

ESCUELA POLITÉCNICA NACIONAL

**FACULTAD DE INGENIERÍA ELÉCTRICA Y
ELECTRÓNICA**

**ANÁLISIS COMPARATIVO DE GRUPOS DE CARACTERÍSTICAS
PARA LA CLASIFICACIÓN DE SEÑALES SÍSMICAS VT Y LP DEL
VOLCÁN COTOPAXI**

**TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO
REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERA EN
TELECOMUNICACIONES**

MARÍA JOSÉ PROAÑO PORTILLA
maria.proano02@epn.edu.ec

DIRECTOR: Ph.D. FELIPE LEONEL GRIJALVA ARÉVALO
felipe.grijalva@epn.edu.ec

Quito, febrero 2022

CERTIFICACIONES

Yo, MARÍA JOSÉ PROAÑO PORTILLA declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

MARÍA JOSÉ PROAÑO PORTILLA

Certifico que el presente trabajo de integración curricular fue desarrollado por MARÍA JOSÉ PROAÑO PORTILLA, bajo mi supervisión.

Ph.D. FELIPE LEONEL GRIJALVA ARÉVALO
DIRECTOR

DECLARACIÓN DE AUTORÍA

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como los productos resultantes del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

MARÍA JOSÉ PROAÑO PORTILLA

Ph.D. FELIPE LEONEL GRIJALVA ARÉVALO

DEDICATORIA

A mis padres *Alba* y *Carlos* por instruirme con amor y ampararme en cada obstáculo que me ha puesto la vida, sé que sin su esfuerzo nada de esto hubiese sido posible.

A mis hermanos *Karla* y *Jonathan* por brindarme su apoyo incondicional en cada paso, por sus consejos y sobretodo por mostrarme que la vida sin un hermano sería color gris.

A mi hija *Rafaella* por ser el motor que me impulsa cada nuevo día, por sostener mi vida con su sonrisa y por llenarme de verdadero amor.

Con amor

María José Proaño

AGRADECIMIENTOS

Agradezco al Dios Todopoderoso por darme fuerza en la debilidad, fortaleza en la dificultad, por sostener mi mano y susurrarme: ¡Tú puedes!, no te des por vencida.

A mis padres *Alba* y *Carlos* por ser el pilar fundamental en mi vida, por guiar mis pasos, por forjarme con sus valores y enseñanzas pero sobretodo por ser el más vivo ejemplo de responsabilidad, valentía y esfuerzo.

A mis hermanos *Karla* y *Jonathan* por llenar mi vida de complicidad y alegría, por enseñarme a ser mejor cada día, pero sobretodo por hacerme sentir orgullosa de poder tenerlos como hermanos.

A ti hija mía *Rafaella* por ser mi luz en la oscuridad, por acariciarme el alma con tu mirada, pero sobretodo por tener la dicha de poder escuchar esa vocecita diciendome: Mamá.

A mi tutor de tesis *PhD. Felipe Grijalva* por demostrar responsabilidad, paciencia, amor a la enseñanza y dedicar su tiempo en guiarme durante este proceso académico.

A mi compañero y gran amigo *Diego Ríos* por brindarme su amistad en las aulas, por impartir su conocimiento conmigo, por su total ayuda, pero sobretodo por hacerme comprender que en el mundo aún existen excelentes personas.

Gracias infinitas

María José Proaño

ÍNDICE DE CONTENIDO

CERTIFICACIONES.....	I
DECLARACIÓN DE AUTORÍA.....	II
DEDICATORIA.....	III
AGRADECIMIENTO.....	IV
ÍNDICE DE CONTENIDO.....	V
RESUMEN.....	VI
ABSTRACT.....	VIII
1. INTRODUCCIÓN.....	1
1.1. Objetivo general.....	2
1.2. Objetivos específicos.....	2
1.3. Alcance.....	2
1.4. Marco teórico.....	4
1.4.1. Sismicidad volcánica.....	4
1.4.2. Recolección y adquisición de datos en el volcán Cotopaxi.....	6
1.4.3. Aprendizaje automático supervisado.....	7
1.4.4. Máquinas de vector soporte.....	8
1.4.5. Validación cruzada k-fold.....	9
1.4.6. Método de Burg.....	10
1.4.7. Método de Welch.....	11
1.4.8. Análisis de componentes principales PCA.....	11
1.4.9. Métricas de evaluación.....	12
1.4.10. Incrustación de vecinos estocásticos distribuidos en t (t-SNE).....	14

2. METODOLOGÍA	16
2.1. Bases de datos.....	17
2.2. Preprocesamiento	17
2.3. Extracción de características	18
2.3.1. Método de Burg	18
2.3.2. Método de Welch.....	18
2.3.3. Base de datos <i>SeisBenchV1</i>	19
2.4. Reducción de dimensionalidad con PCA.....	22
2.5. Clasificador.....	22
2.6. Comparación	24
3. RESULTADOS, CONCLUSIONES Y RECOMENDACIONES.....	25
3.1. Resultados	25
3.2. Conclusiones.....	36
3.3. Recomendaciones.....	37
4. REFERENCIAS BIBLIOGRÁFICAS.....	38
ANEXOS	42

RESUMEN

El Ecuador posee varios volcanes en estado activo, de los cuales el volcán Cotopaxi ha sido considerado como el más peligroso, es por esto que el Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN) ha destinado una gran cantidad de recursos para monitorearlo permanentemente. Al tener una vigilancia constante, se recopilan grandes cantidades de datos a través de sensores, esta información debe ser sometida a etiquetación y clasificación del tipo de evento suscitado, por lo que esta ardua labor no debería ser ejecutada manualmente pues consumiría cuantiosos recursos y tiempo. Con esto surge la necesidad de implementar métodos de aprendizaje automático, los cuales permiten obtener resultados más rápidos y precisos.

En el presente trabajo se compara los siguientes grupos de características en la tarea de clasificación de señales sísmicas (VT y LP): características espectrales obtenidas directamente de los métodos de estimación de densidad espectral de potencia de Burg y Welch y las características de la base de datos *SeisBenchV1* documentadas en [1]. Ambos grupos de características alimentan un clasificador para posteriormente medir la exactitud con la que se desempeñan en esta tarea. Este análisis se llevó a cabo con y sin reducción de dimensionalidad, usando análisis de componentes principales (PCA). Los resultados demostraron que los datos más relevantes se consiguieron en la estación VC1 en comparación con la estación BREF, de lo que se puede dictaminar que la estación más desafiante a ser analizada es BREF, esto debido a que puede existir más presencia de ruido al momento de captar las señales.

PALABRAS CLAVE: Aprendizaje Automático, señales sísmicas, Cotopaxi, Burg, Welch, PCA.

ABSTRACT

Ecuador has several volcanoes in an active state, of which the Cotopaxi volcano has been considered the most dangerous, this is why that Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN) has allocated a large amount of resources to permanently monitor it. By having constant surveillance, large amounts of data are collected through sensors, this information must be subjected to labeling and classification of the type of event raised, so this arduous work should not be carried out manually as it would consume considerable resources and time. With this arises the need to implement machine learning methods, which allow faster and more accurate results.

In the present work, the following groups of characteristics are compared in the seismic signal classification task (VT and LP): spectral characteristics obtained directly from the Burg and Welch power spectral density estimation methods and the features of the base of *Seis-BenchV1* data documented in [1]. Both sets of features feed a classifier to later measure the accuracy with which it performs on this task. This analysis was carried out with and without dimensionality reduction, using principal component analysis (PCA). The results showed that the most relevant data were obtained in the VC1 station compared to the BREF station, from which it can be ruled that the most challenging station to be analyzed is BREF, this is due to the fact that there may be more presence of noise at the moment to pick up the signals.

KEYWORDS: Machine Learning, seismic signals, Cotopaxi, Burg, Welch, PCA.

1. INTRODUCCIÓN

Tomando en cuenta que alrededor del mundo existe una gran cantidad de volcanes en estado activo y millones de personas que habitan en lugares aledaños a estos, los volcanes simbolizan diversos tipos de peligros para esta población. Datos registrados a partir del año 1500 hasta cerca del año 2018, indican que el número de muertes por actividad de volcanes ascendió a una cifra de 280.000 aproximadamente [2]. Actualmente, unos 800 millones de individuos habitan dentro de una zona de alrededor de 100 kilómetros de distancia de un volcán, exponiéndose directamente al impacto mortal que pueden causar este tipo de eventos volcánicos [2]. Siendo de nuestro interés Ecuador, este se encuentra ubicado geográficamente en una zona de alta sismicidad volcánica. Por esta razón, en el año 1983 se conformó el Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN), pues este es el principal ente investigativo orientado al diagnóstico y vigilancia de peligros sísmicos y volcánicos que surjan [3].

Así, uno de los volcanes más estudiado y monitoreado permanentemente por el IGEPN es el volcán Cotopaxi, que se encuentra ubicado a una distancia de 35 km al Noreste de Latacunga y 45 km al Sureste de Quito, sobre la cordillera Oriental. Es considerado uno de los volcanes más peligrosos del mundo, debido a su estilo eruptivo, cobertura glaciar, relieve, erupciones frecuentes reportadas a lo largo de la historia y por la gran cantidad de población aledaña propensa a sus peligros y amenazas [4]. Hoy en día habitan alrededor de 300.000 personas cerca de zonas amenazadas en caso de erupciones. Adicionalmente, la caída de ceniza generada durante una erupción del volcán Cotopaxi, podría afectar a una gran parte de la región Sierra y Costa del Ecuador [4]. Es por esto que se ha destinado gran parte de recursos para la realización de un monitoreo ininterrumpido de este volcán. Por consiguiente, el análisis de señales receptadas de la actividad sísmica de un volcán es de vital importancia, pues mediante el estudio de las mismas, se podría detectar y predecir a tiempo posibles desastres naturales [1].

Por lo tanto este trabajo se desarrolla mediante una aplicación de aprendizaje automático, pues este además propone un enfoque orientado al análisis comparativo de características extraídas directamente de los métodos de estimación de densidad espectral de potencia de Burg y Welch para realizar esto se utilizó la base de datos *MigSigV1* proporcionada por el IGEPN, documentada en [1] y las características ya extraídas en un trabajo anterior [1], las mismas que son analizadas en dominio de tiempo, frecuencia y escala de la base de datos

SeisBenchV1, entregada de igual manera por el IGEPN . Ambos grupos de características alimentan un clasificador para posteriormente medir la exactitud con la que se desempeñan en esta tarea. Este análisis se llevó a cabo con y sin reducción de dimensionalidad, usando análisis de componentes principales (PCA).

1.1. Objetivo general

Analizar comparativamente grupos de características para la clasificación de señales sísmicas VT y LP del volcán Cotopaxi.

1.2. Objetivos específicos

- Conocer los conceptos básicos de sismicidad volcánica y aprendizaje automático.
- Implementar un clasificador para analizar la exactitud de clasificación orientado a diferentes características.
- Evaluar el desempeño del clasificador en la tarea de clasificación para las diferentes características implementadas.

1.3. Alcance

Como primer paso se realizará una investigación orientada a la adquisición de conocimientos de aprendizaje supervisado y sus aplicaciones, así como también al estudio de sismicidad volcánica. Siguiendo el objetivo general, en el cual se menciona la comparación entre grupos de características de señales sísmicas volcánicas. Las características anteriormente nombradas serán:

1. Las 84 características en tiempo, frecuencia y escala documentadas en [1].
2. Las características espectrales obtenidas directamente de los métodos de estimación de densidad espectral de potencia de Welch y Burg.

Este análisis comparativo se llevará a cabo con y sin reducción de dimensionalidad usando análisis de componentes principales (PCA).

Previamente se realizará el preprocesamiento de la base de datos, la cual consta de 1187 eventos registrados en dos estaciones de monitoreo, que son VC1 y BREF. A continuación se detalla las tareas a realizar en el preprocesamiento:

- Recorte de las señales.
- Normalización de la frecuencia de muestreo.
- Filtrado de las señales con filtro butterworth.
- Retirar la media y normalizar la amplitud al valor RMS.

Posterior a esto, se realizará la extracción de las diferentes características para las señales, las cuales serán sometidas al proceso de normalización. Una vez normalizadas las señales, se podrá realizar o no, reducción de dimensionalidad, usando análisis de componentes principales (PCA). Para luego ser enviadas al clasificador SVM, el cual retornará la exactitud en el trabajo de clasificación de etiquetas VT y LP, a través de una matriz de confusión y el ploteo de la curva ROC. De las cuales se obtendrán diferentes métricas, como son: exactitud, AUC y F1score. Es importante mencionar también que se evaluará en un esquema de k fold cross validation para las diferentes características y se realizarán test estadísticos.

1.4. Marco teórico

En el siguiente apartado se expondrá en la sección 1.4.1 conceptos ligados a sismicidad volcánica. También una breve explicación acerca de la adquisición y recopilación de datos en el Volcán Cotopaxi 1.4.2. Luego en la sección 1.4.3 se explicará acerca del aprendizaje automático supervisado. Seguidamente en la sección 1.4.4 se comenta acerca del clasificador SVM (Support Vector Machine), luego en la sección 1.4.5 , se detalla la técnica de validación cruzada k-fold. A continuación se definirán los conceptos para el método de Burg en la sección 1.4.6 y el método de Welch en la sección 1.4.7. Después en la sección 1.4.8 se muestra una explicación relacionada al análisis de componentes principales (PCA). Finalmente en la sección 1.4.9 se nombran las métricas de evaluación consideradas, además en la sección 1.4.10 se explica la definición y uso del t-SNE (Incrustación de vecinos estocásticos distribuidos en t).

