



**Mejora del desempeño de un clasificador mediante la selección de características de las
señales de microsismos del volcán Cotopaxi**

Morales Zapata, Steve Alexander

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. Bernal Oñate, Carlos Paúl

16 de febrero del 2024

Certificado Antiplagio



Plagiarism and AI Content Detection Report

MoralesSteveTesisP.pdf

Scan details

Scan time:
February 15th, 2024 at 16:56 UTC

Total Pages:
54

Total Words:
13304

Plagiarism Detection



Types of plagiarism		Words
Identical	4.2%	559
Minor Changes	1.1%	146
Paraphrased	1.4%	180
Omitted Words	0%	0

AI Content Detection



Text coverage		Words
AI text	0%	0
Human text	100%	13304

[Learn more](#)

Plagiarism Results: (85)

Control Estadístico de Calidad y Seis Sigma 1.2%

<https://www.uv.mx/personal/ermeneses/files/2018/05/6-control-estadistico-de-la-calidad-y-seis-sigma-gutierr...>

Humberto Gutiérrez Pulido

CONTROL ESTADÍSTICO DE CALIDAD Y SEIS SIGMA CONTROL ESTADÍSTICO DE CALIDAD Y SEIS SIGMA Segunda edición Humberto Gutiérrez Pulido Centr...

PROBABILIDAD y ESTADÍSTICA Guía de Estudio 1%

<https://rdu.unc.edu.ar/bitstream/handle/11086/15055/guia%20pyes%202020%201er%20semestre.pdf?seque...>

mdimitroff

FACULTAD DE CIENCIAS EXACTAS, FÍSICAS Y NATURALES UNIVERSIDAD NACIONAL DE CÓRDOBA PROBABILIDAD y ESTADÍSTICA Guía de Estudio 2020 AUTOR...

52. Tratamiento estadístico de datos con aplicaciones en R -web-.pdf?seq... 1%

<https://repositoryinst.uniguajira.edu.co/bitstream/handle/uniguajira/309/52.%20tratamiento%20estadistico%...>

Luis Alberto Díaz Chávez Jairo Rafael Rosado Vega Tratamiento Estadístico de Datos con Aplicaciones en R ISBN 978-958-8942-58-2 9 78958...

ESPE_56bb69fa8f3cc30751367d90da9aac37 0.9%

https://rraae.cedia.edu.ec/record/espe_56bb69fa8f3cc30751367d90da9aac37

Skip to content My Title ...



Certified by
Copleaks

About this report
help.copleaks.com

copleaks.com
in f o t



Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Certificación

Certifico que el trabajo de titulación: **“Mejora del desempeño de un clasificador mediante la selección de características de las señales de microsismos del volcán Cotopaxi”** fue realizado por el señor **Morales Zapata, Steve Alexander**; el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 16 de febrero de 2024



Bernal Oñate, Carlos Paúl

C. C. 170977563-7



Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Responsabilidad de Autoría

Yo, **Morales Zapata, Steve Alexander**, con cédula de ciudadanía n°172234539-2, declaro que el contenido, ideas y criterios del trabajo de titulación: **Mejora del desempeño de un clasificador mediante la selección de características de las señales de microsismos del volcán Cotopaxi**, 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, 16 de febrero de 2024

Morales Zapata, Steve Alexander

C.C.: 172234539-2



Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Autorización de Publicación

Yo **Morales Zapata, Steve Alexander**, con cédula de ciudadanía n°172234539-2, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **Mejora del desempeño de un clasificador mediante la selección de características de las señales de microsismos del volcán Cotopaxi** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 16 de febrero de 2024

Morales Zapata, Steve Alexander

C.C.: 172234539-2

Dedicatoria

Dedico este trabajo a mi amada familia, Mayra, Angelita y Víctor, cuyo constante apoyo y amor incondicional han sido mi mayor fortaleza durante este arduo camino. Este logro no solo es mío, sino también de ustedes, quienes han estado siempre a mi lado, en constante apoyo.

Steve Alexander Morales Zapata

Agradecimiento

En primer lugar, agradezco a Dios por ser la guía en mi vida, y permitirme culminar esta etapa.

Agradezco profundamente a mi familia: Mayra, Angelita y Víctor, por su incondicional comprensión, sabios consejos y constante apoyo. Su presencia ha sido el pilar fundamental en mi vida, brindándome fortaleza y motivación en todo momento. Sin ustedes, este logro no habría sido posible. Gracias por estar siempre a mi lado.

A la Universidad de las Fuerzas Armadas ESPE y los profesores que la conforman, su compromiso y dedicación en compartir sus conocimientos han sido fundamentales en mi formación académica y personal.

Agradezco a Angie y Carlos V. por su constante apoyo y preocupación en la realización de este trabajo.

A mis amigos, les doy las gracias por estar siempre presentes y por los momentos compartidos que han hecho de esta etapa universitaria una experiencia inolvidable.

Agradezco a mi amigo y tutor Bernal Oñate Carlos Paúl, y cotutor Lara Cueva Román por su constante apoyo, orientación y motivación durante todo el proceso de este trabajo.

Steve Alexander Morales Zapata

Índice de Contenido

Dedicatoria.....	6
Agradecimiento.....	7
Resumen	17
Abstract.....	18
Capítulo I	19
Definición del Proyecto	19
Introducción.....	19
Antecedentes	19
Justificación e Importancia	22
Alcance	23
Objetivos	24
Objetivo General.....	24
Objetivos Específicos.....	24
Capítulo II	25
Marco Teórico.....	25
Volcán Cotopaxi	25
Eventos microsismos.....	25
Eventos Volcanotectónico (VT).....	25
Eventos Largo Periodo (LP).....	25
Eventos Híbridos (HB)	26

Rupturas de glaciar (IC)	26
Eventos Regionales (RG)	26
Escala ERB	26
Aprendizaje Automático.....	26
Aprendizaje supervisado.....	27
Características	27
Características lineales	28
Características psico acústicas	29
Algoritmos de clasificación	29
Autoencoder	29
L2WeightRegularization y sparsityRegularization.....	30
SparsityProportion	31
ScaleData	31
Matriz de confusión	31
Métricas de clasificación.....	32
K-nearest neighbor (KNN)	33
Decision Tree (DT)	33
Selección de características	34
Mutual Information (MI)	34
Statistical Dependency (SD)	35
Sequential Forward Selection (SFS)	36

	10
Sequential Forward Floating Selection (SFFS)	36
Capítulo III	37
Materiales y Métodos	37
Materiales.....	37
Metodología	37
Base de datos	38
Preprocesamiento	39
Procesamiento	40
Evaluación.....	41
Interfaz gráfica.....	41
Pestaña Ejecutable	41
Pestaña Informativo	44
Funcionamiento de la Interfaz.....	44
Capítulo IV	46
Pruebas y Resultados.....	46
Pruebas con Autoencoder Simple	46
Pruebas con KNN.....	50
Pruebas con Decision Tree	51
Pruebas de reducción de características	53
Pruebas con MI.....	53
Pruebas con SD.....	56

	11
Pruebas con SFS.....	59
Pruebas con SFFS.....	61
Prueba de pesos	63
Pruebas de interfaz gráfica.....	66
Fase 1.....	67
Fase 2.....	70
Fase 3.....	73
Resultados	78
Resultado de pruebas	78
Fase 1.....	79
Fase 2.....	79
Fase 3.....	80
Resultados de los clasificadores.....	81
Autoencoder	81
KNN	83
DT.....	85
Resultados de las mejores características.....	87
Autoencoder	87
KNN	87
DT.....	88
Comparación con trabajo previo.....	89

Capítulo V	91
Conclusiones y Recomendaciones	91
Conclusiones.....	91
Recomendaciones.....	91
Trabajos Futuros	92
Referencias.....	93

Índice de Tablas

Tabla 1 Características en el dominio del tiempo, de la frecuencia y de la escala	28
Tabla 2 Características psico acústicas	29
Tabla 3 Métricas de clasificación	32
Tabla 4 Información de la base de datos MicSigV1	38
Tabla 5 Prueba con parámetros con mejores resultados en (Minango, 2022)	46
Tabla 6 Prueba con los mejores resultados de (Minango, 2022)	49
Tabla 7 Prueba con los mejores parámetros obtenidos con las características psico acústicas	49
Tabla 8 Prueba KNN 13 vecinos.....	51
Tabla 9 Prueba KNN con 3 vecinos	51
Tabla 10 Prueba DT con 10 número máximo de divisiones	52
Tabla 11 Prueba DT con 3 número máximo de divisiones	53
Tabla 12 Prueba de selección de características con MI	54
Tabla 13 Prueba de Autoencoder con MI.....	54
Tabla 14 Prueba de KNN con MI	55
Tabla 15 Prueba de DT con MI.....	55
Tabla 16 Prueba de selección de características con SD	57
Tabla 17 Prueba de Autoencoder con SD.....	57
Tabla 18 Prueba KNN con SD	58
Tabla 19 Prueba DT con SD.....	58
Tabla 20 Prueba de selección de características con SFS	60
Tabla 21 Prueba de Autoencoder con SFS.....	60
Tabla 22 Prueba de KNN con SFS	61
Tabla 23 Prueba de DT con SFS	61

Tabla 24	Prueba de selección de características con SFFS	62
Tabla 25	Prueba de Autoencoder con SFFS	62
Tabla 26	Prueba de KNN con SFFS	62
Tabla 27	Prueba de DT con SFFS.....	63
Tabla 28	Características con peso mayor a 0.1 con MI	63
Tabla 29	Características con peso mayor a 0.8 con MI	64
Tabla 30	<i>Características con peso mayor a 0.8 con MI</i>	64
Tabla 31	Prueba de Autoencoder con 49 características seleccionadas con MI	65
Tabla 32	Prueba de Autoencoder con 21 características seleccionadas con MI	65
Tabla 33	Prueba de Autoencoder con 28 características seleccionadas con MI	66
Tabla 34	Mejores resultados de la Fase 1	79
Tabla 35	Mejores resultados de Fase 2.....	80
Tabla 36	Mejores resultados Fase 3.....	80
Tabla 37	Características con mejor resultado de Autoencoder	87
Tabla 38	Características con mejor resultado de KNN.....	88
Tabla 39	Características con mejor resultado de DT	89
Tabla 40	Resultado de ERB 256 bandas de (Minango, 2022)	90
Tabla 41	Mejores resultados obtenidos	90

Índice de Figuras

Figura 1 Descripción de aprendizaje supervisado	27
Figura 2 Matriz de confusión 3x3	31
Figura 3 Árbol de decisión diagrama	34
Figura 4 Diagrama de bloques que describe el proyecto.....	37
Figura 5 Preprocesamiento	39
Figura 6 Diagrama de flujo del preprocesamiento	39
Figura 7 Diagrama de bloques del procesamiento	40
Figura 8 Diagrama de bloques de Evaluación	41
Figura 9 Pestaña ejecutable.....	43
Figura 10 Pestaña informativo.....	43
Figura 11 Diagrama de flujo de interfaz gráfica parte 1	44
Figura 12 Diagrama de flujo de interfaz gráfica parte 2	45
Figura 13 Pruebas Realizadas al variar el número de neuronas y épocas	47
Figura 14 Prueba al variar las funciones de transferencia.	47
Figura 15 Prueba de porcentaje de acierto de combinaciones de L2WeightRegularization y SparsityRegularization	48
Figura 16 Relación entre el sparsity proportion y el porcentaje de acierto	48
Figura 17 Pruebas al variar el número de neuronas y épocas.....	50
Figura 18 Árbol de decisión con 10 número máximo de divisiones.....	52
Figura 19 Árbol de decisión con 3 número máximo de divisiones.....	53
Figura 20 Prueba de DT con MI	56
Figura 21 Prueba de DT con SD	59
Figura 22 Pruebas Fase 1 con Autoencoder	67
Figura 23 Pruebas Fase 1 con KNN.....	68

Figura 24	Pruebas Fase 1 con DT.....	69
Figura 25	Pruebas Fase 2 con Autoencoder	70
Figura 26	Pruebas Fase 2 con KNN.....	71
Figura 27	Pruebas Fase 2 con DT.....	72
Figura 28	Pruebas Fase 3 con Autoencoder y SD.....	73
Figura 29	Pruebas Fase 3 con Autoencoder y MI.....	74
Figura 30	Pruebas Fase 3 con KNN y SD	75
Figura 31	Pruebas Fase 3 con KNN y MI	76
Figura 32	Pruebas Fase 3 con DT y SD	77
Figura 33	Pruebas Fase 3 con DT y MI.....	78
Figura 34	Resultado con los mejores parámetros obtenidos de Autoencoder	81
Figura 35	Matriz de confusión de clasificación del Autoencoder.....	82
Figura 36	Resultado con los mejores parámetros obtenidos de KNN.....	83
Figura 37	Matriz de confusión de clasificación de KNN	84
Figura 38	Resultado con los mejores parámetros obtenidos de DT.....	85
Figura 39	Matriz de confusión de clasificación de DT	86
Figura 40	Árbol de decisión en la clasificación de DT.....	86

Resumen

Se realizó la mejora del desempeño de un clasificador mediante la selección de características de las señales de microsismos del volcán Cotopaxi, las que se identifican las mejores características mediante el empleo de escalas líneas y psicoacústicas, con métodos de selección de características como: Mutual Information (MI), Statistical Dependency (SD), Sequential Forward Selection (SFS), Sequential Forward Floating Selection (SFFS) con la clasificación de técnicas de aprendizaje supervisado como: K-Nearest Vecinos (KNN) y Decision Tree (DT); y no supervisado como Autoencoder. Se clasifican tres clases de eventos: largo período (LP), volcano tectónico (VT) u Otros, la cual contiene los eventos regionales (RG), híbridos (HB) y deslizamiento de glaciares (IC), estas señales, de la base de datos del volcán Cotopaxi del Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN).

Se implementó un sistema en el software MATLAB® en el cual se engloban todas las fases del proyecto, el cual permite el análisis de características y la evaluación del clasificador en el cual se visualizan los resultados de una manera detallada y gráfica. El mejor resultado de este clasificador es obtenido con el clasificador KNN y el selector MI, el cual tiene un porcentaje de clasificación de 95.77%, con una precisión del 95.84%, Especificidad de 91.81%, Sensibilidad de 95.65% y un Ber de 0.06, se tienen las mejores características: D2 Peak-to-peak in time-domain, D2 Max. peak in freq.-domain, D2 RMS in time-domain, Percentage of energy for D2, Density of peaks above RMS, Freq. of max. peak in 10–20 Hz Band y Spectral Decrease; las cuales son las que se muestran en los selectores.

Palabras clave: Escalas lineales, escalas psicoacústicas, selección de características, aprendizaje supervisado, aprendizaje no supervisado.

Abstract

The performance of a classifier was improved by selecting characteristics of the microseism signals of the Cotopaxi volcano, identifying the best characteristics through the use of line and psychoacoustic scales, with characteristic selection methods such as Mutual Information (MI), Statistical Dependency (SD), Sequential Forward Selection (SFS), Sequential Forward Floating Selection (SFFS), and being able to classify with supervised learning techniques such as: K-Nearest Ne These signals are classified into three types of events: long period (LP), volcano tectonic (VT), and Others, which include Regional Events (RG), Hybrid (HB), and glacial sliding (IC), based on data of Coptopaxi volcano database from the Instituto Geofísico de la Escuela Politécnica Nacional (IGEPN).

A system was developed using MATLAB® software that have all aspects of the project and allows for the study of attributes as well as the assessment of the classifier, with the findings visualized in a thorough and graphic manner. This classifier performs best when combined with the KNN classifier and the MI selector, with a classification percentage of 95.77%, accuracy of 95.84%, specificity of 91.81%, sensitivity of 95.65%, and Ber of 0.06. Having the Best Features: D2 Peak-to-peak in time-domain, D2 Max. peak in freq.-domain, D2 RMS in time-domain, Percentage of energy for D2, Density of peaks above RMS, Freq. of max. peak in 10–20 Hz Band and Spectral Decrease

Keywords: Linear scales, psychoacoustic scales, feature selection, supervised learning, unsupervised learning.

Capítulo I

Definición del Proyecto

Introducción

Debido a su potencial actividad, el volcán Cotopaxi, considerado uno de los más peligrosos de Ecuador, y está en constante monitoreo, esto depende del análisis de las señales, de los diferentes fenómenos físicos, químicos, etc. emitidos por este volcán; es esencial comprender estas señales de manera automatizada. El aprendizaje automático supervisado, fue utilizado para poder realizar este proyecto. Los métodos de selección de características son procesos cruciales que determinan qué partes del conjunto de datos son más relevantes para el análisis.

La investigación utiliza un método experimental cuantitativo. El objetivo principal es determinar las técnicas de selección de características más efectivas a través de un análisis comparativo. Se busca no solo determinar cuáles son las mejores características, sino también comprender cómo varios enfoques de selección de características afectan el rendimiento de los clasificadores mediante el software MATLAB®, el cual se utiliza principalmente para evaluar el desempeño, además que nos permite las herramientas para la implementación del sistema, además de observar de manera detallada y gráfica.

Esta investigación tiene como objetivo mejorar la capacidad de comprensión de las señales sísmicas del volcán Cotopaxi con el uso del aprendizaje automático y la selección de las mejores características.

Antecedentes

Según (R. A. Lara-Cueva, Benítez, et al., 2016), propone una nueva forma de caracterizar los eventos sísmicos volcánicos que se basa en estimadores espectrales y de máxima entropía convencionales. El análisis se basó en los datos de los eventos sísmicos del

volcán Cotopaxi; mediante el análisis de su densidad espectral de potencia obtuvo que los eventos de Largo periodo (LP) se produjeron en frecuencias inferiores a 10 Hz después de detectar el evento, los eventos Vulcano Tectónicos (VT) se concentraron entre 0 y 20 Hz, y los eventos producidos por rayos u otros fenómenos se produjeron entre 15 y 35 Hz. El detector diseñado tiene una sensibilidad del 93.8%, lo que permite detectar en tiempo real los lugares de eventos volcánicos específicos.

Un sistema de detección automática de alta confiabilidad en tiempo real, realizó (R. A. Lara-Cueva, Benítez, et al., 2016) el cual es un sistema de reconocimiento automático de sismicidad volcánica que utiliza la extracción de características, la clasificación de eventos y la detección de eventos para reducir el tiempo de procesamiento. Se comparó dos técnicas de selección de características comunes: información mutua y dependencia estadística, cada una de las cuales utiliza algoritmos de clasificación apropiados y apropiados, como k Vecinos más cercanos (k-NN) y Árboles de decisión (DT). Para el análisis se utilizó de señales de actividad sísmica registrada en el Volcán Cotopaxi en Ecuador entre 2009 y 2010, con estas señales se analizó con una ventana de segmentación de 15 segundos, una matriz de características en el dominio de frecuencia y un clasificador DT, que dieron los mejores resultados, con una sensibilidad y precisión de detección del 99%.

En (R. A. Lara-Cueva, Moreno, et al., 2016), se propone la implementación de detectores de eventos LP y VT en tiempo real basados en algoritmos de detección de actividad de voz. La principal ventaja de este detector es que puede localizar los puntos finales de un evento sísmico. Para las pruebas, se utilizó una base de datos que contiene 436 eventos sísmicos (LP y VT) registrados en el volcán Cotopaxi, los resultados mostraron un porcentaje de acierto del 95.2 % y una BER de 0.005, lo que indica un rendimiento comparable al de los algoritmos de detección de eventos anteriormente creados para el mismo conjunto de datos, pero con mucha menos complejidad computacional.

