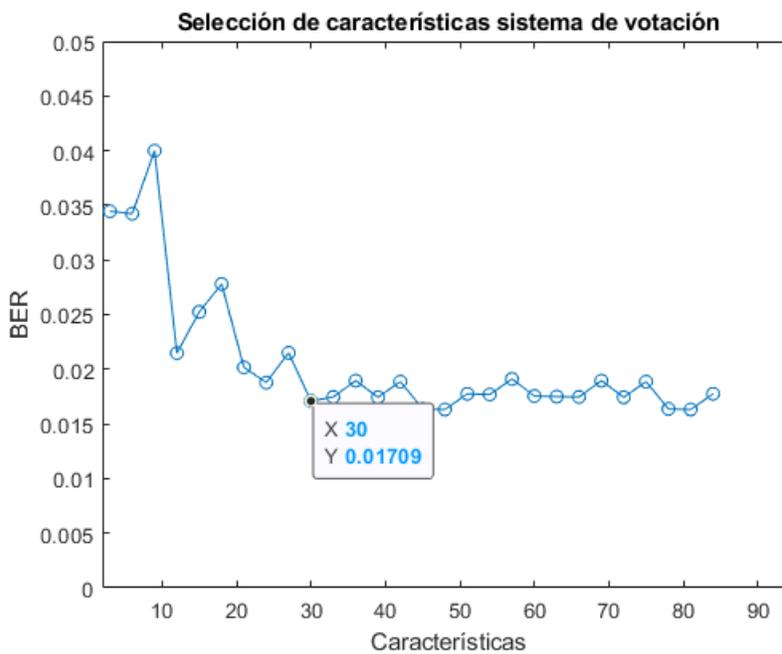


Figura 36

Análisis del BER para la selección de características usando método de filtro.



Como se observa en la Tabla 11, se obtienen las métricas de rendimiento con las 30 primeras características al remover los componentes DC, donde se obtienen los algoritmos de clasificación con las mismas características en cada modelo y un sistema de votación entre los tres métodos mencionados.

Tabla 11

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 30

Características.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Métricas de Rendimiento 30 Características					
kNN	97,74	93,70	98,53	97,44	0,020
DT	96,22	89,03	98,57	95,32	0,031
SVM	98,24	95,23	98,56	97,98	0,017
Votación	97,94	94,44	98,55	97,71	0,018

Para las 257 características PSD se obtiene el siguiente orden: 'f18', 'f19', 'f22', 'f23', 'f20', 'f24', 'f25', 'f16', 'f17', 'f21', 'f15', 'f26', 'f29', 'f27', 'f30', 'f32', 'f28', 'f31', 'f14', 'f13', 'f12', 'f33', 'f36', 'f34', 'f35', 'f37', 'f39', 'f38', 'f11', 'f40', 'f41', 'f42', 'f10', 'f46', 'f47', 'f44', 'f48', 'f45', 'f49', 'f43', 'f9', 'f50', 'f51', 'f53', 'f8', 'f52', 'f55', 'f54', 'f56', 'f57', 'f7', 'f58', 'f59', 'f60', 'f74', 'f61', 'f73', 'f75', 'f62', 'f63', 'f68', 'f77', ..., 'f252'; con el orden ya establecido se aplica un modelo kNN con $k=30$, el modelo DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, donde se aumentan el número de características. Para determinar cuáles son las mejores características se evalúan las métricas de rendimiento del sistema de votación. Este proceso se observa en la Figura 37, donde se determinan que las mejores características son: 'f18', 'f19', 'f22', 'f23', 'f20', 'f24', 'f25', 'f16', 'f17', 'f21', 'f15', 'f26', 'f29', 'f27', 'f30', 'f32', 'f28', 'f31', 'f14', 'f13', 'f12', 'f33', 'f36', 'f34', 'f35', 'f37', 'f39', 'f38', 'f11', 'f40', 'f41', 'f42', 'f10',

'f46', 'f47', 'f44', 'f48', 'f45', 'f49', 'f43', 'f9', 'f50', 'f51', 'f53', 'f8', 'f52', 'f55', 'f54', 'f56', 'f57', 'f7', 'f58', 'f59', 'f60', 'f74', 'f61', 'f73', 'f75', 'f62', 'f63', 'f68', 'f77'; las cuales corresponden a las 100 primeras características obtenidas por el método de filtro. En la Figura 38 se observa como disminuye el BER a medida que se aumenta las características.

Figura 37

(a) Métricas de rendimiento para la selección de características PSD usando método de filtro, (b) Frecuencias de interés selección de características método de filtro.

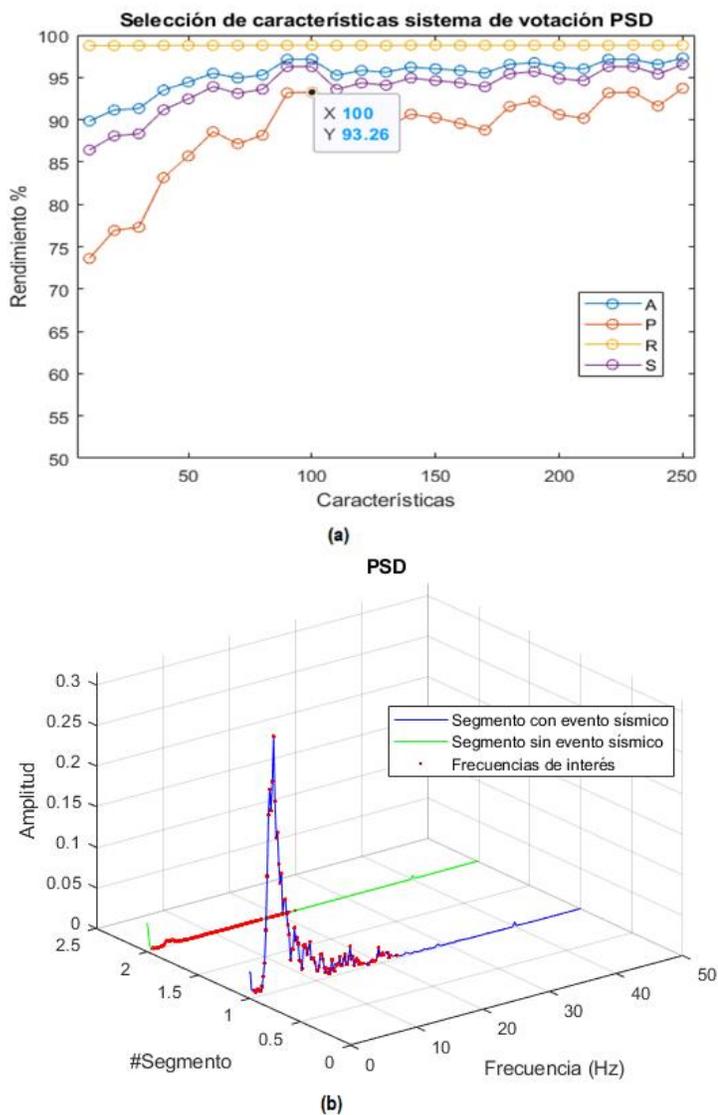
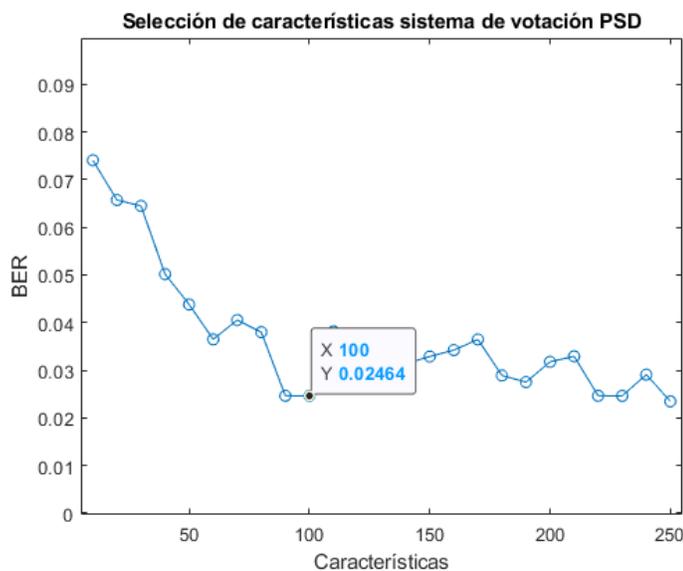


Figura 38

Análisis del BER para la selección de características PSD usando método de filtro.



Como se observa en la Tabla 12, se obtienen las métricas de rendimiento con las 100 primeras características con el método de filtro de las 257 características PSD al remover los componentes DC, donde se obtienen los algoritmos de clasificación, y un sistema de votación entre los tres métodos mencionados.

Tabla 12

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 100

Características PSD.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Métricas de Rendimiento 100 Características PSD					
kNN	98,27	96,49	98,80	98,00	0,016
DT	79,51	49,52	98,72	74,68	0,133
SVM	94,25	85,18	98,78	95,30	0,030
Votación	95,45	88,52	98,78	93,85	0,033

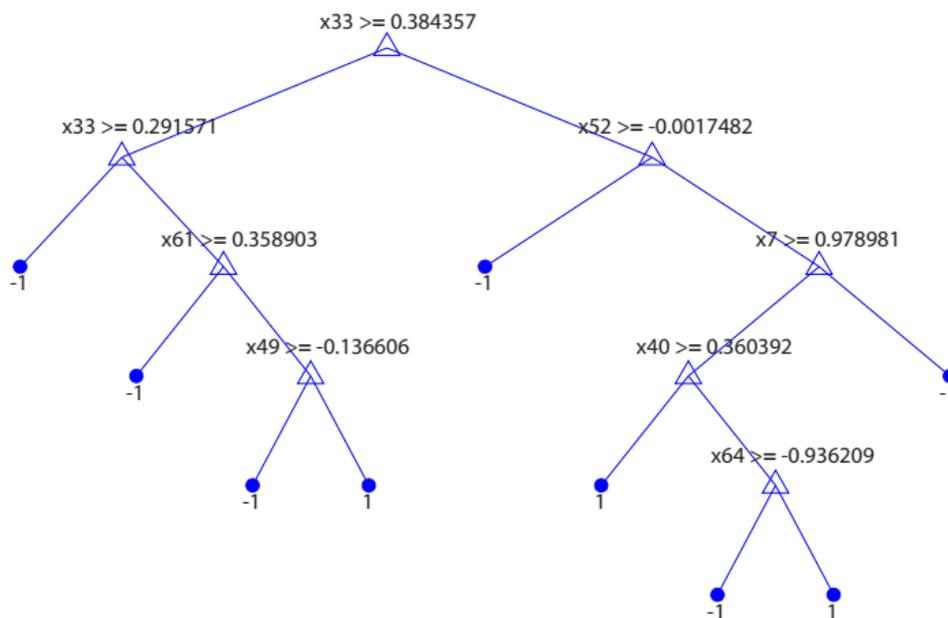
- **Embebido**

Este método se aplica para obtener los niveles del árbol de decisión, donde se obtienen las características con mayor relevancia. Este proceso se la realiza tanto con las 84 características y 257 características PSD.

Para las 84 características el software MATLAB® al generar el algoritmo de aprendizaje supervisado DT selecciona las características con mayor relevancia y genera un árbol de decisión como se observa en la Figura 39. Este método indica que las mejores características son: 'f33', 'f52', 'f61', 'f49', 'f40', 'f12', 'f7', de las cuales 2 son en tiempo, 1 en frecuencia y 4 en escala.

Figura 39

Representación de árbol de decisión considerando 84 características.



Como se observa en la Tabla 13, se obtienen las métricas de rendimiento con las 7 primeras características obtenidas con los niveles de poda del método embebido de las 84 características al remover los componentes DC, se obtiene un modelo kNN con k=20, el modelo

DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, y un sistema de votación entre los tres métodos mencionados.

Tabla 13

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 7

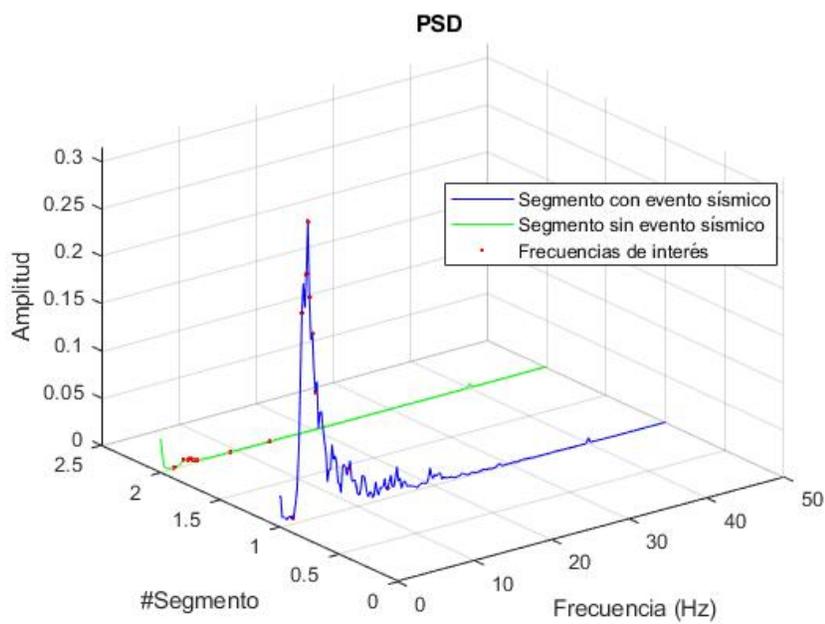
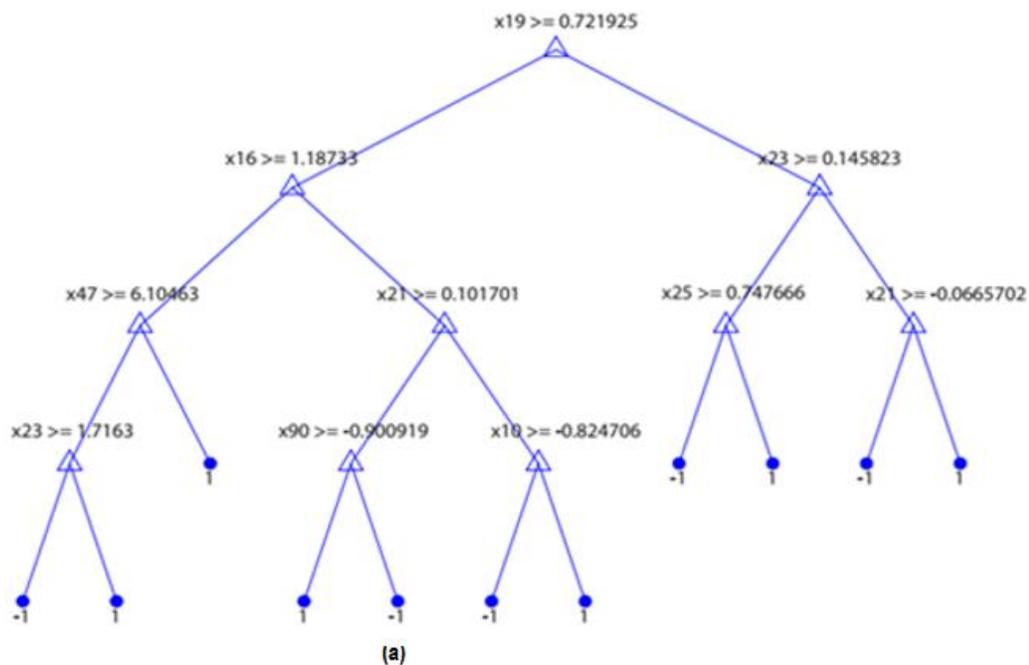
Características.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Métricas de Rendimiento 7 Características					
kNN	97,79	94,23	98,66	97,42	0,019
DT	95,00	85,45	98,60	93,63	0,038
SVM	97,38	92,76	98,60	96,90	0,022
Votación	98,17	95,33	98,62	98,00	0,017

Para las 257 características PSD, se genera el modelo de árbol de decisión como se observa en la Figura 40. Este método indica que las mejores características son: 'f19', 'f16', 'f23', 'f47', 'f20', 'f25', 'f21', 'f90', 'f73', 'f10'. Las frecuencias de interés que representa cada característica son: 'f10' (corresponde a la frecuencia de 1,95 Hz), 'f16' (corresponde a la frecuencia de 3,12 Hz), 'f19' (corresponde a la frecuencia de 3,70 Hz), 'f20' (corresponde a la frecuencia de 3,90 Hz), 'f21' (corresponde a la frecuencia de 4,10 Hz), 'f23' (corresponde a la frecuencia de 4,50 Hz), 'f47' (corresponde a la frecuencia de 9,15 Hz), 'f73' (corresponde a la frecuencia de 14,20 Hz) y 'f90' (corresponde a la frecuencia de 17,50 Hz).