1.4.1. Sismicidad volcánica

Se dice que la sismología de volcanes está enfocada en el estudio de terremotos de origen volcánico, donde un terremoto volcánico se define como un evento que ocurre en o cerca de un volcán [5]. Los volcanes producen una gran variedad de señales sísmicas que están relacionadas con el transporte de fluidos (agua, magma, gas) y su interacción con roca sólida, que por lo general ocurren en cantidades mayores antes o durante las erupciones. Los eventos más recurrentes que se asocian a las señales sísmicas son: Volcano-Tectónicos (VT, Volcano Tectonic Events), Período Largo (LP, Long Period Events), Híbridos (HB, Hybrid), Temblores (TR, Tremors), EXP quakes [6]. A continuación, se describen algunos de los eventos sísmicos más comunes e importantes, las cuales serán usadas en este estudio:

- **Volcano-Tectónicos (VP, Volcano Tectonic):** este tipo de eventos se generan debido a la presión de magma que fragmenta la roca sólida, posee frecuencias relativamente altas [7]. Considerado uno de los primeros signos de reactivación volcánica, su espectro de frecuencia puede alcanzar los 15 Hz [6].
- **Periodo Largo (LP, Long Period):** estos eventos están relacionados al movimiento de fluidos magma y/o gases, tienen baja frecuencia. El contenido espectral se limita a rangos de 0.5 a 5 Hz, mantienen una duración por debajo de los 90s. El inicio de un evento LP suele ser emergente, resultando difícil la determinación del momento exacto de su comienzo [8], [9].

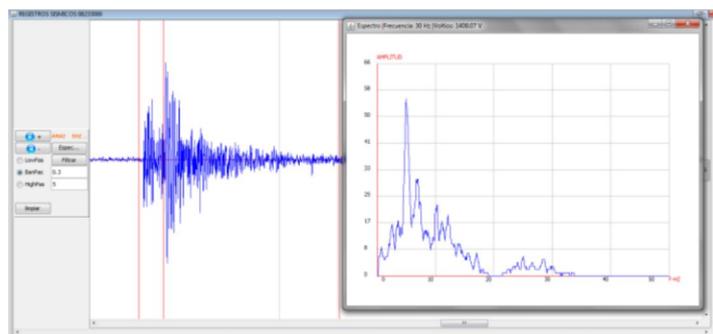


Figura 1.1: Evento volcánico-tectónico VT (señal sísmica y espectro con frecuencia >10 Hz).
Fuente: [8]

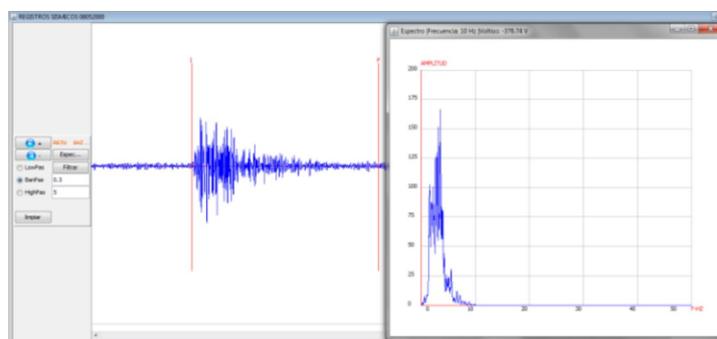


Figura 1.2: Evento periodo-largo LP (señal sísmica y espectro entre rango de 0 a 10 Hz).
Fuente: [8]

Cabe destacar que los tipos de eventos mencionados no son los únicos existentes, hay otras señales sísmicas que son frecuentes pero no se encuentran en todos los volcanes. Por otro lado, hoy en día se recopilan grandes volúmenes de datos sísmicos volcánicos, con el fin de dar seguimiento a largo plazo la actividad volcánica [6]. La información adquirida por los sensores (sismómetros) de las estaciones no es procesada, por lo que se requiere de un tratamiento y análisis minucioso con el fin de detectar y clasificar el tipo de evento. Estas tareas complejas son ejecutadas manualmente por especialistas en vulcanología, consumiendo gran cantidad de tiempo y recursos, a lo que se sumarían errores de observación. No es de asombro que los expertos modifiquen su comprensión de las señales con el tiempo, lo que haría que la interpretación de datos evolucione [6]. Es por esto que se considera de gran interés crear métodos más eficientes y prácticos, haciendo uso de la inteligencia artificial. De manera que la clasificación de señales sísmicas permitirá realizar la predicción de desastres naturales, con la intención de salvaguardar la seguridad pública y contrarrestar el impacto social y económico que se pueda generar.

1.4.2. Recolección y adquisición de datos en el volcán Cotopaxi

El Volcán Cotopaxi está ubicado en los Andes del Ecuador sus coordenadas son: latitud $0^{\circ}41'05''$ Sur y longitud $78^{\circ}25'54.8''$ Oeste. Es uno de los volcanes más activos y peligrosos en el Ecuador. Desde el año 1533 se han manifestado al menos 36 erupciones de este volcán. En los años 1534, 1742 1744, 1766-1768 y 1877 se reportaron grandes erupciones explosivas [10], la erupción más reciente ocurrió entre agosto y noviembre del 2015. Es por esto que el IGEPN ha desplegado una red de sensores destinada para la recopilación de información relevante, en esta se registran varios tipos de actividad sismo-volcánica [1]. Cabe destacar que esta red de alta resolución está conformada por seis estaciones sísmicas de banda ancha (BB) y seis estaciones sísmicas de período corto (SP) distribuidas permanentemente en los costados del volcán, ver figura 1.3. Además cuentan con tres sismómetros que interactúan en un rango de respuesta de frecuencia entre 0,03-50 Hz para estaciones banda ancha y rango de 1-50 Hz para estaciones de período corto.

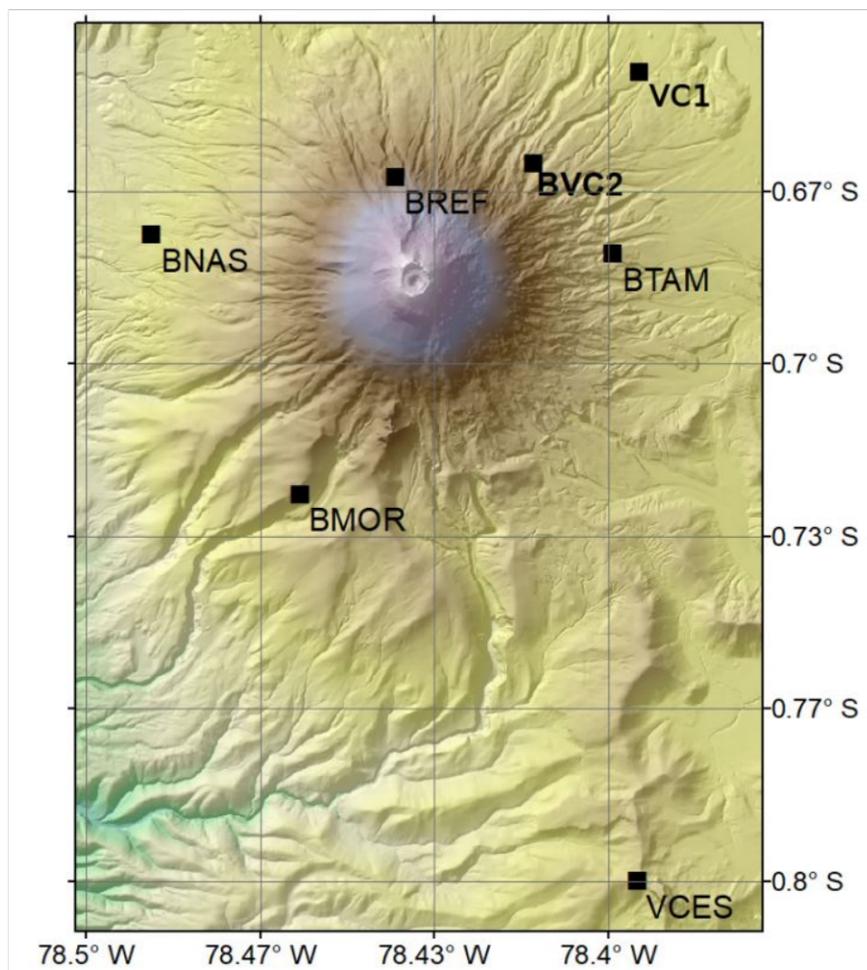


Figura 1.3: Algunas estaciones de la red sismológica del Volcán Cotopaxi. Fuente [1]

Es importante señalar que la base de datos proporcionada por el IGEPN, la cual es usada en este trabajo, recopiló señales sísmicas específicamente de VC1 (estación SP) y BREF (estación BB), las cuales trabajan a una frecuencia de muestreo de 100Hz y 50Hz respectivamente. Así, en el período de experimentación se recopilaron un total de 1187 eventos sísmicos en las estaciones VC1 y BREF, 1044 muestras de eventos de período largo (LP), 101 eventos de tipo volcano-tectónico, 8 muestras de eventos híbridos, 27 muestras regionales y 7 terremotos [1].

1.4.3. Aprendizaje automático supervisado

Una derivación del Aprendizaje Automático (*Machine Learning*), es el aprendizaje automático supervisado. El aprendizaje automático ha logrado un gran éxito en varias aplicaciones utilizando aprendizaje supervisado. Este construye modelos predictivos aprendiendo de un gran conjunto de datos de entrenamiento, cada ejemplo de entrenamiento está etiquetado es decir pertenece a un evento, objeto o clase. Un evento de entrenamiento se constituye de dos partes: un vector de características que describe el evento y una etiqueta que indica la salida, es decir los valores que queremos predecir [11]. Es importante mencionar que la salida que se etiqueta inicialmente para cada evento de entrenamiento se define como salida objetivo, mientras que la calculada a través del algoritmo de aprendizaje se denomina salida calculada. El aprendizaje supervisado tiene como finalidad minimizar la clasificación errónea entre la salida objetivo y la salida calculada [12].

A continuación se describe brevemente el proceso de aprendizaje que siguen los ejemplos de entrenamiento etiquetados en el algoritmo de aprendizaje supervisado. Para iniciar se debe recopilar el conjunto de datos, si un experto en el campo está disponible podría sugerir que atributos o características proporcionan más información. En caso de no ser así, la recolección de datos pasa a ser de "fuerza bruta", lo que indica que se podría medir todo lo disponible con la esperanza de que las características correctas tanto informativas como relevantes puedan aislarse. Sin embargo un conjunto de datos recolectado por este método, no sería adecuado porque en la mayoría de casos se podría encontrar ruido, por lo que sería necesario implementar un preprocesamiento [13]. Como segundo punto, los parámetros se inicializan al azar y se acondicionan los ejemplos de entrenamiento etiquetados. Luego se calculan los valores de salida y los parámetros se actualizan para reducir el error entre las salidas objetivo y salidas calculadas. Después esta sucesión de pasos se itera hasta minimizar el error o alcanzar convergencia en los parámetros. Finalmente los parámetros se optimizan para minimizar la tasa de clasificación errónea o el error [12].

El aprendizaje supervisado se subdivide en dos tipos: clasificación y regresión.

- **Clasificación:** es el proceso de asignar una o algunas de las clases predefinidas a cada evento. La clasificación más simple es conocida como clasificación binaria en la que cada elemento se clasifica en una de las dos categorías ya sea positiva o negativa. La clasificación binaria se puede expandir a una clasificación múltiple al establecer más clases [12]. Ver figura 1.4.
- **Regresión:** se define como el proceso de estimar un valor o valores de salida continuos para los datos de entrada. Existen dos tipos de regresión: la regresión univariante donde se estima un valor de salida y para la regresión multivariante se generan múltiples valores de salida para los datos de entrada [12]. Ver figura 1.4.

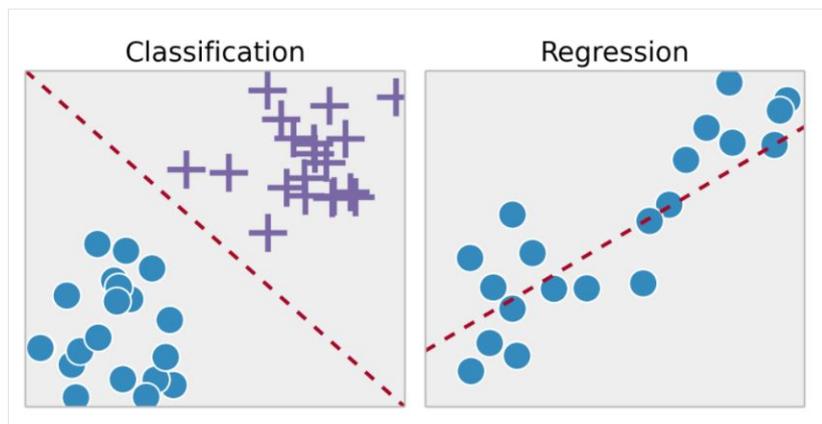


Figura 1.4: Clasificación y Regresión. Adaptada de [14]

1.4.4. Máquinas de vector soporte

Máquinas de vector soporte (Support Vector Machines) corresponden a la técnica de aprendizaje supervisado más reciente y popular. El concepto de las máquinas de vector soporte se liga a la idea de un "margen", es decir, SVM dibuja márgenes como límite entre las clases en el conjunto de datos proporcionado. Por otro lado, los puntos de datos que se encuentran cerca de los márgenes se conocen como puntos de vector de soporte y la solución se representa como combinación lineal de solo estos puntos. Cabe señalar que los márgenes se definen como la distancia entre dos vectores de apoyo separados por un hiperplano. Mientras mayor sea la distancia entre los vectores de soporte, mayor serán las posibilidades de que los puntos se clasifiquen correctamente y consecuentemente minimizar el error [13]. Ver figura 1.5