En (R. Lara-Cueva et al., 2017), presenta un sistema automático que reconoce señales de microsismos, como eventos LP y VT, así como señales de origen no volcánico, como relámpagos y ruido de fondo basándose en algoritmos de aprendizaje automático. Se empleó una base de datos de registros sísmicos recopilados por una red sísmica establecida en el volcán Cotopaxi. El método toma en cuenta la energía de los coeficientes de la transformada wavelet como una característica clave para diferenciar eventos en sismogramas volcánicos. La etapa de detección pudo detectar eventos con ruido de fondo con una precisión del 98%, mientras que la etapa de clasificación pudo detectar eventos con ruido de fondo con una precisión del 90%.

Los eventos sísmicos LP y VT son los más importantes a seguir entre los diferentes eventos sísmicos registrados, ya que su aumento de ocurrencia puede ayudar a pronosticar posibles erupciones, mostrados en (Venegas et al., 2019). El valor de salida de una función de predicción después de la capacitación del clasificador con múltiples pares de entrada y salida válidos es un problema de separación de datos en diferentes clases. El problema que analiza en este trabajo tiene dos valores de salida discretos (clases): eventos sísmicos LP y VT. Como resultado, se puede representar como un problema de clasificación de dos clases. El mejor modelo de clasificación fue el clasificador árbol de decisión con 100 predictores. La prueba estadística de Wilcoxon (Divine et al., 2013) demostró que la estrategia sugerida fue efectiva para proporcionar modelos competitivos para la clasificación de eventos sísmicos volcánicos. En general, los modelos de clasificación obtuvieron puntuaciones de 0.92 a 0.94, que eran puntuaciones razonables para abordar el problema examinado.

Un clasificador desarrollado por (Minango, 2022), con la capacidad de detectar tres categorías de eventos: LP, VT u Otros; el último incluye eventos regionales (RG), híbridos (HB) y deslizamiento de glaciares (IC). La investigación utiliza técnicas de aprendizaje supervisado como k-Vecino más cercano (k-NN, por sus siglas en inglés k-Nearest Vecinos) y Árbol de

decisiones (DT). Además, utiliza técnicas de aprendizaje no supervisado como el Autoencoder (Michelucci, 2022). Debido a que las señales de sismo volcánico de la base de datos se muestran a 50 y 100 Hz, es necesario sobre muestrearlas a 100 Hz. Luego, se elimina la media de las señales y la componente DC que los elementos de medición pudieron introducir para evitar sesgos en la red. Finalmente, se redimensionó la cantidad de muestras de las diferentes señales de sismo volcánico. La etapa de procesamiento recibe cinco matrices específicas correspondientes a los eventos LP, VT, RG, IC y HB. Las características espectrales de cada matriz se extrajeron y se les realizó un análisis de componentes principales para determinar cuáles características contenían más información para entrenar los modelos de clasificación. Como resultados se tiene que el modelo DT presentó una exactitud del 96.29% con un Ber del 0.042, mientras que el modelo k-NN y el autoencoder alcanzaron una exactitud del 98.41% con un Ber del 0.018.

Justificación e Importancia

Los fenómenos naturales según (Martín Neira & Lilian C., 2021) son eventos que pueden provocar daños, de tal manera puede perjudicar a la vida de personas si están en contacto con estos. Uno de estos fenómenos naturales puede ser las erupciones volcánicas. Menciona (Brown, 2018) que unos 60 volcanes erupcionan alrededor del planeta cada año, se considera que algunos lo hacen con regularidad mientras que otros de forma inesperada. Ecuador se encuentra en una zona alta de actividad por lo que es uno de los mayores peligros que puede presentarse. Uno de los volcanes que se encuentran en el territorio es el Cotopaxi, es uno de los más grandes y que se encuentra activo, esto es un peligro debido a que podría afectar a zonas altamente pobladas. La última erupción según (Bolívar Cáceres et al., 2008) fue en 1877 el cual produjo la destrucción parcial de Latacunga y el valle de los Chillos. Es un peligro para la población cercana, por esto este volcán ha estado en constante investigación.

El poder analizar las señales que este volcán produce, puede ser de gran ayuda para determinar un posible peligro, estos fenómenos provocan un sin número de señales, estas pueden ser interpretadas de varias maneras y no de forma correcta, para esto sería necesario un clasificador de señales. El clasificador debería de tener la capacidad de poder entender estas señales, para esto es necesario del aprendizaje automático o mejor conocido como *machine learning*. El uso de los métodos de selección de características las cuales según (Peluffo-Ordóñez et al., 2009) generan matrices de datos y mediante la información redundante poder llegar a una clasificación.

Los clasificadores se basan en aprendizaje supervisado y no supervisados, la diferencia entre estos es que para los supervisados se toman valores de características y al mismo tiempo una etiqueta de clase, mientras que para los no supervisados no tienen etiquetas de clase (Rostami et al., 2021).

Entre algunos de los métodos encontramos los de filtro que las variables son seleccionadas en base a criterios estadísticos, en los métodos envolventes, se van añaden o se eliminan variables de un conjunto para mejorar el resultado, y los embebidos, que se logra de la combinación de los dos anteriores con lo mejor de cada uno.

Por lo mencionado anteriormente, este trabajo se enfoca en mejorar clasificadores el cual permita identificar las mejores características en un software, el cual permita de manera automática ver que parámetros son los más representativos.

Alcance

El presente proyecto tiene como objetivo ayudar a la sociedad con un clasificador de eventos sismo volcánicos, se utilizan señales recopiladas por el IGEPN de diversas estaciones en el volcán Cotopaxi. Las señales serán sometidas a un proceso de extracción de

características, luego serán procesadas para lograr ser clasificadas. Este clasificador tiene la ayuda de selectores para mejorar el resultado, así mismo este clasificador permite seleccionar los parámetros.

Objetivos

Objetivo General

Mejorar del desempeño de un clasificador mediante la selección de características de las señales de microsismos del volcán Cotopaxi.

Objetivos Específicos

- Analizar el estado del arte de las metodologías utilizadas en la obtención de características.
- Identificar las mejores características mediante el empleo de escalas lineales y psicoacústicas, con diferentes métodos de selección de características, con diferentes técnicas.
- Evaluar el desempeño del clasificador de señales de microsismos del volcán Cotopaxi y mostrarlos en un sistema en MATLAB®.

Capítulo II

Marco Teórico

Volcán Cotopaxi

Es un volcán con un proceso activo desde el 2015, y finalizado en febrero del 2023, con una altura de 5897 metros desde el nivel del mar, es uno de los volcanes más peligrosos del mundo por la destrucción que este provoca, esto se atribuye a que forma lahares de gran tamaño, actualmente es uno de los volcanes con mayor vigilancia. (Instituto Geofísico - EPN, 2014)

Eventos microsismos

Se tiene en cuenta que los eventos que ocurren dentro de un volcán, pueden proporcionar información sobre su actividad si se registran y evalúan adecuadamente. Los eventos se han clasificado principalmente en las siguientes categorías después de analizar su forma, frecuencia y posible origen los cuales se detallan a continuación (Malfante et al., 2018):

Nota. Los eventos microsismos término utilizado en este proyecto es actualmente llamado micro terremotos,

Eventos Volcanotectónico (VT)

La presión provocada por el movimiento del magma rompe una roca sólida lo cual provoca un VT. Este tipo de evento genera un movimiento de frecuencia moderadamente alta, con frecuencia hasta 15 Hz.

Eventos Largo Periodo (LP)

Los cambios de presión dentro de las fisuras y canales llenos de magma causan eventos LP con una frecuencia entre 0.5 y 5 Hz. A diferencia de los eventos VT, los eventos LP pueden mostrar el movimiento del magma y el crecimiento de la presión dentro de un volcán, con esto se puede predecir posibles erupciones.

Eventos Híbridos (HB)

Un evento LP puede provocar un evento VT, y viceversa, una señal sísmica HB combina elementos de ambas categorías. En ciertas ocasiones estos eventos son generados ya que se genera un evento VT y puede desencadenar un evento LP y viceversa, de ese modo se genera una combinación de ambos eventos.

Rupturas de glaciación (IC)

Es un evento que su nombre indica provocado por ruptura de glaciares, esto provoca la formación de grietas y crestas en el hielo, lo que puede generar ondas sísmicas las que son registradas para su análisis.

Eventos Regionales (RG)

Estos eventos no están relacionados directamente con el volcán, en otras palabras, estos eventos se encuentran alejados, pero son de gran intensidad lo cual son detectados.

Escala ERB

La capacidad de detectar sinusoides en presencia de ruido determina la anchura de banda rectangular equivalente, también conocida como escala ERB por sus siglas en inglés Equivalent Rectangular Bandwidth (Rehman et al., 2015). Para convertir de Hz a ERB:

$$ERB = 21.3 \log_{10} \left(1 + \frac{f}{228.7} \right) \quad (1)$$

Aprendizaje Automático

El objetivo del campo de la informática conocido como *machine learning* es crear algoritmos que se basan en un conjunto de datos de un evento específico. La forma en que un algoritmo aprende a interpretar el fenómeno es mediante diferentes tipos, entre los cuales se encuentran el aprendizaje supervisado, el aprendizaje no supervisado, el aprendizaje semisupervisado y el aprendizaje por refuerzo, que son las cuatro formas principales de

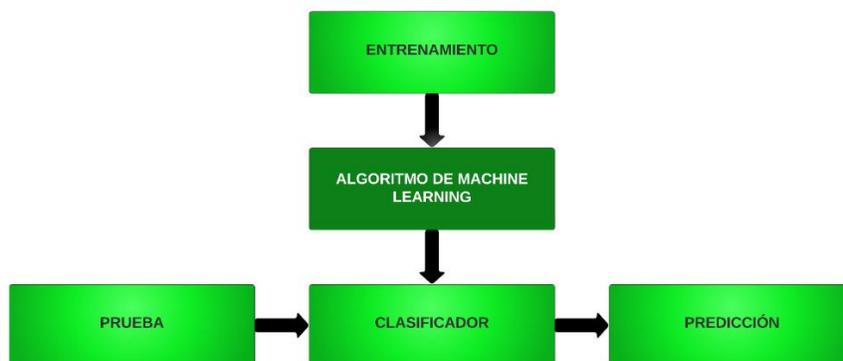
aprendizaje. El tipo de datos que se quieren predecir determina el tipo de algoritmo a usar. En este caso se centra en el uso del aprendizaje supervisado (Carniel & Raquel Guzmán, 2020).

Aprendizaje supervisado

Según (Nasteski, 2017) el uso de conjuntos de datos etiquetados define el aprendizaje supervisado. Estos conjuntos de datos están destinados a entrenar o supervisar algoritmos para clasificar datos. El modelo puede medir su precisión y aprender con el tiempo con entradas y salidas etiquetadas como se muestra en la Figura 1. El objetivo generalmente es que la máquina aprenda el sistema de clasificación que hemos creado, por lo que es crucial tener en cuenta algunos aspectos importantes. Para empezar, debemos proporcionar a la máquina un conjunto de datos de entrenamiento lo más variado y completo posible.

Figura 1

Descripción de aprendizaje supervisado



Nota. Referenciado de (Nasteski, 2017)

Características

En la Tabla 1 se muestran las características en el dominio del tiempo, en el dominio de la frecuencia, en el dominio de la escala; mientras que en la Tabla 2 se muestran las características psico acústicas. Estas características forman un total 98 divididas de la siguiente manera:

- Características en el dominio del tiempo: f1-f13 un total de 13
- Características en el dominio de la frecuencia: f14-f34 un total de 21
- Características en el dominio de la escala: f35-f84 un total de 50
- Características psico acústicas: f85-f98 un total de 14

Características lineales

Tabla 1

Características en el dominio del tiempo, de la frecuencia y de la escala

ID	Nombre de la característica	ID	Nombre de la característica	ID	Nombre de la característica
Time-domain					
f1	Mean	f28	Peak-to-RMS ratio	f56	Percentage of energy for A6
f2	Standard deviation	f29	Power	f57	Percentage of energy for D1
f3	Variance	f30	Density of peaks above RMS	f58	Percentage of energy for D2
f4	Entropy	f31	2nd highest peak value	f59	Percentage of energy for D3
f5	Kurtosis	f32	Freq. of 2nd highest peak	f60	Percentage of energy for D4
f6	Multiscale entropy (MSE)	f33	3rd highest peak value	f61	Percentage of energy for D5
f7	Time to reach the maximum peak	f34	Freq. of 3rd highest peak	f62	Percentage of energy for D6
f8	RMS value	Scale-domain		f63	A6 RMS in time-domain
f9	Peak-to-peak value	f35	A6 Max. peak in freq.-domain	f64	D1 RMS in time-domain
f10	Peak-to-RMS ratio	f36	D1 Max. peak in freq.-domain	f65	D2 RMS in time-domain
f11	Energy	f37	D2 Max. peak in freq.-domain	f66	D3 RMS in time-domain
f12	Zero-crossing rate	f38	D3 Max. peak in freq.-domain	f67	D4 RMS in time-domain
f13	Density of peaks above RMS	f39	D4 Max. peak in freq.-domain	f68	D5 RMS in time-domain
Frequency-domain					
f14	Frequency of maximum peak	f40	D5 Max. peak in freq.-domain	f69	D6 RMS in time-domain
f15	Bandwidth of 90% energy	f41	D6 Max. peak in freq.-domain	f70	A6 Peak-to-peak in time-domain
f16	Entropy	f42	A6 Freq. of max. peak	f71	D1 Peak-to-peak in time-domain
f17	Mean	f43	D2 Freq. of max. peak	f72	D2 Peak-to-peak in time-domain
f18	Standard deviation	f44	D3 Freq. of max. peak	f73	D3 Peak-to-peak in time-domain
f19	Variance	f45	D4 Freq. of max. peak	f74	D4 Peak-to-peak in time-domain
f20	Energy	f46	D5 Freq. of max. peak	f75	D4 Peak-to-peak in time-domain
f21	Kurtosis	f47	D6 Freq. of max. peak	f76	D6 Peak-to-peak in time-domain
f22	Multiscale entropy	f48	A6 Mean in freq.-domain	f77	A6 Peak-to-RMS ratio in time-domain
f23	Maximum peak in 10–20 Hz band	f49	D1 Mean in freq.-domain	f78	D1 Peak-to-RMS ratio in time-domain
f24	Freq. of max. peak in 10–20 Hz Band	f50	D2 Mean in freq.-domain	f79	D2 Peak-to-RMS ratio in time-domain
f25	Maximum peak in 20–30 Hz band	f51	D3 Mean in freq.-domain	f80	D3 Peak-to-RMS ratio in time-domain
f26	Freq. of max. Peak in 20–30 Hz Band	f52	D4 Mean in freq.-domain	f81	D4 Peak-to-RMS ratio in time-domain
f27	RMS value	f53	D5 Mean in freq.-domain	f82	D5 Peak-to-RMS ratio in time-domain
		f54	D6 Mean in freq.-domain	f83	D6 Peak-to-RMS ratio in time-domain
		f55	Mean energy of components	f84	Mean energy of wavelet coefficients

En la Tabla 1, se muestran las características en el dominio del tiempo, de la frecuencia y de la escala, las cuales dan un total de 84. Estas características son obtenidas en referencia a (Pérez et al., 2020). Para la tabla 2 estas se hacen referencia de (Minango, 2022), en el cual utiliza las 11 primeras desde f85 a f95, para f96, f97 y f98 fueron agregadas a la versión 2023 de MATLAB®.

Características psico acústicas

Tabla 2

Características psico acústicas

ID	Nombre de la característica	ID	Nombre de la característica
f85	Kurtosis	f92	Spread
f86	Slope	f93	Centroid
f87	Crest	f94	Entropy
f88	Flux	f95	RolloffPoint
f89	Skewness	f96	shorTimeEnergy
f90	Decrease	f97	harmonicRatio
f91	Flatness	f98	zeroCrossRate

Algoritmos de clasificación

Autoencoder

Es un tipo de red neuronal, su principal uso es para codificar una señal a su entrada, y luego decodificarla para que la entrada al momento de volver a construirla, tenga una semejanza o sea lo más similar posible a la original (Bank et al., 2021). Se tiene el número de neuronas que se especifica un valor entero positivo para el número de neuronas en la capa oculta. El número de épocas es un número entero positivo, las épocas o iteraciones de entrenamiento se podría también definir como un parámetro que establece la cantidad de veces que el algoritmo de aprendizaje funcionará en todo el conjunto de datos de entrenamiento. El número de épocas, lo que permite que el algoritmo de aprendizaje funcione hasta que el error

del modelo se haya reducido lo suficiente. Las funciones de transferencia de Encoder y Decoder se tiene:

Logistic sigmoid function (logsig)

$$f(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

Positive saturating linear transfer function (satlin)

$$f(z) = \begin{cases} 0, & \text{si } z \leq 0 \\ z, & \text{si } 0 < z < 1 \\ 1, & \text{si } z \geq 1 \end{cases} \quad (3)$$

Se agrega una función más al Decoder el cual es el linear transfer function (purelin)

$$f(z) = z \quad (4)$$

L2WeightRegularization y sparsityRegularization

Es el coeficiente para la regularización L2, que toma la función de pérdida (lossFunction), en conjunto con (msespars), se usa como la función de error cuadrático medio para el entrenamiento del autoencoder que representa la siguiente ecuación:

$$E = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K (x_{kn} - \hat{x}_{kn})^2 + \lambda * \Omega_{weights} + \beta * \Omega_{sparsity} \quad (5)$$

Donde λ es el coeficiente de regularización L2 y β es el coeficiente de escasez.

SparsityProportion

La proporción de dispersión es un parámetro del regularizador de dispersión, esto permite que, con un valor bajo, las neuronas en la capa oculta se especialicen en dar un mayor grado de escasez.

ScaleData

Permite cambiar la escala de los datos de entrada, para que el rango de los valores de entrada coincida con el rango de la función de transferencia del decodificador.

Matriz de confusión

Es utilizada según (Caelen, 2017) para la visualización del comportamiento de los modelos de aprendizaje automático, la cual es una matriz cuadrada donde sus filas muestran la clase real mientras que sus columnas sus predicciones de tamaño $L \times L$, donde L es tamaño de número de clases, para este estudio se utilizó 3×3 . Con la ayuda de una comparación completa entre los valores reales y previstos, podemos evaluar diferentes métricas de rendimiento de nuestros modelos. Con esta información, podemos tomar decisiones y modificar los algoritmos para obtener los mejores resultados.

Figura 2

Matriz de confusión 3x3

		PREDICCIONES				
		A	B	C		
REAL	A	AA	AB	AC	TP: True Positive	
	B	BA	BB	BC	FP: False Positive	
	C	CA	CB	CC	FN: False Negative	
					TN: True Negative	

En la Figura 2, se muestra una Matriz de confusión 3×3 , la cual es utilizada en este trabajo debido al número de clases. En el cual se tiene los diferentes tipos de celdas los cuales son dados referente a cada una de las predicciones obtenidas:

- **TP:** Es el número de clases positivas que han sido correctamente clasificadas.
- **FP:** Es el número de clases negativas clasificados como positivos.
- **FN:** Es el número de clases positivas clasificadas incorrectamente.
- **TN:** Es el número de clases negativas clasificadas correctamente.