Figura 40

(a) Representación de árbol de decisión considerando 257 características PSD, (b) Frecuencias de interés selección de características método embebido.



(b)

Como se observa en la Tabla 12, se obtienen las métricas de rendimiento con las 10 primeras características obtenidas con los niveles de poda del método embebido de las 257 características PSD al remover los componentes DC, se obtiene un modelo kNN con $k=30$, el modelo DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, y un sistema de votación entre los tres métodos mencionados.

Tabla 14

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 10 Características PSD.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Métricas de Rendimiento 10 Características PSD					
kNN	93,88	84,28	98,77	91,66	0,048
DT	80,59	52,33	98,74	75,68	0,128
SVM	97,38	92,76	98,60	96,90	0,022
Votación	88,60	70,67	98,75	84,86	0,082

- **Envoltura**

Al obtenerlos modelos de SVM con las 84 características y 257 características PSD, se pueden obtener las mejores características respectivamente. Cada modelo que se ha conseguido por medio del software MATLAB[®] proporciona un α_k que permite dar orden a las características con mayor relevancia como se observa en la ecuación (18).

Para las 84 características se obtiene el siguiente orden: 'f57', 'f15', 'f40', 'f39', 'f60', 'f33', 'f31', 'f30', 'f12', 'f52', 'f68', 'f9', 'f8', 'f38', 'f84', 'f3', 'f11', 'f20', 'f51', 'f61', 'f29', 'f13', 'f41', 'f54', 'f6', 'f34', 'f69', 'f76', 'f44', 'f55', 'f58', 'f2', 'f5', 'f32', 'f28', 'f80', 'f75', 'f81', 'f53', 'f83', 'f82', 'f73', 'f63', 'f47', 'f4', 'f42', 'f72', 'f24', 'f66', 'f25', 'f77', 'f46', 'f79', 'f14', 'f37', 'f10', 'f59', 'f43', 'f65', 'f7', 'f35', 'f78', 'f56', 'f74', 'f50', 'f26', 'f45', 'f70', 'f23', 'f48', 'f64', 'f67', 'f1',

'f62', 'f71', 'f36', 'f16', 'f17', 'f18', 'f19', 'f21', 'f22', 'f27', 'f49'; con el orden ya establecido se aplican algoritmos de clasificación indicados, donde se aumentan el número de características. Para determinar cuáles son las mejores características se evalúan las métricas de rendimiento del sistema de votación. Este proceso se puede observar en la Figura 41, donde se determina que las mejores características son: 'f57', 'f15', 'f40', 'f39', 'f60', 'f33', 'f31', 'f30', 'f12', 'f52', 'f68', 'f9', 'f8', 'f38', 'f84', 'f3', 'f11', 'f20', 'f51', 'f61', 'f29'; las cuales corresponden a las 21 primeras características obtenidas por el método de envoltura de las cuales 5 son en tiempo, 5 en frecuencia y 11 en escala. En la Figura 42 se observa como disminuye el BER a medida que se aumentan las características.

Figura 41

Métricas de rendimiento para la selección de características PSD usando método de envoltura.

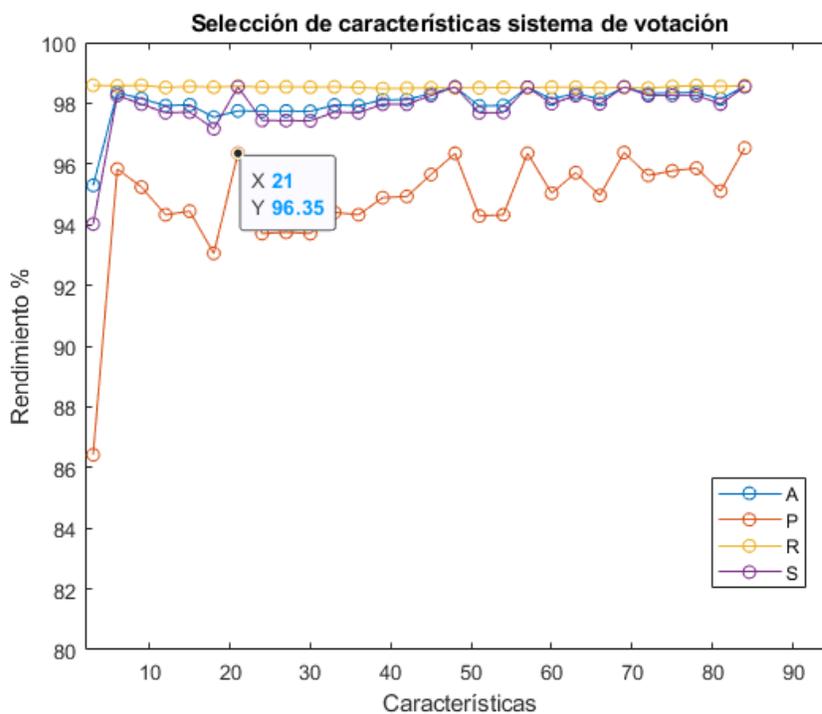
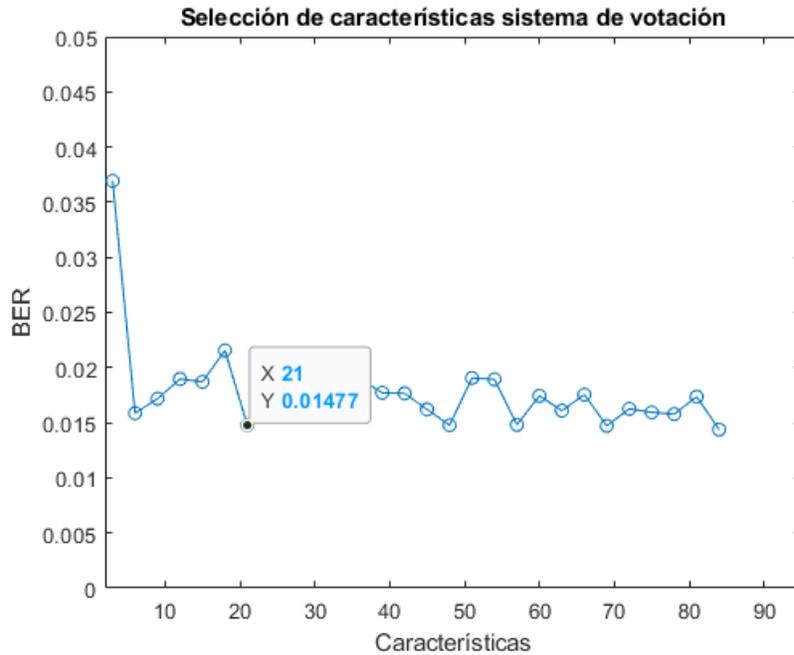


Figura 42

Análisis del BER para la selección de características PSD usando método de envoltura.



Como se observa en la Tabla 15, se obtienen las métricas de rendimiento con las 21 primeras características del orden obtenido con el método de envoltura de las 84 características al remover los componentes DC, se obtiene un modelo kNN con $k=20$, DT con una validación cruzada de 10 y SVM con un $\nu = 0.5$, y un sistema de votación entre los tres métodos mencionados.

Tabla 15

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 21

Características.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Métricas de Rendimiento 21 Características					
kNN	98,15	95,23	98,59	97,97	0,017
DT	98,10	95,17	98,57	97,91	0,018
SVM	97,75	93,79	98,55	97,43	0,020
Votación	98,15	95,20	98,58	97,98	0,017

Para las 257 características PSD se obtiene el siguiente orden 'f22', 'f15', 'f25', 'f19', 'f32', 'f29', 'f28', 'f30', 'f33', 'f13', 'f12', 'f17', 'f35', 'f34', 'f11', 'f38', 'f41', 'f42', 'f27', 'f44', 'f16', 'f9', 'f47', 'f49', 'f10', 'f8', 'f51', 'f26', 'f48', 'f14', 'f54', 'f53', 'f57', 'f58', 'f56', 'f74', 'f59', 'f37', 'f73', 'f61', 'f75', 'f77', 'f43', 'f62', 'f39', 'f76', 'f50', 'f68', 'f18', 'f70', 'f6', 'f64', 'f82', 'f65', 'f69', 'f71', 'f60', 'f36', 'f24', 'f46', 'f31', 'f109', 'f4', 'f81', 'f110', 'f96', 'f55', 'f23', 'f102', 'f67', 'f78', 'f66', 'f7', 'f84', 'f115', 'f108', 'f52', 'f40', 'f5', 'f103', 'f88', 'f101', 'f93', 'f142', 'f135', 'f97', 'f85', 'f119', 'f120', 'f20', 'f128', 'f147', 'f90', 'f181', 'f130', 'f251', 'f99', 'f72', 'f2', 'f116', 'f79', 'f114', 'f80', 'f107', 'f134', 'f121', 'f136', 'f92', 'f232', 'f86', 'f122', 'f91', 'f105', 'f151', 'f118', 'f94', 'f98', 'f45', 'f111', 'f237', 'f139', 'f183', 'f159', ..., 'f164'; con el orden ya establecido se aplican los algoritmos de clasificación indicados, donde se aumentan el número de características. Para determinar cuáles son las mejores características se evalúa las métricas de rendimiento del sistema de votación. Este proceso se observa en la Figura 43 donde se determinan que las mejores características son: 'f22', 'f15', 'f25', 'f19', 'f32', 'f29', 'f28', 'f30', 'f33', 'f13', 'f12', 'f17', 'f35', 'f34', 'f11', 'f38', 'f41', 'f42', 'f27', 'f44', 'f16', 'f9', 'f47', 'f49', 'f10', 'f8', 'f51', 'f26', 'f48', 'f14', 'f54', 'f53', 'f57', 'f58', 'f56', 'f74', 'f59', 'f37', 'f73', 'f61', 'f75', 'f77', 'f43', 'f62', 'f39', 'f76',

'f50', 'f68', 'f18', 'f70', 'f6', 'f64', 'f82', 'f65', 'f69', 'f71', 'f60', 'f36', 'f24', 'f46', 'f31', 'f109', 'f4', 'f81', 'f110', 'f96', 'f55', 'f23', 'f102', 'f67', 'f78', 'f66', 'f7', 'f84', 'f115', 'f108', 'f52', 'f40', 'f5', 'f103', 'f88', 'f101', 'f93', 'f142', 'f135', 'f97', 'f85', 'f119', 'f120', 'f20', 'f128', 'f147', 'f90', 'f181', 'f130', 'f251', 'f99', 'f72', 'f2', 'f116', 'f79', 'f114', 'f80', 'f107', 'f134', 'f121', 'f136', 'f92', 'f232', 'f86', 'f122', 'f91', 'f105', 'f151', 'f118', 'f94', 'f98', 'f45', 'f111', 'f237', 'f139', 'f183', 'f159'; las cuales corresponden a las 100 primeras características obtenidas por el método de envoltura. En la Figura 44 se observa como disminuye el BER a medida que se aumentan las características.

Figura 43

(a) Métricas de rendimiento para la selección de características PSD usando método de envoltura, (b) Frecuencias de interés selección de características método de envoltura.

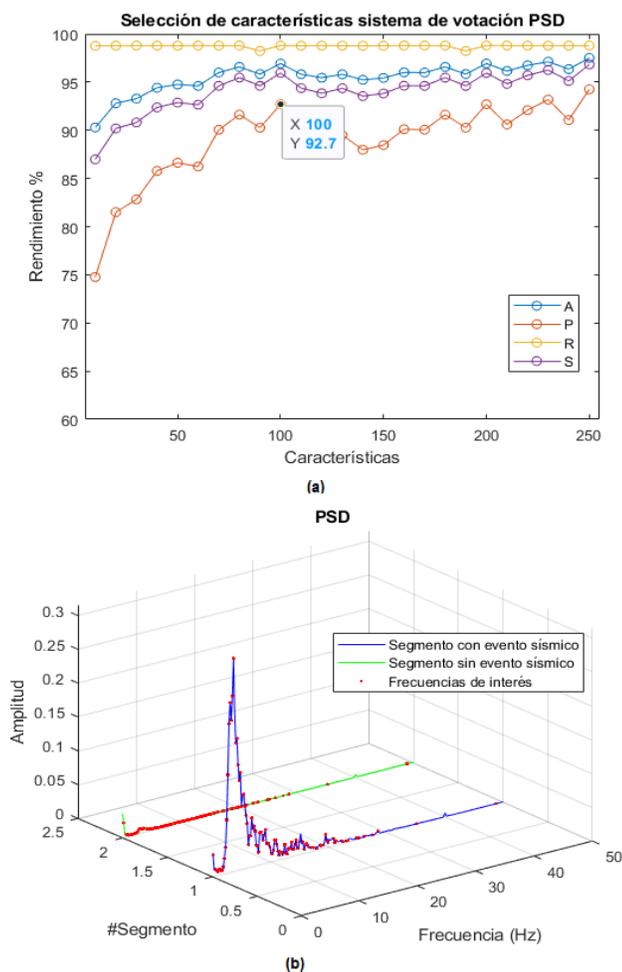
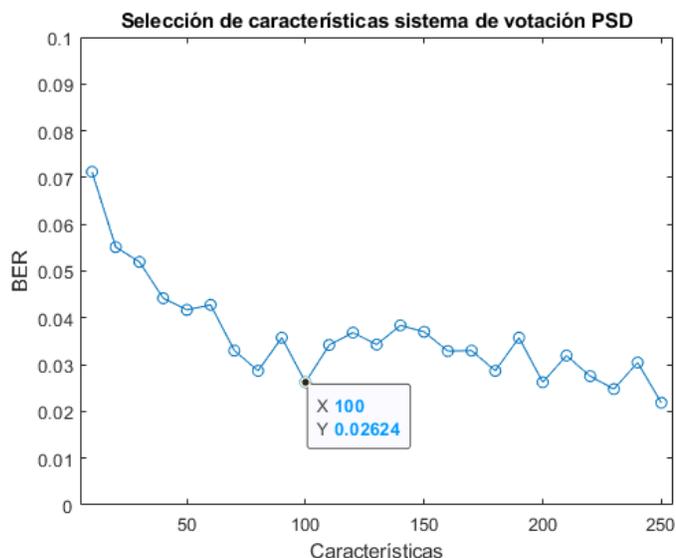


Figura 44

Análisis del BER para la selección de características PSD usando método de envoltura.



Como se observa en la Tabla 16, al obtener las medidas de rendimiento con las 100 primeras características del orden obtenido con el método de envoltura de las 257 características PSD al remover los componentes DC, se obtiene un modelo kNN con $k=30$, el modelo DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, y un sistema de votación entre los tres métodos mencionados.

Tabla 16

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 100 Características PSD.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Métricas de Rendimiento 100 Características PSD					
kNN	97,68	94,82	98,80	97,07	0,020
DT	80,48	52,17	98,73	75,55	0,128
SVM	93,90	84,21	98,77	91,73	0,047
Votación	95,08	87,50	98,78	93,35	0,039

- **Mejores características**

Se realiza una comparación de las características obtenidas en cada método, donde se toman las características que se repiten para determinar cuales se usaran en los modelos finales del sistema de detección de microterremotos.