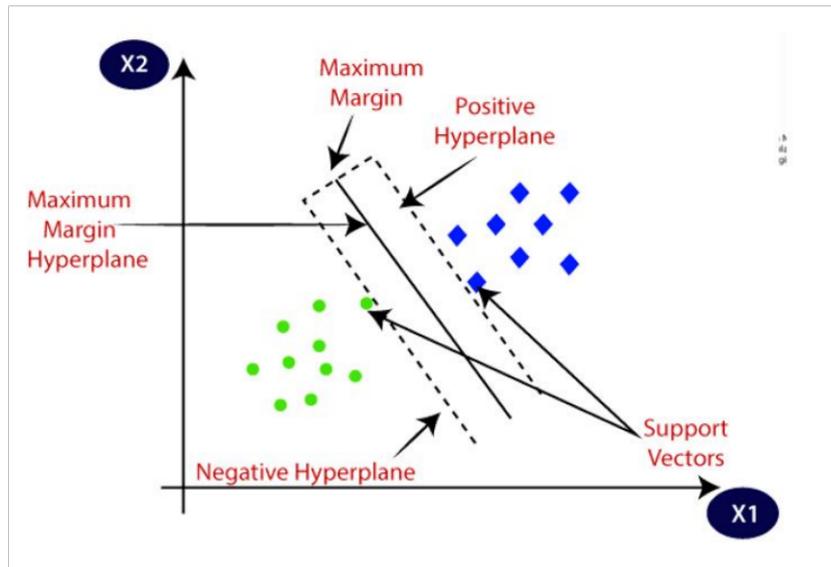


Figura 1.5: Visualización de características en SVM. Adaptada de [15]

Ahora bien se llama máquina de vector de soporte lineal, si todos los datos de entrenamiento están separados en sus categorías por un hiperplano óptimo en su propio espacio, cada ejemplo de entrenamiento se representa como un vector bidimensional [12]. Si los datos de entrenamiento no se pueden dividir linealmente, surge como solución máquinas de vectores de soporte no lineal. En este proceso se plantea el mapeo de los datos en un espacio de mayor dimensión y ahí poder definir un hiperplano de separación. A este espacio de dimensiones mayores se lo nombra como espacio de características. Pues con un espacio de características de suficiente dimensionalidad, cualquier conjunto de datos de entrenamiento puede convertirse en separable [13]. Con esto se introduce un concepto importante denominado kernels, son funciones que permiten que los productos internos sean calculados directamente en el espacio de características, sin necesidad de realizar el mapeo antes mencionado. Una vez creado el hiperplano, se usa kernel para mapear nuevos puntos en el espacio de características [13]. Finalmente se diría que la mayoría de problemas reales involucran datos no separables, donde no se podría fijar un hiperplano que separe con éxito las clases positivas de las negativas, con lo que se tendría que aplicar máquinas de vectores de soporte no lineal [12] [13].

1.4.5. Validación cruzada k-fold

La validación cruzada k-fold es un método estadístico usado para la evaluación predictiva del rendimiento de un modelo de aprendizaje automático. Empieza dividiendo los datos en k segmentos o pliegues iguales o casi iguales, de esto se elige un pliegue para probar y los pliegues k-1 restantes se usan para entrenar el modelo, por tanto el entrenamiento y pruebas

se realizan en k iteraciones. Para cada iteración se deja un pliegue para probar y se entrena el modelo con los k sobrantes, la precisión del modelo resulta del promedio de la precisión obtenida en cada repetición. Es importante mencionar que los datos se reorganizan antes de dividirse en k segmentos con el fin de que cada pliegue sea un buen representante del todo. Lo más decisivo en aplicar validación cruzada k -fold es la elección de la relación de división de datos para entrenamiento y prueba. A esta relación se la denota como tamaño de validación, un error al elegir el tamaño de validación podría reflejar un ajuste excesivo (overfitting) o insuficiente (underfitting) [16].

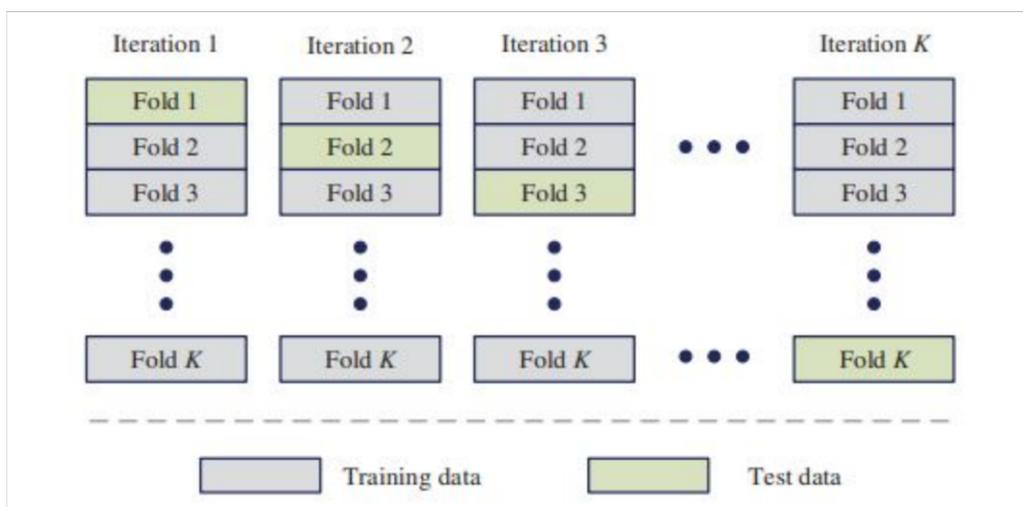


Figura 1.6: Método de validación cruzada k -fold. Fuente [17]

1.4.6. Método de Burg

Con el objetivo de detectar cambios en la potencia de la señal, es importante realizar una estimación de la densidad espectral de potencia PSD . Para esto se presenta la técnica de análisis espectral de máxima entropía de Burg, la cual ajusta un modelo de predicción lineal autorregresivo (AR) de orden especificado a la señal de entrada. Luego se enfoca en minimizar (usando mínimos cuadrados) la media aritmética de los errores de predicción hacia delante y hacia atrás, a continuación la densidad espectral se calcula en base a los coeficientes obtenidos en la respuesta del filtro autorregresivo (AR). Los parámetros del filtro cumplen con la recursividad de Levinson, esto para formar una técnica rápida y recursiva [18] [19] [20]. Burg es elegido por muchos debido a su alta precisión, convergencia más rápida para pequeñas señales y flexibilidad en la elección de diferente orden en el modelo autorregresivo (AR) [19]. Más información de AR en [21].

1.4.7. Método de Welch

El método de Welch permite obtener la estimación de la densidad espectral de potencia. Welch es un método no paramétrico que surgió al mejorar el método de Barlett, pues se obtienen mejores resultados en cuanto a precisión, reducción de varianza, reducción de ruido en los espectros de potencia estimados y también se obtiene una relación señal a ruido más alta, pero a cambio de estos beneficios se reduce la resolución de frecuencia. Para poder lograr esto cambia dos aspectos del anterior método, primero los segmentos de datos pueden superponerse y segundo se aplica una ventana a cada segmento de datos antes de calcular el periodograma [22], [23], [24].

El procedimiento que sigue es dividir los datos de la serie en segmentos sucesivos que posiblemente estarán superpuestos, luego calculará un periodograma modificado de cada segmento y promediará las estimaciones de densidad espectral de potencia, obteniendo la estimación de densidad espectral de potencia de Welch. El hecho de que se realice superposición entre segmentos tendería a introducir información redundante, pero esto se podría reducir a través de la utilización de ventanas, algunas ventanas son: Hamming, Hanning y Blackman. Además, la aplicación de ventanas espectrales conduce a una reducción del efecto fuga. Incluyendo esta herramienta al proceso, el método de Welch abre una ventana de los segmentos de datos antes de calcular el periodograma, este proceso de ventanas en los segmentos es lo que hace que Welch tenga un periodograma modificado. La figura 1.7 muestra el diagrama de flujo que sigue Welch [22], [23], [24].

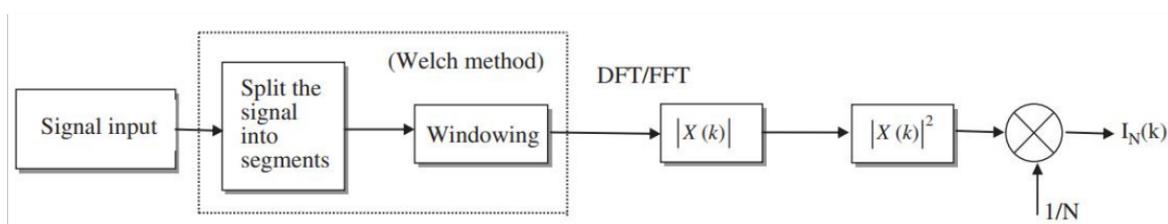


Figura 1.7: Diagrama de flujo del método Welch con ventana. Fuente [24]

1.4.8. Análisis de componentes principales PCA

Actualmente varias técnicas de medición recopilan muestras con muchas variables para ser analizadas, es por esto que en un conjunto de datos al obtener cientos o miles de variables se puede perder la pista de que variables se tiene exactamente. Con esta problemática, surge la reducción de dimensionalidad que permite comprimir datos, así como utilizar menos recursos computacionales y también permitirá acelerar los algoritmos de aprendizaje [25].

El análisis de componentes principales (PCA) es una técnica que analiza una tabla de datos, en la cual las muestras se describen a través de múltiples variables dependientes interrelacionadas. PCA tiene como fin extraer información importante donde se encuentra la máxima variación del conjunto de datos y expresa esta nueva información como un conjunto de variables ortogonales llamadas componentes principales [26]. El proceso para calcular las nuevas variables llamadas componentes principales involucra un estudio de álgebra lineal, pues estas se obtienen como combinaciones lineales de las variables originales. Es así como el primer componente principal posee la mayor varianza posible, el segundo componente se calcula bajo la restricción de ortogonalidad con el primer componente. Luego los demás componentes se calculan de la misma manera, los valores de estas nuevas variables se denominan puntajes de factoro [26]. Es importante mencionar que cuando el número de variables es mayor que el número de muestras, PCA puede reducir la dimensionalidad máximo al número de muestras, sin perder información [27].

PCA puede explicarse también de manera geométrica como las proyecciones de las observaciones sobre las componentes principales. Lo que PCA hace es encontrar una superficie de menor dimensión sobre la cual proyectar los datos. Esta superficie será elegida de manera que al calcular error cuadrático de proyección sea de valor mínimo [25] [27]. A continuación en la figura 1.8 se observa la aplicación de análisis de componentes principales bidimensional (2D) a una dimensión (1D).

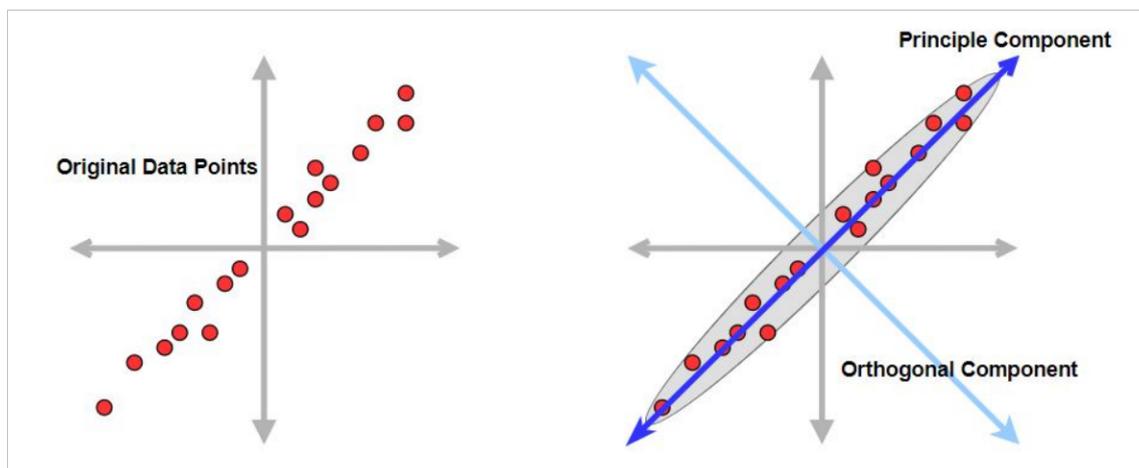


Figura 1.8: Aplicación de PCA de 2D a 1D. Fuente [28]

1.4.9. Métricas de evaluación

Con el fin de evaluar el rendimiento del clasificador es necesario la adquisición de métricas de evaluación. De cada experimento realizado se obtiene una matriz de confusión, la cual imprime diferentes valores, los cuales se muestran en la figura 1.9 [29].

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Figura 1.9: Matriz de confusión y sus valores. Fuente [30]

Donde las abreviaturas corresponden a:

- TP (true positives): número de verdaderos positivos clasificados por el modelo.
- TN (true negatives): número de verdaderos negativos clasificados por el modelo.
- FP (false positives): número de falsos positivos clasificados por el modelo.
- FN (false negatives): número de falsos negativos clasificados por el modelo.

Con estos valores se puede obtener el valor de la métrica accuracy (exactitud) y simbolizada con ACC, la cual mide el porcentaje de casos que el modelo ha clasificado correctamente. Es un valor entre 0 y 1, mientras más alto se dice que es mejor. Se la puede calcular con la siguiente fórmula:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1.1)$$

Otra métrica de evaluación obtenida a partir de la matriz de confusión es el F-score, que resulta de la siguiente ecuación:

$$F - score = \frac{TP}{TP * \frac{1}{2} (FP + FN)} \quad (1.2)$$

Esta métrica proporciona una puntuación única que equilibra los valores entre precisión y recuperación de los resultados obtenidos. En matemática es la media armónica de precisión y recuperación [30].