Métricas de clasificación

Evaluar el rendimiento, es una de las fases cruciales para determinar si un modelo de aprendizaje automático es exitoso, para lo cual se tienen las siguientes métricas descritas en Tabla 3 mostrada a continuación:

Tabla 3

Métricas de clasificación

Métricas	Descripción	Fórmula
Exactitud	Proporción de clasificaciones correctamente predichas sobre el total de instancias	$A = \frac{\sum_{i=1}^k M_{ii}}{\sum_{i=1}^k \sum_{j=1}^k M_{ij}} \times 100 \quad (6)$
Precisión	La precisión determina la fracción de predicciones positivas que fueron precisas.	$P = \frac{\sum_{i=1}^k p_i}{k} \text{ con } p_i = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100 \quad (7)$
Sensibilidad (Recall)	Porcentaje de casos positivos que se encuentran bien clasificados	$R = \frac{\sum_{i=1}^k r_i}{k} \text{ con } r_i = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100 \quad (8)$
Especificidad	Es conocida como la tasa negativa verdadera o en otras palabras porcentaje de casos negativos que fueron bien clasificados.	$S = \frac{\sum_{i=1}^k s_i}{k} \text{ con } s_i = \frac{N_{TN}}{N_{TN} + N_{FP}} \times 100 \quad (9)$
Ber	La proporción promedio de muestras mal clasificadas en cada clase se calcula mediante la tasa de error balanceada, también conocida como Ber, es menos sesgado hacia las clases mayoritarias.	$BER = \frac{\sum_{i=1}^k ber_i}{k} \text{ con } ber_i = 1 - \frac{R + S}{2 \times 100} \quad (10)$

En la tabla 13, se observa las métricas de clasificación donde M es la matriz de confusión; p_i, r_i, s_i, ber_i , representan los valores individuales de cada una de estas métricas para cada evento con $i = j = \{1, 2, \dots, m\}$, para nuestro caso se tiene que $m = 3$, por las tres clases. Para $N_{TP}, N_{FP}, N_{FN}, N_{FP}$, son los valores utilizados de la matriz de confusión mostrados en la Figura 2.

K-nearest neighbor (KNN)

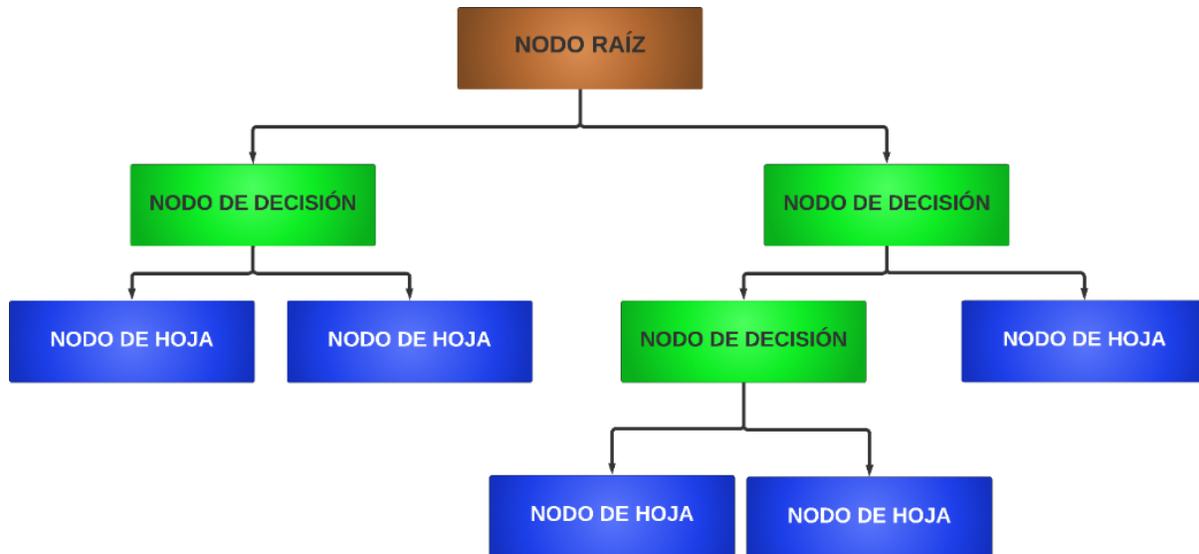
El algoritmo de vecinos más cercanos k , también conocido como KNN o k -NN, es un clasificador de aprendizaje supervisado, que utiliza la proximidad para hacer clasificaciones o predicciones sobre la agrupación de un punto de datos. Se utiliza principalmente como un algoritmo de clasificación, aunque también puede usarse para problemas de clasificación o regresión. Se basa en la suposición de que puntos similares se pueden encontrar cerca unos de otros (Abu Alfeilat et al., 2019).

Decision Tree (DT)

Según (Taha Jijo & Mohsin Abdulazeez, 2021) el árbol de decisión, un método de aprendizaje supervisado, puede usarse para resolver problemas de clasificación y regresión. Es un clasificador estructurado con varias divisiones para el proceso de toma de decisiones, forma nodos internos para las características de un conjunto de datos, y nodos externos para el resultado de la clasificación como se observa en la Figura 3 a continuación:

Figura 3

Árbol de decisión diagrama



Nota. Referenciado de (Taha Jijo & Mohsin Abdulazeez, 2021)

Selección de características

El objetivo de las técnicas de selección de características en el aprendizaje automático es determinar el conjunto de características más óptimo para que se puedan crear modelos optimizados de los fenómenos estudiados, de (Pohjalainen et al., 2013a), se tienen los siguientes algoritmos de selección de características:

Mutual Information (MI)

MI permite poder explicar el conocimiento compartido entre variables, es decir, la medida en que la incertidumbre de una variable se reduce al conocer la otra. Si X y Y son dos variables aleatorias, entonces la siguiente definición se aplica a si se tiene información mutua:

$$MI(X, Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (11)$$

Donde x y y representan las muestras de las variables X y Y , respectivamente, $p(x)$ y $p(y)$ se refieren a las probabilidades.

Statistical Dependency (SD)

Determina si los valores de una función dependen de etiquetas de clase asociadas o coinciden por coincidencia. Para que cada una de las características generalmente contenga la misma cantidad de muestras en todo el conjunto de datos. En lugar de utilizar una escala de cuantificación uniforme tradicional, se seleccionan de esta manera para proporcionar cierta validez estadística a la ocurrencia de varios niveles de cuantización. La ecuación 6 a continuación proporciona una evaluación estadística de la dependencia entre los valores de características discretizadas y las etiquetas de clase z .

$$SD(X, Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \frac{p(x, y)}{p(x)p(y)} \quad (12)$$

Tanto los métodos SD como MI producen una puntuación y clasificación de características, de las cuales se puede elegir un número predeterminado de características con los valores más altos.

Según (Pohjalainen et al., 2013) la selección automática de características se puede formular como el problema de encontrar el mejor subconjunto posible S de características a partir de un conjunto inicial, y posiblemente muy grande, de características F (es decir, $S \subset F$). Dado que el objetivo final es realizar la clasificación de muestras de datos, se podría definir el subconjunto óptimo de características como aquel que proporciona la mejor capacidad de clasificación en la tarea dada, medida por una función de criterio $G(S, D, M) = c$, donde D denota el conjunto de datos utilizado y M denota el modelo de clasificación.

Sequential Forward Selection (SFS)

El enfoque básico de selección de características es construir un subconjunto del número requerido de características incrementalmente, inicia con el conjunto vacío (enfoque de abajo a arriba) o para comenzar el conjunto completo de características y eliminar las características redundantes, el conjunto de características se actualiza iterativamente al incluirlas, en cada paso, la característica f que da como resultado la puntuación máxima $G(S, D, M)$. Por tanto, el conjunto de características de tamaño D viene dado por:

$$S_d = S_{d-1} \cup \underset{f}{\operatorname{argmax}} G(S_{d-1} \cup f, D, M) \quad (13)$$

Sequential Forward Floating Selection (SFFS)

El SFFS según (Pudil et al., 1994) es básicamente un procedimiento de búsqueda ascendente que incluye nuevas características mediante la aplicación del procedimiento SFS básico a partir del conjunto de características actual, seguido de una serie de exclusiones condicionales sucesivas de la peor característica en el conjunto recientemente actualizado, siempre que se pueda realizar una mejora adicional.

Capítulo III

Materiales y Métodos

En este capítulo se muestra los materiales necesarios para la realización de este proyecto, como segundo punto se describe la metodología con la cual se llevó a cabo.

Materiales

El presente proyecto hace uso del software MATLAB® en su edición R2023a, aprovecha los paquetes de herramientas especializadas en aprendizaje profundo y procesamiento de señales. La elección de esta versión en particular responde a la necesidad de acceder a características específicas que solo se encuentran disponibles desde esta versión.

Metodología

En este proyecto se utilizará una investigación tipo experimental cuantitativa, el cual permite obtener la información necesaria para poder determinar los métodos de selección de características principales más útiles para mejorar los clasificadores de señales, el cual está basado en el diagrama de bloques que se muestra en la Figura 4
Diagrama de bloques que describe el proyecto a continuación.

Figura 4

Diagrama de bloques que describe el proyecto



En la Figura 4, se muestra el proceso de este proyecto, para esto se ingresan los valores de una base de datos, ingresa a un preprocesamiento para pasar a un procesamiento, se evalúan, se vuelve a procesar hasta que la evaluación sea la mejor. Este proceso es descrito a continuación.

Base de datos

Este proyecto consta de una base de datos obtenida de (IGEPN, 2024), esta base cuenta con datos del volcán Cotopaxi, con el nombre de MicSigV1. Cuenta con 1187 señales, que de manera visual fueron etiquetados a que evento corresponden y con un total de: 1044 LP, 101 VT, 27 REG, 8 HB y 7 IC

Esta base de datos se encuentra en formato mat, y consta con la siguiente información mostrada en la Tabla 4:

Tabla 4

Información de la base de datos MicSigV1

Componente	Descripción
Red	La Red en la que se encuentra, denominada "EC".
Estación	La estación en la cual fue identificado el evento.
Frecuencia de muestreo	La frecuencia a la que fue muestreada, se encuentran valores de 50 y 100 Hz.
Año	El año en el que fue identificado el evento
Mes	El mes en el que fue identificado el evento
Tipo	Describe cual es el tipo de evento detectado
Duración	Cuál es la duración del evento.
Punto de inicio	Punto de inicio del evento
Punto de fin	Punto de fin del evento.
Data	Se encuentra el evento.

De los datos mostrados en la Tabla 4, se analiza su frecuencia, su tipo y el punto de fin del evento para que la Data sea ingresada al preprocesamiento.

Preprocesamiento

Figura 5

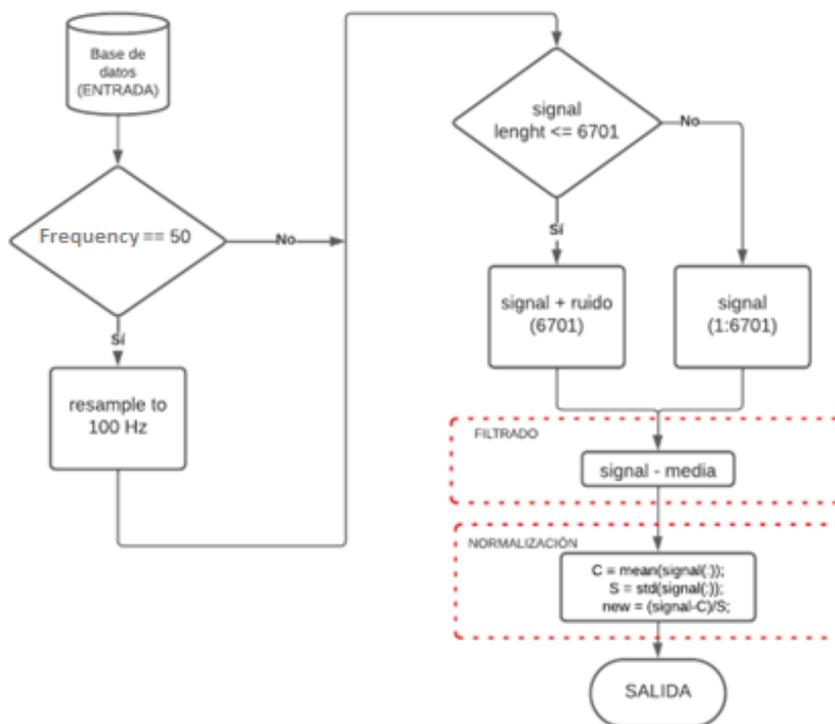
Preprocesamiento



En la Figura 5 se muestra el preprocesamiento, el cual está determinado por una entrada, que es la base de datos antes descrita. Se procede con la edición de la señal, la cual se toma en cuenta el diagrama de flujo de la Figura 6.

Figura 6

Diagrama de flujo del preprocesamiento



En la Figura 6 se presenta un diagrama de flujo que ilustra el proceso de preprocesamiento. En esta fase, se inicia en ver cuál es la frecuencia de muestreo de la señal.

Si esta frecuencia es igual a 50 Hz, se implementa una técnica de remuestreo para ajustarla a 100 Hz. Luego, se estandariza el tamaño de todas las señales para garantizar uniformidad.

Este proceso incorpora una condición: si el tamaño de la señal es menor o igual a 6701 (valor de referencia considerado debido a que las señales tienden a concluir eventos cercanos a este punto), se introduce ruido en la señal. En caso contrario, se realiza un recorte para mantener la integridad de la señal al eliminar el ruido excesivo. Para el proceso de filtrado, se retira la componente DC de la señal; y finalmente se procede a realizar la normalización, se obtiene la media y la desviación estándar. Con esto se tiene una nueva base de datos lista para poder ser procesada.

Procesamiento

Figura 7

Diagrama de bloques del procesamiento



En la Figura 7, se presenta el diagrama de bloques del proceso de procesamiento de señales, feature extraction implica la obtención de características mediante llamar a la función *calculate_features* la cual contiene el cálculo necesario para la obtención de las mismas. Una vez obtenidas estas características, se realiza una fase de normalización, con el objetivo de evitar que los datos presenten una dispersión excesiva, lo que facilita su manejo y análisis.

Posteriormente, se lleva a cabo una preparación adicional de los datos. En esta etapa, la totalidad de los datos se divide en dos conjuntos: uno destinado al entrenamiento del modelo, el cual constituye el 70% de los datos, y otro destinado a la prueba del modelo, que abarca el 30% restante. Esta división facilita la evaluación del rendimiento del modelo en datos

que no ha visto durante el entrenamiento, y contribuye así a una evaluación más robusta y representativa del clasificador o modelo utilizado.

Evaluación

Figura 8

Diagrama de bloques de Evaluación



En este punto con los datos en matrices de entrenamiento y prueba, se procede a ingresarlos a un selector, es aquí donde selecciona las mejores características, y se procede a ingresarlas a un clasificador, como se muestra en la Figura 8. En el clasificador se determina el porcentaje de acierto y se observa como fue el desempeño con las características ingresadas mediante métricas.

Interfaz gráfica

La interfaz gráfica consta de dos pestañas, como se muestran en las Figuras 9 y 10. La primera pestaña describe un ejecutable el cual permite ejecutar con los parámetros deseados el clasificador, y la pestaña de Informativo, el cual al seleccionar parámetros determinados se podrá observar los resultados.

Pestaña Ejecutable

En la Figura 9, en la interfaz se observa varios puntos los cuales se encuentran enumerados y se describen de la siguiente manera

1. Se encuentran el menú de selección de selectores, entre los cuales tenemos las opciones de MI, SD, SFS, SFFS y Ninguno (No tener selector).

2. Se puede seleccionar entre las opciones del menú de clasificadores entre los cuales tenemos Autoencoder, KNN y DT.
3. En el Spinner, puedes elegir el número que deseas ingresar. Si seleccionas Autoencoder, introduce la cantidad de neuronas; si eliges KNN, ingresa el número de vecinos; y si se opta por DT, introduce el valor deseado de divisiones.
4. Este Spinner se encuentra asociado con el Selector, pero se habilita para MI y SD, debido a que SFS y SFFS, su tiempo de procesamiento es demasiado alto. Ingresamos el valor deseado de Q (valor de cuantización) que tenga el selector. Este valor se refiere al número de divisiones que se utilizan para discretizar las variables involucradas en el cálculo del selector.
5. Ingresamos el valor del peso deseado de las características que tenga el selector. Este se encuentra disponible solo para MI y SD de igual manera que el apartado anterior.
6. Contamos con un interruptor (switch) que posibilita la generación de valores aleatorios para conjuntos de entrenamiento y prueba. Esto nos ofrece la flexibilidad de tener variabilidad en el análisis de los datos en cada ejecución.
7. El botón ejecutar permite tomar los valores seleccionados e ingresados y ejecutar el programa.
8. Se muestra una tabla con las métricas de desempeño, las cuales se tienen exactitud, precisión, sensibilidad, especificidad y Ber. Estas son analizadas sus valores por cada clase de las señales, además de un general. Para estos se utiliza un promedio ponderado para cada una de las clases.
9. Una tabla en la cual se muestra las características tomadas por el selector.

Figura 9

Pestaña ejecutable



ESPE

 UNIVERSIDAD DE LAS FUERZAS ARMADAS

 INNOVACIÓN PARA LA EXCELENCIA

Clasificador de señales del volcán Cotopaxi

Steve Alexander Morales Zapata

Ejecutable Informativo

Selectores: **1** Clasificadores: **2**

MI Autoencoder
 SD KNN
 SFS DT
 SFFS DT
 Ninguno

Ingrese el número de neuronas

3

Ingrese el valor de Q de MI

4

Ingrese el peso de las características de MI

5 Off On **6**

7 Train y Test Aleatorio

Tipo	Exactitud	Presición	Sensibilidad	Especificidad	Ber
VT	97.1831	90.0000	79.4118	97.8462	0.1137
LP	95.4930	98.4026	96.5517	73.8095	0.1482
Otros	96.6197	8.3333	50.0000	99.7085	0.2515
Overall	94.6479	94.4909	93.4422	76.7720	0.1489

8

No	#Característica	Característica
1	72	w_t_peak2peak_D2
2	37	w_f_maxval_D2
3	24	f_peak_1020_pos
4	30	f_PeaksAboveRMSDensity_fun

9

Figura 10

Pestaña informativo



ESPE

 UNIVERSIDAD DE LAS FUERZAS ARMADAS

 INNOVACIÓN PARA LA EXCELENCIA

Clasificador de señales del volcán Cotopaxi

Steve Alexander Morales Zapata

Ejecutable Informativo

Selectores: Clasificadores:

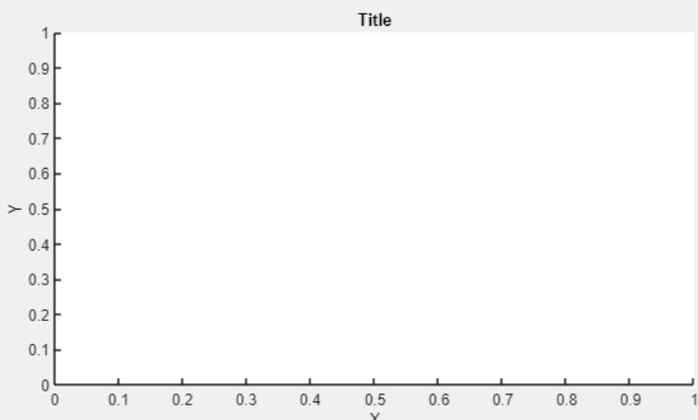
MI Autoencoder
 SD KNN
 SFS DT
 SFFS DT
 Ninguno

Ingrese el numero de Q

Ingrese el número de neuronas

Ingrese el número de peso de selector

Title



Pestaña Informativo

Esta interfaz, mostrada en la Figura 10, se distingue del modo ejecutable en el sentido de que los valores seleccionados aquí no se utilizan para ejecutar el programa directamente, sino en valores ya precargados de los resultados más destacados obtenidos.