Para las 84 características se determina que las mejores son: 'f11', 'f33', 'f31', 'f52', 'f40', 'f61', 'f12', 'f68', 'f9', 'f8', 'f60', 'f57', 'f39'. Se obtienen un total de 13 características de las cuales 4 son en tiempo, con 2 en frecuencia y 7 en escalas, con las cuales se obtiene un modelo kNN con $k=20$, el modelo DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, y se calcula las métricas de rendimiento como se observa en la Tabla 17. Estos son los modelos finales para la detección de microterremotos.

Tabla 17

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con las 13 mejores Características.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Métricas de Rendimiento 13 Características					
kNN	98,15	95,20	98,59	98,00	0,017
DT	94,11	82,63	98,57	92,50	0,044
SVM	97,95	94,56	98,58	97,71	0,018
Votación	98,15	95,23	98,59	98,00	0,017

Para las 257 características PSD se determinan que las mejores son: 'f22', 'f15', 'f25', 'f19', 'f32', 'f29', 'f28', 'f30', 'f33', 'f13', 'f12', 'f17', 'f35', 'f34', 'f11', 'f38', 'f41', 'f42', 'f27', 'f44', 'f16', 'f9', 'f47', 'f49', 'f10', 'f8', 'f51', 'f26', 'f48', 'f14', 'f54', 'f53', 'f57', 'f58', 'f56', 'f74', 'f59', 'f37', 'f73', 'f61', 'f75', 'f77', 'f43', 'f62', 'f39', 'f76', 'f50', 'f68', 'f18', 'f70', 'f6', 'f64', 'f82', 'f65', 'f69', 'f71', 'f60', 'f36', 'f24', 'f46', 'f31', 'f109', 'f4', 'f81', 'f110', 'f96', 'f55', 'f23', 'f102', 'f67',

'f78', 'f66', 'f7', 'f84', 'f115', 'f108', 'f52', 'f40', 'f5', 'f103', 'f88', 'f101', 'f93', 'f142', 'f135', 'f97', 'f85', 'f119', 'f120', 'f20', 'f128', 'f147', 'f90', 'f181', 'f130', 'f251', 'f99', 'f72', 'f2', 'f116', 'f79', 'f114', 'f80', 'f107', 'f134', 'f121', 'f136', 'f92', 'f232', 'f86', 'f122', 'f91', 'f105', 'f151', 'f118', 'f94', 'f98', 'f45', 'f111', 'f237', 'f139', 'f183', 'f159'. Se obtienen un total de 100 características obtenidas con el método embebido como se observa en la Figura 43, con las cuales obtiene un modelo kNN con $k=30$, el modelo DT con una validación cruzada de 10 y el modelo SVM con un $\nu = 0.5$, y se calculan las métricas de rendimiento como se observa en la Tabla 18. Estos son los modelos finales para la detección de microterremotos.

Tabla 18

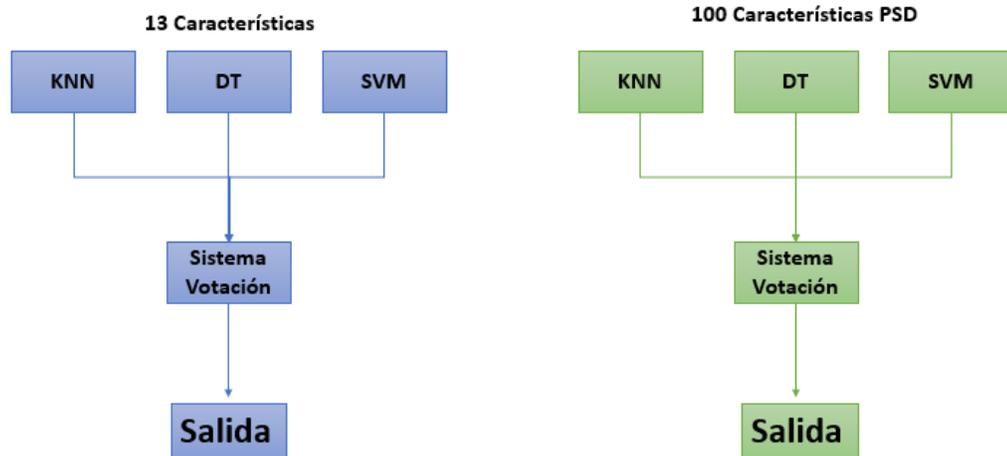
Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 100 mejores Características PSD.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Métricas de Rendimiento 100 Características PSD					
kNN	98,27	96,49	98,80	98,00	0,016
DT	79,51	49,52	98,72	74,68	0,133
SVM	94,25	85,18	98,78	95,30	0,030
Votación	95,45	88,52	98,78	93,85	0,033

Se unifica los modelos obtenidos con los dos distintos tipos de características para obtener un sistema de votación. Primero, se obtiene un sistema de votación con 5 modelos; posteriormente un sistema con 3 modelos que tienen las mejores métricas de desempeño; y finalmente se elige un sistema de votación que será unificados en la interfaz gráfica para el sistema de reconocimiento. Se obtienen 6 modelos para la detección de microterremotos como se observa en la Figura 45.

Figura 45

Modelos finales sistema de detección de microterremotos



- **Modelos de detección unificados**

Se seleccionan 5 y 3 modelos de detección de microterremotos obtenidos entre las 84 características y 257 características PSD. Se implementa un sistema de votación como se observa en la Figura 46 y Figura 47. Finalmente se obtienen las métricas de rendimiento del sistema de votación.

Figura 46

Sistema de votación con 5 modelos sistema de detección de microterremotos

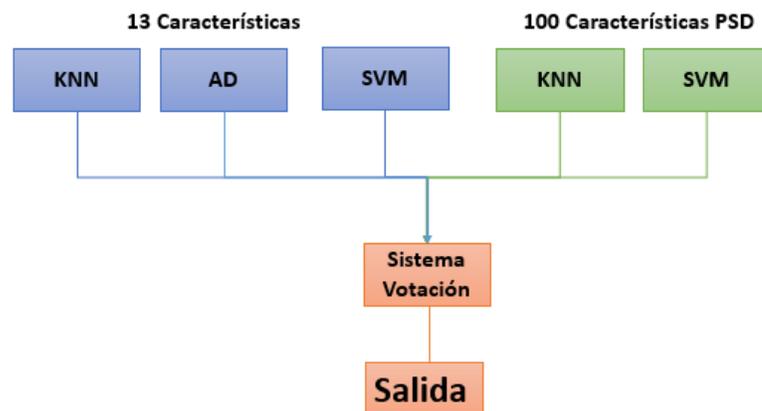
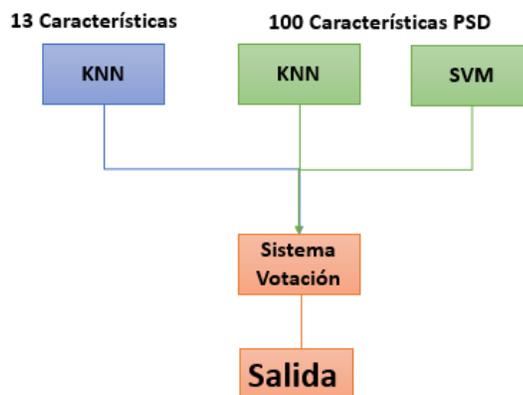


Figura 47

Sistema de votación con 3 modelos sistema de detección de microterremotos



Como se observa en la Figura 46, se seleccionan los modelos con las mejores métricas de rendimiento unificando los 5 modelos para obtener un sistema de votación entre todos. Posteriormente como se observa en la Figura 47, se realiza otra selección de 3 modelos para obtener las métricas de rendimiento. Finalmente se selecciona el sistema de votación unificado con los mejores resultados.

Tabla 19

Métricas de rendimiento sistema de votación con 5 y 3 modelos sistema de detección de microterremotos.

A(%)	P(%)	R(%)	S(%)	BER
Métricas de rendimiento sistema de votación 5 modelos				
98,55	96,62	98,62	98,42	0,014
Métricas de rendimiento sistema de votación 3 modelos				
98,74	97,60	98,78	98,71	0,012

En la Tabla 19, se puede observar que el sistema de votación unificado con 3 modelos alcanza una exactitud del 98,74 % y un BER de 0,012. Este sistema de votación será unificado en la interfaz gráfica.

Clasificación

Para el entrenamiento y obtención de modelos se realizó primero la obtención de las métricas de rendimiento con todas las características, posteriormente se seleccionan las características y se obtienen los modelos finales para la clasificación de microterremotos. A continuación, se detallan los resultados obtenidos en cada proceso realizado.

Modelos de clasificación con todas las características

Se obtiene las medidas de rendimiento con las 84 características y 257 características PSD. Primero se realiza la clasificación de 82 eventos y finalmente con 563 eventos entre LP y VT.

- **84 características**

Como se observa en la Figura 48 y Figura 49, se muestra la matriz de confusión de la clasificación con un modelo kNN con $k=35$, el modelo DT por defecto de MATLAB® con nivel máximo de podamiento de 10 y el modelo SVM con un $\nu = 0.5$, en los distintos escenarios. Las predicciones positivas (+1) son los eventos VT clasificados correctamente y las negativas (-1) son los eventos LP clasificados correctamente.

Figura 48

Matriz de confusión clasificación 82 microterremotos con 84 características.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	1	40	41 LP

Figura 49

Matriz de confusión clasificación 563 microterremotos con 84 características.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	29	493	522 LP

Como se observa en la Tabla 20, se presentan las métricas de rendimiento de las dos pruebas del sistema de clasificación con todas las características. En la clasificación de 82 microterremotos se obtiene una exactitud del 89,00 % y una precisión del 97,02 %, en el caso de la clasificación de 563 eventos la exactitud aumenta al 93,43 % y una precisión disminuye al 52,22 %. La disminución en la precisión se debe a que los eventos LP y VT no están en equilibrio existiendo más eventos del tipo LP que se clasifica como se observa en la Figura 49.

Tabla 20

Métricas de rendimiento sistema de votación para la clasificación de 82 y 563 microterremotos con 84 características.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Métricas de Rendimiento clasificación de 82 eventos					
Votación	89,00	97,01	80,50	97,60	0,110
Métricas de Rendimiento clasificación de 563 eventos					
Votación	93,42	52,22	80,48	94,40	0,120

- **257 características PSD**

Como se observa en la Figura 50 y Figura 51, se muestra la matriz de confusión de la clasificación de eventos con un modelo kNN con $k=30$, el modelo DT por defecto de MATLAB® con nivel máximo de podamiento de 10 y el modelo SVM con un $\nu = 0.5$, en los distintos escenarios. Las predicciones positivas (+1) son los eventos VT clasificados correctamente y las negativas (-1) son los eventos LP clasificados correctamente.

Figura 50

Matriz de confusión clasificación 82 microterremotos con 257 características PSD.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	28	13	41 VT
	Negativos	1	40	41 LP

Figura 51

Matriz de confusión clasificación 563 microterremotos con 257 características PSD.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	28	13	41 VT
	Negativos	25	497	522 LP

Como se observa en la Tabla 21, se presentan las medidas de rendimiento de las dos pruebas del sistema de clasificación con todas las 257 características PSD. En la clasificación de 82 microterremotos se obtiene una exactitud del 82,92 % y una precisión del 96,55 %, en el caso de la clasificación de 563 eventos la exactitud aumenta al 93,25 % y una precisión disminuye al 52,83 %.

Tabla 21

Métricas de rendimiento sistema de votación para la clasificación de 82 y 563 microterremotos con 257 características PSD.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Métricas de Rendimiento clasificación de 82 eventos					
Votación	82,92	96,55	68,30	97,56	0,170
Métricas de Rendimiento clasificación de 563 eventos					
Votación	93,25	52,83	68,30	95,21	0,180

Selección de características

Se presentan las métricas de rendimiento de la selección de características con la clasificación de 82 y 563 microterremotos. Los métodos usados para la selección de

características son: filtro, embebido y envoltura. Este proceso se realizó con las 84 características y las 257 características PSD.

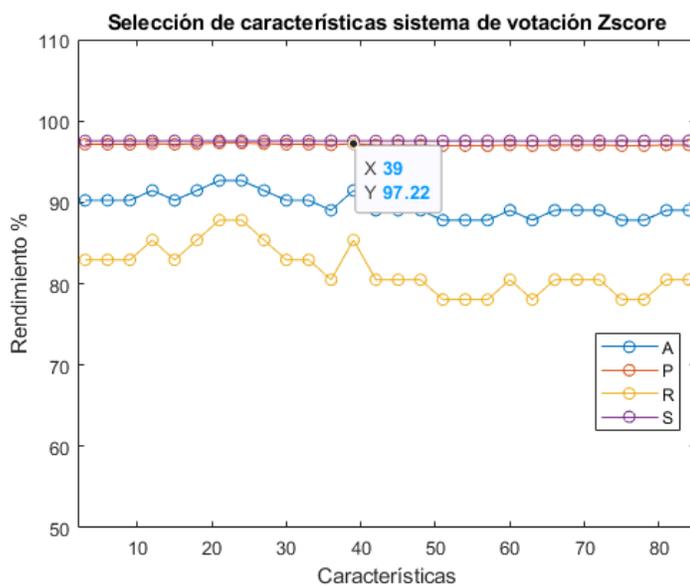
- **Filtro**

Este método permite dar un orden a las características que tienen mejor correlación de cada segmento. Este proceso se lo realiza tanto con las 84 características y 257 características PSD.

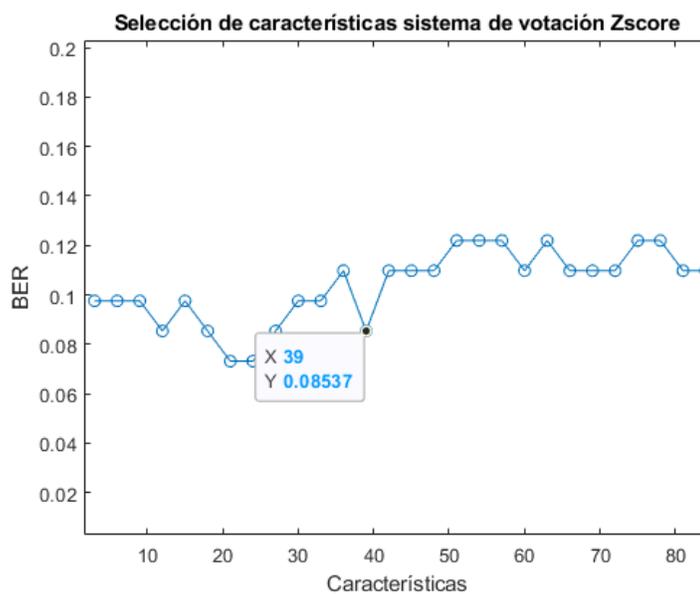
Para las 84 características se obtiene el siguiente orden: 'f72', 'f30', 'f23', 'f37', 'f65', 'f58', 'f60', 'f61', 'f75', 'f68', 'f25', 'f79', 'f40', 'f28', 'f15', 'f7', 'f31', 'f59', 'f66', 'f74', 'f39', 'f33', 'f14', 'f71', 'f50', 'f32', 'f41', 'f73', 'f34', 'f82', 'f78', 'f44', 'f4', 'f67', 'f69', 'f26', 'f38', 'f76', 'f22', 'f12', 'f45', 'f36', 'f51', 'f9', 'f24', 'f62', 'f13', 'f21', 'f17', 'f81', 'f83', 'f64', 'f27', 'f57', 'f16', 'f55', 'f11', 'f20', 'f70', 'f42', 'f49', 'f35', 'f63', 'f6', 'f46', 'f47', 'f5', 'f52', 'f29', 'f10', 'f3', 'f48', 'f19', 'f84', 'f56', 'f18', 'f43', 'f77', 'f2', 'f8', 'f80', 'f53', 'f1', 'f54'; con el orden ya establecido se aplica un modelo kNN con $k=35$, el modelo DT por defecto de MATLAB® con nivel máximo de podamiento de 10 y el modelo SVM con un $\nu = 0.5$, donde se aumenta el número de características. Para determinar cuáles son las mejores características se evalúan las métricas de rendimiento del sistema de votación. Este proceso se puede observar en la Figura 52, donde se determina que las mejores características son: 'f72', 'f30', 'f23', 'f37', 'f65', 'f58', 'f60', 'f61', 'f75', 'f68', 'f25', 'f79', 'f40', 'f28', 'f15', 'f7', 'f31', 'f59', 'f66', 'f74', 'f39', 'f33', 'f14', 'f71', 'f50', 'f32', 'f41', 'f73', 'f34', 'f81', 'f78', 'f44', 'f4', 'f67', 'f69', 'f26', 'f38'; las cuales corresponden a las 39 primeras características obtenidas por el método de filtro de las cuales 2 son en tiempo, 15 en frecuencia y 22 en escala. En la Figura 53 se observa como disminuye el BER a medida que se aumentan las características.