Por otro lado se tiene la gráfica de la curva característica de operación del receptor (ROC), nos indica que tan bien realizó el modelo la distinción entre dos umbrales. Es decir, representa la relación entre la tasa de verdaderos positivos TPR (sensibilidad) y la tasa de falsos positivos ($1 - \text{especificidad}$). Todos los valores positivos por encima del umbral serán verdaderos positivos, mientras que los valores negativos por encima del umbral serán falsos positivos, puesto que se predicen incorrectamente como positivos [31]. De aquí se obtiene otra métrica de evaluación denominada área bajo la curva (AUC), donde, un valor de AUC muy cercano a '1' infiere una clasificación exitosa, por lo contrario, un valor de AUC cercano a '0' refleja una clasificación errada [31] [32]. A continuación en la figura 1.10 se muestra las diferentes curvas ROC que se pueden obtener dependiendo de que tan bien se realizó la clasificación.

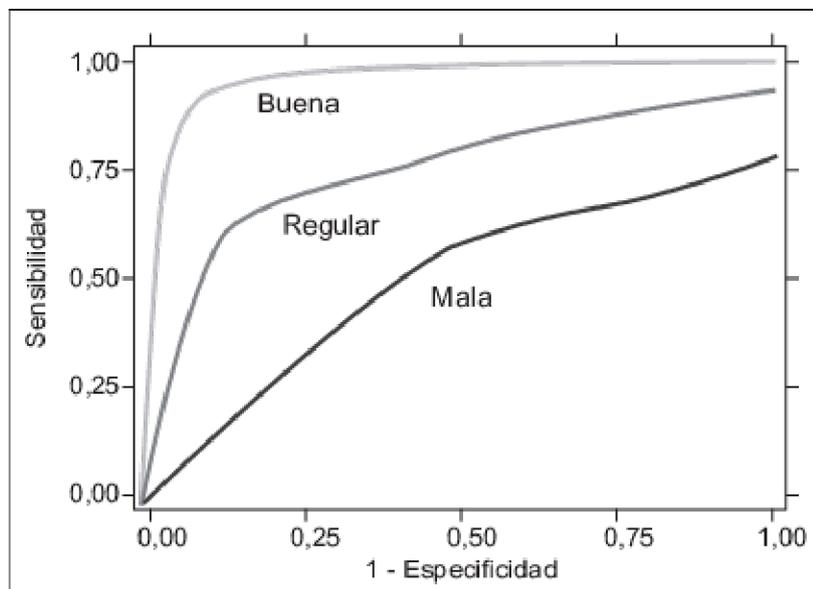


Figura 1.10: Diferentes resultados para curvas ROC. Fuente [33]

1.4.10. Incrustación de vecinos estocásticos distribuidos en t (t-SNE)

t-SNE es una técnica de visualización de datos de alta dimensión, tiene como finalidad situar a los datos similares resultantes en una ubicación específica de un plano 2D o 3D [34]. De igual manera los demás métodos de reducción de dimensionalidad se esfuerzan por tener en su mapeo la cercanía de muestras similares colocando las muestras diferentes a una distancia mayor, pero t-SNE es no lineal por lo que produce grupos visualmente más atractivos en relación a otros métodos [35]. En el ámbito del aprendizaje automático se ha vuelto popular puesto que tiene una capacidad casi increíble para crear mapas bidimensionales con alta fiabilidad a partir de grupos de datos con cientos o miles de dimensiones. Además de ser

flexible debido a que puede encontrar una estructura que para otras técnicas de reducción no es posible [36].

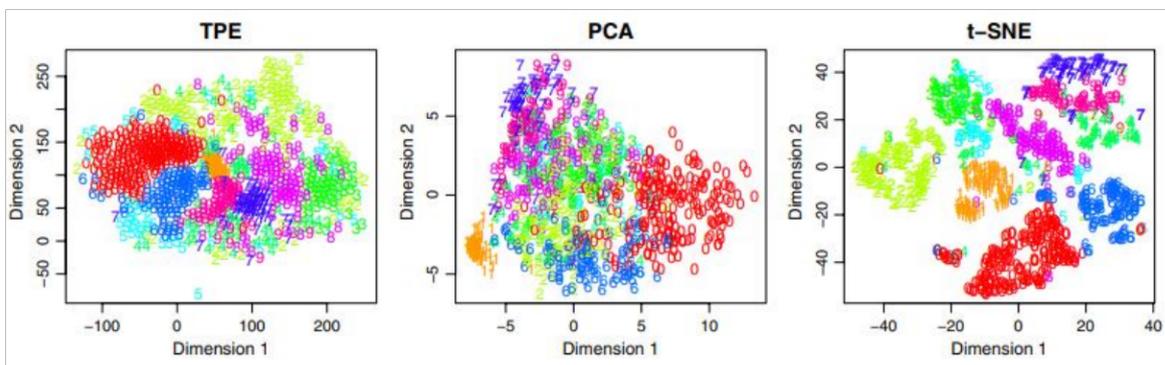


Figura 1.11: Visualización de dígitos escritos a mano, mostrados por diferentes técnicas TPE, PCA y t-SNE. Fuente [37]

2. METODOLOGÍA

En el siguiente capítulo se detalla los procedimientos que se implementaron para la realización del presente componente. En la sección 2.1 se expone de manera general información acerca de las bases de datos, las cuales contienen señales sísmicas volcánicas a ser analizadas. A continuación, en la sección 2.2 se explica el preprocesamiento y acondicionamiento de la base de datos, en la sección 2.3 se muestra la extracción de características realizada para los diferentes eventos sísmicos en estudio. Consecuentemente, en la sección 2.4 se detalla el uso de reducción de dimensionalidad a través de análisis de componentes principales (PCA). Más adelante, en la sección 2.5 se expone el desarrollo y parámetros considerados para el clasificador. Finalmente en la sección 2.6 se detalla la comparación entre las diferentes características a ser analizadas. En la siguiente figura 2.1 se muestran las diferentes etapas a ejecutar durante el proceso.

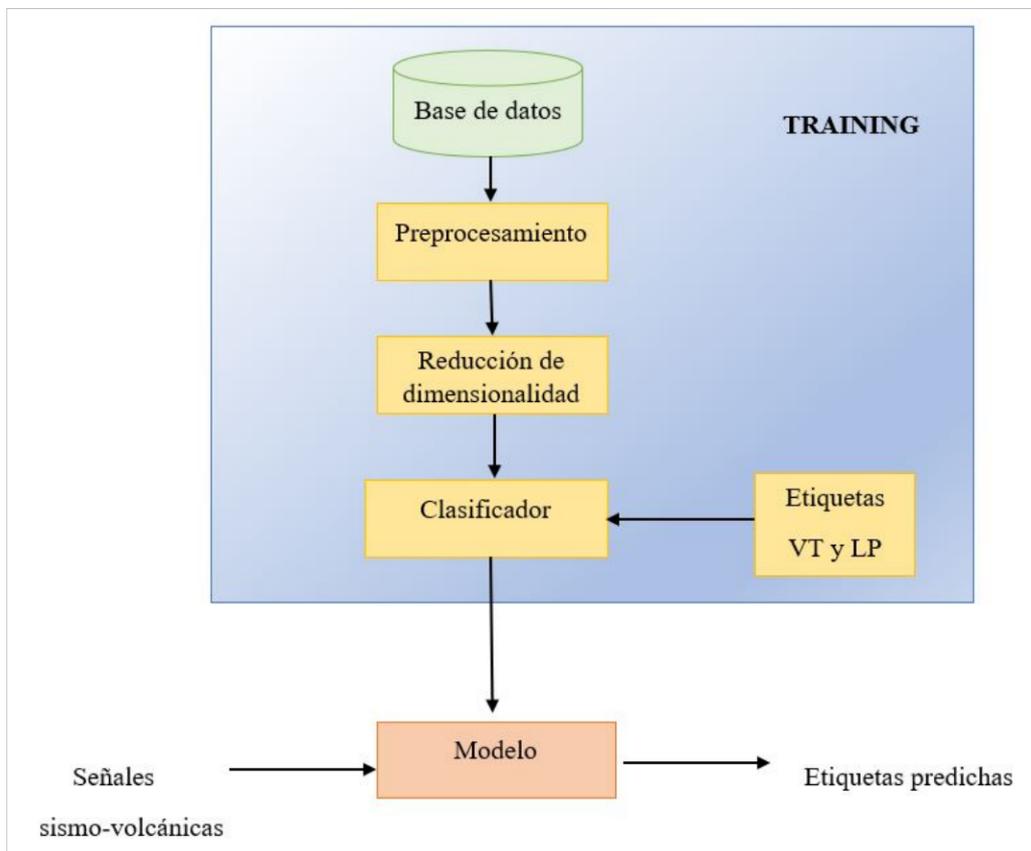


Figura 2.1: Diagrama de bloques de las etapas de este proceso.

2.1. Bases de datos

Las bases de datos utilizadas, fueron extraídas del primer repositorio público ecuatoriano de sismicidad volcánica denominado ESeismic, que contiene señales sísmicas del volcán Cotopaxi supervisadas por expertos [1]. Por un lado, tenemos el conjunto de datos *MicSigV1* que consta de 1187 eventos volcánicos sin procesar, recolectados de dos estaciones sísmicas, una de período corto establecida como VC1 y otra de banda ancha conocida como BREF [1]. Conviene destacar que los 1187 sucesos volcánicos se distribuyen de la siguiente manera: 1044 muestras para LP, 101 muestras VT, híbridos 8 muestras, 27 muestras para eventos regionales REG y 7 eventos de terremotos ICE. Por otra parte se utiliza el conjunto de datos *SeisBenchV1*, este es un repositorio público proporcionado por el Instituto Geofísico de la Escuela Politécnica Nacional, que contiene 84 características distintivas para la evaluación comparativa extraídas del conjunto de datos *MicSigV1*. Entre las 84 características encontramos: 13 características en el dominio del tiempo, 21 características en frecuencia y 50 características en el dominio de escala [1].

2.2. Preprocesamiento

Para preparar la base de datos *MicSigV1*, la cual contiene señales de tiempo sin procesar, se realiza un preprocesamiento antes de la extracción de características. El proceso seguido se detalla a continuación:

1. Se recortan las señales indicando punto de inicio y fin, seguido a esto, se suprimen valores NaN existentes en los eventos.
2. Se realiza normalización de las señales a una frecuencia de muestreo con un valor de 100 Hz.
3. Se aplica un filtro pasabajos Butterworth de orden 8, con frecuencia de corte $F_c=15$, con el fin de retener contenido espectral significativo de los eventos.
4. Se retira la media y se normaliza la amplitud al valor eficaz RMS de cada señal.

Después de la ejecución de estos pasos se tiene la señal sismo-volcánica acondicionada y se procederá a realizar la extracción de características, las cuales se detallan en la siguiente sección.

2.3. Extracción de características

Para la extracción de características se enfocó en el uso de métodos que involucren análisis de densidad espectral de potencia, como es el caso del método de Burg y Welch. Además de estos métodos, se utilizó la base de datos *SeisBenchV1*, la cual es una recopilación de 84 características ya obtenidas y listas para ser usadas.

2.3.1. Método de Burg

Involucra estimación de densidad espectral de potencia autorregresiva. Para implementarlo se utilizó la sintaxis establecida por Matlab, la cual se detalla a continuación:

```
pxx = pburg(x,order,nfft)
```

En el código 1 se muestra la implementación del método de Burg:

```
1 % Fft de las señales con pburg
2 % n=40
3 % nfft=1024
4     for j=1:f_senales
5         signals_pburg(j,:) = pburg(signals_normalized{j,:},40,1024);
6     end
7     [~,c_signals_pburg]=size(signals_pburg);
```

Listing 1: Implementación del método de Burg

Donde **order** es el orden del modelo autorregresivo establecido en 40, para **nfft** se utiliza 1024 puntos en la transformada discreta de Fourier (DFT). De lo cual se obtiene **pxx** que tiene una longitud $(nfft/2 + 1)$ [38].

2.3.2. Método de Welch

La estimación de densidad espectral de potencia de Welch, involucra un promedio de periodogramas a lo largo del tiempo. Matlab nos proporciona la siguiente sintaxis:

```
[pxx,f] = pwelch(x>window,noverlap,f,fs)
```

En el código 2 se indica la implementación del método de Welch:

```
1 % Fft de las señales con pwelch
2 % nfft=1024
3 % Fs=100
```

```

4   for j=1:f_senales
5       signals_pwelch(j,:) = pwelch(signals_normalized{j,:},hamming(1024),floor
        (0.5*1024),1024,Fs);
6   end
7   [~,c_signals_pwelch]=size(signals_pwelch);

```

Listing 2: Implementación del método de Welch

Para el indicador **window** se asignó con *hamming*(1024), así la señal se dividirá en segmentos de longitud de ventana. El parámetro **noverlap** representa el número de muestras superpuestas asignadas con un porcentaje del 50% de la longitud de la ventana. Luego para **f** que representa el valor de nfft se establecen 1024 puntos en la transformada discreta de Fourier (DFT) que se utilizarán en la estimación de la densidad espectral de potencia. Y para **fs** frecuencia de muestreo, $f_s = 100$, que representa el número de muestras por unidad de tiempo [39].

Luego de ejecutar estos métodos, se elimina la componente DC obtenida del paso previo realizado. Además, se expresa en *dB* la amplitud de cada señal. Es así como se obtiene una matriz de vectores con $C = 512$ columnas. Posteriormente se buscan valores infinitos en la matriz y se los sustituye con un valor de -100, con el objetivo de evitar errores y conflictos, más adelante, se realiza una normalización de tipo $z - score$ para estandarizar los datos.