Funcionamiento de la Interfaz

La interfaz cumple el proceso del diagrama de flujo de que se muestra en la Figura 11 y 12, en el cual si se selecciona la pestaña Ejecutable cumple con la condición inicial, con esto se procede a determinar el selector, luego el clasificador y finalmente enviar las métricas a la interfaz. Si se selecciona la pestaña Informativa se determina el selector, luego el clasificador y se envía los valores ya cargados.

Figura 11

Diagrama de flujo de interfaz gráfica parte 1

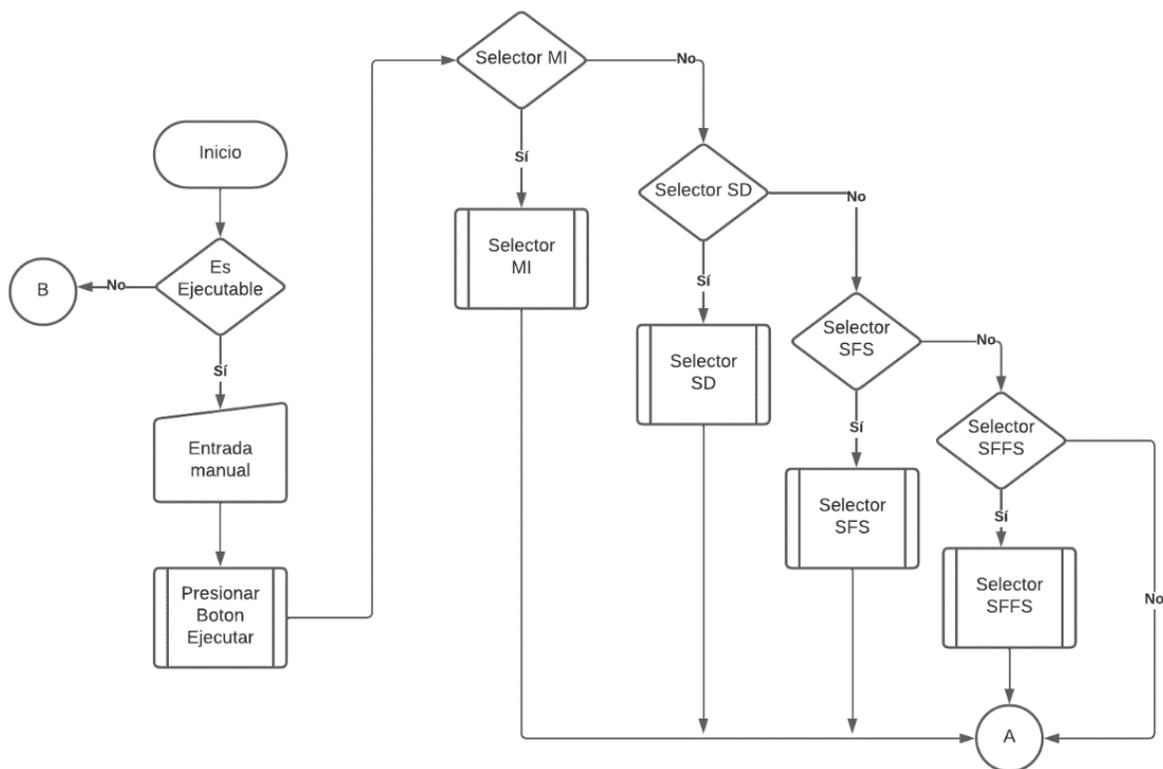
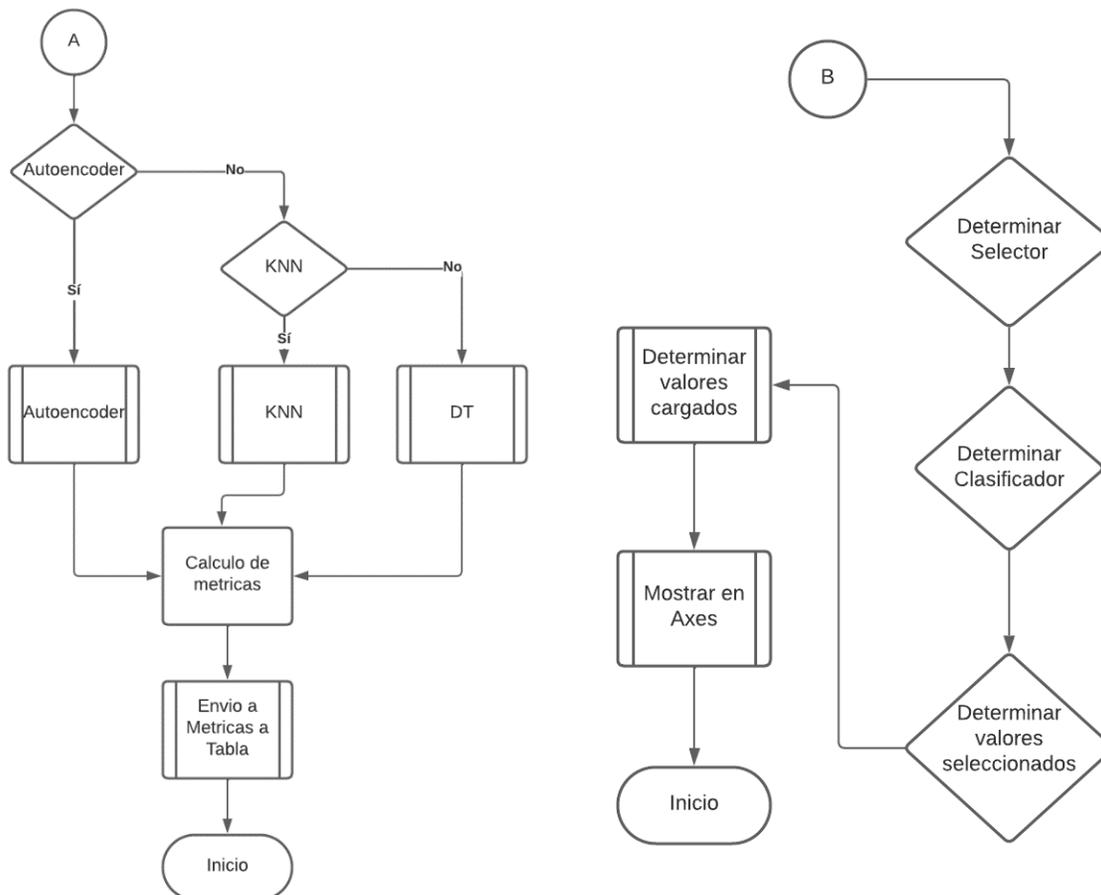


Figura 12*Diagrama de flujo de interfaz gráfica parte 2*

Capítulo IV

Pruebas y Resultados

En este capítulo se muestra las pruebas realizadas para determinar los parámetros en los cuales se tienen los mejores resultados, luego se realizan pruebas con los mejores resultados, prueba de los selectores y pruebas de la interfaz gráfica y finalmente se tienen los resultados.

Inicialmente solo con las características psicoacústicas, se procede a realizar las pruebas con todas las características, y luego las pruebas con las mejores características.

Pruebas con Autoencoder Simple

Estas pruebas se realizan con las 14 características psico acústicas, las cuales los mejores resultados en (Minango, 2022) con los parámetros y resultados se muestran en la Tabla 5 a continuación:

Tabla 5

Prueba con parámetros con mejores resultados en (Minango, 2022)

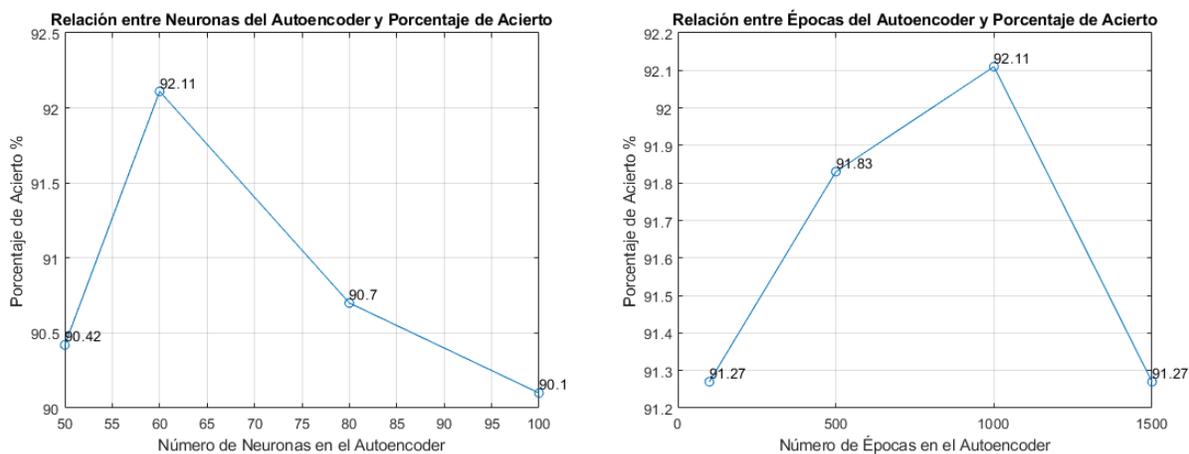
Parámetros de Autoencoder Simple								
Neuronas	Épocas	ETF	DTF	L2WR	SR	SP	SD	
70	1000	logsig	satlin	0.004	1	0.05	FALSE	
		Predicción			Métricas			
	Clase	VT	LP	Otros	Recall%	SP%	A%	Ber
	VT	16	3	4	69.57	95.8	94.1	0.173
Observación	LP	14	306	5	94.15	76.7	92.7	0.146
	Otros	0	4	3	42.86	97.4	96.3	0.299
	P%	53.33	97.76	25.00	Overall Average (%)		91.55	

Nota. Como se puede observar en Tabla 3, se tiene un promedio general de acierto del 91.55%

Con los valores iniciales, se varia un solo parámetro para ver con cual se tiene mejores resultados en la clasificación.

Figura 13

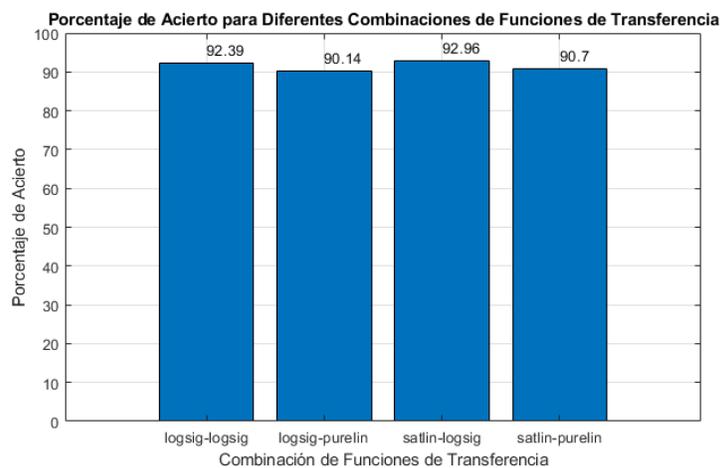
Pruebas Realizadas al variar el número de neuronas y épocas



En la Figura 13, se puede observar que se realizaron las pruebas al variar el número de neuronas la cual con 60 da un porcentaje de acierto del 92.11%. Por otro lado, se observa que al variar el número de épocas se observa que con 1000 se obtiene el mejor resultado.

Figura 14

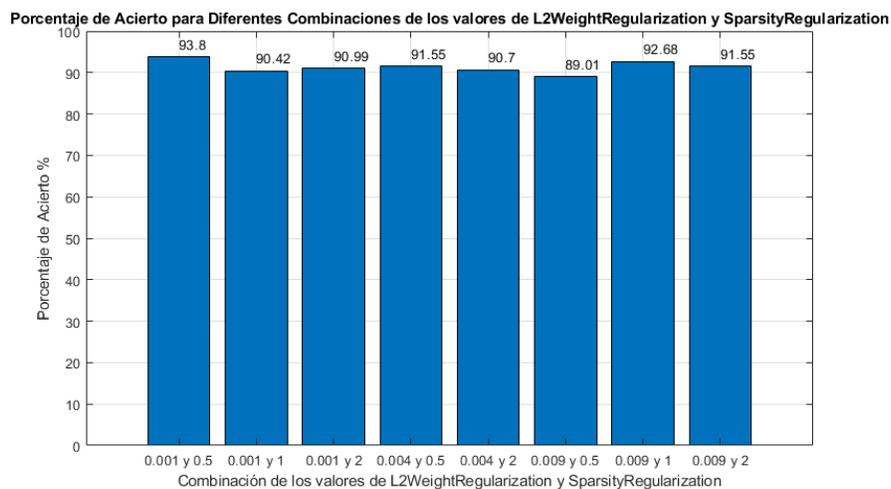
Prueba al variar las funciones de transferencia.



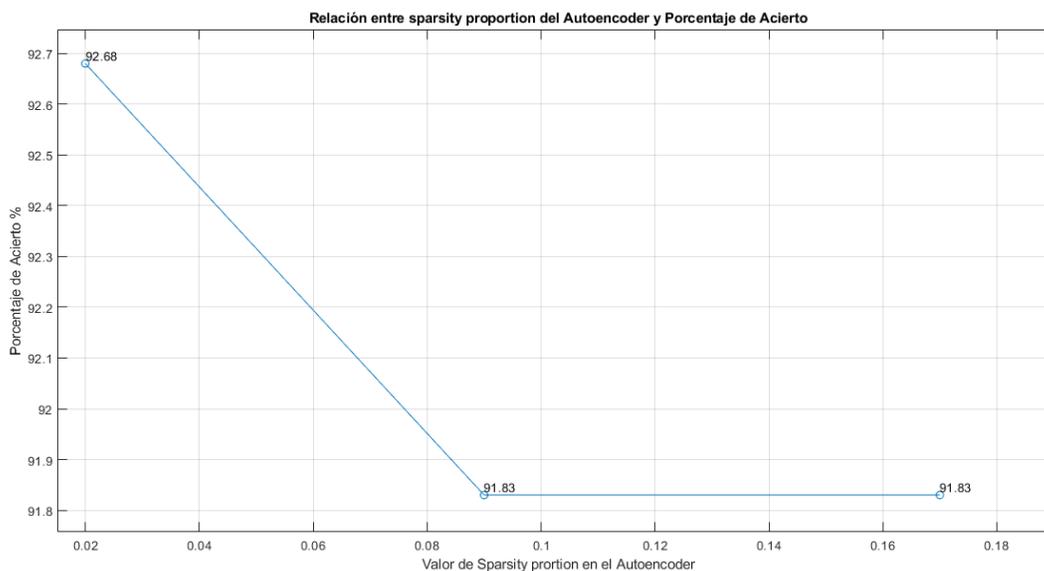
Esta prueba que se muestra en la Figura 14, se observa que el mejor resultado en clasificación de las funciones de activación es satlin-logsig, con un porcentaje de clasificación del 92.96%.

Figura 15

Prueba de porcentaje de acierto de combinaciones de L2WeightRegularization y SparsityRegularization

**Figura 16**

Relación entre el sparsity proportion y el porcentaje de acierto



En la Figura 15, se muestra la prueba al variar los valores de L2WeightRegularization y SparsityRegularization con 0.001 y 0.5 respectivamente, con un porcentaje de clasificación del

93.8%. En la Figura 16 se varía el valor de sparsity proportion y se obtiene el mejor resultado de clasificación de 92.68% con el valor de 0.02.

Se realiza las pruebas con las 98 características, se toman inicialmente los valores obtenidos en (Minango, 2022), el cual se muestra en la Tabla 6 a continuación.

Tabla 6

Prueba con los mejores resultados de (Minango, 2022)

Parámetros de Autoencoder Simple							
Neuronas	Épocas	ETF	DTF	L2WR	SR	SP	SD
70	1000	logsig	satlin	0.004	1	0.05	FALSE
		Predicción			Métricas		
	Clase	VT	LP	Otros	Recall%	SP%	A% Ber
	VT	13	5	4	59.09	94.9	92.7 0.23
Observación	LP	17	304	5	93.25	69	91.3 0.189
	Otros	0	4	3	42.86	97.4	96.3 0.299
	P%	43.33	97.12	25.00	Overall Average (%)		90.14

Tabla 7

Prueba con los mejores parámetros obtenidos con las características psico acústicas

Parámetros de Autoencoder Simple							
Neuronas	Épocas	ETF	DTF	L2WR	SR	SP	SD
60	1000	satlin	satlin	0.004	1	0.05	FALSE
		Predicción			Métricas		
	Clase	VT	LP	Otros	Recall%	SP%	A% Ber
	VT	18	4	2	75	96.4	94.9 0.143
Observación	LP	12	305	8	93.85	73.3	92.1 0.164
	Otros	0	4	2	33.33	97.1	96.1 0.348
	P%	60	97.44	16.67	Overall Average (%)		91.55

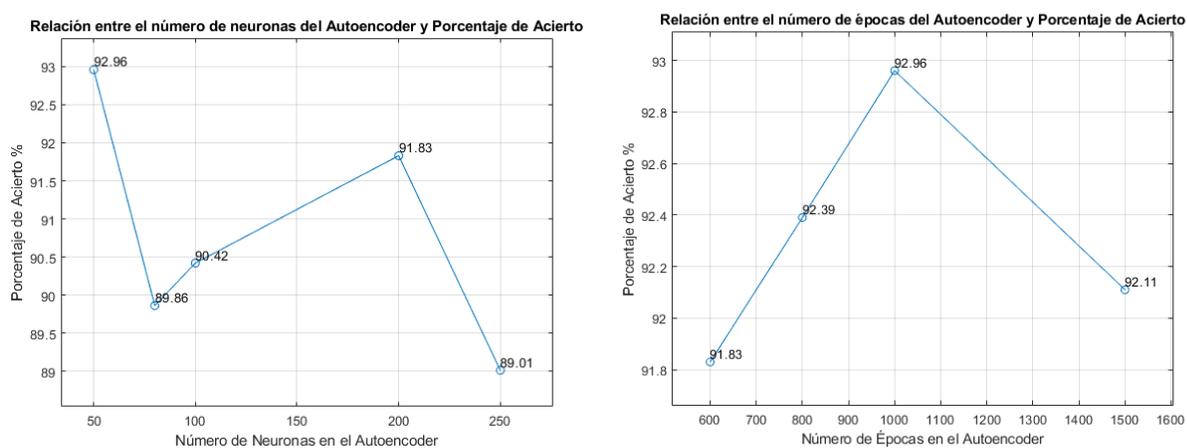
En la Tabla 6 se muestra el resultado de la prueba, con un porcentaje de clasificación del 90.14%, se procede a colocar los mejores parámetros que se obtuvieron con las

características psico acústica que se muestra en la Tabla 7, el cual el resultado es un porcentaje de clasificación del 91.55%.