Figura 52

Métricas de rendimiento para la selección de características del sistema de clasificación usando el método de filtro

**Figura 53**

Análisis del BER para la selección de características del sistema de clasificación usando el método de correlación



Como se observa en la Tabla 22, se obtienen las métricas de rendimiento con las 39 primeras características del orden obtenido con el método de filtro de las 84 características al remover los componentes DC, se obtienen los algoritmos de clasificación mencionados y un sistema de votación entre los tres métodos. Se obtiene la matriz de confusión de cada sistema de votación como se observa en la Figura 54 y Figura 55.

Tabla 22

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 39 Características.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Clasificación 82 eventos con 39 características					
kNN	91,46	100,00	82,93	100,00	0,085
DT	82,93	90,90	73,17	92,68	0,170
SVM	84,14	96,67	70,73	97,67	0,158
Votación	90,24	100,00	80,49	100,00	0,097
Clasificación 563 eventos con 39 características					
kNN	90,48	41,98	82,93	91,00	0,130
DT	91,12	43,47	73,17	92,53	0,170
SVM	94,49	60,42	70,73	96,36	0,160
Votación	95,20	63,46	80,49	96,36	0,115

Figura 54

Matriz de confusión clasificación 82 microterremotos con 39 características.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	0	41	41 LP

Figura 55

Matriz de confusión clasificación 563 microterremotos con 39 características.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	19	503	522 LP

Para las 257 características PSD se obtiene el siguiente orden: ‘f17’, ‘f19’, ‘f16’, ‘f18’, ‘f20’, ‘f21’, ‘f15’, ‘f22’, ‘f14’, ‘f13’, ‘f12’, ‘f23’, ‘f24’, ‘f49’, ‘f45’, ‘f80’, ‘f30’, ‘f50’, ‘f44’, ‘f46’, ‘f86’, ‘f39’, ‘f38’, ‘f81’, ‘f79’, ‘f43’, ‘f11’, ‘f85’, ‘f76’, ‘f51’, ‘f72’, ‘f73’, ‘f41’, ‘f48’, ‘f75’, ‘f77’, ‘f74’, ‘f87’, ‘f42’, ‘f47’, ‘f84’, ‘f82’, ‘f68’, ‘f78’, ‘f35’, ‘f31’, ‘f27’, ‘f71’, ‘f36’, ‘f40’, ‘f56’, ‘f52’, ‘f26’, ‘f55’, ‘f83’, ‘f34’, ‘f10’, ‘f57’, ‘f54’, ‘f53’, ‘f28’, ‘f67’, ‘f91’, ‘f25’, ‘f32’, ‘f65’, ‘f88’, ‘f92’, ‘f9’, ‘f58’, ‘f37’, ‘f93’, ‘f70’, ‘f96’, ‘f29’, ‘f66’, ‘f33’, ‘f113’, ‘f95’, ‘f103’, ‘f62’, ‘f64’, ‘f94’, ‘f60’, ‘f69’, ‘f61’, ‘f59’, ‘f63’, ‘f108’, ‘f111’, ‘f102’, ‘f117’, ‘f90’, ‘f107’, ‘f116’, ‘f98’, ‘f104’, ‘f110’, ‘f89’, ‘f109’, ‘f106’, ‘f105’, ‘f97’, ‘f115’, ‘f8’, ‘f112’, ‘f99’, ‘f101’, ‘f114’, ‘f118’, ‘f100’, ‘f119’, ‘f120’, ‘f7’, ‘f124’,..., ‘f257’; con el orden ya establecido se aplica un modelo kNN con $k=30$, el modelo DT por defecto de MATLAB® con nivel máximo de podamiento de 10 y el modelo SVM con un $\nu = 0.5$, a medida que aumentan el número de características. Para determinar cuáles son las mejores características se evalúan las métricas de rendimiento del sistema de votación. Este proceso se observa en la Figura 56 donde se determinan las mejores características que son: f17’, ‘f19’, ‘f16’, ‘f18’, ‘f20’, ‘f21’, ‘f15’, ‘f22’, ‘f14’, ‘f13’, ‘f12’, ‘f23’, ‘f24’, ‘f49’, ‘f45’, ‘f80’, ‘f30’, ‘f50’, ‘f44’, ‘f46’, ‘f86’, ‘f39’, ‘f38’, ‘f81’, ‘f79’, ‘f43’, ‘f11’, ‘f85’, ‘f76’, ‘f51’, ‘f72’, ‘f73’, ‘f41’, ‘f48’, ‘f75’, ‘f77’, ‘f74’, ‘f87’, ‘f42’, ‘f47’, ‘f84’, ‘f82’, ‘f68’, ‘f78’, ‘f35’, ‘f31’, ‘f27’, ‘f71’, ‘f36’, ‘f40’; las cuales corresponden a las 50 primeras características obtenidas por el método de filtro. En la Figura 57 se observa como disminuye el BER a medida que se aumentan las características.

Figura 56

(a) Métricas de rendimiento para la selección de características PSD sistema de clasificación usando método de correlación, (b) Frecuencias de interés selección de características método de filtro.

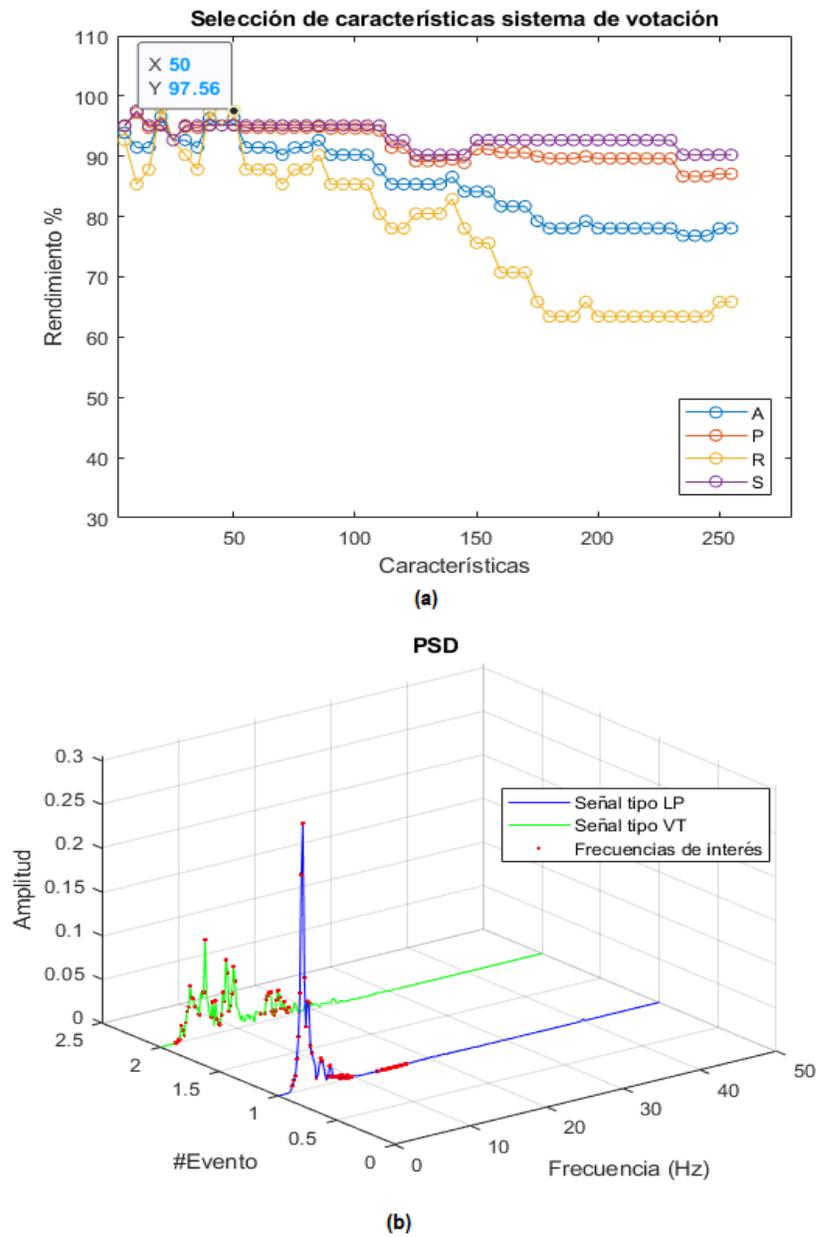
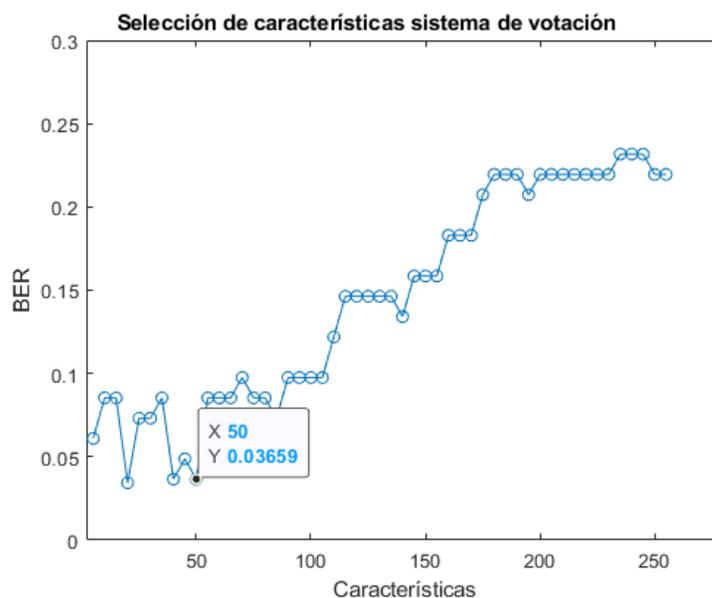


Figura 57

Análisis del BER para la selección de características PSD sistema de clasificación usando método de correlación



Como se observa en la Tabla 23, al obtener las medidas de rendimiento con las 50 primeras características del orden obtenido con el método de filtro de las 257 características PSD al remover los componentes DC, se obtienen los algoritmos de clasificación mencionados y un sistema de votación entre los tres métodos. Se obtiene la matriz de confusión de cada sistema de votación como se observa en la Figura 58 y Figura 59.

Tabla 23

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 50

Características PSD.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Clasificación 82 eventos con 50 características PSD					
kNN	87,81	80,39	100,00	75,60	0,120
DT	69,51	83,33	48,78	90,24	0,300
SVM	91,43	97,22	85,36	97,56	0,080

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 50

Características PSD.

Votación	91,46	94,73	87,80	95,12	0,080
Clasificación 563 eventos con 50 características PSD					
kNN	56,12	14,23	100,00	52,68	0,230
DT	85,61	25,00	48,78	88,50	0,310
SVM	92,36	48,61	85,36	92,11	0,108
Votación	87,92	36,36	87,80	87,93	0,120

Figura 58

Matriz de confusión clasificación 82 microterremotos con 50 características PSD.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	36	5	41 VT
	Negativos	2	39	41 LP

Figura 59

Matriz de confusión clasificación 563 microterremotos con 50 características PSD.

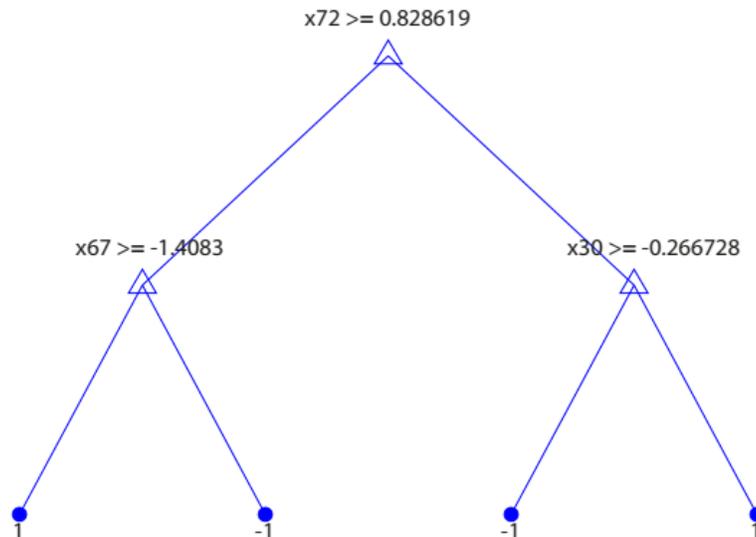
		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	36	5	41 VT
	Negativos	63	459	522 LP

- **Embebido**

Para las 84 características el software MATLAB® al obtener el algoritmo de aprendizaje supervisado DT seleccionan las características con mayor relevancia y genera un árbol de decisión como se observa en la Figura 60. Este método indica que las mejores características son: 'f30' en el dominio de la frecuencia (Densidad de pico RMS) y 'f67' (D4 RMS en el dominio del tiempo), 'f72' (D2 Pico a pico en el dominio del tiempo) en escala.

Figura 60

Representación de árbol de decisión considerando 84 características sistema de clasificación.



Como se observa en la Tabla 24, se obtienen las métricas de rendimiento con las 3 primeras características obtenidas con los niveles de poda del método embebido de las 84 características al remover los componentes DC, se aplica un modelo kNN con $k=35$, el modelo DT por defecto de MATLAB® con nivel máximo de podamiento de 10 y el modelo SVM con un $\nu = 0.5$, y un sistema de votación entre los tres métodos mencionados. Se obtiene la matriz de confusión de cada sistema de votación como se observa en la Figura 61 y Figura 62.

Tabla 24

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 3

Características.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Clasificación 82 eventos con 3 características					
kNN	90,25	100,00	80,49	100,00	0,098
DT	82,93	90,91	73,17	92,68	0,171
SVM	90,25	100,00	80,49	100,00	0,098
Votación	90,25	100,00	80,49	100,00	0,098
Clasificación 563 eventos con 3 características					
kNN	92,90	50,76	80,49	93,87	0,129
DT	91,12	43,44	73,17	92,53	0,171
SVM	93,96	55,93	80,49	95,02	0,122
Votación	93,96	55,93	80,49	95,02	0,122

Figura 61

Matriz de confusión clasificación 82 microterremotos con 3 características.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	0	41	41 LP

Figura 62

Matriz de confusión clasificación 563 microterremotos con 3 características.