2.3.3. Base de datos *SeisBenchV1*

Antes de empezar a describir esta base de datos, vale la pena destacar que el proceso de extracción de las 84 características ya fue realizado y documentado en [1]. Es decir este proceso de extracción es independiente de este proyecto, pues la base de datos *SeisBenchV1* fue proporcionada por el Instituto Geofísico de la Escuela Politécnica Nacional, pues es un repositorio público, que puede ser usado para los fines pertinentes.

Ahora bien *SeisBenchV1* abarca un conjunto de 84 características obtenidas de los mismos registros de eventos sísmológicos pertenecientes a la base de datos *MicSigV1*. Para el dominio de tiempo, se normalizaron las señales por su valor absoluto máximo, teniendo como resultado 13 características que van desde la f1 a f13 [1]. Mas adelante se calculan 21 características en dominio de frecuencia que van desde f14 a f34, usando estimación de densidad espectral de potencia con una ventana rectangular empleando 1024 puntos FFT. Todas las características en dominio de frecuencia se extrajeron del espectro de potencia de magnitud logarítmica en dB excepto f28 y f30 calculadas a partir del espectro de potencia de magnitud. Posteriormente, las características del dominio de escala se calcularon a

partir de una descomposición wavelet utilizando la familia db10 Daubechies hasta el nivel 6, formando 50 características (es decir, f35 a f84) [1]. Después de cargar esta base de datos, de la misma forma como para las características anteriores, se realiza una normalización *z - score* con el fin de estandarizar los datos.

A continuación se muestra la tabla 2.1 extraída de [1], donde se detallan cada una de las características anteriormente mencionadas, en sus diferentes dominios (tiempo, frecuencia, escala).

Tabla 2.1: Resumen de 84 características

Dominio del tiempo		Dominio de escala	
ID	Nombre de la característica	ID	Nombre de la característica
f1	Promedio	f42	Frecuencia A6 de pico máximo
f2	Desviación estándar	f43	Frecuencia D2 de pico máximo
f3	Varianza	f44	Frecuencia D3 de pico máximo
f4	Entropía	f45	Frecuencia D4 de pico máximo
f5	Kurtosis	f46	Frecuencia D5 de pico máximo
f6	Entropía multiescala	f47	Frecuencia D6 de pico máximo
f7	Tiempo para alcanzar el pico máximo	f48	Media A6 en el dom. de la frec.
f8	Valor RMS (media cuadrática)	f49	Media D1 en el dom. de la frec.
f9	Valor pico - pico	f50	Media D2 en el dom. de la frec.
f10	Relación pico - RMS	f51	Media D3 en el dom. de la frec.
f11	Energía	f52	Media D4 en el dom. de la frec.
f12	Tasa de cruce por cero	f53	Media D5 en el dom. de la frec.
f13	Densidad del pico encima de RMS	f54	Media D6 en el dom. de la frec.
Dominio de la frecuencia		f55	Energía media de los componentes
f14	Frecuencia del pico máximo	f56	Porcentaje de energía para A6
f15	Ancho de banda del 90 % de energía	f57	Porcentaje de energía para D1
f16	Entropía	f58	Porcentaje de energía para D2
f17	Promedio	f59	Porcentaje de energía para D3
f18	Desviación estándar	f60	Porcentaje de energía para D4
f19	Varianza	f61	Porcentaje de energía para D5
f20	Energía	f62	Porcentaje de energía para D6
f21	Kurtosis	f63	A6 RMS en el dom. del tiempo
f22	Entropía multiescala	f64	D1 RMS en el dom. del tiempo
f23	Pico máximo en la banda 10-20 Hz	f65	D2 RMS en el dom. del tiempo
f24	Frec. de pico máx. banda 10-20 Hz	f66	D3 RMS en el dom. del tiempo
f25	Pico máximo banda 20-30 Hz	f67	D4 RMS en el dom. del tiempo
f26	Frec. de pico máx. banda 20-30 Hz	f68	D5 RMS en el dom. del tiempo
f27	Valor RMS	f69	D6 RMS en el dom. del tiempo
f28	Relación pico - RMS	f70	A6 pico-pico en el dom. del tiempo
f29	Potencia	f71	D1 pico-pico en el dom. del tiempo
f30	Densidad del pico encima de RMS	f72	D2 pico-pico en el dom. del tiempo
f31	2do valor pico más alto	f73	D3 pico-pico en el dom. del tiempo
f32	Frec. del 2do pico más alto	f74	D4 pico-pico en el dom. del tiempo
f33	3er valor pico más alto	f75	D5 pico-pico en el dom. del tiempo
f34	Frec. del 3er pico más alto	f76	D6 pico-pico en el dom. del tiempo
Dominio de escala		f77	A6 relación pico-RMS dom. tiempo
f35	Pico máximo A6 dom. de la frec.	f78	D1 relación pico-RMS dom. tiempo
f36	Pico máximo D1 dom. de la frec.	f79	D2 relación pico-RMS dom. tiempo
f37	Pico máximo D2 dom. de la frec.	f80	D3 relación pico-RMS dom. tiempo
f38	Pico máximo D3 dom. de la frec.	f81	D4 relación pico-RMS dom. tiempo
f39	Pico máximo D4 dom. de la frec.	f82	D5 relación pico-RMS dom. tiempo
f40	Pico máximo D5 dom. de la frec.	f83	D6 relación pico-RMS dom. tiempo
f41	Pico máximo D6 dom. de la frec.	f84	Energía media coef. de ondículas

2.4. Reducción de dimensionalidad con PCA

Como parte del proceso se aplica reducción de dimensionalidad, enfocado en análisis de componentes principales (PCA), este método encuentra componentes principales que capturan la máxima cantidad de variación en los datos. Es por esto, que se puede guardar un valor de varianza alta usando una cantidad pequeña de componentes principales [40]. Para implementar reducción de dimensionalidad se utilizó una función propia del software Matlab, la misma que retorna una matriz que contiene: los coeficientes y las puntuaciones de las componentes principales obtenidas a partir del conjunto de datos proporcionados, así como el porcentaje retenido de varianza y la media estimada para cada una de las componentes. Para este caso se eligieron únicamente las componentes que llegan al 99% de la varianza. Es importante mencionar que el análisis se realizó con y sin reducción de dimensionalidad. A continuación se muestra el segmento de código 3 implementado para obtener PCA:

```
1 %% PCA
2 [X_nor_small_pca_coef, X_nor_small_pca_pre,~,~, X_nor_small_pca_varianza,
3 X_nor_small_pca_mu] = pca(X_nor_small);
4
5 % Se eligen los componentes que llegan al 99% de la varianza
6 sum_explained = 0;
7 numcomp = 0;
8 while sum_explained <= 99
9     numcomp = numcomp + 1;
10    sum_explained = sum_explained + X_nor_small_pca_varianza(numcomp);
11 end
12 X_nor_small_pca = X_nor_small_pca_pre(:,1:numcomp);
13 X_nor_small_pca_coef = X_nor_small_pca_coef(:,1:numcomp);
```

Listing 3: Implementación de PCA

2.5. Clasificador

Como paso previo para alimentar al clasificador se compone una tabla seleccionando solamente los datos y etiquetas de los eventos VT y LP, tanto para la estación VC1 como para BREF, cada etiqueta está relacionada tomando valores como el '1' y '0' respectivamente. En este caso se utiliza un clasificador máquinas de vector soporte de tipo lineal. Para la programación del mismo se utilizan funciones establecidas por el software Matlab.

A continuación en el segmento de código 4 se muestra las funciones que nos permiten implementar el clasificador:

```
1 %% Función kfoldds
2
3     k_folds = 5;
4     [~,c_labels] =size(dataset_input);
5     AllLabels = table2array(dataset_input(:,c_labels));
6     cv = cvpartition(AllLabels,'Kfold',k_folds);
7     table_labels_postprobs = [];
8
9 %% Se setean los valores para la optimización
10
11     VariableDescriptions = hyperparameters('fitcsvm',predictors,response);
12     VariableDescriptions(1).Range = [1e-4,1e6]; % Se modifica BoxConsraint
13     VariableDescriptions(2).Range = [1e-4,1e6]; % Se modifica KernelScale
14     VariableDescriptions(5).Optimize = true; % Se usa optimización de
15     Standardize
16 %% Predict para obtener LabelsPredicted y PostProbs
17
18     PosteriorSVMModel = fitPosterior(bstSVMmdl,predictors,response);
19     [PredictedLabels,PostProbs] = predict(PosteriorSVMModel,testData);
20     Truelabels = table2array(testData(:,c_labels));
```

Listing 4: Implementación de funciones para el clasificador SVM

- `c = cvpartition(group,'Kfold',k)`.- Permite la partición aleatoria de un grupo de datos, estableciendo conjuntos de prueba y entrenamiento, pues aplica validación cruzada (cross-validation) esto es para validar el modelo estadístico. Se asigna una valor de $k = 5$, el cual representa el número de pliegues en la partición, donde k debe ser menor que el número total de observaciones. Cada submuestra contiene aproximadamente el mismo número de observaciones y mismas proporciones de clase [41].
- `VariableDescriptions = hyperparameters(FitFcnName,predictors,response)`.- Retorna variables predeterminadas para la función de ajuste dada. El parámetro *FitFcnName* es el nombre de la función de ajuste, en este caso se estableció como *'fitcsvm'* la cual entrena un modelo de máquina de vectores de soporte (SVM) para la clasificación (binaria) de una y dos clases en un conjunto de datos predictores de baja o moderada

dimensión. Con esto se realiza optimización de parámetros de SVM lineal, para obtener una clasificación más acertada de las señales VT y LP [42].

- `fitPosterior(SVMModel,Name,Value)`.- Permite ajustar probabilidades posteriores para el clasificador de máquina de vectores de soporte (SVM) [43].
- `predict(Mdl,X)`.- Devuelve un vector de etiquetas de clase pronosticadas, donde '*Mdl*' es el modelo de clasificación entrenado y '*X*' son los predictores a clasificar, especificados como una matriz [44] .

Finalmente se obtienen las predicciones que son mostradas a través de la matriz de confusión, así como también al graficar la curva ROC.

2.6. Comparación

Para el presente proyecto se implementaron seis scripts en Matlab, cabe mencionar que cada script ejecuta todo el proceso detallado en la figura 2.1. Así, el primer script realiza la extracción de la densidad espectral de potencia con el método de Burg como primer característica, incluyendo reducción de dimensionalidad PCA. Para el segundo script se aplica el mismo método de Burg pero no se implementa reducción de dimensionalidad. En el caso del tercer script se obtiene la siguiente característica a través del método de Welch, añadiendo PCA. Por otro lado para el cuarto script se realiza la misma extracción mencionada con anterioridad, pero sin PCA. Por consiguiente, el quinto script utiliza la extracción de las 84 características impartidas por el repositorio público, añadiendo el análisis de reducción de componentes principales. Por último en el sexto script se emplean las 84 características sin usar PCA. Aquí es importante mencionar que se realiza el procedimiento establecido con cada método tanto para la estación VC1 como para BREF.

Con el fin de obtener resultados que sean estadísticamente significativos, es decir que es improbable que hayan resultado debido al azar, se ejecutan todos los scripts mencionados veinte veces, cambiando en cada repetición el valor de la semilla del generador de números aleatorios de control '*rng*' proporcionado por Matlab, puesto que se trabaja con las dos estaciones de manera independiente, se obtienen resultados para VC1 y BREF. Luego con los valores obtenidos de cada métrica se realizan test estadísticos, para cada par de métodos, teniendo en cuenta que un resultado menor a 0,05 nos indica una significancia estadística.

3. RESULTADOS, CONCLUSIONES Y RECOMENDACIONES

3.1. Resultados

La tabla 3.1 presenta los resultados que se obtuvieron al promediar los valores obtenidos en la ejecución de las 20 repeticiones para cada una de las métricas de evaluación, las cuales son: exactitud (ACC), AUC y F1 score, esto de todos los métodos propuestos y enfocados en la estación VC1. Además de los valores de promedio, se añadió el resultado del cálculo de la desviación estándar. Cabe mencionar que los símbolos: \bar{X} representa el promedio y σ simboliza la desviación estándar.

Tabla 3.1: Resultados de las métricas de evaluación para estación VC1

MÉTODO	EXACTITUD		AUC		F1 score	
	\bar{X}	σ	\bar{X}	σ	\bar{X}	σ
SeisBenchV1 con PCA	0.973750	0.012883	0.985581	0.016565	0.985254	0.007183
SeisBenchV1 sin PCA	0.975833	0.010611	0.981553	0.027797	0.986414	0.005950
Burg con PCA	0.972813	0.015680	0.980800	0.017034	0.984715	0.008737
Burg sin PCA	0.977708	0.012605	0.986953	0.014867	0.987384	0.007118
Welch con PCA	0.961250	0.017866	0.980933	0.016707	0.978201	0.010002
Welch sin PCA	0.968542	0.015280	0.984331	0.018231	0.982207	0.008647

Los resultados muestran que para la métrica *exactitud*, el mayor valor corresponde al método de Burg sin análisis de componentes principales (PCA), con un valor de $ACC = 0,977708$ y una $\sigma = 0,012605$. Con respecto al mejor valor obtenido de *AUC*, se mantiene en el mismo método, con un valor de $AUC = 0,986953$ y una $\sigma = 0,014867$. De la misma manera para la métrica *F1 score* el valor más alto se consigue en el método de Burg sin análisis de componentes principales (PCA), con valores de $F1score = 0,987384$ y una $\sigma = 0,007118$.