Se varían los parámetros, para poder ver que cuales son los mejores con el aumento de características, se inicia con el número de neuronas y épocas mostrados en la Figura 17.

Figura 17

Pruebas al variar el número de neuronas y épocas



Con 50 neuronas y 1000 épocas se obtiene el mejor resultado de clasificación con un porcentaje del 92.96%, como se muestra en la Figura 17. Se procede a realizar pruebas con otros clasificadores.

Pruebas con KNN

El resultado mostrado en la Tabla 8 muestra que con 13 vecinos se obtiene un porcentaje de clasificación de 91.5, con 3 vecinos como se observa en la Tabla 9, se obtiene un de clasificación del 91.8%; con esto se observa que el número de vecinos no presenta una variación muy alta en la clasificación.

Tabla 8*Prueba KNN 13 vecinos*

Parámetros KNN (13 Vecinos)								
	Clase	Predicción			Métricas			
		VT	LP	Otros	Recall%	SP%	A%	Ber
Observación	VT	16	14	0	53.33	98.2	94.4	0.243
	LP	3	307	3	98.08	50	92.4	0.26
	Otros	3	7	2	16.67	99.1	96.3	0.421
	P%	72.73	93.60	40.00	Overall Average (%)			91.5

Tabla 9 Prueba*KNN con 3 vecinos*

Parámetros KNN (3 Vecinos)								
	Clase	Predicción			Métricas			
		VT	LP	Otros	Recall%	SP%	A%	Ber
Observación	VT	18	11	1	60	97.8	94.6	0.211
	LP	5	304	4	97.12	59.5	92.7	0.217
	Otros	2	6	4	33.33	98.5	96.3	0.341
	P%	72	94.70	44.44	Overall Average (%)			91.8

Pruebas con Decision Tree

Estas pruebas se observan en la Tabla 10 y 11, las cuales se realizaron pruebas con 10 y 3 divisiones, los cuales tiene un porcentaje de clasificación del 91.3. en las Figuras 18 y 19 se puede observar el proceso de decisión que realiza.

Tabla 10

Prueba DT con 10 número máximo de divisiones

Parámetros Decision Tree (10 Divisiones)								
Observación	Predicción			Métricas				
	Clase	VT	LP	Otros	Recall%	SP%	A%	Ber
	VT	21	9	0	70	97.5	95.2	0.162
	LP	6	303	4	96.81	54.8	91.8	0.242
	Otros	2	10	0	0	98.8	95.5	0.506
	P%	72.41	94.10	0.00	Overall Average (%)			91.3

Figura 18

Árbol de decisión con 10 número máximo de divisiones

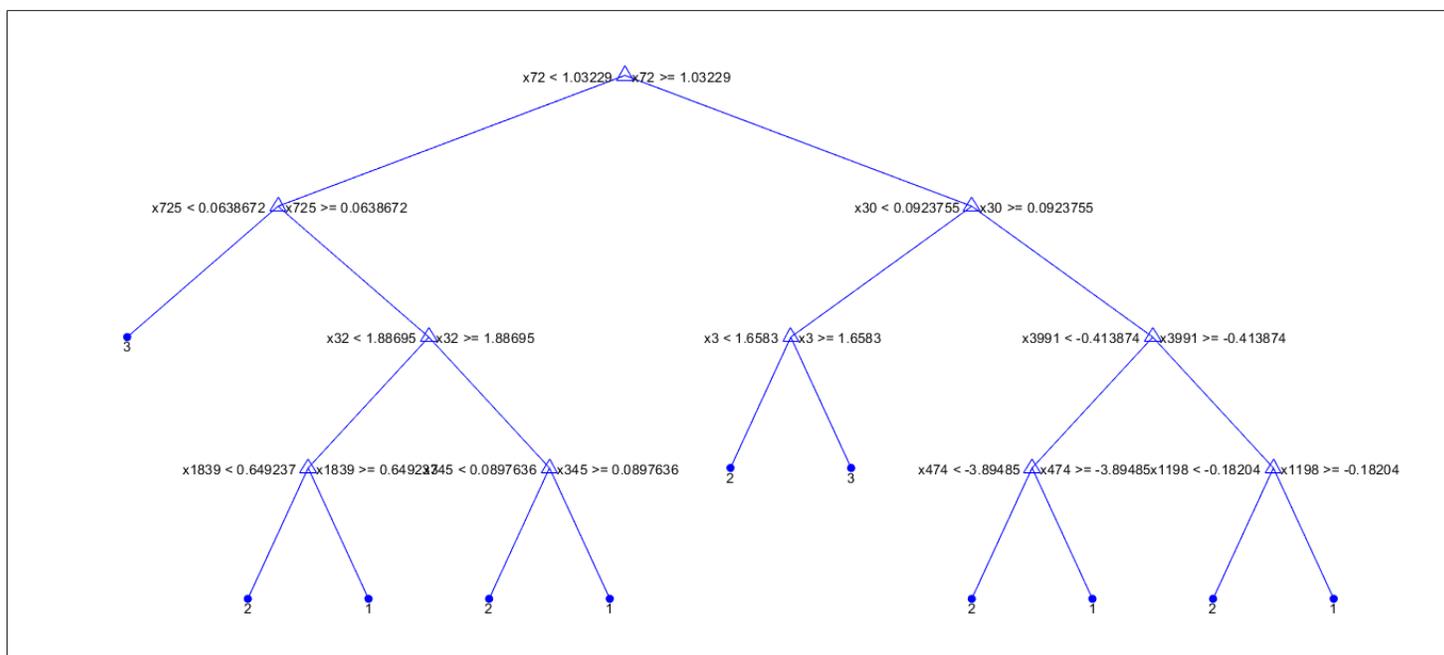


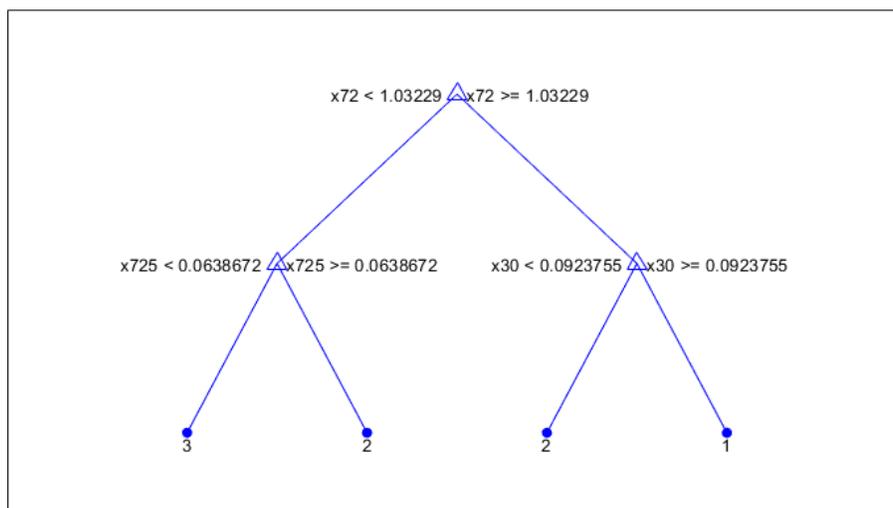
Tabla 11

Prueba DT con 3 número máximo de divisiones

Parámetros Decision Tree (3 Divisiones)								
Observación	Clase	Predicción			Métricas			
		VT	LP	Otros	Recall%	SP%	A%	Ber
VT	VT	23	7	0	76.67	96.6	94.9	0.134
LP	LP	8	301	4	96.17	61.9	92.1	0.21
Otros	Otros	3	9	0	0	98.8	95.5	0.506
P%		67.65	94.95	0.00	Overall Average (%)			91.3

Figura 19

Árbol de decisión con 3 número máximo de divisiones



Pruebas de reducción de características

Pruebas con MI

La prueba realizada con el selector MI, el cual permite obtener las mejores características, con esto se obtiene las 100 mejores características para poder realizar las pruebas con los clasificadores. Las características son mostradas en la Tabla 12.

En la Tabla 13 se observa la prueba con el Autoencoder, la cual tiene un porcentaje de clasificación del 89.86, estas características no son tan relevantes para la clasificación.

Tabla 12*Prueba de selección de características con MI*

°	Feature	Number	Feature	Number	Feature	Number	Feature	Number	Feature
1	72	21	2814	41	1492	61	2812	81	834
2	37	22	1482	42	1820	62	829	82	1837
3	65	23	1484	43	1489	63	1479	83	1817
4	58	24	2805	44	1823	64	1807	84	2157
5	30	25	1493	45	2813	65	166	85	1803
6	24	26	1826	46	2809	66	833	86	1805
7	1814	27	2808	47	1488	67	3146	87	820
8	1812	28	2815	48	2816	68	2811	88	1506
9	1813	29	1811	49	3148	69	2145	89	1486
10	1483	30	1494	50	2152	70	821	90	158
11	1824	31	2819	51	831	71	167	91	3468
12	1815	32	2817	52	1480	72	3138	92	1829
13	1816	33	2806	53	2155	73	1819	93	1809
14	2807	34	1485	54	3145	74	2153	94	1498
15	1490	35	1491	55	2803	75	3469	95	3149
16	1821	36	1810	56	1487	76	171	96	1505
17	1825	37	2818	57	169	77	1808	97	827
18	1822	38	1818	58	2804	78	1827	98	2156
19	1481	39	159	59	2820	79	3150	99	3151
20	1495	40	828	60	1496	80	1804	100	2830

Tabla 13*Prueba de Autoencoder con MI*

Parámetros de Autoencoder Simple							
Neuronas	Épocas	ETF	DTF	L2WR	SR	SP	SD
50	1000	satlin	satlin	0.004	1	0.05	FALSE
		Predicción			Métricas		
	Clase	VT	LP	Otros	Recall%	SP%	A% Ber
	VT	15	8	1	62.5	95.5	93.2 0.21
Observación	LP	15	300	7	93.17	60.6	90.1 0.231
	Otros	0	5	4	44.44	97.7	96.3 0.289
	P%	50	95.85	33.33	Overall Average (%)		89.86

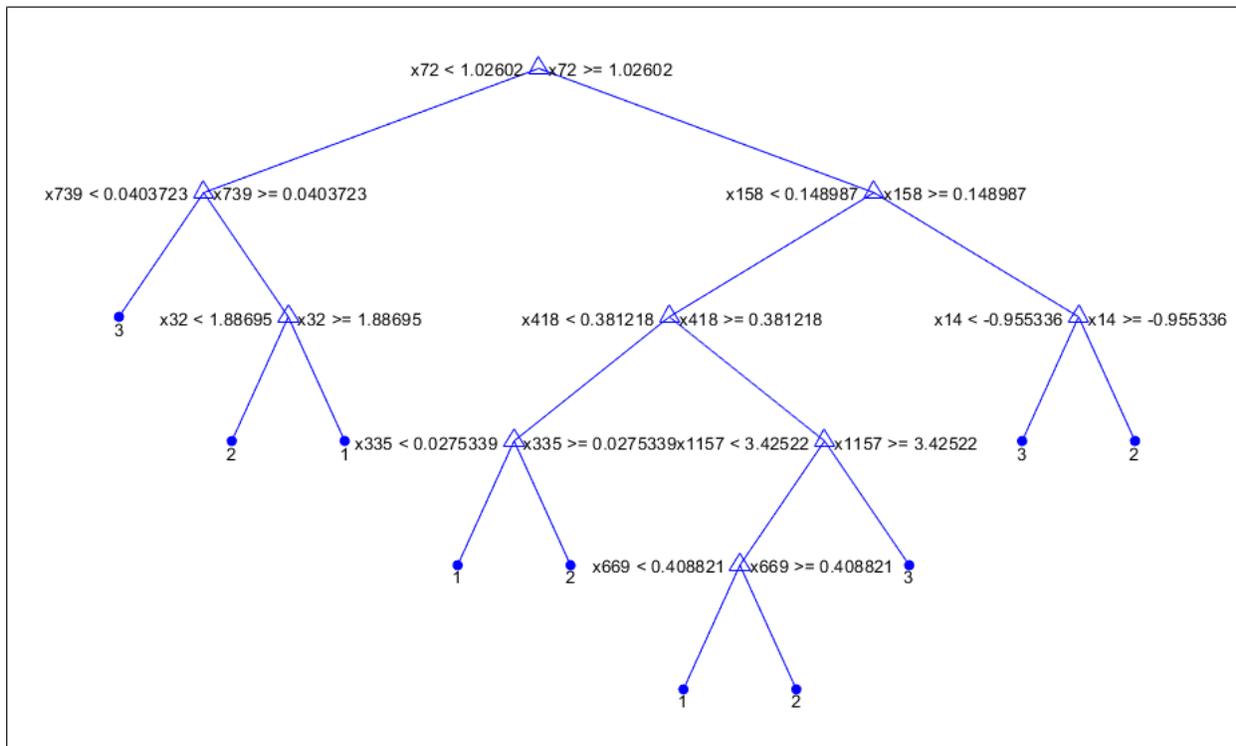
Tabla 14*Prueba de KNN con MI*

Parámetros KNN (3 Vecinos)								
	Clase	Predicción			Métricas			
		VT	LP	Otros	Recall%	SP%	A%	Ber
Observación	VT	17	11	2	56.67	97.5	94.1	0.229
	LP	6	304	3	97.12	54.8	92.1	0.241
	Otros	2	8	2	16.67	98.5	95.8	0.424
	P%	68	94.12	28.57	Overall Average (%)			91.0

Tabla 15*Prueba de DT con MI*

Parámetros Decision Tree (10 Divisiones)								
	Clase	Predicción			Métricas			
		VT	LP	Otros	Recall%	SP%	A%	Ber
Observación	VT	12	18	0	40	97.5	92.7	0.312
	LP	5	307	1	98.08	42.9	91.5	0.295
	Otros	3	6	3	25	99.7	97.2	0.376
	P%	60	92.75	75.00	Overall Average (%)			90.7

En la Tabla 14, se tiene la prueba realizada con KNN con un porcentaje de clasificación de 91. La prueba realizada con DT mostrada en la Tabla 15, se observa un porcentaje de clasificación de 90.7, en la Figura 20; se muestra el proceso de decisión que tiene este clasificador.

Figura 20*Prueba de DT con MI****Pruebas con SD***

Esta prueba se muestra en la Tabla 16, en la cual se seleccionan 100 primeras características, estas ingresan a los clasificadores y se obtienen los resultados mostrados en las Tablas 17, 18 y 19, las cuales corresponden a pruebas con Autoencoder, KNN y DT respectivamente. En la Tabla 17 se obtiene un porcentaje de clasificación de 89.86, en la Tabla 18 un porcentaje de clasificación de 90.7, y en la Tabla 19 un porcentaje de 92. En la Figura 21 se muestra el proceso de selección que tiene DT en esta prueba.

Tabla 16*Prueba de selección de características con SD*

Number	Feature								
1	72	21	1826	41	1492	61	1809	81	159
2	37	22	1484	42	1805	62	2813	82	1819
3	65	23	2807	43	2816	63	2157	83	3136
4	58	24	2819	44	833	64	1489	84	821
5	30	25	2806	45	1480	65	1479	85	2812
6	24	26	1822	46	2804	66	2809	86	3137
7	1814	27	1821	47	3469	67	2820	87	2144
8	1813	28	1490	48	3148	68	2803	88	1487
9	1812	29	1494	49	2815	69	1496	89	3481
10	1824	30	2817	50	1820	70	2152	90	1800
11	1815	31	2818	51	831	71	1806	91	1827
12	1483	32	1803	52	2155	72	169	92	158
13	2805	33	2808	53	1818	73	2145	93	820
14	1481	34	1807	54	31	74	3145	94	3151
15	1495	35	1810	55	828	75	1817	95	1802
16	1825	36	1804	56	3468	76	1488	96	2811
17	1816	37	2814	57	1491	77	3138	97	829
18	1482	38	1485	58	3150	78	1486	98	1478
19	1493	39	1823	59	171	79	2143	99	166
20	1811	40	1808	60	3467	80	834	100	2158

Tabla 17*Prueba de Autoencoder con SD*

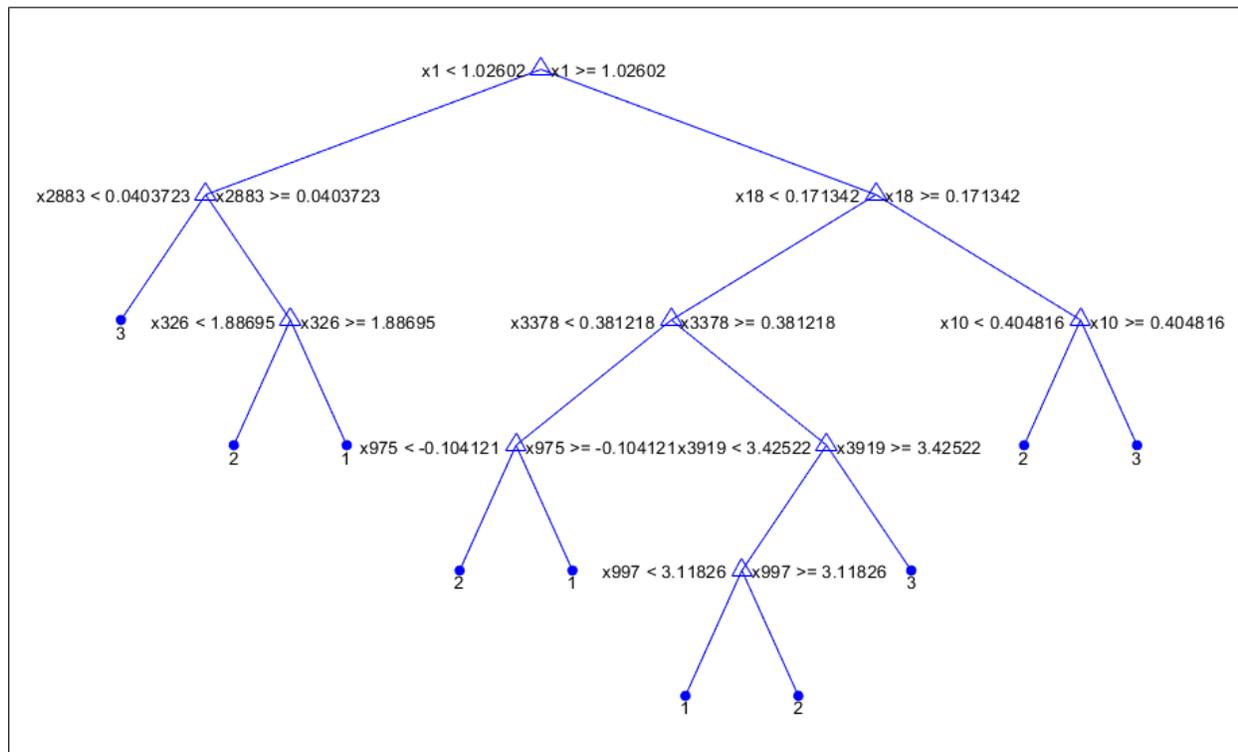
Parámetros de Autoencoder Simple								
Neuronas	Épocas	ETF	DTF	L2WR	SR	SP	SD	
50	1000	satlin	satlin	0.004	1	0.05	FALSE	
		Predicción			Métricas			
	Clase	VT	LP	Otros	Recall%	SP%	A%	Ber
	VT	15	8	1	62.5	95.5	93.2	0.21
Observación	LP	15	300	7	93.17	60.6	90.1	0.231
	Otros	0	5	4	44.44	97.7	96.3	0.289
	P%	50	95.85	33.33	Overall Average (%)		89.86	

Tabla 18*Prueba KNN con SD*

Parámetros KNN (3 Vecinos)								
	Clase	Predicción			Métricas			
		VT	LP	Otros	Recall%	SP%	A%	Ber
Observación	VT	19	10	1	63.33	96.9	94.1	0.199
	LP	9	302	2	96.49	52.4	91.3	0.256
	Otros	1	10	1	8.333	99.1	96.1	0.463
	P%	65.52	93.79	25.00	Overall Average (%)			90.7

Tabla 19*Prueba DT con SD*

Parámetros Decision Tree (10 Divisiones)								
	Clase	Predicción			Métricas			
		VT	LP	Otros	Recall%	SP%	A%	Ber
Observación	VT	19	11	0	63.33	98.2	95.2	0.193
	LP	5	306	2	97.76	57.1	93	0.225
	Otros	1	7	4	33.33	99.4	97.2	0.336
	P%	76	94.44	66.67	Overall Average (%)			92.7

Figura 21*Prueba de DT con SD****Pruebas con SFS***

Esta prueba se realiza la selección de características y es mostrada en la Tabla 20, esta cuenta con una selección de 51 características, pruebas con autoencoder son mostradas en la Tabla 21, y tiene un porcentaje de clasificación de 90.7, en la Tabla 22 con KNN se tiene un porcentaje de clasificación de 90.4, y con DT se muestra en la Tabla 23, la cual tiene un porcentaje de clasificación de 89.9.