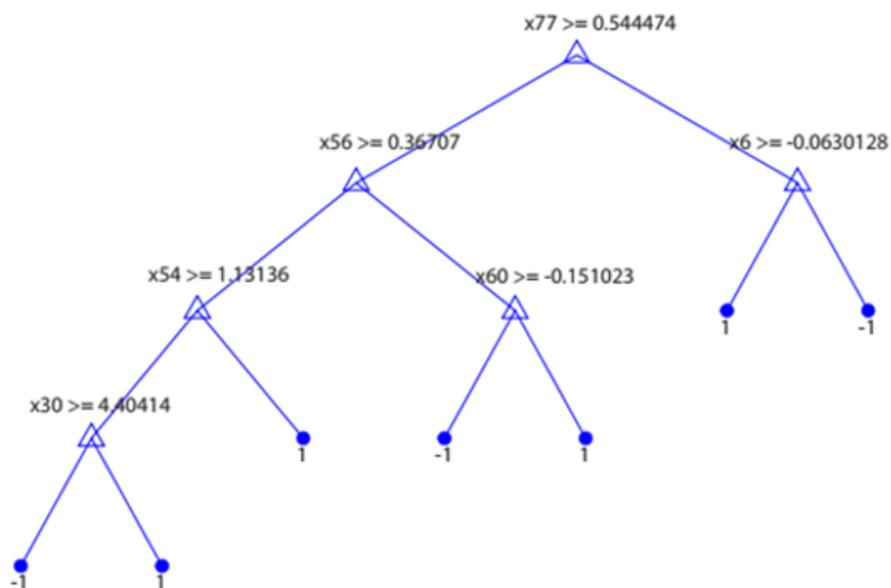
		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	26	496	522 LP

Para las 257 características PSD, se obtiene el modelo de árbol de decisión como se observa en la Figura 63. Este método indica que las mejores características son: 'f77', 'f56', 'f6', 'f54', 'f60', 'f30'. Las frecuencias de interés que representa cada característica son: 'f6' (corresponde a la frecuencia de 1,16 Hz), 'f30' (corresponde a la frecuencia de 5,84 Hz), 'f54' (corresponde a la frecuencia de 10,50 Hz), 'f56' (corresponde a la frecuencia de 10,90 Hz), 'f60' (corresponde a la frecuencia de 11,67 Hz) y 'f77' (corresponde a la frecuencia de 15,00 Hz).

Como se observa en la Tabla 25, se obtienen las métricas de rendimiento con las 6 primeras características obtenidas con los niveles de poda del método embebido de las 257 características PSD al remover los componentes DC, se aplica un modelo kNN con $k=30$, el modelo DT por defecto de MATLAB® con nivel máximo de podamiento de 10 y el modelo SVM con un $\nu = 0.5$, y un sistema de votación entre los tres métodos mencionados. Se obtiene la matriz de confusión de cada sistema de votación como se observa en la Figura 64 y Figura 65.

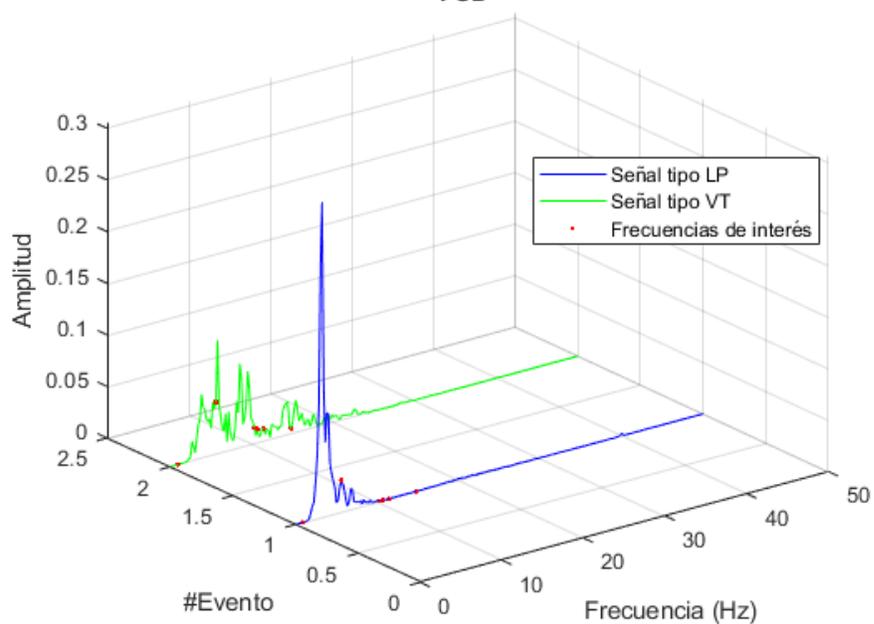
Figura 63

(a) Representación de árbol de decisión considerando 257 características PSD sistema de clasificación, (b) Frecuencias de interés selección de características método embebido.



(a)

PSD



(b)

Tabla 25

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 6

Características PSD.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Clasificación 82 eventos con 6 características PSD					
kNN	80,50	96,30	63,41	97,57	0,195
DT	84,15	93,75	73,17	95,12	0,158
SVM	82,93	88,57	75,61	90,24	0,170
Votación	84,14	96,67	70,73	97,56	0,158
Clasificación 563 eventos con 6 características PSD					
kNN	90,41	40,00	63,42	92,52	0,220
DT	83,66	27,03	73,17	84,48	0,212
SVM	80,28	23,48	75,61	80,65	0,218
Votación	87,74	33,72	70,73	89,08	0,200

Figura 64

Matriz de confusión clasificación 82 microterremotos con 6 características PSD.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	29	12	41 VT
	Negativos	1	40	41 LP

Figura 65

Matriz de confusión clasificación 563 microterremotos con 6 características PSD.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	29	12	41 VT
	Negativos	57	465	522 LP

- **Envoltura**

Para las 84 características se obtiene el siguiente orden: 'f22', 'f72', 'f30', 'f60', 'f41', 'f65', 'f75', 'f61', 'f28', 'f58', 'f37', 'f73', 'f25', 'f12', 'f74', 'f17', 'f71', 'f27', 'f15', 'f44', 'f24', 'f36', 'f40', 'f23', 'f31', 'f66', 'f49', 'f26', 'f50', 'f13', 'f64', 'f14', 'f34', 'f68', 'f32', 'f33', 'f45', 'f39', 'f82', 'f16', 'f9', 'f18', 'f59', 'f67', 'f79', 'f19', 'f46', 'f21', 'f81', 'f69', 'f57', 'f51', 'f6', 'f76', 'f42', 'f4', 'f35', 'f78', 'f5', 'f29', 'f10', 'f43', 'f83', 'f3', 'f77', 'f84', 'f55', 'f11', 'f20', 'f2', 'f8', 'f38', 'f54', 'f47', 'f63', 'f62', 'f70', 'f52', 'f80', 'f56', 'f48', 'f53', 'f1', 'f7'; con el orden ya establecido se aplica un modelo kNN con $k=35$, el modelo DT por defecto de MATLAB[®] con nivel máximo de podamiento de 10 y el modelo SVM con un $\nu = 0.5$, a medida que aumentan el número de características. Para determinar cuáles son las mejores características se evalúan las métricas de rendimiento del sistema de votación. Este proceso se puede observar en la Figura 66, donde se determinan que las mejores características son: 'f2' (Pico máximo 10-20 Hz RMS), 'f30' (Densidad de pico RMS) en el dominio de la frecuencia; y 'f41' (D6 máximo pico en frecuencia), 'f60' (Porcentaje de energía para D4), 'f65' (D2 RMS en el dominio del tiempo), 'f72' (D2 Pico a pico en el dominio del tiempo) en escala. Las cuales corresponden a las 6 primeras características obtenidas por el método de envoltura, 2 en el dominio de la frecuencia y 4 en escala. En la Figura 67 se observa como disminuye el BER a medida que se aumentan las características.

Figura 66

Métricas de rendimiento para la selección de características PSD sistema de clasificación usando el método de envoltura

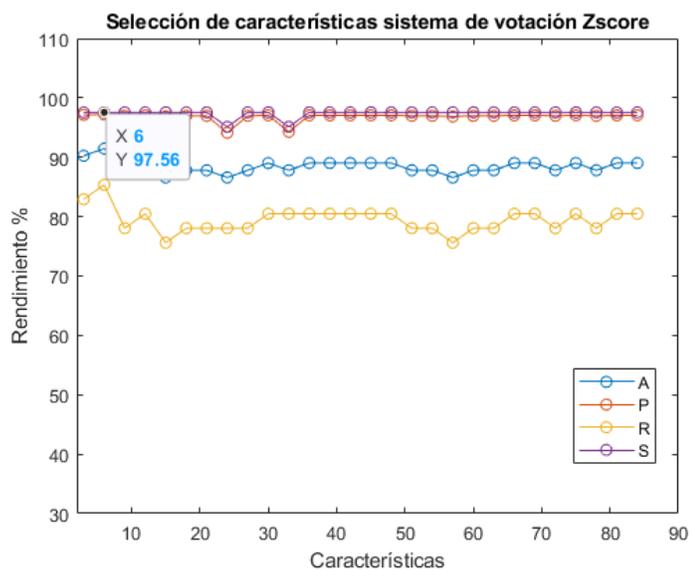
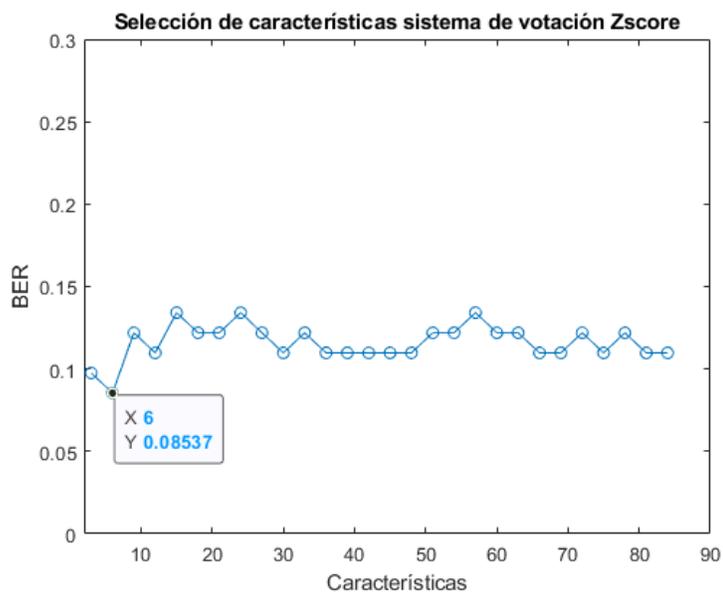


Figura 67

Análisis del BER para la selección de características PSD sistema de clasificación usando el método de envoltura



Como se observa en la Tabla 26, se obtienen las métricas de rendimiento con las 6 primeras características del orden obtenido con el método de envoltura de las 84 características al remover los componentes DC, se obtiene los algoritmos de clasificación mencionados y un sistema de votación entre los tres métodos. Se obtiene la matriz de confusión de cada sistema de votación como se observa en la Figura 68 y Figura 69.

Tabla 26

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 6 Características.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Clasificación 82 eventos con 6 características					
kNN	56,09	100,00	12,19	100,00	0,439
DT	91,46	92,50	90,24	92,68	0,085
SVM	91,46	97,22	85,36	97,56	0,085
Votación	92,68	100,00	85,36	100,00	0,073
Clasificación 563 eventos con 6 características					
kNN	93,60	100,00	12,19	100,00	0,439
DT	83,30	29,13	90,24	82,75	0,135
SVM	92,54	49,30	85,37	93,10	0,108
Votación	93,60	53,85	85,36	94,25	0,102

Figura 68

Matriz de confusión clasificación 82 microterremotos con 6 características.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	35	6	41 VT
	Negativos	0	41	41 LP

Figura 69

Matriz de confusión clasificación 563 microterremotos con 6 características.

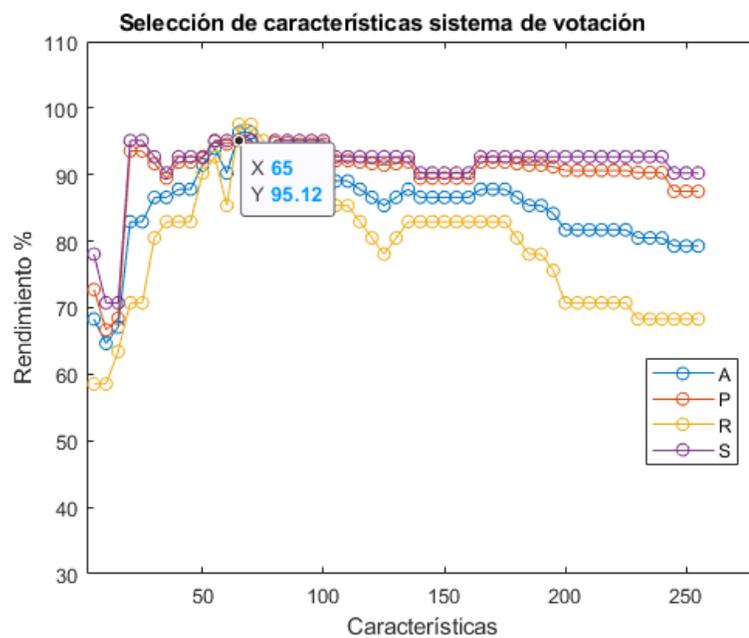
		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	35	6	41 VT
	Negativos	30	492	522 LP

Para las 257 características PSD se obtiene el siguiente orden: 'f83', 'f82', 'f73', 'f18', 'f78', 'f17', 'f56', 'f58', 'f80', 'f57', 'f79', 'f74', 'f77', 'f62', 'f2', 'f81', 'f59', 'f63', 'f54', 'f16', 'f1', 'f3', 'f75', 'f61', 'f53', 'f55', 'f72', 'f85', 'f66', 'f84', 'f51', 'f50', 'f15', 'f88', 'f64', 'f60', 'f19', 'f45', 'f104', 'f76', 'f100', 'f52', 'f20', 'f89', 'f14', 'f101', 'f108', 'f109', 'f107', 'f98', 'f10', 'f46', 'f86', 'f49', 'f105', 'f106', 'f43', 'f90', 'f114', 'f11', 'f87', 'f99', 'f21', 'f12', 'f48', 'f13', 'f70', 'f103', 'f112', 'f44', 'f97', 'f67', 'f69', 'f102', 'f4', 'f71', 'f9', 'f22', 'f42', 'f116', 'f91', 'f8', 'f119', 'f118', 'f65', 'f94', 'f92', 'f115', 'f113', 'f95', 'f7', 'f117', 'f93', 'f5', 'f110', 'f96', 'f47', 'f6', 'f121', 'f39', 'f38', 'f111', 'f41', 'f120', 'f135', 'f68', 'f134', 'f122', 'f123', 'f136', 'f148', 'f124', 'f133', 'f125', 'f24', 'f40', 'f23', 'f137', 'f36', 'f149', 'f147', 'f144', 'f150', 'f35', 'f189', 'f201', 'f184', 'f141', 'f30', 'f126', 'f146', 'f140', 'f139',...'f6'; con el orden ya establecido se aplica un modelo kNN con k=30, el modelo DT

por defecto de MATLAB® con nivel máximo de podamiento de 10 y el modelo SVM con un $\nu = 0.5$, a medida que aumentan el número de características. Para determinar cuáles son las mejores características se evalúan las métricas de rendimiento del sistema de votación. Este proceso se observa en la Figura 70, donde se determina que las mejores características son: 'f83', 'f82', 'f73', 'f18', 'f78', 'f17', 'f56', 'f58', 'f80', 'f57', 'f79', 'f74', 'f77', 'f62', 'f2', 'f81', 'f59', 'f63', 'f54', 'f16', 'f1', 'f3', 'f75', 'f61', 'f53', 'f55', 'f72', 'f85', 'f66', 'f84', 'f51', 'f50', 'f15', 'f88', 'f64', 'f60', 'f19', 'f45', 'f104', 'f76', 'f100', 'f52', 'f20', 'f89', 'f14', 'f101', 'f108', 'f109', 'f107', 'f98', 'f10', 'f46', 'f86', 'f49', 'f105', 'f106', 'f43', 'f90', 'f114', 'f11', 'f87', 'f99', 'f21', 'f12', 'f48'. Las cuales corresponden a las 65 primeras características obtenidas por el método de correlación. En la Figura 71 se observa como disminuye el BER a medida que se aumentan las características.

Figura 70

(a) Métricas de rendimiento para la selección de características PSD sistema de clasificación usando el método de envoltura, (b) Frecuencias de interés selección de características método de envoltura.