Para la estación BREF la tabla 3.2 muestra los resultados que se obtienen al promediar los valores adquiridos en la ejecución de las 20 repeticiones para cada una de las métricas de evaluación, las cuales son: exactitud (ACC), AUC y F1 score para todos los métodos planteados. Los resultados muestran que para la métrica *exactitud*, el mayor valor corresponde a las características de la base de datos *SeisBenchV1* con análisis de componentes principales (PCA), con un valor de $ACC = 0,953711$ y una $\sigma = 0,009820$. Con respecto al mejor valor obtenido de *AUC*, se mantiene en el mismo método, con un valor de $AUC = 0,945566$ y una $\sigma = 0,045978$. Asimismo para la métrica *F1 score* el valor más alto se consigue en *SeisBenchV1* con análisis de componentes principales (PCA), con valores de $F1score = 0,987384$ y una $\sigma = 0,007118$.

Tabla 3.2: Resultados de las métricas de evaluación para estación BREF

MÉTODO	EXACTITUD		AUC		F1 score	
	\bar{X}	σ	\bar{X}	σ	\bar{X}	σ
SeisBenchV1 con PCA	0.953711	0.009820	0.945566	0.045978	0.975296	0.005188
SeisBenchV1 sin PCA	0.949023	0.012759	0.919354	0.063112	0.972677	0.006974
Burg con PCA	0.871278	0.074103	0.821936	0.083214	0.926764	0.048391
Burg sin PCA	0.906466	0.058263	0.898063	0.044121	0.947822	0.036369
Welch con PCA	0.896767	0.061467	0.853229	0.072059	0.942489	0.037556
Welch sin PCA	0.932857	0.022367	0.914095	0.037613	0.963864	0.012451

También en base a los resultados se puede agregar que entre las estaciones bajo análisis que son VC1 y BREF, la estación VC1 presenta mejores valores de clasificación con respecto a la estación BREF, esto se puede ligar al hecho de que BREF es la estación más desafiante analizada, puesto que la mayoría de las señales sin procesar son mucho más ruidosas.

A continuación en las tablas 3.3, 3.4 y 3.5 se muestran los resultados obtenidos en los test estadísticos para cada métrica (exactitud, AUC, F1 score) respectivamente, esto para la estación VC1. Ahora bien, las columnas representan el primer argumento a ser considerado en la comparación y por lo tanto las filas serán el segundo argumento. Además se detallan las abreviaciones usadas para los distintos métodos aplicados: B CON PCA: Burg con PCA, B SIN PCA: Burg sin PCA, 84 CON PCA: 84 características *SeisBenchV1* documentadas en [1] con PCA, 84 SIN PCA: 84 características *SeisBenchV1* documentadas en [1] sin PCA, W CON PCA: Welch con PCA, W SIN PCA: Welch sin PCA.

Tabla 3.3: Resultados obtenidos de test estadísticos para métrica EXACTITUD en VC1

VC1-EXACTITUD-TTEST						
	B CON PCA	B SIN PCA	84 CON PCA	84 SIN PCA	W CON PCA	W SIN PCA
B CON PCA	-	0.010598	0.360855	0.122063	0.999660	0.931144
B SIN PCA	0.989402	-	0.904812	0.775770	0.999998	0.999582
84 CON PCA	0.639145	0.095188	-	0.085674	0.999881	0.971246
84 SIN PCA	0.877937	0.224230	0.914326	-	0.999934	0.994598
W CON PCA	0.00034	0.000002	0.000119	0.000066	-	0.000067
W SIN PCA	0.068856	0.000418	0.028754	0.005402	0.999933	-

* Los valores marcados en negrilla representan significancia estadística al ser menores a 0.05

Tabla 3.4: Resultados obtenidos de test estadísticos para métrica AUC en VC1

VC1-AUC-TTEST						
	B CON PCA	B SIN PCA	84 CON PCA	84 SIN PCA	W CON PCA	W SIN PCA
B CON PCA	-	0.039952	0.188984	0.462710	0.487799	0.203011
B SIN PCA	0.960048	-	0.605033	0.762373	0.903363	0.693895
84 CON PCA	0.811016	0.394967	-	0.775499	0.852262	0.585124
84 SIN PCA	0.537290	0.237627	0.224501	-	0.538670	0.362521
W CON PCA	0.512201	0.096637	0.147738	0.461330	-	0.214492
W SIN PCA	0.796989	0.306105	0.414876	0.637479	0.785508	-

* Los valores marcados en negrilla representan significancia estadística al ser menores a 0.05

Tabla 3.5: Resultados obtenidos de test estadísticos para F1 score en VC1

VC1-F1 SCORE-TTEST						
	B CON PCA	B SIN PCA	84 CON PCA	84 SIN PCA	W CON PCA	W SIN PCA
B CON PCA	-	0.011355	0.355072	0.118253	0.999706	0.940015
B SIN PCA	0.988645	-	0.899250	0.760020	0.999998	0.999627
84 CON PCA	0.644928	0.100750	-	0.084548	0.999901	0.975872
84 SIN PCA	0.881747	0.239980	0.915452	-	0.999943	0.995221
W CON PCA	0.000294	0.000002	0.000099	0.000057	-	0.000064
W SIN PCA	0.059985	0.000373	0.024128	0.004779	0.999936	-

* Los valores marcados en negrilla representan significancia estadística al ser menores a 0.05

Ahora examinaremos a detalle los valores estadísticamente significativos para cada métrica. En la columna $c = 2$ y fila $f = 7$, para la métrica *exactitud*, el método de Burg con PCA es estadísticamente significativo en comparación con el método de Welch con PCA. En el caso de la métrica *AUC*, el método de Burg con PCA no es estadísticamente significativo con respecto al método de Welch con PCA. Luego para la métrica *F1 score*, el método de Burg con PCA si es estadísticamente significativo en relación con el método de Welch con PCA. En consecuencia podemos decir que existe mayoría entre las diferentes métricas para poder concluir que el método de Burg con PCA es estadísticamente significativo en comparación con el método de Welch con PCA. A esto podemos correlacionar los valores obtenidos en la tabla 3.1 y observar que el método de Burg con PCA arroja mejores resultados en sus métricas en relación al método de Welch con PCA.

Prosiguiendo el análisis en la columna $c = 3$ y fila $f = 3$, para la métrica *exactitud*, el método de Burg sin PCA es estadísticamente significativo con respecto al método de Burg con PCA. Ahora con énfasis en la métrica *AUC*, el método de Burg sin PCA tiene significancia estadística referente al método de Burg con PCA. De igual modo para la métrica *F1 score*, el método de Burg sin PCA es estadísticamente significativo en relación con el método de Burg con PCA. Al tener unanimidad en los resultados, se puede concluir que el método de Burg sin PCA es estadísticamente significativo en comparación con el método de Burg con PCA. Añadiendo a esto en la tabla 3.1 se puede constatar que para las diferentes métricas en el método de Burg sin PCA se obtienen mejores resultados respecto al método de Burg con PCA.

Continuando en la columna $c = 3$ y fila $f = 7$, para la métrica *exactitud*, el método de Burg sin PCA tiene significancia estadística referente al método de Welch con PCA. Luego para la métrica *AUC*, el método de Burg sin PCA no es estadísticamente significativo con respecto al método de Welch con PCA. En el caso de la métrica *F1 score*, el método de Burg sin PCA tiene significancia estadística en relación al método de Welch con PCA. De lo que se

concluye que por resultados mayoritarios el método de Burg sin PCA es estadísticamente significativo en cuanto al método de Welch con PCA. Ahora relacionando esto con la tabla 3.1 se evidencia que para las métricas analizadas, los valores devueltos por el método de Burg sin PCA son mejores en relación al método de Welch con PCA. Así mismo se evidencia para el caso de la columna $c = 3$ y fila $f = 8$, la mayoría de resultados permiten concluir que el método de Burg sin PCA es estadísticamente significativo con respecto al método de Welch sin PCA, añadiendo el análisis de los resultados de la tabla 3.1 se constata que el método de Burg sin PCA entrega mejores valores que el método de Welch sin PCA.

Consideremos ahora la columna $c = 4$ y fila $f = 7$, para la métrica *exactitud*, la extracción de las 84 características de *SeisBenchV1* con PCA tiene significancia estadística en comparación con el método de Welch con PCA. En el caso de la métrica *AUC*, la extracción de las 84 características de *SeisBenchV1* con PCA no es estadísticamente significativa con respecto al método de Welch con PCA. Luego para la métrica *F1 score*, la extracción de las 84 características de *SeisBenchV1* con PCA tiene significancia estadística en relación con el método de Welch con PCA. En consecuencia podemos decir que existe mayoría entre las diferentes métricas para poder concluir que la extracción de las 84 características de *SeisBenchV1* con PCA es estadísticamente significativa en comparación con el método de Welch con PCA. A esto podemos correlacionar los valores obtenidos en la tabla 3.1 y observar que la extracción de las 84 características de *SeisBenchV1* con PCA arroja mejores resultados en las métricas en relación al método de Welch con PCA. De la misma forma en la columna $c = 4$ y fila $f = 8$, la mayoría de resultados permiten concluir que la extracción de las 84 características de *SeisBenchV1* con PCA tiene significancia estadística con respecto al método de Welch sin PCA, añadiendo el análisis de los resultados de la tabla 3.1 se constata que la extracción de las 84 características de *SeisBenchV1* con PCA entrega mejores valores que el método de Welch sin PCA.

A continuación en la columna $c = 5$ y fila $f = 7$, para la métrica *exactitud*, la extracción de las 84 características de *SeisBenchV1* sin PCA es estadísticamente significativa en comparación con el método de Welch con PCA. En el caso de la métrica *AUC*, la extracción de las 84 características de *SeisBenchV1* sin PCA no es estadísticamente significativa con respecto al método de Welch con PCA. Luego para la métrica *F1 score*, la extracción de las 84 características de *SeisBenchV1* sin PCA tiene significancia estadística en relación con el método de Welch con PCA. En conclusión podemos decir que existe mayoría entre las diferentes métricas para poder declarar que la extracción de las 84 características de *Seis-*

BenchV1 sin PCA es estadísticamente significativa en comparación con el método de Welch con PCA. A esto podemos correlacionar los valores obtenidos en la tabla 3.1 y observar que la extracción de las 84 características de *SeisBenchV1* sin PCA arroja mejores resultados en sus métricas en relación al método de Welch con PCA. Similarmente en la columna $c = 5$ y fila $f = 8$, la mayoría de resultados permiten concluir que la extracción de las 84 características de *SeisBenchV1* sin PCA tiene significancia estadística con respecto al método de Welch sin PCA, añadiendo el análisis de los resultados de la tabla 3.1 se constata que la extracción de las 84 características de *SeisBenchV1* sin PCA entrega mejores valores que el método de Welch sin PCA.

Finalmente en la columna $c = 7$ y fila $f = 7$, para la métrica *exactitud*, el método de Welch sin PCA tiene significancia estadística referente al método de Welch con PCA. Luego para la métrica *AUC*, el método de Welch sin PCA no es estadísticamente significativo con respecto al método de Welch con PCA. En el caso de la métrica *F1 score*, el método de Welch sin PCA tiene significancia estadística en relación al método de Welch con PCA. De lo que se concluye que por resultados mayoritarios el método de Welch sin PCA es estadísticamente significativo en cuanto al método de Welch con PCA. Luego relacionando esto con la tabla 3.1 se evidencia que para las métricas analizadas, los valores retornados por el método de Welch sin PCA son mejores con respecto al método de Welch con PCA.

Por otro lado para la estación BREF se exponen las tablas 3.6, 3.7 y 3.8 en las cuales se muestran los resultados obtenidos en los test estadísticos para cada métrica (*exactitud*, *AUC*, *F1 score*) respectivamente. Ahora bien, las columnas representan el primer argumento a ser considerado en la comparación y por lo tanto las filas serán el segundo argumento. Además se detallan las abreviaciones usadas para los distintos métodos aplicados: B CON PCA: Burg con PCA, B SIN PCA: Burg sin PCA, 84 CON PCA: 84 características *SeisBenchV1* documentadas en [1] con PCA, 84 SIN PCA: 84 características *SeisBenchV1* documentadas en [1] sin PCA, W CON PCA: Welch con PCA, W SIN PCA: Welch sin PCA.

Tabla 3.6: Resultados obtenidos de test estadísticos para EXACTITUD en BREF

BREF-EXACTITUD-TTEST						
	B CON PCA	B SIN PCA	84 CON PCA	84 SIN PCA	W CON PCA	W SIN PCA
B CON PCA	-	0.016342	0.000040	0.000143	0.067429	0.000670
B SIN PCA	0.983658	-	0.001392	0.003431	0.710120	0.040882
84 CON PCA	0.999960	0.998608	-	0.898658	0.999761	0.999445
84 SIN PCA	0.999857	0.996569	0.101342	-	0.999587	0.994441
W CON PCA	0.932571	0.289880	0.000239	0.000413	-	0.013957
W SIN PCA	0.999330	0.959118	0.000555	0.005559	0.986043	-

* Los valores marcados en negrilla representan significancia estadística al ser menores a 0.05

Tabla 3.7: Resultados obtenidos de test estadísticos para AUC en BREF

BREF-AUC-TTEST						
	B CON PCA	B SIN PCA	84 CON PCA	84 SIN PCA	W CON PCA	W SIN PCA
B CON PCA	-	0.000329	0.000045	0.000751	0.078088	0.000244
B SIN PCA	0.999671	-	0.001704	0.141603	0.992406	0.097795
84 CON PCA	0.999955	0.998296	-	0.910594	0.999733	0.988373
84 SIN PCA	0.999249	0.858397	0.089406	-	0.994020	0.638569
W CON PCA	0.921912	0.007594	0.000267	0.005980	-	0.003294
W SIN PCA	0.999756	0.902205	0.011627	0.361431	0.996706	-

* Los valores marcados en negrilla representan significancia estadística al ser menores a 0.05

Tabla 3.8: Resultados obtenidos de test estadísticos para F1 score en BREF

BREF-F1 SCORE-TTEST						
	B CON PCA	B SIN PCA	84 CON PCA	84 SIN PCA	W CON PCA	W SIN PCA
B CON PCA	-	0.019673	0.000127	0.000344	0.076083	0.001322
B SIN PCA	0.980327	-	0.002212	0.004961	0.688763	0.043034
84 CON PCA	0.999873	0.997788	-	0.904900	0.999543	0.999396
84 SIN PCA	0.999656	0.995039	0.095100	-	0.999293	0.993841
W CON PCA	0.923917	0.311237	0.000457	0.000707	-	0.015511
W SIN PCA	0.998678	0.956966	0.000604	0.006159	0.984489	-

* Los valores marcados en negrilla representan significancia estadística al ser menores a 0.05

Consideremos ahora los valores estadísticamente significativos para cada métrica. En la columna $c = 3$ y fila $f = 3$, para la métrica *exactitud*, el método de Burg sin PCA es estadísticamente significativo en comparación con el método de Burg con PCA. En el caso de la métrica *AUC*, el método de Burg sin PCA es estadísticamente significativo con respecto al método de Burg con PCA. Luego para la métrica *F1 score*, el método de Burg sin PCA tiene significancia estadística en relación al método de Burg con PCA. En consecuencia podemos decir que existe unanimidad entre las diferentes métricas para poder concluir que el método de Burg sin PCA tiene significancia estadística en comparación con el método de Burg con PCA. A esto podemos correlacionar los valores obtenidos en la tabla 3.2 y observar que el método de Burg sin PCA arroja mejores resultados en sus métricas en relación al método de Burg con PCA.