Tabla 20*Prueba de selección de características con SFS*

Number	Feature	Number	Feature	Number	Feature
1	72	18	1265	35	1080
2	1475	19	1316	36	1081
3	492	20	4375	37	1085
4	2711	21	416	38	1088
5	2795	22	417	39	1094
6	2691	23	418	40	1090
7	2632	24	419	41	1105
8	2836	25	420	42	1108
9	1277	26	421	43	1351
10	4311	27	422	44	1102
11	3879	28	423	45	430
12	4695	29	424	46	440
13	4519	30	427	47	1095
14	4277	31	615	48	1109
15	3915	32	425	49	1406
16	1275	33	429	50	1124
17	2507	34	1078	51	431

Tabla 21*Prueba de Autoencoder con SFS*

Parámetros de Autoencoder Simple								
Neuronas	Épocas	ETF	DTF	L2WR	SR	SP	SD	
50	1000	satlin	satlin	0.004	1	0.05	FALSE	
		Predicción			Métricas			
	Clase	VT	LP	Otros	Recall%	SP%	A%	Ber
	VT	18	12	0	60	96.3	93.2	0.218
Observación	LP	12	297	5	94.59	61	90.7	0.222
	Otros	0	4	7	63.64	98.5	97.5	0.189
	P%	60	94.89	58.33	Overall Average (%)		90.7	

Tabla 22*Prueba de KNN con SFS*

Parámetros KNN (3 Vecinos)								
Observación	Clase	Predicción			Métricas			
		VT	LP	Otros	Recall%	SP%	A%	Ber
	VT	14	15	1	46.67	96.9	92.7	0.282
	LP	10	300	3	95.85	52.4	90.7	0.259
	Otros	0	5	7	58.33	98.8	97.5	0.214
	P%	58.33	93.75	63.64	Overall Average (%)			90.4

Tabla 23*Prueba de DT con SFS*

Parámetros Decision Tree (10 Divisiones)								
Observación	Clase	Predicción			Métricas			
		VT	LP	Otros	Recall%	SP%	A%	Ber
	VT	13	17	0	43.33	97.5	93	0.296
	LP	7	304	2	97.12	38.1	90.1	0.324
	Otros	1	9	2	16.67	99.4	96.6	0.42
	P%	61.9	92.12	50.00	Overall Average (%)			89.9

Pruebas con SFFS

Esta prueba se realiza con los resultados mostrados en la Tabla 24, la cual permite observar las características de SFFS, Estas características ingresan al Autoencoder y se observa el resultado en la Tabla 25, con un porcentaje de 89.86, se realiza la prueba con KNN, y se obtiene un porcentaje de 91.3 mostrada en la Tabla 26, y el porcentaje de clasificación de DT es de 91.3 mostrada en la Tabla 27.

Tabla 24*Prueba de selección de características con SFFS*

Number	Feature	Number	Feature	Number	Feature
1	72	9	4463	17	4008
2	492	10	2458	18	4255
3	1480	11	686	19	4114
4	335	12	1245	20	4224
5	1277	13	3983	21	1305
6	507	14	2821	22	1393
7	1116	15	3889	23	1078
8	1253	16	1265		

Tabla 25*Prueba de Autoencoder con SFFS*

Parámetros de Autoencoder Simple								
Neuronas	Épocas	ETF	DTF	L2WR	SR	SP	SD	
50	1000	satlin	satlin	0.004	1	0.05	FALSE	
	Predicción			Métricas				
	Clase	VT	LP	Otros	Recall%	SP%	A%	Ber
	VT	15	8	0	65.22	95.5	93.5	0.197
Observación	LP	14	301	9	92.9	61.3	90.1	0.229
	Otros	1	4	3	37.5	97.4	96.1	0.325
	P%	50	96.17	25.00	Overall Average (%)		89.86	

Tabla 26*Prueba de KNN con SFFS*

Parámetros Decision Tree (10 Vecinos)								
	Predicción			Métricas				
	Clase	VT	LP	Otros	Recall%	SP%	A%	Ber
	VT	20	10	0	66.67	97.2	94.6	0.181
Observación	LP	9	299	5	95.53	59.5	91.3	0.225
	Otros	0	7	5	41.67	98.5	96.6	0.299
	P%	68.97	94.62	50.00	Overall Average (%)		91.3	

Tabla 27*Prueba de DT con SFFS*

Parámetros Decision Tree (10 Divisiones)								
Observación	Clase	Predicción			Métricas			
		VT	LP	Otros	Recall%	SP%	A%	Ber
	VT	18	12	0	60	97.5	94.4	0.212
	LP	8	301	4	96.17	54.8	91.3	0.245
	Otros	0	7	5	41.67	98.8	96.9	0.297
	P%	69.23	94.06	55.56	Overall Average (%)			91.3

Prueba de pesos

Este experimento se inicia sin considerar las características psicoacústicas, con esto el resto de características suman un total de 84. En esta fase, se realiza un filtrado al seleccionar únicamente aquellos valores que superan el umbral de 0.1. Con este criterio, se logra identificar y retener un conjunto de 21 características con MI. Para una comprensión más detallada, se presenta la Tabla 28 que exhibe las características seleccionadas en este proceso.

Tabla 28*Características con peso mayor a 0.1 con MI*

Features	Weights	Features	Weights	Features	Weights
72	0.26111	15	0.152614	39	0.131857
37	0.227736	68	0.147615	71	0.124101
65	0.224041	67	0.143155	40	0.121239
58	0.222199	61	0.142364	14	0.118067
30	0.215972	60	0.141186	59	0.11283
24	0.213929	33	0.133587	66	0.112023
31	0.155035	75	0.132079	25	0.100488

En la Tabla 29 se muestran las características obtenidas de la selección con MI que poseen valores mayores a 0.8, de las características psico acústicas. Mostradas en la Tabla 30,

con un total de 14 características. Se procede a realizar una comparación con los resultados a ingresar a un clasificador que son mostrados en la Tabla 31 y 32.

Tabla 29

Características con peso mayor a 0.8 con MI

Features	Weights	Features	Weights	Features	Weights	Features	Weights
2722	0.09438929	2392	0.08548363	76	0.08352541	2060	0.08148612
1399	0.09331912	1727	0.08521766	2386	0.08324773	737	0.08125461
2721	0.09265455	1731	0.08503065	3054	0.08313277	1401	0.08123596
1729	0.09043479	2389	0.08482114	2061	0.08307738	2380	0.08121038
1398	0.09042028	1391	0.08479087	2715	0.08301194	2390	0.08115759
1730	0.09032792	2394	0.08468513	2385	0.08274234	1732	0.08104223
2723	0.08996703	65	0.08459269	1714	0.08234961	1396	0.08062149
1728	0.08923663	1722	0.08450126	1392	0.0822571	2718	0.08057575
1397	0.08809508	736	0.08446187	2391	0.08220132	1407	0.08023441
68	0.08791875	1395	0.08433543	2719	0.08205017	2720	0.08003391
75	0.0874148	2724	0.08417661	4375	0.08190466		
74	0.08624632	3053	0.08401603	64	0.08170733		
1400	0.08568799	1738	0.08364701	1724	0.0816987		

Tabla 30

Características con peso mayor a 0.8 con MI

Features	Weights	Features	Weights	Features	Weights	Features	Weights
72	0.26111	15	0.152614	39	0.131857	32	0.099578
37	0.227736	68	0.147615	71	0.124101	28	0.092275
65	0.224041	67	0.143155	40	0.121239	34	0.091267
58	0.222199	61	0.142364	14	0.118067	73	0.086993
30	0.215972	60	0.141186	59	0.11283	74	0.086103
24	0.213929	33	0.133587	66	0.112023	44	0.086059
31	0.155035	75	0.132079	25	0.100488	79	0.084062

En la Tabla 31, se muestra el resultado de clasificación con referencia a las características de la Tabla 29. Tiene un resultado de clasificación del 84.79%, este un resultado

el cual no tiene un alto porcentaje de clasificación con las otras pruebas, se observa que estas características no son tan relevantes para definir cada tipo de señal.

Tabla 31

Prueba de Autoencoder con 49 características seleccionadas con MI

Parámetros de Autoencoder Simple									
Neuronas	Épocas	ETF	DTF	L2WR	SR	SP	SD		
50	1000	satlin	satlin	0.004	1	0.05	FALSE		
				Predicción		Métricas			
		Clase	VT	LP	Otros	Recall%	SP%	A%	Ber
		VT	7	14	0	33.33	93.1	89.6	0.368
Observación	LP		23	294	12	89.36	26.9	84.8	0.419
	Otros		0	5	0	0	96.6	95.2	0.517
	P%		23.33	93.93	0.00	Overall Average (%)		84.79	

Tabla 32

Prueba de Autoencoder con 21 características seleccionadas con MI

Parámetros de Autoencoder Simple									
Neuronas	Épocas	ETF	DTF	L2WR	SR	SP	SD		
50	1000	satlin	satlin	0.004	1	0.05	FALSE		
				Predicción		Métricas			
		Clase	VT	LP	Otros	Recall%	SP%	A%	Ber
		VT	23	10	0	69.7	97.8	95.2	0.162
Observación	LP		5	299	7	96.14	68.2	92.7	0.178
	Otros		2	4	5	45.45	98	96.3	0.283
	P%		76.67	95.53	41.67	Overall Average (%)		92.11	

En la Tabla 32 se observa el resultado de clasificación de las características mostradas en la Tabla 28. Se obtiene un porcentaje de clasificación del 92.11%. Se realiza esta prueba debido a que se tiene un bajo porcentaje en el resultado de la Tabla 28, se observa que con

menos características se obtiene un mejor resultado, en este caso con un número mayor de peso.

Tabla 33

Prueba de Autoencoder con 28 características seleccionadas con MI

Parámetros de Autoencoder Simple									
Neuronas	Épocas	ETF	DTF	L2WR	SR	SP	SD		
50	1000	satlin	satlin	0.004	1	0.05	FALSE		
		Predicción			Métricas				
		Clase	VT	LP	Otros	Recall%	SP%	A%	Ber
		VT	21	6	2	72.41	97.2	95.2	0.152
Observación		LP	9	302	6	95.27	71.1	92.7	0.168
		Otros	0	5	4	44.44	97.7	96.3	0.289
		P%	70	96.49	33.33	Overall Average (%)		92.11	

En la Tabla 33, se muestra el resultado de clasificación con las características de la Tabla 30, se tiene un 92.11%.

Pruebas de interfaz gráfica

Para esta prueba se varían los parámetros para encontrar los mejores resultados, entre los cuales, se realiza en tres fases. La Fase 1 consta de realizar las pruebas sin feature selection por lo cual seleccionamos Ninguno, este con el fin de poder determinar con todas las características que parámetro de cada uno de los clasificadores se obtiene mejores resultados, además de realizar las pruebas de los selectores SFS y SFFS, ya que estos no se tomaron en cuenta para la toma de pesos. Con los mejores resultados se procede a determinar las pruebas se varía los niveles de cuantización Q denominándole Fase 2, y finalmente la Fase 3, se varía el número de características referente a los pesos obtenido en feature selection, específicamente MI y SD. Respectivamente las pruebas tomaran el orden de autoencoder, KNN y DT; y en el caso de los selectores SD y MI.

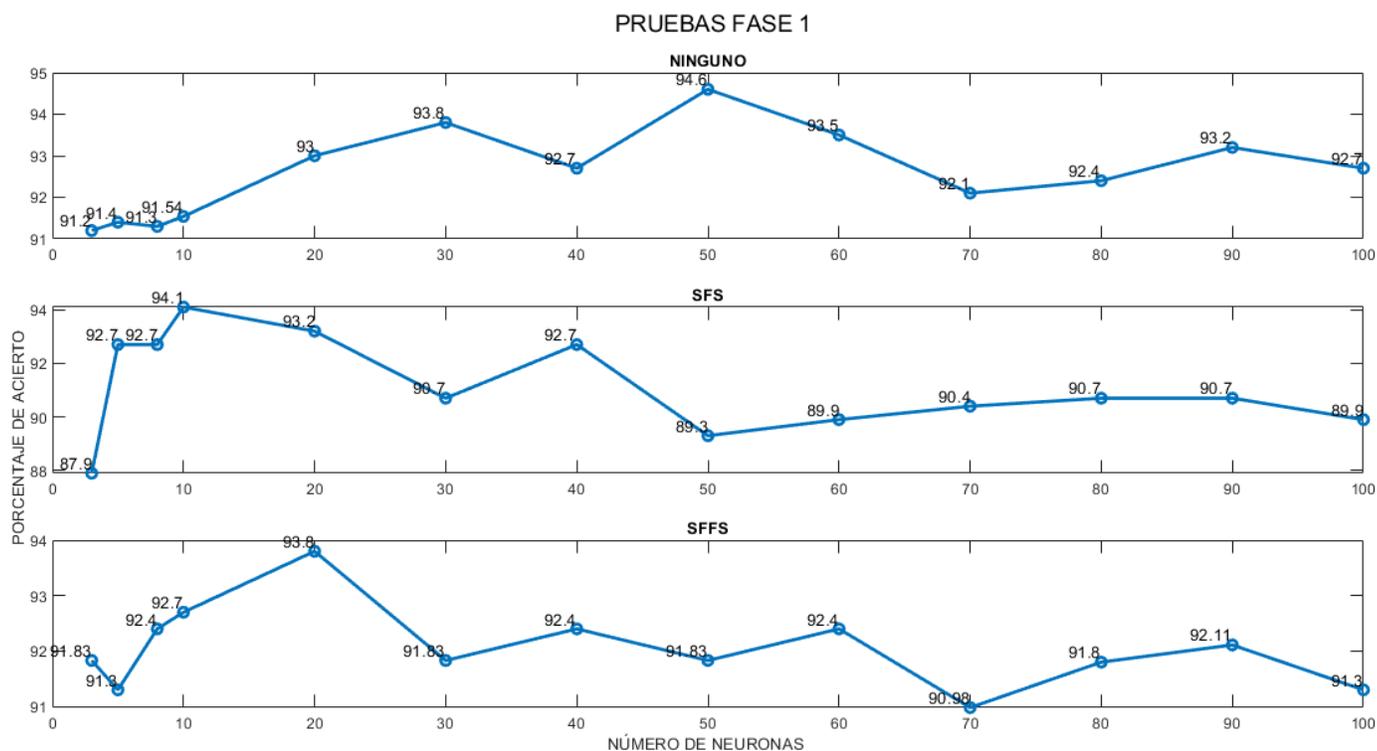
Las pruebas de la Fase 1, son mostradas en las Figuras 22, 23 y 24, en las cuales se detalla el clasificador utilizado, y la variación de los parámetros.

Fase 1

Autoencoder

Figura 22

Pruebas Fase 1 con Autoencoder

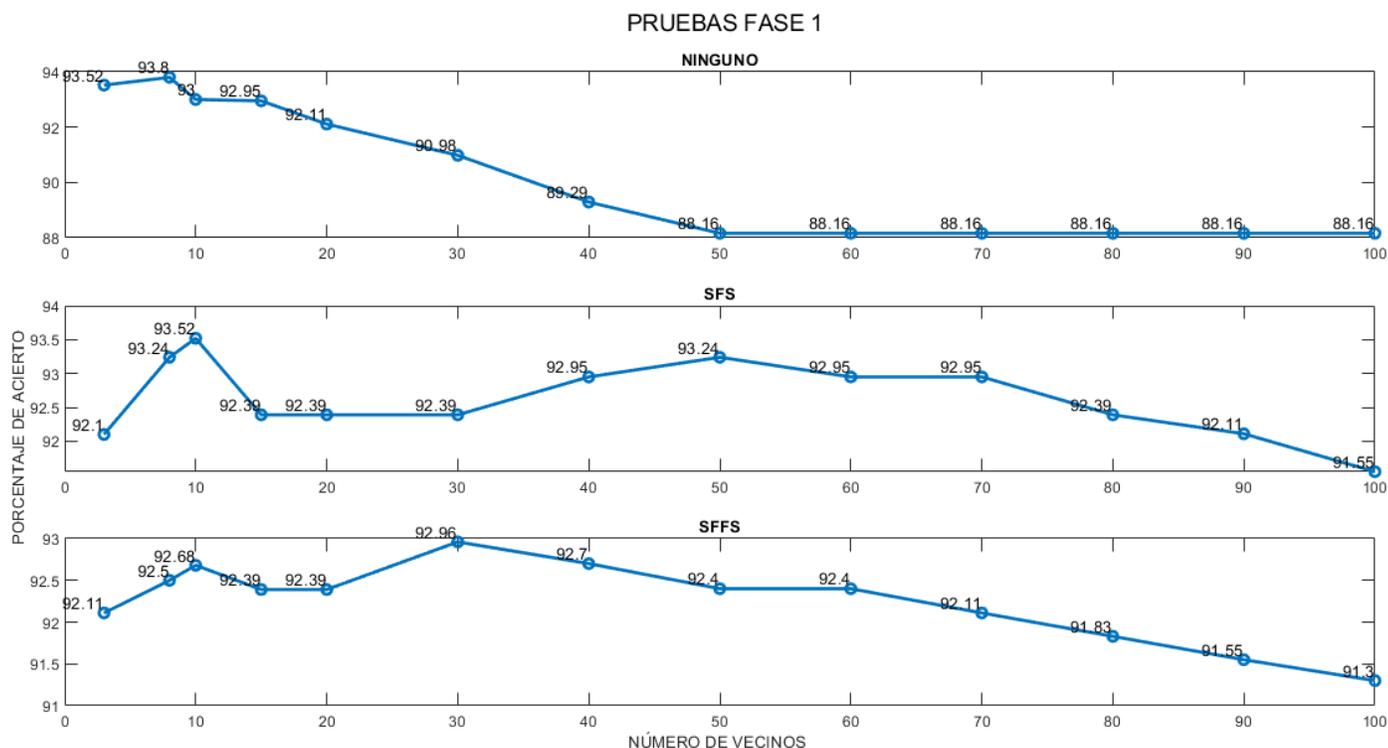


En la Figura 22, se muestra la prueba realizada con Autoencoder donde no se tiene ningún selector, y con los selectores SFS y SFFS. Con ningún selector se tiene el mejor resultado a las 50 neuronas, además observamos que su gráfico es ascendente, toma su máximo porcentaje de acierto en 50 neuronas y tiende a disminuir. Para SFS tiende al mismo comportamiento, toma su máximo a las 10; y de igual para SFFS, a las 20 neuronas.