PSD

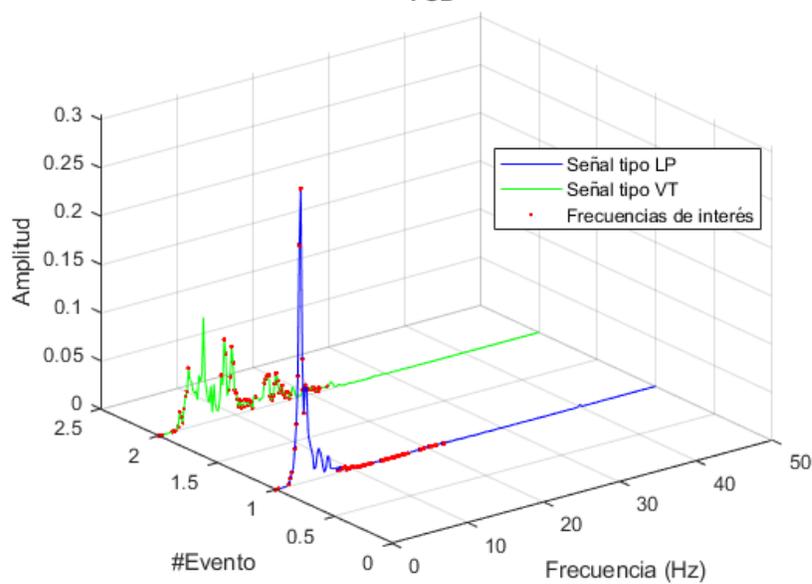
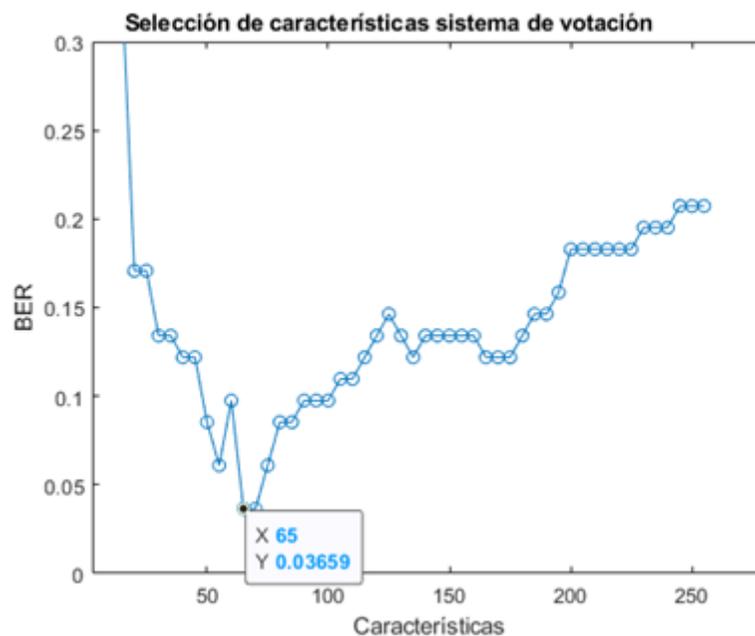


Figura 71

Análisis del BER para la selección de características PSD sistema de clasificación usando el método de envoltura



Como se observa en la Tabla 25, al obtener las métricas de rendimiento con las 65 primeras características del orden obtenido con el método de envoltura de las 257 características PSD al remover los componentes DC, se obtienen los algoritmos de clasificación mencionados y un sistema de votación entre los tres métodos. Se obtiene la matriz de confusión de cada sistema de votación como se observa en la Figura 72 y Figura 73.

Tabla 27

Métricas de rendimiento de algoritmos kNN, DT, SVM y un sistema de votación con 65

Características PSD.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Clasificación 82 eventos con 65 características PSD					
kNN	81,70	100,00	63,41	100,00	0,183
DT	82,93	96,55	68,30	97,56	0,171
SVM	81,70	84,21	78,05	85,37	0,183
Votación	85,36	100,00	70,73	100,00	0,146
Clasificación 563 eventos con 65 características PSD					
kNN	93,42	54,16	63,41	95,78	0,204
DT	87,21	32,19	68,20	88,70	0,215
SVM	87,39	34,04	78,05	88,12	0,169
Votación	91,83	46,03	70,73	93,50	0,178

Figura 72

Matriz de confusión clasificación 82 microterremotos con 65 características PSD.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	29	12	41 VT
	Negativos	0	41	41 LP

Figura 73

Matriz de confusión clasificación 563 microterremotos con 65 características PSD.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	29	12	41 VT
	Negativos	34	488	522 LP

- **Mejores características**

Para las 84 características se determina que las mejores son: 'f72', 'f30', 'f23', 'f37', 'f65', 'f58', 'f60', 'f61', 'f75', 'f68', 'f25', 'f79', 'f40', 'f28', 'f15', 'f7', 'f31', 'f59', 'f66', 'f74', 'f39', 'f33', 'f14', 'f71', 'f50', 'f32', 'f41', 'f73', 'f34', 'f81', 'f78', 'f44', 'f4', 'f67', 'f69', 'f26', 'f38'. Se obtienen un total de 39 características como se observa en la Figura 52 de las cuales 2 son en tiempo, 15 en frecuencia y 22 en escala, se calculan los algoritmos de aprendizaje supervisado para obtener los modelos finales para la clasificación de microterremotos y se calculan las métricas de rendimiento que se observa en la Tabla 26. Se obtiene la matriz de confusión como se observa en la Figura 74 y Figura 75, se aplica un modelo kNN con un $k=35$, el modelo DT por defecto de MATLAB® con nivel máximo de poda de 10 y el modelo SVM con un $\nu = 0.5$, y un sistema de votación entre los tres métodos mencionados. Estos modelos son los finales usados para la clasificación de microterremotos.

Tabla 28

Métricas de rendimiento sistema de votación con 39 Características

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Clasificación 82 eventos con 39 características					
Votación	90,24	100,00	80,49	100,00	0,097
Clasificación 563 eventos con 39 características					
Votación	95,20	63,46	80,49	96,36	0,115

Figura 74

Matriz de confusión clasificación 82 microterremotos con 39 características.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	0	41	41 LP

Figura 75

Matriz de confusión clasificación 563 microterremotos con 39 características.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	19	503	522 LP

Para las 257 características PSD se determina que las mejores son: 'f17', 'f19', 'f16', 'f18', 'f20', 'f21', 'f15', 'f22', 'f14', 'f13', 'f12', 'f23', 'f24', 'f49', 'f45', 'f80', 'f30', 'f50', 'f44', 'f46', 'f86', 'f39', 'f38', 'f81', 'f79', 'f43', 'f11', 'f85', 'f76', 'f51', 'f72', 'f73', 'f41', 'f48', 'f75', 'f77', 'f74',

'f87', 'f42', 'f47', 'f84', 'f82', 'f68', 'f78', 'f35', 'f31', 'f27', 'f71', 'f36', 'f40'. Se obtiene un total de 50 características que se observa en la Figura 56, con las cuales se calculan los algoritmos de aprendizaje supervisado, donde se calculan las métricas de rendimiento que se pueden observar en la Tabla 29 y se obtiene la matriz de confusión como se observa en la Figura 76 y Figura 77, se aplica un modelo kNN con un $k=30$, el modelo DT por defecto de MATLAB® con nivel máximo de podamiento de 10 y el modelo SVM con un $\nu = 0.5$, y un sistema de votación entre los tres métodos mencionados. Estos modelos son los finales usados para la clasificación de microterremotos.

Tabla 29

Métricas de rendimiento sistema de votación con 50 Características PSD.

Algoritmo	A(%)	P(%)	R(%)	S(%)	BER
Clasificación 82 eventos con 50 características PSD					
Votación	91,46	94,73	87,80	95,12	0,080
Clasificación 563 eventos con 50 características PSD					
Votación	87,92	36,36	87,80	87,93	0,120

Figura 76

Matriz de confusión clasificación 82 microterremotos con 50 características PSD.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	36	5	41 VT
	Negativos	2	39	41 LP

Figura 77

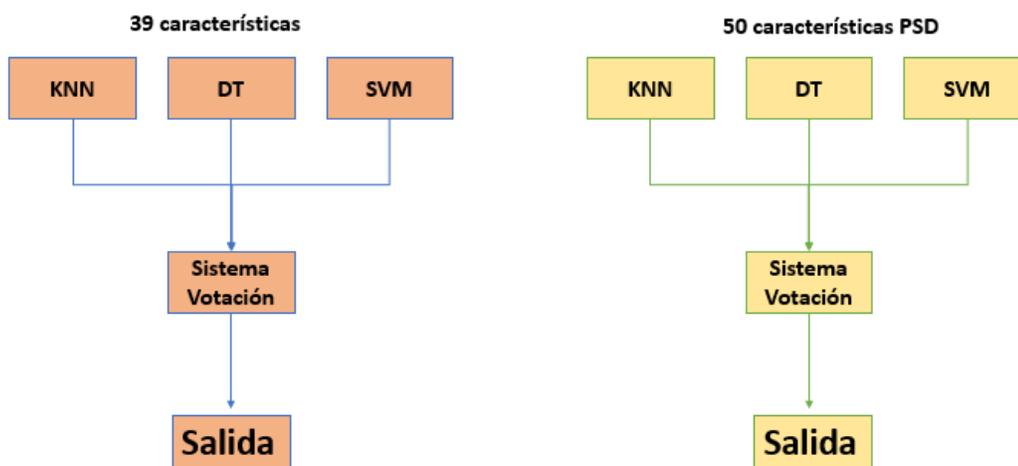
Matriz de confusión clasificación 563 microterremotos con 50 características PSD.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	36	5	41 VT
	Negativos	63	459	522 LP

Se obtienen 6 modelos para la clasificación de microterremotos como se observa en la Figura 78. Se unifican los modelos obtenidos con los dos distintos tipos de características para obtener un sistema de votación. Primero, se obtiene un sistema de votación con 5 modelos; posteriormente un sistema con 3 modelos que tienen las mejores métricas de desempeño; y finalmente se elige un sistema de votación que será unificadas en la interfaz gráfica para el sistema de reconocimiento.

Figura 78

Modelos finales sistema de clasificación de microterremotos



- **Modelos de clasificación unificados**

Se seleccionan 5 y 3 modelos de detección de microterremotos obtenidos entre las 84 características y 257 características PSD. Se implementa un sistema de votación con se observa en la Figura 79 y Figura 80. Finalmente se obtienen las métricas de rendimiento del sistema de votación.

Figura 79

Sistema de votación con 5 modelos sistema de clasificación de microterremotos

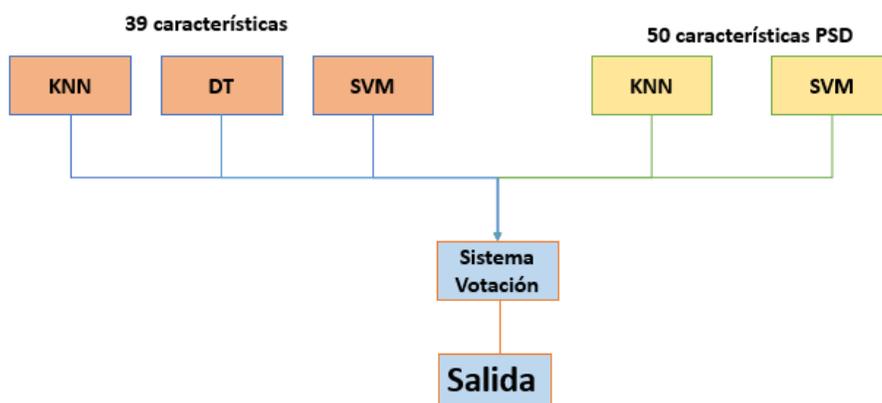
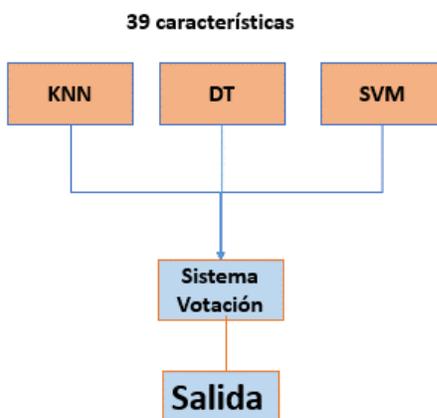


Figura 80

Sistema de votación con 3 modelos sistema de clasificación de microterremotos



Como se observa en la Figura 79, se seleccionan los modelos con las mejores métricas de rendimiento al unificar los 5 modelos para obtener un sistema de votación entre todos. Posteriormente como se observa en la figura 80, se realiza otra selección de 3 modelos para obtener las métricas de rendimiento. Finalmente se selecciona el sistema de votación unificado con los mejores resultados.

Se obtienen las métricas de rendimiento del sistema de votación como se observa en la Tabla 30, y la matriz de confusión correspondiente para la prueba realizada como se observa en la Figura 81, Figura 82, Figura 83 y Figura 84.

Tabla 30

Métricas de rendimiento sistema de votación con 5 y 3 modelos sistema de clasificación de microterremotos.

Eventos	A(%)	P(%)	R(%)	S(%)	BER
Sistema de clasificación con 5 modelos					
82 eventos	90,24	100,00	80,49	100,00	0,097
563 eventos	92,71	50,00	80,49	93,67	0,143
Sistema de clasificación con 3 modelos					
82 eventos	90,24	100,00	80,49	100,00	0,097
563 eventos	95,20	63,46	80,49	96,36	0,115

Figura 81

Matriz de confusión clasificación 82 microterremotos 5 modelos de clasificación.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	0	41	41 LP

Figura 82

Matriz de confusión clasificación 563 microterremotos 5 modelos de clasificación.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	33	489	522 LP

Figura 83

Matriz de confusión clasificación 82 microterremotos 3 modelos de clasificación.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	0	41	41 LP

Figura 84

Matriz de confusión clasificación 563 microterremotos 3 modelos de clasificación.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	33	8	41 VT
	Negativos	19	503	522 LP

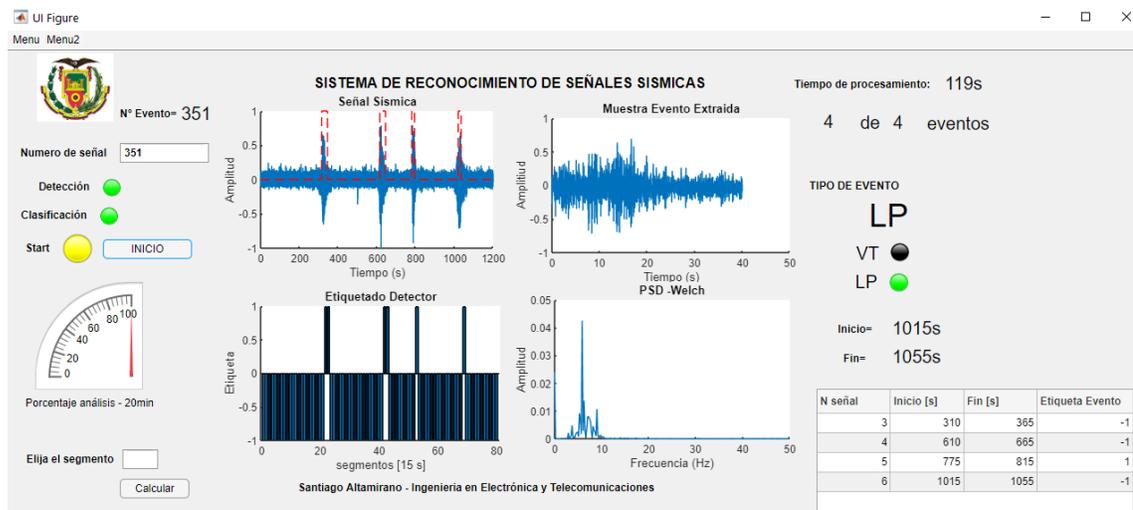
En la Tabla 30, se puede observar que el sistema de votación unificado con 3 modelos alcanza una exactitud del 95,20 % y un BER de 0,115. Este sistema de votación será unificado en la interfaz gráfica.