Siguiendo el análisis en la columna $c = 4$ y fila $f = 3$, para la métrica *exactitud*, la extracción de las 84 características de *SeisBenchV1* con PCA es estadísticamente significativa con respecto al método de Burg con PCA. Ahora con énfasis en la métrica *AUC*, la extracción de las 84 características de *SeisBenchV1* con PCA tiene significancia estadística referente al método de Burg con PCA. De igual modo para la métrica *F1 score*, la extracción de las 84 características de *SeisBenchV1* con PCA es estadísticamente significativa en relación con el método de Burg con PCA. Al tener concordancia en todos los resultados, se puede concluir que la extracción de las 84 características de *SeisBenchV1* con PCA es estadísti-

camente significativa en comparación con el método de Burg con PCA. Añadiendo a esto en la tabla 3.2 se puede constatar que para las diferentes métricas en la extracción de las 84 características de *SeisBenchV1* con PCA se obtienen mejores resultados respecto al método de Burg con PCA. Así mismo se evidencia para el caso de la columna $c = 4$ y fila $f = 4$, la mayoría de resultados permiten concluir que la extracción de las 84 características de *SeisBenchV1* con PCA es estadísticamente significativa con respecto al método de Burg con PCA, añadiendo el análisis de los resultados de la tabla 3.2 se evidencia que la extracción de las 84 características de *SeisBenchV1* con PCA entrega mejores valores que el método de Burg con PCA.

Continuando en la columna $c = 4$ y fila $f = 7$, para la métrica *exactitud*, la extracción de las 84 características de *SeisBenchV1* con PCA tiene significancia estadística referente al método de Welch con PCA. Luego para la métrica *AUC*, la extracción de las 84 características de *SeisBenchV1* con PCA es estadísticamente significativa con respecto al método de Welch con PCA. En el caso de la métrica *F1 score*, la extracción de las 84 características de *SeisBenchV1* con PCA tiene significancia estadística en relación al método de Welch con PCA. De lo que se concluye que por resultados unánimes la extracción de las 84 características de *SeisBenchV1* con PCA es estadísticamente significativa en cuanto al método de Welch con PCA. Ahora relacionando esto con la tabla 3.2 se evidencia que para las métricas analizadas, los valores devueltos por la extracción de las 84 características de *SeisBenchV1* con PCA son mejores en relación al método de Welch con PCA. Así mismo se evidencia para el caso de la columna $c = 4$ y fila $f = 8$, la mayoría de resultados permiten concluir que la extracción de las 84 características de *SeisBenchV1* con PCA es estadísticamente significativa con respecto al método de Welch sin PCA, añadiendo el análisis de los resultados de la tabla 3.2 se constata que la extracción de las 84 características de *SeisBenchV1* con PCA entrega mejores valores que el método de Welch sin PCA.

Ahora veamos la columna $c = 5$ y fila $f = 3$, para la métrica *exactitud*, la extracción de las 84 características de *SeisBenchV1* sin PCA tiene significancia estadística en comparación con el método de Burg con PCA. En el caso de la métrica *AUC*, la extracción de las 84 características de *SeisBenchV1* sin PCA es estadísticamente significativa con respecto al método de Burg con PCA. Luego para la métrica *F1 score*, la extracción de las 84 características de *SeisBenchV1* sin PCA tiene significancia estadística en relación con el método de Burg con PCA. En consecuencia podemos decir que existe coincidencia entre las diferentes métricas para poder concluir que la extracción de las 84 características de *SeisBenchV1* sin PCA

es estadísticamente significativa en comparación con el método de Burg con PCA. A esto podemos correlacionar los valores obtenidos en la tabla 3.2 y observar que la extracción de las 84 características de *SeisBenchV1* sin PCA arroja mejores resultados en las métricas en relación al método de Burg con PCA. De la misma forma en la columna $c = 5$ y fila $f = 4$, los resultados permiten deducir que la extracción de las 84 características de *SeisBenchV1* sin PCA tiene significancia estadística con respecto al método de Burg sin PCA, añadiendo el análisis de los resultados de la tabla 3.2 se constata que la extracción de las 84 características de *SeisBenchV1* sin PCA entrega mejores valores que el método de Burg sin PCA.

A continuación en la columna $c = 5$ y fila $f = 7$, para la métrica *exactitud*, la extracción de las 84 características de *SeisBenchV1* sin PCA es estadísticamente significativa en comparación con el método de Welch con PCA. En el caso de la métrica *AUC*, la extracción de las 84 características de *SeisBenchV1* sin PCA es estadísticamente significativa con respecto al método de Welch con PCA. Luego para la métrica *F1 score*, la extracción de las 84 características de *SeisBenchV1* sin PCA tiene significancia estadística en relación con el método de Welch con PCA. En conclusión podemos decir que existe unanimidad entre las diferentes métricas para poder declarar que la extracción de las 84 características de *SeisBenchV1* sin PCA es estadísticamente significativa en comparación con el método de Welch con PCA. A esto podemos correlacionar los valores obtenidos en la tabla 3.2 y observar que la extracción de las 84 características de *SeisBenchV1* sin PCA arroja mejores resultados en sus métricas en relación al método de Welch con PCA. Similarmente en la columna $c = 5$ y fila $f = 8$, la mayoría de resultados permiten concluir que la extracción de las 84 características de *SeisBenchV1* sin PCA tiene significancia estadística con respecto al método de Welch sin PCA, añadiendo el análisis de los resultados de la tabla 3.2 se constata que la extracción de las 84 características de *SeisBenchV1* sin PCA entrega mejores valores que el método de Welch sin PCA.

Luego en la columna $c = 7$ y fila $f = 3$, para la métrica *exactitud*, el método de Welch sin PCA tiene significancia estadística referente al método de Burg con PCA. Luego para la métrica *AUC*, el método de Welch sin PCA es estadísticamente significativo con respecto al método de Burg con PCA. En el caso de la métrica *F1 score*, el método de Welch sin PCA tiene significancia estadística en relación al método de Burg con PCA. De lo que se concluye que por resultados unánimes el método de Welch sin PCA es estadísticamente significativo en cuanto al método de Burg con PCA. Después relacionando esto con la tabla

3.2 se evidencia que para las métricas analizadas, los valores retornados por el método de Welch sin PCA son mejores con respecto al método de Burg con PCA.

De la misma manera en la columna $c = 7$ y fila $f = 4$, la mayoría de resultados permiten concluir que el método de Welch sin PCA tiene significancia estadística con respecto al método de Burg sin PCA, añadiendo el análisis de los resultados de la tabla 3.2 se constata que el método de Welch sin PCA entrega mejores valores que el método de Burg sin PCA.

En último lugar para la columna $c = 7$ y fila $f = 7$, para la métrica *exactitud*, el método de Welch sin PCA tiene significancia estadística referente al método de Welch con PCA. Luego para la métrica *AUC*, el método de Welch sin PCA es estadísticamente significativo con respecto al método de Welch con PCA. En el caso de la métrica *F1 score*, el método de Welch sin PCA tiene significancia estadística en relación al método de Welch con PCA. De lo que se concluye que por resultados coincidentes el método de Welch sin PCA es estadísticamente significativo en cuanto al método de Welch con PCA. Luego relacionando esto con la tabla 3.2 se evidencia que para las métricas analizadas, los valores retornados por el método de Welch sin PCA son mejores con respecto al método de Welch con PCA.

También se obtuvieron gráficas t-SNE, las mismas que permiten visualizar los datos similares en una ubicación específica del plano en este caso bidimensional. Estas gráficas se implementaron antes del ingreso al clasificador y se pudo observar las siguientes particularidades, las mismas se muestra en la figura 3.1 y figura 3.2.

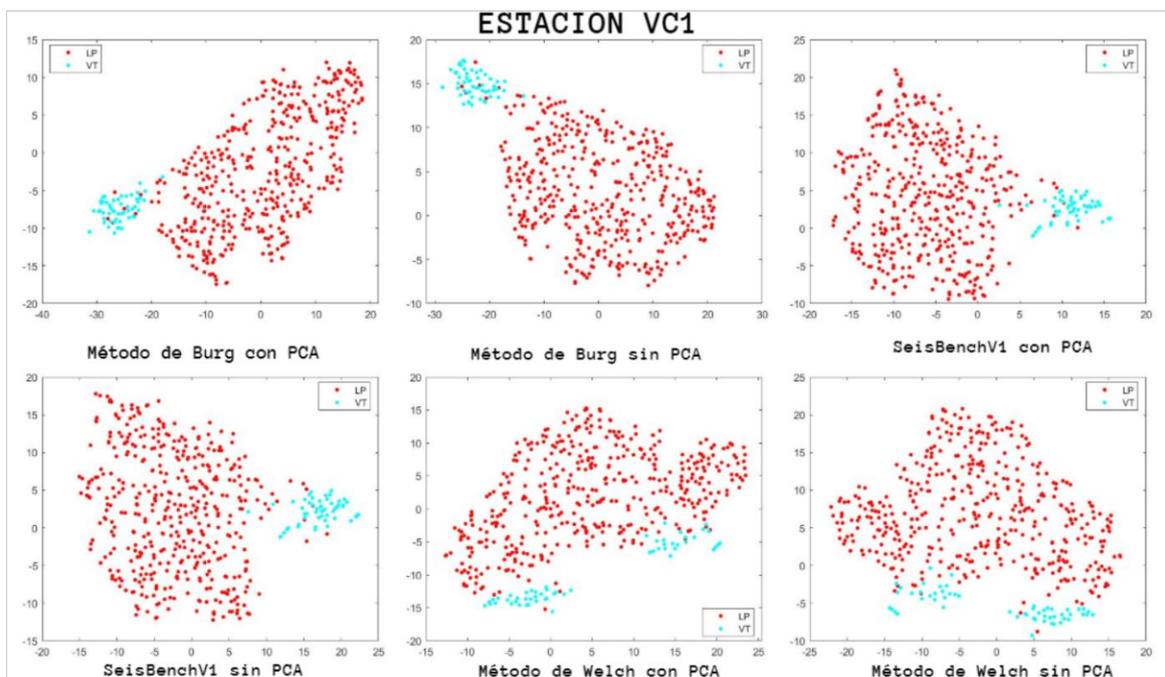


Figura 3.1: t-SNE para cada característica extraída en estación VC1

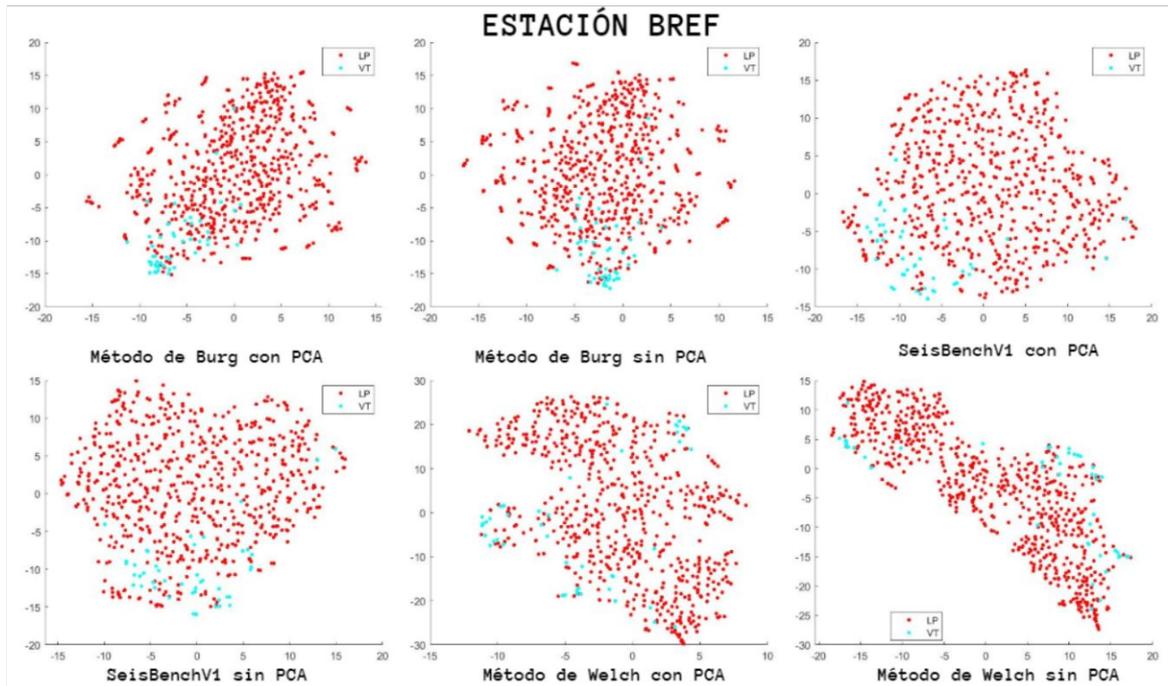


Figura 3.2: t-SNE para cada característica extraída en estación BREF

En la figura 3.1 se muestra las gráficas de los t-SNE ejecutados antes de ingresar al clasificador, los mismos que fueron realizados para cada característica extraída en la estación VC1. De igual manera en la figura 3.2 se observa el conjunto de gráficos t-SNE realizados antes del ingreso al clasificador, los mismos que fueron efectuados para cada característica extraída en la estación BREF. Se puede observar en las leyendas de cada gráfico, que los eventos VT se indican a través de las marcas color *cian*, por otro lado, para el caso de los eventos LP se establecen indicadores en color *rojo*. Ahora bien, se puede notar que en la estación BREF las muestras de los eventos VT Y LP están juntas, es decir, se dificulta separar unas de otras, esto debido a que la estación BREF es más ruidosa y por tanto incrementará el ruido en las señales.