KNN

Figura 23

Pruebas Fase 1 con KNN

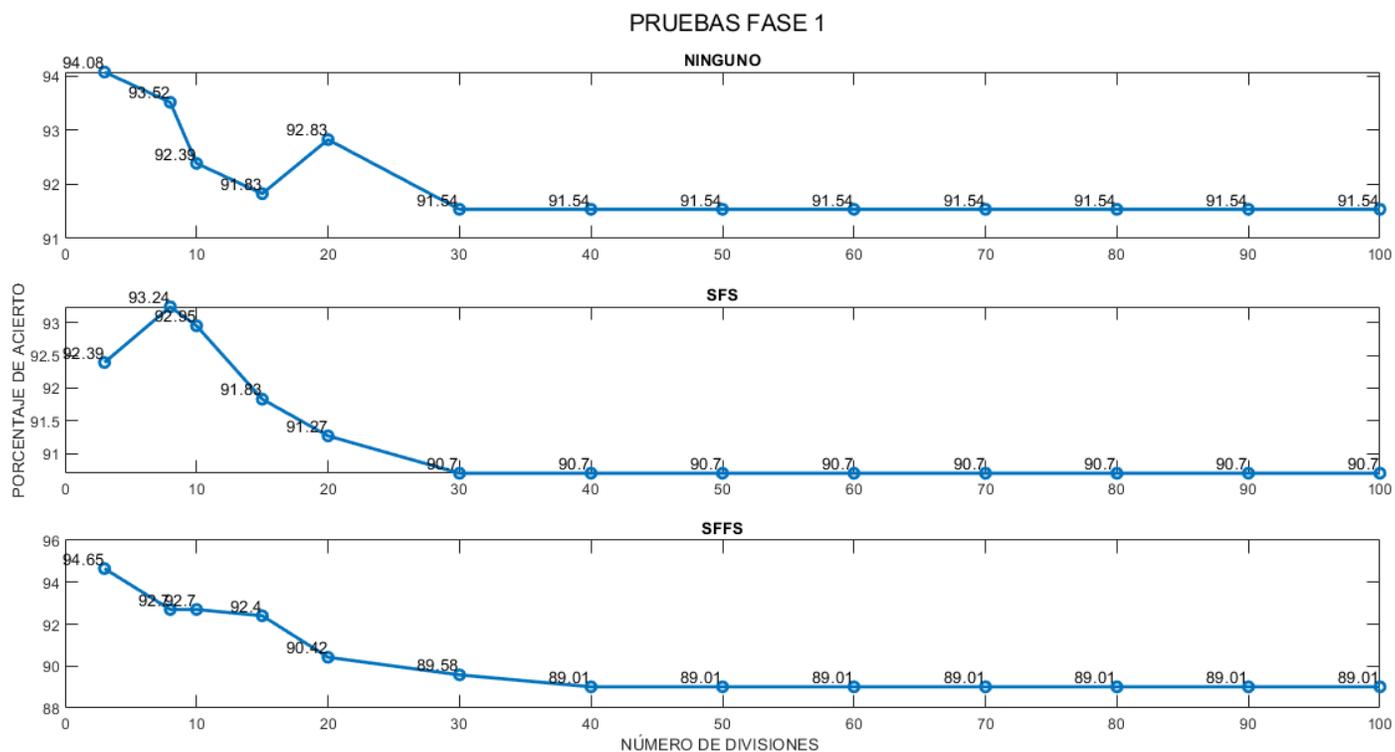


Para la prueba con KNN de la Fase 1, es mostrada en la Figura 23, para esto se observa que con Ninguno tiende a disminuir su porcentaje de acierto, el mayor porcentaje de acierto es de 93.80 y lo obtiene con 8 vecinos. Para SFS de igual manera con 8 vecinos obtiene un porcentaje de acierto del 93.24. Para SFFS, un porcentaje de acierto de 92.96 con 30 vecinos.

DT

Figura 24

Pruebas Fase 1 con DT



La prueba de Fase 1 con DT, es mostrada en la Figura 24, tiende a disminuir al aumentar los valores de divisiones. Con ningún clasificador se tiene que con 3 divisiones un porcentaje de acierto del 94.08. Para SFS, un porcentaje de acierto de 963.24 con 8 divisiones. Para SFFS con 3 divisiones un porcentaje de acierto del 94.65.

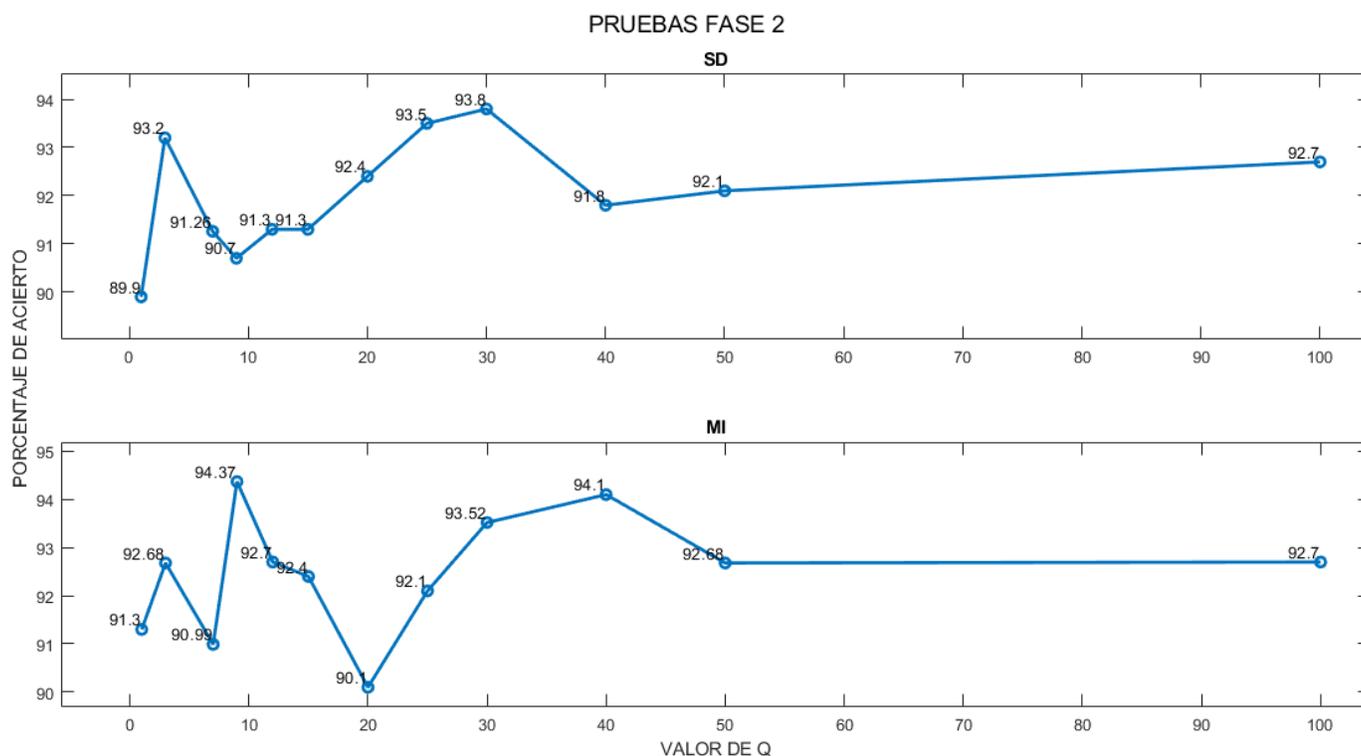
Para las pruebas de la Fase 2, son mostrados en la Figuras 25, 26 y 27 con su respectivo clasificador y se varía los valores de Q de cada selector, de base se utiliza los mejores resultados de la Fase 1.

Fase 2

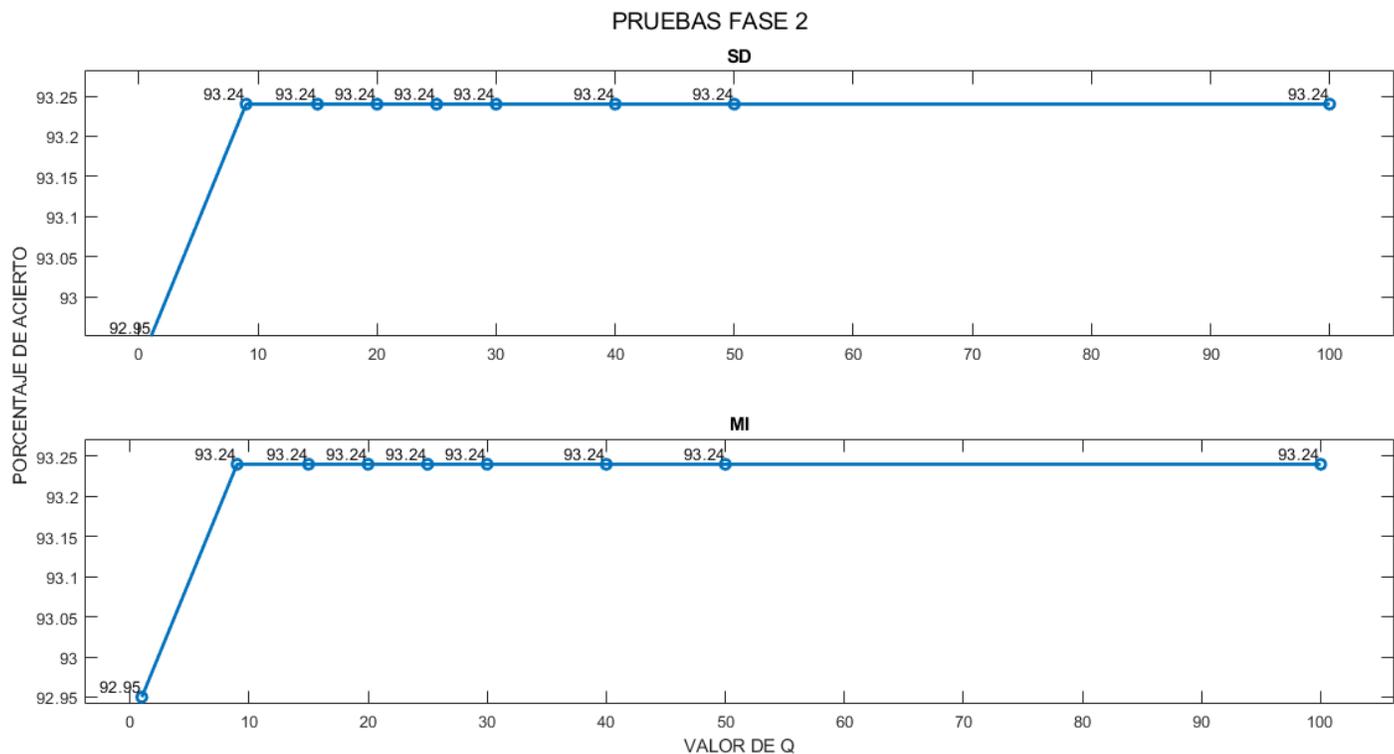
Autoencoder

Figura 25

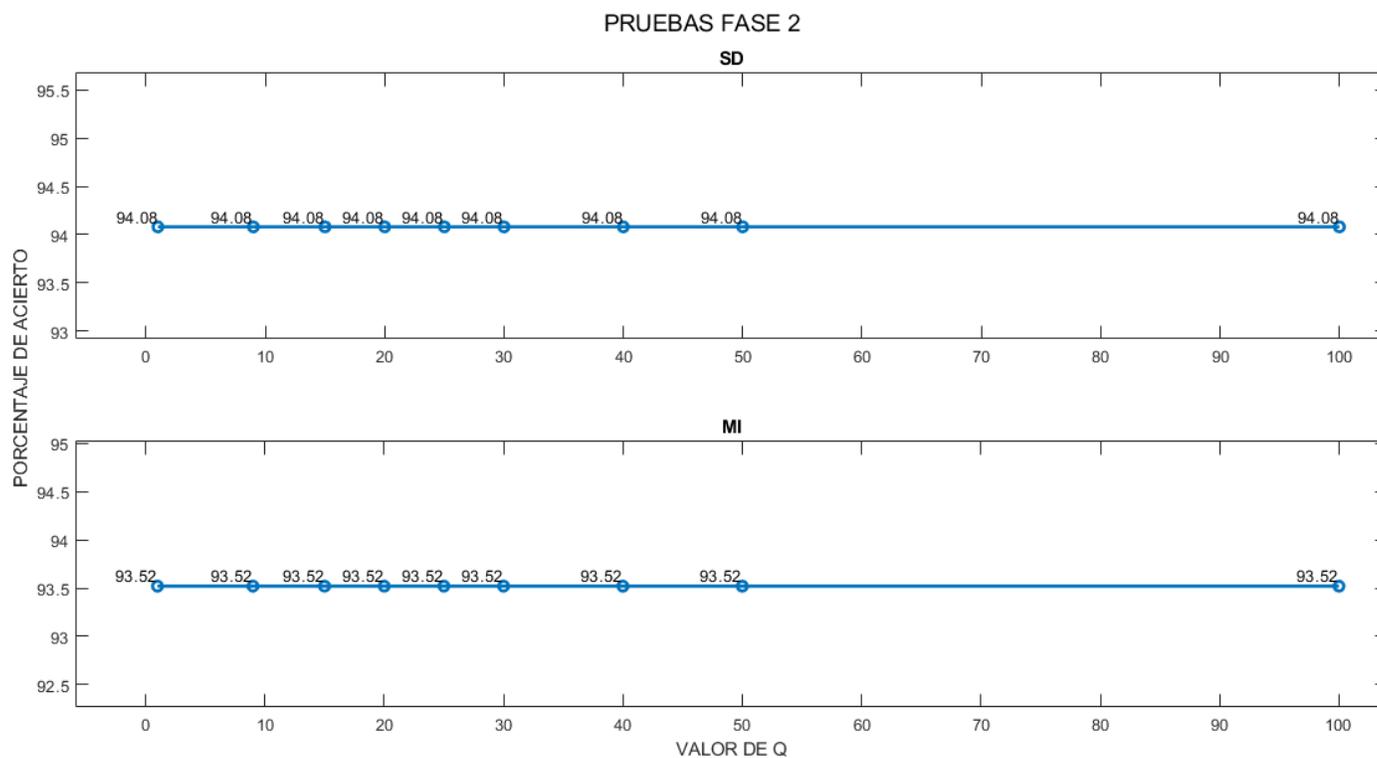
Pruebas Fase 2 con Autoencoder



La prueba de Fase 2 con Autoencoder es mostrada en la Figura 25, en la cual se observa una variación al cambiar Q, para MI el mejor resultado es con Q valor de 30 con porcentaje de acierto de 93.8, para SD con valor de Q igual a 9 un porcentaje de acierto de 94.37.

KNN**Figura 26***Pruebas Fase 2 con KNN*

Para KNN se observa en la Figura 26, que, al variar los niveles de Q, se obtiene valores constantes, menos con un nivel de cuantización. El mejor porcentaje de clasificación con MI y SD es de 93.24. Se observa que no influye en la clasificación el valor de Q.

DT**Figura 27***Pruebas Fase 2 con DT*

Para MI que me observa en la Figura 27 que, al variar los niveles de Q, se obtiene valores constantes. El mejor porcentaje de clasificación con MI es de 93.52 mientras que con SD es de 94.06. Se observa que no influye en la clasificación el valor de Q.

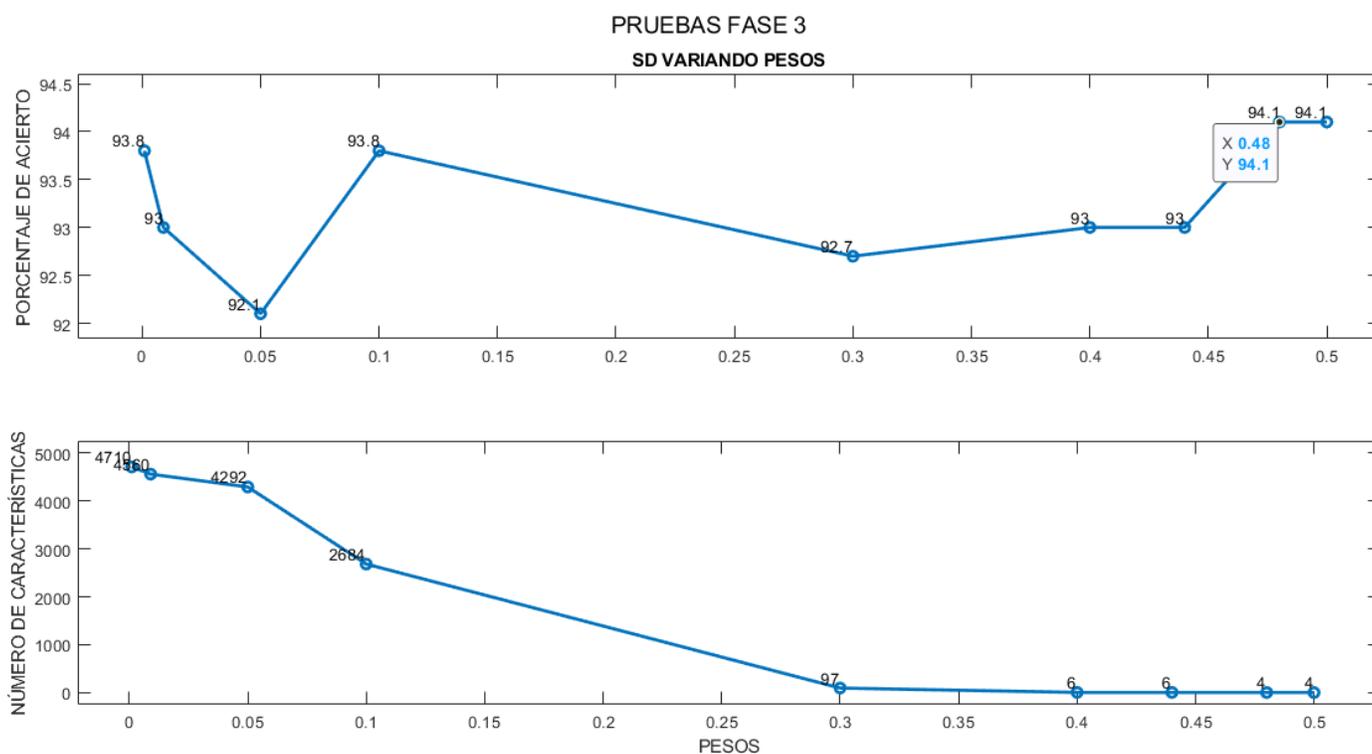
Para las pruebas de la Fase 3, son mostrados a continuación, con su respectivo clasificador y selector, se varía los valores de pesos de cada selector, se toma como base los mejores resultados de la Fase 1 y 2.