Sistema de Reconocimiento en tiempo real.

Este sistema es la unificación de los modelos obtenidos tanto para la detección y clasificación respectivamente. La interfaz gráfica diseñada en *appdesigner* es la presentación final de la herramienta que realiza el reconocimiento de microterremotos en tiempo real y se puede observar en la Figura 85. El sistema de reconocimiento de microterremotos implementado en este trabajo de investigación tarda en la detección y clasificación de una señal de 20 minutos un promedio de 33 segundos.

Figura 85

Interfaz gráfica sistema de reconocimiento de microterremotos del volcán Cotopaxi



Evaluación del sistema de reconocimiento de microterremotos en tiempo real con los expertos del IGEPN.

Se realiza una comparación en el reconocimiento de microterremotos del volcán Cotopaxi obtenido por los expertos del IGEPN de la estación de referencia (BREF) y el sistema de reconocimiento implementado en este trabajo de investigación. El reconocimiento se realiza en las fechas 21 y 22 de diciembre del 2020, donde el IGEPN realiza tanto la detección y la clasificación de los eventos registrados y proporcionan los resultados obtenidos de su monitorización para la comparación con los resultados del sistema de reconocimiento obtenido en este trabajo de investigación.

En la fecha del 21 de diciembre los expertos del IGEPN detectaron 8 eventos microsísmicos, donde 3 son eventos del tipo LP, 3 eventos del tipo VT, 1 evento del tipo ICQ y 1 evento HYB. En la fecha del 22 de diciembre los expertos del IGEPN detectaron 60 eventos microsísmicos, donde 25 son eventos del tipo LP, 32 eventos del tipo VT y 3 evento HYB.

Al procesar las bases de datos de las fechas indicadas el sistema de reconocimiento en la fecha 21 de diciembre se detectaron un total de 108 eventos microsísmicos, donde 56 son eventos del tipo LP y 52 eventos del tipo VT. En la fecha del 22 de diciembre se detectaron un total de 122 eventos microsísmicos, donde 58 son eventos del tipo LP y 64 eventos del tipo VT.

Comparación

Se realiza la comparación del reconocimiento realizado por los expertos del IGEPN y el sistema de reconocimiento en las fechas previamente mencionadas. La comparación se realiza solo en una estación, pero la recomendación de los expertos del IGEPN es que, la comparación para determinar si un microterremoto es efectivamente un evento microsísmico se lo debe realizar con tres estaciones o más. Se obtienen las métricas de rendimiento con ayuda de una matriz de confusión para el reconocimiento de los microterremotos en las determinadas fechas.

En la fecha del 21 de diciembre el sistema de reconocimiento detectó 108 eventos, donde están incluidos todos los eventos microsísmicos detectados por los expertos de IGEPN. Se obtienen la matriz de confusión tanto para detección y clasificación como se observa en la Figura 86, Figura 87 y Figura 88 respectivamente y se presenta en la Tabla 31.

Figura 86

Matriz de confusión detección todos los microterremotos 21 de diciembre 2020.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	3	49	52 VT
	Negativos	53	3	56 LP

Figura 87

Matriz de confusión detección 8 eventos recopilados 21 de diciembre 2020.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	8	0	8 EV
	Negativos	0	0	0 EV

Figura 88

Matriz de confusión Clasificación de microterremotos 21 de diciembre 2020.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	2	1	3VT
	Negativos	0	3	3 LP

Tabla 31

Métricas de rendimiento detección y clasificación de microterremotos 21 de diciembre.

A(%)	P(%)	R(%)	S(%)	BER
Sistema de detección 108 eventos				
5,55	5,36	5,77	5,36	0,94
Sistema de detección 8 eventos				
100	100	100	100	0
Sistema de clasificación				
83,33	100	66,67	100	0,16

En la fecha del 22 de diciembre el sistema de reconocimiento detecto 122 eventos, donde están incluidos 42 eventos de los 60 eventos microsísmicos detectados por los expertos de IGEPN. Se obtienen la matriz de confusión tanto para detección y clasificación como se observa en la Figura 89, Figura 90 y Figura 91 respectivamente y se presenta en la Tabla 32.

Figura 89

Matriz de confusión detección de microterremotos 22 de diciembre 2020.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	18	46	64 VT
	Negativos	42	16	58 LP

Figura 90

Matriz de confusión detección 60 eventos recopilados 21 de diciembre 2020.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	42	10	52 EV
	Negativos	8	0	8 EV

Figura 91

Matriz de confusión Clasificación de microterremotos 22 de diciembre 2020.

		Predicción		Total Eventos
		Positivos	Negativos	
Observación	Positivos	18	4	22VT
	Negativos	4	16	20 LP

Tabla 32

Métricas de rendimiento detección y clasificación de microterremotos 22 de diciembre .

A(%)	P(%)	R(%)	S(%)	BER
Sistema de detección 122 eventos				
27,87	30,00	28,12	27,58	0,72
Sistema de detección 60 eventos				
70,00	84,00	80,77	0,00	0,59
Sistema de clasificación				
80,95	81,82	81,82	80	0,19

Se puede observar en la Tabla 31 y Tabla 32, que las métricas de rendimiento con respecto a la detección y clasificación son extremadamente bajas, esto se debe que el sistema de reconocimiento detecta más eventos microsísmicos que los encontrados expertos del IGEPN en las fechas determinadas. Adicionalmente se debe considerar que el reconocimiento de microterremotos está realizado en una sola estación.

Comparación con trabajos relacionados

El presenta trabajo de investigación es comparado con el sistema desarrollado en (Lara Cueva, 2015) denominado Monitorización Volcánica en Tiempo Real Usando Redes de Sensores Inalámbricas en el volcán Cotopaxi, el mismo que realiza el reconocimiento de microterremotos en tiempo real. Al igual que el sistema de reconocimiento desarrollado en este trabajo, se realiza la extracción de características, selección de características en el dominio del tiempo, frecuencia y escala, donde se aplican algoritmos de aprendizaje supervisado kNN con un k=20,

DT con una validación cruzada de 10 y SVM con un $\nu = 0.5$, para la detección y algoritmos kNN con un $k=30$, DT con nivel máximo de categoría de 10 y SVM con un $\nu = 0.5$, para la clasificación. En el caso de la detección el sistema de Monitorización Volcánica obtuvo las siguientes métricas de rendimiento, en la etapa de detección se obtiene una precisión del 99% y en la etapa de clasificación entre eventos LP y VT una precisión del 97%.

El sistema de Monitorización Volcánica en Tiempo Real Usando Redes de Sensores Inalámbricas realiza una extracción de 81 características y posteriormente una selección de características, donde al aplicar el método de envoltura para la selección de características, en la clasificación se determina que 15 características son las de mayor relevancia como se observa en la Tabla 33.

Tabla 33

Características de interés (Lara Cueva, 2015)

Características de interés	
Característica	Descripción
Tiempo	
f12	Z_t
Frecuencia	
f14	ψ_f
f21	MSE_f
f22	U_{10-20}
f27	RMS_f
f29	E_f^{RMS}

Características de interés trabajo (Lara Cueva, 2015)

Escala	
f40	$V_w^{\delta 5}$
f42	$\mu_w^{\delta 5}$
f54	$\mu_w^{\delta 1}$
f57	$E_{\%w}^{\delta 5}$
f60	$E_{\%w}^{\delta 2}$
f69	$P2P_w^{\delta 5}$
f79	$P2P_w^{\delta 2}$
f80	$P2P_w^{\delta 2,RMS}$
f81	$P2P_w^{\delta 1}$

En el caso de la selección de características realizadas en este trabajo se obtuvieron 39 características ('f72', 'f30', 'f23', 'f37', 'f65', 'f58', 'f60', 'f61', 'f75', 'f68', 'f25', 'f79', 'f40', 'f28', 'f15', 'f7', 'f31', 'f59', 'f66', 'f74', 'f39', 'f33', 'f14', 'f71', 'f50', 'f32', 'f41', 'f73', 'f34', 'f81', 'f78', 'f44', 'f4', 'f67', 'f69', 'f26', 'f38'), de las cuales las seis de las características seleccionadas se encuentran en común en las dos trabajos investigaciones ('f14', 'f40', 'f60', 'f69', 'f78', 'f81') los cuales son: una en frecuencia (ψ_f) y cinco en el dominio de la escala ($\mu_w^{\delta 5}$, $P2P_w^{\delta 5}$, $P2P_w^{\delta 2}$, $P2P_w^{\delta 2,RMS}$, $P2P_w^{\delta 1}$), por este motivos los resultados en la etapa de clasificación son similares. Adicionalmente el análisis espectral realizado en el trabajo de (Lara Cueva, 2015), donde la selección de características con método embebido DT reflejan que 3 frecuencias son características claves de interés 'f16' $\geq -0,42$ (correspondiente a 3 Hz), 'f21' $\geq -0,58$ (correspondiente a 4 Hz), y 'f36' $\geq 0,47$ (correspondiente a 6,8 Hz), para la clasificación de eventos LP y VT, en el presente trabajo las frecuencias de interés que se observan en la Figura

63, donde 6 son características claves de interés ' $f_{77}' \geq -0,544$ (correspondiente a 15 Hz), ' $f_{56}' \geq 0,367$ (correspondiente a 10,8 Hz), ' $f_6' \geq -0,063$ (correspondiente a 1,2 Hz), ' $f_{54}' \geq 1,131$ (correspondiente a 10,5 Hz), ' $f_{60}' \geq -0,151$ (correspondiente a 11,7 Hz), y ' $f_{30}' \geq 4,404$ (correspondiente a 5,8 Hz), se observa que en el trabajo de (Lara Cueva, 2015) las frecuencia de interés están por debajo de 7 Hz mientras que las obtenidas en este trabajo de investigación abarcan ciertas frecuencias desde 1,2 Hz hasta 15 Hz.

Adicionalmente se compara la detección de microsismos obtenida en el trabajo de (Rosero, 2019), donde se aplica una detección basada en algoritmos de Actividad de Voz en señales sísmicas del volcán Cotopaxi con duración de 20 minutos. Las métricas de rendimiento alcanzan una exactitud del 96,38%, precisión del 100% y un BER de 0,011. Estos resultados son similares a los obtenidos en este trabajo de investigación, donde se llega a obtener una exactitud del 97,60 %, precisión del 98,72% y un BER de 0,012, con estos resultados se observa que los dos detectores aplicando distintos tipos de metodologías cumplen con los requerimientos del IGEPN y podría ser implementados como una herramienta de ayuda.

Capítulo V

Conclusiones y Recomendaciones

En esta investigación se ha enfocado a la obtención de modelos de aprendizaje supervisado para la detección y clasificación de microterremotos. Para la obtención de los modelos se han utilizado distintas técnicas para el cálculo de características con las cuales se define una metodología clara de procesamiento de un evento microsísmico. Las características calculadas son una mezcla entre análisis en el dominio del tiempo, frecuencia, escala y otro tipo de caracterización al aplicar técnicas espectrales, lo que ha permitido obtener mayor variedad de modelos de aprendizaje supervisado que se implementaron en este trabajo de investigación. En cuanto a la detección de microterremotos, se obtuvo una exactitud del 97,60 %, precisión del 98,72% y un BER de 0,012 al aplicar un sistema de votación entre modelos de características y estimación espectral, se debe considerar que este sistema es capaz de detectar eventos del tipo LP, VT, TRE y HYB. En el caso de la clasificación de 563 eventos se obtuvo una exactitud del 95,20 %, precisión del 63,46% y un BER de 0,115 donde se debe considerar que este sistema es capaz de clasificar eventos del tipo LP y VT. Finalmente, estos dos sistemas son unificados en una interfaz gráfica para realizar el proceso de reconocimiento de microterremotos en tiempo real.

La interfaz gráfica implementada servirá como una herramienta de apoyo para los expertos del IGEPN en la detección y clasificación de microterremotos del Volcán Cotopaxi en tiempo real, por el grupo de investigación SRASI de la Universidad de las Fuerzas Armadas – ESPE. Los principales aportes de este trabajo de investigación son las métricas de rendimiento que se ajustan con los requerimientos del IGEPN, la facilidad de uso y procesamiento en tiempo

real. El sistema de reconocimiento tarda en promedio 33 segundos en realizar el proceso de detección y clasificación de microterremotos en señales de 20 minutos.

Al comparar el reconocimiento realizado por los expertos del IGEPN y el sistema de reconocimiento de microterremotos obtenido en esta investigación se puede observar, que el 21 de diciembre el sistema de reconocimiento obtiene una exactitud 100 % y una precisión de 100 % en la detección, y una exactitud 83,33 % y una precisión de 100 % en la clasificación. El 22 de diciembre el sistema de reconocimiento obtiene una exactitud 70 % y una precisión de 100 % en la detección, y una exactitud 80,95 % y una precisión de 81,82 % en la clasificación. Estos resultados se deben a que el sistema de reconocimiento detecta un mayor número de evento sísmicos que los reconocidos por los expertos del IGEPN y se está referenciado a una solo estación, se debe analizar los eventos detectados para verificar si la detección realizada por el sistema de reconociendo realizo el proceso manera correcta. Adicionalmente se pudo constatar al existir un evento sísmico de un periodo de duración largo el sistema de reconocimiento divide al evento y lo procesa por segmentos separados.

Conclusiones

El estudio del estado del arte realizado permitió recopilar distintos criterios aplicados en investigaciones previas para el análisis de microterremotos como extracción de características, selección de características, algoritmos de aprendizaje supervisado, métricas de rendimiento y evitó escoger técnicas que disminuyan las métricas de rendimiento esperadas.

La extracción de características de un microterremoto en el dominio de tiempo, frecuencia y escala facilitaron la implementación de técnicas de aprendizaje supervisado y obtención de modelos para el reconocimiento de microterremotos.

Las técnicas de aprendizaje supervisado kNN, DT y SVM, implementadas en esta investigación funcionaron correctamente al realizar el proceso de detección y clasificación de señales microsísmicas del volcán Cotopaxi, donde se obtuvieron 6 modelos para la detección y 6 modelos para la clasificación de microterremotos en tiempo real.

Las métricas de rendimiento obtenidas en el sistema de reconocimiento arrojan resultados aceptables en la etapa de detección, que se ajustan con los requerimientos del IGEPN. En el caso del detector el IGEPN requiere un BER de 0.010, donde el sistema obtenido arroja un BER de 0.012 y alcanza el requerimiento deseado. En el caso del clasificador el sistema arroja un BER de 0.115 que está por encima del requerimiento del IGEPN, pero se debe considerar que las pruebas realizadas son en escenarios desproporcionados de números de eventos sísmicas LP y VT no iguales.

El desarrollo e implementación de un sistema de reconocimiento de microterremotos en tiempo real es una herramienta que facilita el proceso de reconocimiento de distintos tipos de eventos sísmicos que se producen en el volcán Cotopaxi, que puede ser usado por los expertos del IGEPN para la prevención y alerta temprana hacia una posible erupción, salvaguardando vidas y recursos de las poblaciones aledañas al Volcán Cotopaxi.

Se logró implementar el sistema de reconocimiento de microterremotos del Volcán Cotopaxi en tiempo real en el IGEPN, donde se utiliza el sistema como una herramienta de ayuda para los expertos de IGEPN.

Trabajos Futuros

El presente trabajo de investigación permite clasificar microterremotos LP y VT, se propone desarrollar un sistema de reconocimiento de microterremotos que permita clasificar eventos tipo LP, VT, TR y HYB.

Nuestro grupo de investigación está interesado en verificar las características que se repiten en los modelos de detección y clasificación, para obtener nuevos modelos de aprendizaje supervisado y observar como varían las métricas de rendimiento.

Implementar el sistema de reconocimiento de microterremotos en tres o más estaciones que se encuentran a lo largo del volcán Cotopaxi.