A manera de síntesis se puede decir que en la estación VC1, los resultados obtenidos del promedio \bar{X} y desviación estándar σ de los experimentos realizados con $rng = [1 - 20]$ indicados en la tabla 3.1, la extracción de características que obtuvo los mejores resultados en todas sus métricas (exactitud, AUC y F1-score) corresponde al método de Burg sin análisis de componentes principales (PCA). Por otro lado para la estación BREF, los resultados obtenidos del promedio \bar{X} y desviación estándar σ de los experimentos realizados con $rng = [1 - 20]$ indicados en la tabla 3.2, muestran que los mejores resultados de todas las métricas (exactitud, AUC y F1-score) se alcanzó en la extracción de las las 84 características de *SeisBenchV1* con análisis de componentes principales (PCA).

Con respecto a los test estadísticos en la estación VC1 se presentaron nueve valores estadísticamente significativos ($< 0,05$) en la mayoría de las métricas evaluadas, en cambio en la estación BREF se presentaron doce valores estadísticamente significativos ($< 0,05$) en la mayoría de las métricas evaluadas, por tanto se podría decir que la estación BREF es más desafiante debido a que es una estación ruidosa en comparación con VC1. A esto se puede añadir que según las gráficas de los tSNE para VC1 (fig 3.1) y tSNE para BREF (fig3.2), en BREF las muestras de los eventos VT Y LP están juntas, es decir, se dificulta separar unas de otras, esto debido a que la estación BREF es más ruidosa y por tanto incrementará el ruido en las señales.

3.2. Conclusiones

Se realizó el análisis comparativo de un grupo de características, las cuales fueron extraídas de señales sismo-volcánicas registradas en el volcán Cotopaxi. Para la extracción de características se usó métodos de análisis de densidad espectral de potencia como son el método de Burg y el método de Welch, así como también las ochenta y cuatro características obtenidas de la base de datos *SeisBenchV1* [1]. Obteniendo como resultado que los mejores valores de las métricas usadas (exactitud = 0.977708, AUC= 0.986953 , F1-Score= 0.987384), luego de haber pasado por el clasificador, se obtuvieron con el método de Burg sin análisis de componentes principales (PCA) para la estación VC1. Lo que no sucede con la estación BREF pues se obtuvieron mejores resultados en las métricas (exactitud= 0.953711, AUC=0.945566, F1-score= 0.975296) luego de pasar por el clasificador con la extracción de las ochenta y cuatro características de *SeisBenchV1* con análisis de componentes principales (PCA). Siguiendo el hilo de esto se analizó que los resultados más relevantes obtenidos se consiguieron en la estación VC1 en comparación con la estación BREF, de lo que se puede inferir que la estación más desafiante a ser analizada es BREF, debido a que existe más presencia de ruido al momento de captar las señales. Por otro lado se implementó un clasificador de máquinas de vector soporte SVM lineal, el cual fue optimizado variando diferentes parámetros con el fin de obtener una clasificación más acertada de las señales VT y LP. Concluyendo este análisis es importante mencionar que en la actualidad en el campo de la vulcanología obtener un proceso automatizado para realizar esta importantísima actividad como es la clasificación de señales sísmo-volcánicas es de gran utilidad, puesto que permitiría ahorrar tiempo y recursos.

3.3. Recomendaciones

- En cuanto a la recopilación de datos enfocados en las estaciones fijadas en los alrededores del volcán Cotopaxi, sería considerable aplicar métodos (filtros) para contrarrestar el ruido que se adiciona en los datos sísmicos.
- Por otro lado es importante contar con una base de datos que contenga la cantidad de eventos de manera equilibrada con el fin de evitar resultados erróneos o sesgados debido a la falta de información.
- Se podría llevar este estudio al ámbito del aprendizaje automático no supervisado con el fin de obtener características a partir de datos que no contengan etiquetas, pues en la mayoría de análisis de ingeniería esta tarea de etiquetado y clasificación es ardua y extensa.
- El análisis en el dominio de la frecuencia permite obtener resultados y rasgos más importantes e interesantes de cualquier tipo de evento, es por esto que se debería buscar mejorar técnicas que involucren este tema y así poder aplicarles en varios campos.

4. REFERENCIAS BIBLIOGRÁFICAS

- [1] N. Pérez, D. Benítez, F. Grijalva, R. Lara-Cueva, M. Ruiz, and J. Aguilar, “Eseismic: Towards an ecuadorian volcano seismic repository,” *Journal of Volcanology and Geothermal Research*, 2020.
- [2] S. Brown, “¿ A cuánta gente matan realmente los volcanes?” <https://www.bbc.com/mundo/noticias-44274988>, May 2018. [Online]. Available: <https://www.bbc.com/mundo/noticias-44274988>
- [3] I. Geofísico, “Presentación - Instituto Geofísico - EPN,” <https://www.igepon.edu.ec/nosotros/content/25-nosotros>, Jun. 2014. [Online]. Available: <https://www.igepon.edu.ec/nosotros/content/25-nosotros>
- [4] S. Manosalvas, “Cotopaxi,” <https://www.igepon.edu.ec/cotopaxi/content/9-cotopaxi>, Jun. 2014. [Online]. Available: <https://www.igepon.edu.ec/cotopaxi/content/9-cotopaxi>
- [5] S. McNutt and D. Roman, *The Encyclopedia of Volcanoes (Second Edition)*. Academic Press, 2015.
- [6] M. Malfante, M. Dalla Mura, J. Metaxian, J. Mars, O. Macedo, and A. Inza, “Machine learning for volcano-seismic signals: Challenges and perspectives,” *IEEE Signal Processing Magazine*, vol. 35, pp. 20–30, 2018.
- [7] R. Trombley, *The Forecasting of Volcanic Eruptions*. iUniverse, 2006.
- [8] E. Viracucha and J. de la Bastida, “Sistema informático para el procesamiento y análisis de señales sísmicas de volcanes en el ecuador,” *Revista Politécnica*, vol. 33, ene 2014.
- [9] R. A. Lara-Cueva, D. S. Benítez, E. V. Carrera, M. Ruiz, and J. L. Rojo-Álvarez, “Automatic recognition of long period events from volcano tectonic earthquakes at cotopaxi volcano,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5247–5257, 2016.
- [10] M. Ruiz, B. Guillier, J.-L. Chatelain, H. Yepes, M. Hall, and P. Ramón, “Possible causes for the seismic cotopaxi volcano, ecuador,” *GEOPHYSICAL RESEARCH LETTERS*, vol. 25, no. 13, pp. 2305–230, 1998.
- [11] Z.-H. Zhou, “A brief introduction to weakly supervised learning,” *National Science Review*, 2017.

- [12] T. Jo, *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*. Springer Nature, 2021.
- [13] I. Maglogiannis, *Emerging Artificial Intelligence Applications in Computer Engineering*. IOS Press, 2007.
- [14] D. Soni, "Supervised vs. unsupervised learning," <https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d>.
- [15] E. García-Gonzalo, Z. Fernández-Muñiz, P. J. García Nieto, A. Bernardo Sánchez, and M. Menéndez Fernández, "Hard-rock stability analysis for span design in entry-type excavations with learning classifiers," *Materials*, vol. 9, no. 7, p. 531, 2016.
- [16] S. Yadav and S. Shukla, "Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification," in *2016 IEEE 6th International conference on advanced computing (IACC)*. IEEE, 2016, pp. 78–83.
- [17] M. Li, "Tectonic discrimination of olivine in basalt using data mining techniques based on major elements: a comparative study from multiple perspectives. big earth data, 10 (1), 3," 2019.
- [18] P. Vaitkus and R. Cobbold, "A comparative study and assessment of doppler ultrasound spectral estimation techniques part i: Estimation methods," *Ultrasound in medicine & biology*, vol. 14, no. 8, pp. 661–672, 1988.
- [19] V. Gupta and M. Mittal, "Efficient r-peak detection in electrocardiogram signal based on features extracted using hilbert transform and burg method," *Journal of the Institution of Engineers (India): Series B*, vol. 101, no. 1, pp. 23–34, 2020.
- [20] R.-v. Even, "pburg-psd using burg method."
- [21] K. Roth, I. Kauppinen, P. A. Esquef, and V. Valimaki, "Frequency warped burg's method for ar-modeling," in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No. 03TH8684)*. IEEE, 2003, pp. 5–8.
- [22] P. K. Rahi and R. Mehra, "Analysis of power spectrum estimation using welch method for various window techniques," *International Journal of Emerging Technologies and Engineering*, vol. 2, no. 6, pp. 106–109, 2014.

- [23] S. Villwock and M. Pacas, "Application of the welch-method for the identification of two- and three-mass-systems," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 1, pp. 457–466, 2008.
- [24] D.-J. Jwo, W.-Y. Chang, and I.-H. Wu, "Windowing techniques, the welch method for improvement of power spectrum estimation," *CMC-COMPUTERS MATERIALS & CONTINUA*, 2021.
- [25] A. Y.-T. Ng, "Principal component analysis algorithm," <https://www.coursera.org/lecture/machine-learning/principal-component-analysis-algorithm-ZYIPa>. [Online]. Available: <https://www.coursera.org/lecture/machine-learning/principal-component-analysis-algorithm-ZYIPa>
- [26] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [27] M. Ringnér, "What is principal component analysis?" *Nature biotechnology*, vol. 26, no. 3, pp. 303–304, 2008.
- [28] E. Blurock, "Pca: Principal component analysis | baeldung on computer science, how-published = <https://www.baeldung.com/cs/principal-component-analysis>," Oct. 2020. [Online]. Available: <https://www.baeldung.com/cs/principal-component-analysis>
- [29] J. M. Heras, "Precision, recall, f1, accuracy en clasificación," *IArtificial. net*, vol. 17, 2019.
- [30] J. Mohajon, "Confusion matrix for your multi-class machine learning model," *Towards Data Science*, 2020.
- [31] S. D. Walter, "The partial area under the summary roc curve," *Statistics in medicine*, vol. 24, no. 13, pp. 2025–2040, 2005.
- [32] L. Gonzalez, "Curvas roc y Área bajo la curva (auc), how-published = <https://aprendeia.com/curvas-roc-y-area-bajo-la-curva-auc-machine-learning/>," May 2019. [Online]. Available: <https://aprendeia.com/curvas-roc-y-area-bajo-la-curva-auc-machine-learning/>
- [33] M. E. BURGOS and C. MANTEROLA, "Cómo interpretar un artículo sobre pruebas diagnósticas," *Revista chilena de cirugía*, vol. 62, no. 3, pp. 301–308, 2010.
- [34] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

- [35] M. C. Cieslak, A. M. Castelfranco, V. Roncalli, P. H. Lenz, and D. K. Hartline, “t-distributed stochastic neighbor embedding (t-sne): A tool for eco-physiological transcriptomic analysis,” *Marine genomics*, vol. 51, p. 100723, 2020.
- [36] M. Wattenberg, F. Viégas, and I. Johnson, “How to use t-sne effectively,” *Distill*, vol. 1, no. 10, p. e2, 2016.
- [37] A. D. Shieh, T. B. Hashimoto, and E. M. Airoidi, “Tree preserving embedding,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 41, pp. 16 916–16 921, 2011.
- [38] “Autoregressive power spectral density estimate — burg’s method - matlab pburg - mathworks américa latina,” <https://la.mathworks.com/help/signal/ref/pburg.html>.
- [39] “Welch’s power spectral density estimate - matlab pwelch - mathworks américa latina,” <https://la.mathworks.com/help/signal/ref/pwelch.html>.
- [40] N. Sharma and K. Saroha, “Study of dimension reduction methodologies in data mining,” *International Conference on Computing, Communication Automation*, pp. 133–137, 2015.
- [41] “Partition data for cross-validation - matlab - mathworks américa latina,” <https://la.mathworks.com/help/stats/cvpartition.html>.
- [42] “Variable descriptions for optimizing a fit function - matlab hyperparameters - mathworks américa latina,” <https://la.mathworks.com/help/stats/hyperparameters.html>.
- [43] “Fit posterior probabilities for support vector machine (svm) classifier - matlab fitposterior - mathworks américa latina,” <https://la.mathworks.com/help/stats/classificationsvm.fitposterior.html>.
- [44] “Predict responses of linear regression model - matlab predict - mathworks américa latina,” <https://la.mathworks.com/help/stats/linearmodel.predict.html>.

ANEXOS

Para la elaboración del presente se uso el software Matlab R2021b.

ANEXO I. Código fuente

ORDEN DE EMPASTADO