Verificar el reconocimiento del sistema desarrollado en este trabajo de investigación por los expertos del IGEPN, para comprobar si los eventos microsísmicos detectados son eventos correctamente detectados.

Modificar el sistema de reconocimiento obtenido en este trabajo de investigación para procesar microterremotos que tienen una duración extendida.

Desarrollar un sistema de reconocimiento de microterremotos al unificar técnicas de Machine Learning tradicional y Deep Learning.

Referencias

- Acuña, F. E. (2000). *Análisis Discriminante*. Puerto Rico: Universidad de Puerto Rico en Mayaguez.
- Aguinaga, D. (2018). *Análisis y Caracterización de microsismos producidos en el volcán Cotopaxi mediante el uso de Técnicas Espectrales*. Sangolquí: Universidad de las Fuerzas Armadas ESPE.
- Alesio, S. M. (2015). *Digital signal processing and spectral analysis for scientists: concepts and applications*. Springer.
- aprendelA. (2019). *aprendelA*. Obtenido de Métodos de Selección de Características: <https://aprendeia.com/metodos-de-seleccion-de-caracteristicas-machine-learning/>
- Barquero, R., Lesage, P., Metaxian, J., Creusot, A., & Fernandez, M. (s.f.). La crisis sísmica en el volcán Irazu en 1991. *Revista Geología de Amperio Central*, 5-20.
- Bean, C., De Barros, L., Lokmer, I., Métaixian, J., O'Brien, G., & Murphy, S. (2014). Period seismicity in the shallow volcanic edifice formed from slow-rupture earthquakes. *Nat. Geosci.* Obtenido de <https://doi.org/10.1038/ngeo2027>
- Benitez, M. C. (2007). Continuous HMM-based seismic-event classification at deception island, Antarctica. *EEE Trans. Geosci. Remote Sens*, 138–146.
- Brusil, C. (Febrero de 2020). Obtenido de Análisis comparativo entre aprendizaje supervisado y aprendizaje semi-supervisado para la clasificación de señales sísmicas vulcanológicas del volcán Cotopaxi.: <https://bibdigital.epn.edu.ec/bitstream/15000/20723/1/CD%2010239.pdf>
- Brusil, C., Grijalva, F., Lara-Cueva, R., & Ruiz, M. (2019). SemiSupervised Approach for Microseisms Classification from Cotopaxi Volcano. *IEEE Latin American Conference on Computational Intelligence*, 1-6. Obtenido de <https://doi.org/10.1109/LA-CI47412.2019.9037033>
- Chouet, B. (1996). Its source and use in eruption forecasting Nature. *Long-Period volcano seismicity*, 309-312.
- Corominas, O., & Martí, J. (2011). Estudio comparativos de los planes de actuación frente al riesgo volcánico (Chile, Costa Rica, El Salvador, Ecuador, España, México y Nicaragua). *Revista Geológica de América Central* .
- Cortés, G., Benitez, M., Garca, L., & Álvarez, I. I. (2015). A comparative study of dimensionality reduction algorithms applied to volcano-seismic signals. *EEE Journal of Selected Topics in*

Applied Earth Observations and Remote Sensing 9. Obtenido de <https://doi.org/10.1109/JSTARS.2015.2479300>

- Cusano, P., Palo, M., & West, M. (2015). Long-period seismicity at shishaldin volcano (Alaska) in 2003–2004: indications of an upward migration of the source before a minor eruption. *J. Volcanol. Geotherm.* Obtenido de <https://doi.org/10.1016/j.jvolgeores.2014.12.008>.
- Dragomiretskiy, k., & Zosso, D. (2014). Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, 531-544.
- edureka! (2020). *Mastering Python*. Obtenido de K-Nearest Neighbors Algorithm Using Python: <https://www.edureka.co/blog/k-nearest-neighbors-algorithm/>
- Gonzales, L. (2018). *AprendeIA*. Recuperado el 2020, de Aprendizaje Supervisado: Decision Tree Classification: <https://aprendeia.com/aprendizaje-supervisado-decision-tree-classification/>
- Gonzales, L. (2018). *AprendeIA*. Recuperado el 2020, de Aprendizaje Supervisado: K-Nearest Neighbors: <https://aprendeia.com/aprendizaje-supervisado-k-nearest-neighbors/>
- Gonzales, L. (2019). *AprendeIA*. Recuperado el 2020, de Árboles de Decisión Regresión – Teoría: <https://aprendeia.com/arboles-de-decision-regresion-teoria-machine-learning/>
- Gutiérrez, L., Ramrez, J., Ibañez, J., & Bentez, C. (2011). Volcano-seismic signal detection and classification processing using hidden markov models-application to San cristóbal. *Hidden Markov Models, Theory and Applications*. Obtenido de <https://doi.org/10.5772/14312>.
- Hall, M., Andrade, D., Mothes, P., Troncoso, P., & ...Yepes, H. (2005). Los peligros volcánicos asociados con el Cotopaxi. *En institut de Recherche Pour le D'évolppement(IRD), & IGEPN. Corporacion nacional*.
- Havskov, J., & Ottemoller, L. (2010). Routine Data Processing in Earthquake Seismology. Obtenido de <https://doi.org/10.1007/978-90-481-8697-6>
- Huang, M., Hung, Y., Lee, W., Li, R., & Jiang, B. (2014). SVM-RFE Based Feature Selection and Taguchi Parameters Optimization for Multiclass SVM Classifier. *The Scientific Wold Journal*, 4-5.
- Ibáñez, J., & Carmona, E. (1997). *Sismicidad Volcánica*. Granada: Instituto Andaluz de Geofísica. Universidad de Granada.
- IGEPN. (3 de Agosto de 2015). *Instituto Geofísico de la Escuela Politécnica Nacional*. Obtenido de ALERTA TEMPRANA - INSTITUTO GEOFÍSICO - EPN: <https://igepn.edu.ec/>

- IGEPN. (06 de Febrero de 2017). Obtenido de Volcán Sangay:
<https://www.igepn.edu.ec/images/portal/historicas/volcanes/MP-Sangay.pdf>
- IGEPN. (2017). *Cotopaxi*. Obtenido de Instituto Geofísico de la Escuela Politécnica Nacional:
<https://www.igepn.edu.ec>
- IGEPN. (2019). *¿Cuántos volcanes activos hay en Ecuador?* Obtenido de Instituto Geofísico de la Escuela Politécnica Nacional: <https://www.igepn.edu.ec/preguntas-frecuentes?tmpl=component&faqid=12#:~:text=De%20ellos%2C%20siete%20volcanes%20continentales,decir%2C%20desde%20el%20a%C3%B1o%201532.>
- IGEPN. (2019). *¿Cuántos volcanes activos hay en Ecuador?* Obtenido de Instituto Geofísico de la Escuela Politécnica Nacional: <https://www.igepn.edu.ec/preguntas-frecuentes?tmpl=component&faqid=12#:~:text=De%20ellos%2C%20siete%20volcanes%20continentales,decir%2C%20desde%20el%20a%C3%B1o%201532.>
- IGEPN. (2019). *Instituto Geofísico de la Escuela Politécnica Nacional*. Obtenido de Red de monitoreo en el Cotopaxi: <https://www.igepn.edu.ec/>
- IGEPN. (2020). *Cotopaxi*. Obtenido de Instituto Geofísico de la Escuela Politécnica Nacional:
<https://www.igepn.edu.ec/cotopaxi>
- IGEPN. (2020). *Volcanes más activos del Ecuador*. Obtenido de Instituto Geofísico de la Escuela Politécnica Nacional: <https://www.igepn.edu.ec/>
- Jaramillo, C. (2015). *Caracterización de señales sísmicas del volcán Cotopaxi utilizando estimadores espectrales clásicos y de máxima entropía*. Sangolquí.
- Kailash, A. (2017). *Eight to Late*. Recuperado el 2020, de A gentle introduction to support vector machines using R: <https://eight2late.wordpress.com/2017/02/07/a-gentle-introduction-to-support-vector-machines-using-r/>
- Lara Cueva, R. (Abril de 2015). *Real-time Volcanic Monitoring Using Wireless Sensor*. Madrid: UNIVERSIDAD REY JUAN CARLOS. Obtenido de Real-time Volcanic Monitoring Using Wireless Sensor.
- Lara, Paillacho, & Villalva. (2016). Detección de eventos del volcán Cotopaxi empleando clasificación supervisada. *RECI*.
- Lara, R., Benítez, D., Carrera, V., Ruiz, M., & Rojo Álvarez, R. (2016). Feature selection of seismic waveforms for long period event detection at Cotopaxi Volcan. *Journal of Volcanology and Geothermal Research*.
- Lara-Cueva, R. (2015). *Real-time Volcanic Monitoring Using Wireless Sensor*.

- Lara-Cueva, R., Benítez, D., Carrera, E., Ruíz, M., & Rojo, J. (2016). Automatic Recognition of Long Period Events From Volcano Tectonic Earthquakes at Cotopaxi Volcano. *IEEE*.
- Lara-Cueva, R., Benítez, D., Paillacho, V., Villalva, M., & Rojo-Álvarez, J. (2017). On the use of multi-class support vector machines for classification of seismic signals at Cotopaxi volcano. . *IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, 1-6.
- Lara-Cueva, R., Larco, J., Benítez, D., Pérez, N., Grijalva, F., & Ruíz, M. (2020). On Finding Possible Frequencies for Recognizing Microearthquakes at Cotopaxi Volcano: A Machine Learning based Approach. *JOURNAL OF VOLCANOLOGY AND GEOTHERMAL RESEARCH*.
- Lara-Cueva, R., Moreno, A., Larco, J., & Benítez, D. (2016). Real-time seismic event detection using voice activity detection techniques. *016d. Real-time seismic event detection using voice activity detection techniques. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Obtenido de <https://doi.org/10.1109/JSTARS.2016.2605061>
- Leire, V. (2014). Cinturón de Fuego del Pacífico. *BBC*. Obtenido de GeoEnciclopedia: https://www.bbc.com/mundo/noticias/2014/08/140826_ciencia_cinturon_fuego_pacifico_zona_mas_sismica_mundo_lv
- Martínez, J., Taboada, J., & Iglesias, C. (2012). METHODOLOGY FOR THE DETECTION OF MICROSEISMOLOGY IN A WATER-FILLED MINE HOLE. *Universidad de Vigo*, 1-9.
- Mathworks. (2020). *Máquinas de vectores de soporte para clasificación binaria*. Obtenido de Comprensión de las máquinas de vectores de soporte: <https://in.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html>
- MathWorks. (2020). *Técnica de aprendizaje automático para construir modelos predictivos a partir de datos de entrada y respuesta conocidos*. Obtenido de MathWorks: <https://in.mathworks.com/discovery/supervised-learning.html>
- Microsoft. (Octubre de 2020). *Microsoft*. Recuperado el Noviembre de 2020, de Selección de funciones basadas en filtros: <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/filter-based-feature-selection>
- Molina, J. P. (2018). *SEISMICKNOWLEDGE*. Obtenido de Cinturón de Fuego del Pacífico: <https://sites.ipleiria.pt/seismicknowledge/cinturon-de-fuego-del-pacifico/#:~:text=El%20cintur%C3%B3n%20o%20anillo%20de,en%20las%20zonas%20que%20abarca>.

- Mora, M., & Alvarado, G. (2001). *Red de sismología Nacional (UCR-ICE)*. Recuperado el 2020, de Priemr taller en actualización de sismología volcánica:
<https://rsn.ucr.ac.cr/documentos/educativos/vulcanología/5099-que-es-un-tremor>
- Orozco, M., Garcia, E., Duin, P., & Castelanos, G. (2006). *Dissimilarity-based classification of seismic signals at Nevado*. Obtenido de Earth Sci.:
<http://www.scielo.org.co/scielo.php?script=>
- Ortiz Herazo, H. (2013). Estudio de los Efectos de Sitio para la Construcción de un Índice de Actividad Sísmica en el volcán Cotopaxi. *Escuela Politécnica Nacional*.
- Otero, V. (2008). *Reconocimiento de localizaciones mediante máquina de soporte vectorial*. Madrid: Universidad Carlos III de Madrid.
- Pérez, A. (2018). Obtenido de Detección automática de eventos sísmicos en el volcán Cotopaxi mediante técnicas de Aprendizaje de Máquinas:
<https://repositorio.espe.edu.ec/bitstream/21000/13812/1/T-ESPE-057527.pdf>
- Pérez, N., Benítez, D., Grijalva, F., Lara, R., Ruiz, M., & Aguilar, J. (2019). Sísmica: Hacia un repositorio sísmico volcánico Ecuatoriano. *ELSEVIER*.
- Pérez, N., Venegas, P., Benítez, D., Lara-Cueva, R., & Ruiz, M. (2020). A new volcanic seismic signal descriptor and its application to a data set from the Cotopaxi volcano. *IEEE Trans. Geosci. Remote Sens*, 1-11. Obtenido de <https://doi.org/10.1109/TGRS.2020.2976896>
- Rodríguez, M. (2018). *A Real-Time Microseisms-Detector Based On Voice Activity Detection And Endpoint Detection: An Approach To Cotopaxi Volcano* . Sangolquí.
- Roman, D., & Cashman, K. V. (2018). Top-down precursory volcanic seismicity: Implications for 'stealth' magma ascent and long-term eruption forecasting. . *Front. Earth Sci*.
- Roman, V. (2019). *Ciencia y Datos*. Recuperado el 2020, de Aprendizaje Supervisado: Introducción a la Clasificación y Principales Algoritmos: <https://medium.com/datos-y-ciencia/aprendizaje-supervisado-introducci%C3%B3n-a-la-clasificaci%C3%B3n-y-principales-algoritmos-dadee99c9407>
- Rosero, K. (2019). *Adaptación e Implementación de Algoritmos Voice activity Detection para la detección de eventos microsismicos del volcán Cotopaxi*. Sangolquí: ESPE.
- Sposova, R. (2017). *Desarrollo y evaluación de métodos de selección de características para la predicción de eventos adversos en pacientes polimedcados*. Navarra, España: Universidad Pública de Navarra, E.T.S. de Ingeniería Industrial, Informática y de Telecomunicaciones.

- Tele13. (30 de Abril de 2015). Obtenido de Como son y donde están los 1555 volcanes del mundo: <https://www.t13.cl/noticia/mundo/donde-estan-y-como-son-los-1.555-volcanes-del-mundo>
- Tilling, R., & Punongbayan, R. (3 de Julio de 1993). Obtenido de Los Peligros Volcánicos: <http://cidbimena.desastres.hn/docum/crid/Volcanes/pdf/spa/doc4482/doc4482-contenido.pdf>
- Universidad de Costa Rica. (2011). *Tipos de onada Sísmica*. Obtenido de Laboratorio de Ingeniería Sísmica: <http://www.lis.ucr.ac.cr/7>
- Vaezi, Y., & Van der Baan, M. (2015). Comparison of the STA/LTA and power spectral density methods for microseismic event detection. *Geophys. J. Int.*, 1896-1908.
- Vásconez, F. (30 de Junio de 2020). Obtenido de Desarrollo de un sistema de clasificación de eventos sismo volcánicos usando Librerías de Machine Learning en Python: <http://repositorio.espe.edu.ec/jspui/bitstream/21000/22408/1/T-ESPE-043768.pdf>
- Venegas, P., Pérez, N., Benítez, D., Lara-Cueva, R., & Ruiz, M. (2019). Combining filter-based feature selection methods and gaussian mixture model for the classification of seismic events from Cotopaxi volcano. *EEE Journal of Selected Topics in Applied Earth*. Obtenido de <https://doi.org/10.1109/JSTARS.2019.2916045>
- Viracucha, E., & De la Batida, J. (2014). Sistema Informático para el Procesamiento y Análisis de Señales Sísmicas de Volcanes en el Ecuador. *EPN*.
- Zobin, V. (2017). Introduction to volcanic seismology. *Science Direct*. Obtenido de <https://doi.org/10.1016/C2015-0-00304-5>.