Fase 3

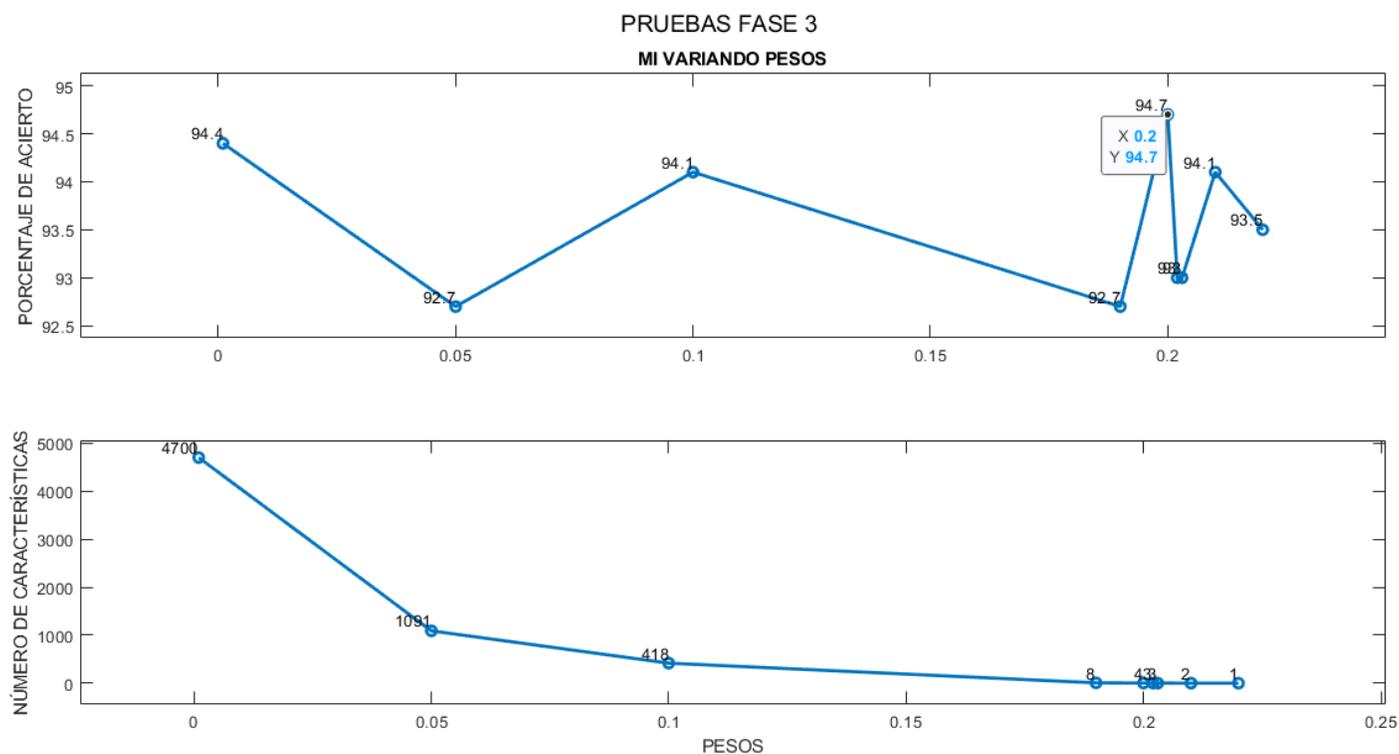
Autoencoder

Figura 28

Pruebas Fase 3 con Autoencoder y SD



Para la prueba de pesos de SD con Autoencoder se observa en la Figura 28, se muestra que se despliega un banner con los valores del mejor resultado de porcentaje de acierto, alcanza un 94.1% con un peso de 0.48, y esto se logra mediante la consideración de un conjunto total de 4 características.

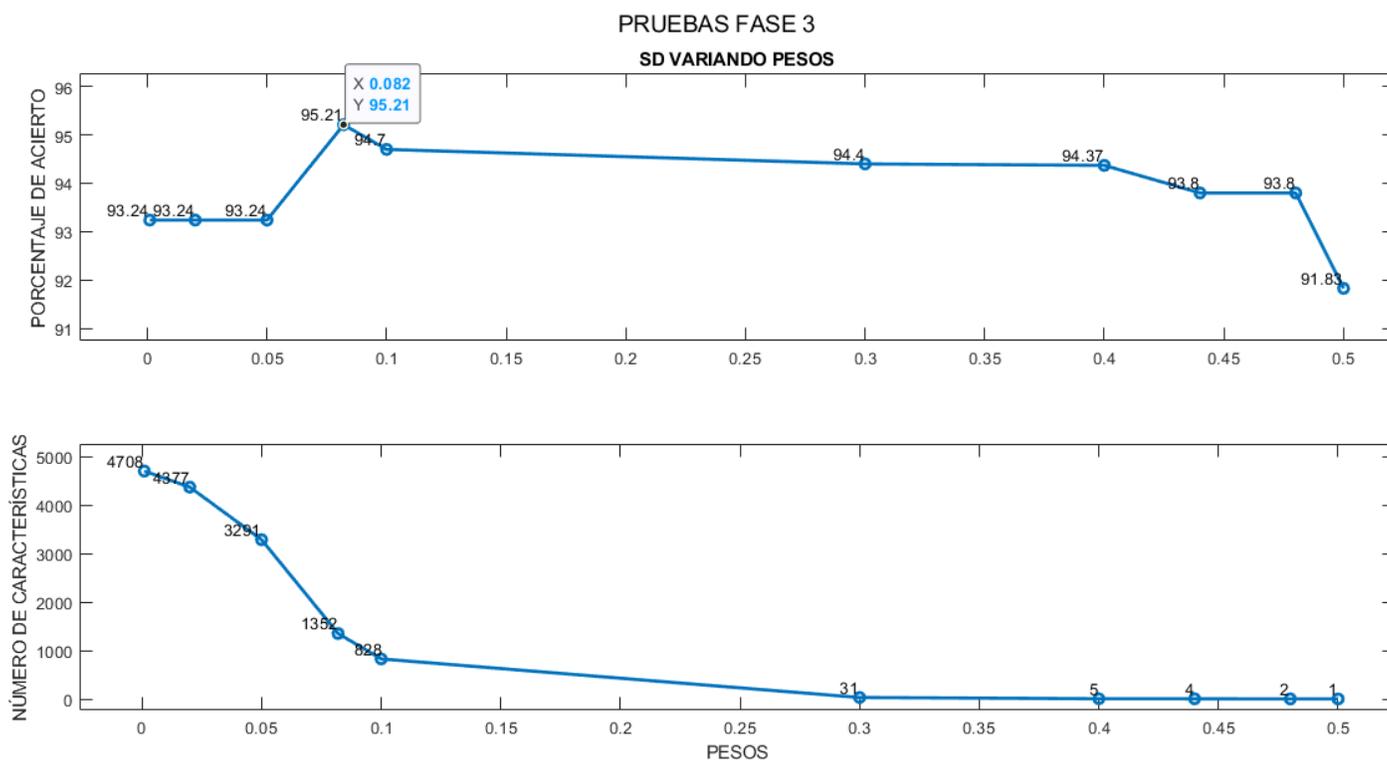
Figura 29*Pruebas Fase 3 con Autoencoder y MI*

Para la prueba de pesos de MI con Autoencoder se observa en la Figura 29, se muestra que se despliega un banner con los valores del mejor resultado de porcentaje de acierto, alcanza un 94.7% con un peso de 0.2, y esto se logra mediante la consideración de un conjunto total de 4 características.

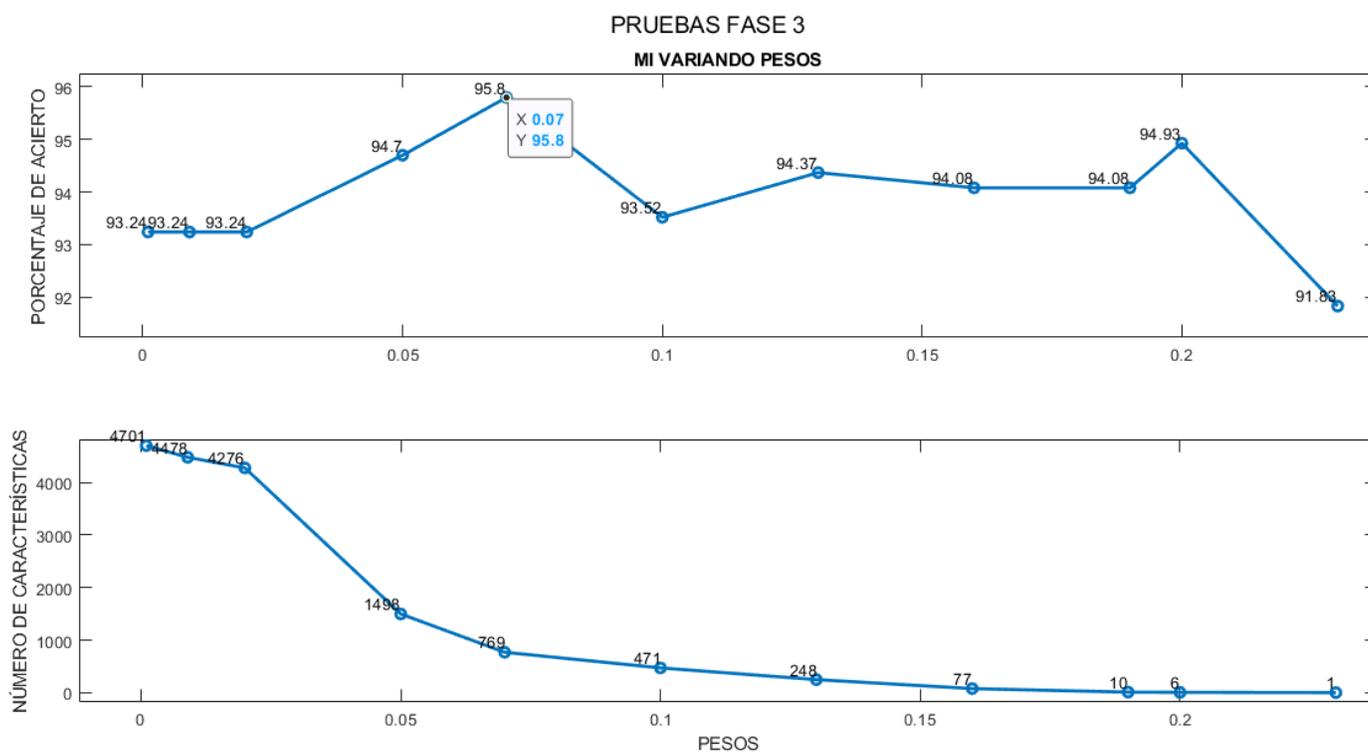
KNN

Figura 30

Pruebas Fase 3 con KNN y SD



Para la prueba de pesos de SD con KNN se observa en la Figura 30, se muestra que se despliega un banner con los valores del mejor resultado de porcentaje de acierto, alcanza un 95.21% con un peso de 0.082, y esto se logra mediante la consideración de un conjunto total de 1352 características.

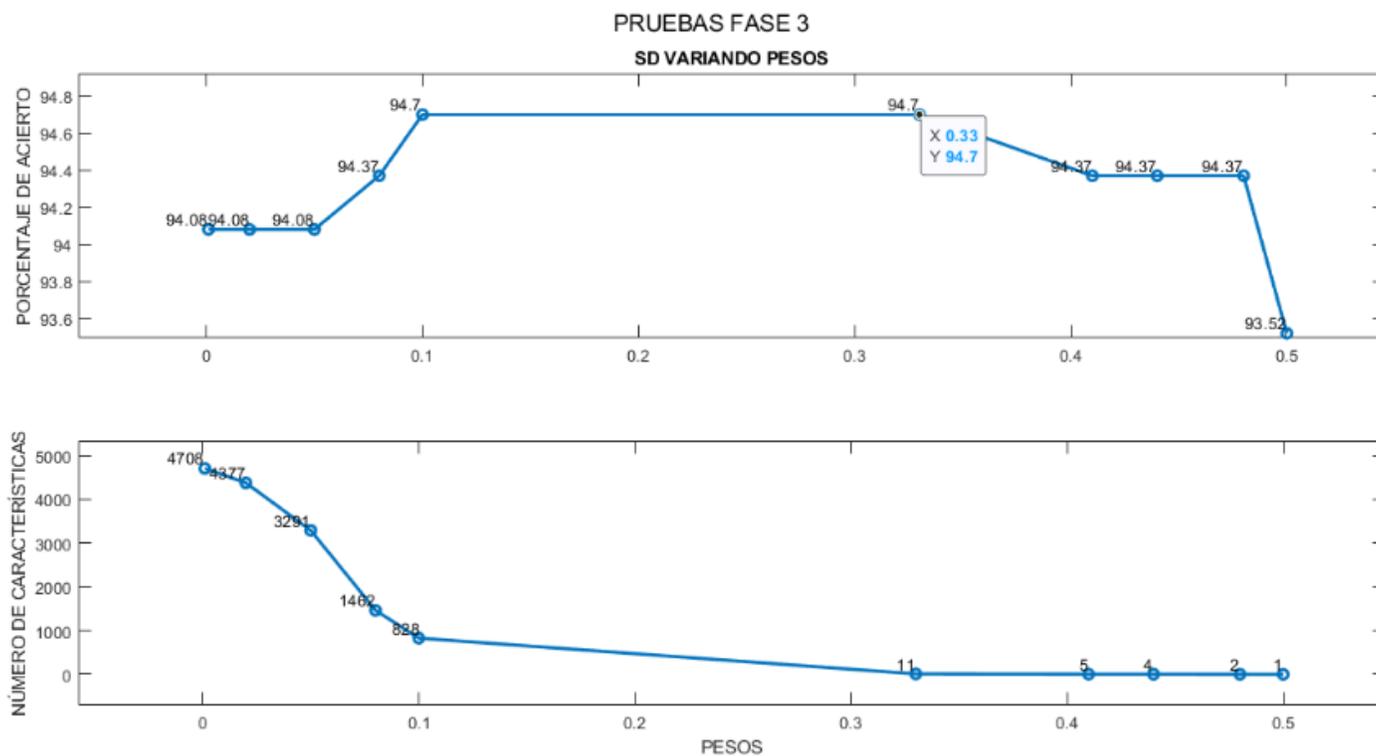
Figura 31*Pruebas Fase 3 con KNN y MI*

Para la prueba de pesos de MI con KNN se observa en la Figura 31, se muestra que se despliega un banner con los valores del mejor resultado de porcentaje de acierto, alcanza un 95.8% con un peso de 0.07, y esto se logra mediante la consideración de un conjunto total de 769 características. El cuál es el mejor resultado obtenido.

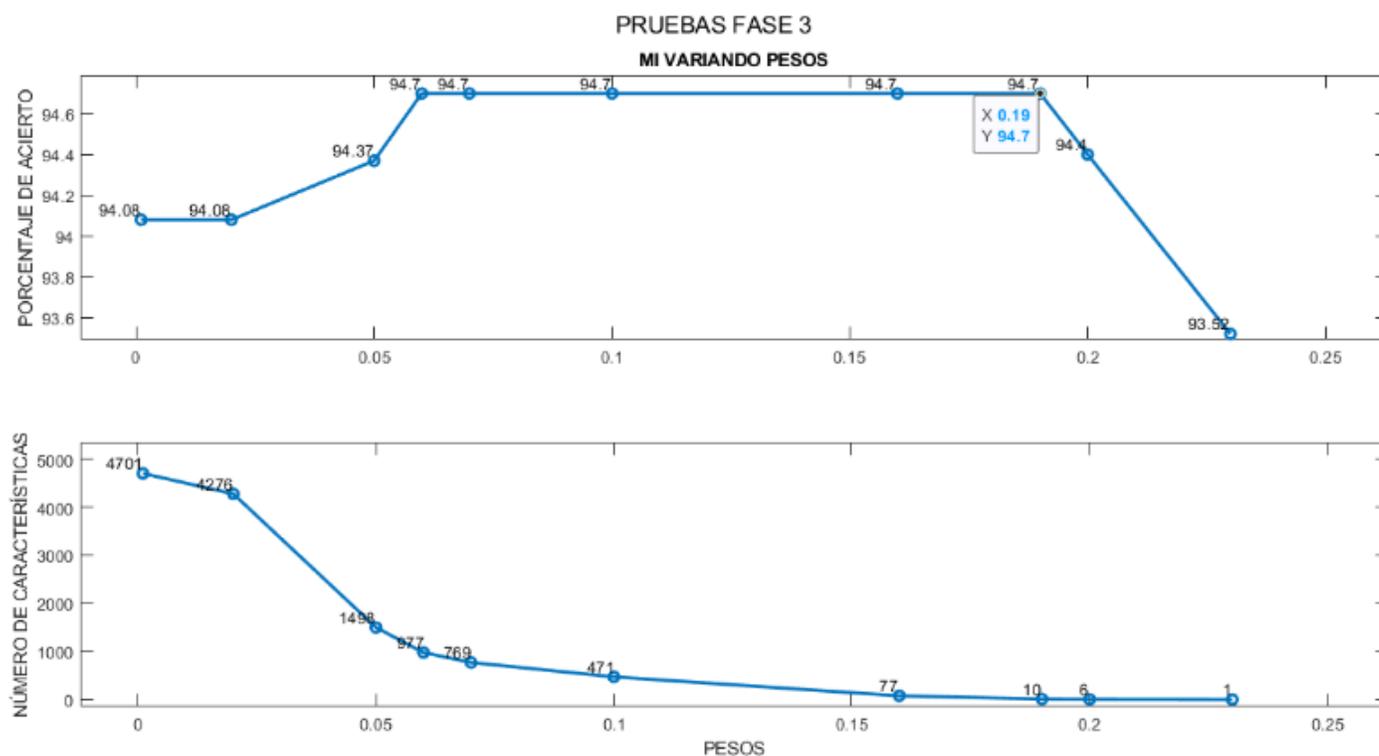
DT

Figura 32

Pruebas Fase 3 con DT y SD



Para la prueba de pesos de SD con DT se observa en la Figura 32, se muestra que se despliega un banner con los valores del mejor resultado de porcentaje de acierto, alcanza un 94.7% con un peso de 0.33, y esto se logra mediante la consideración de un conjunto total de 11 características.

Figura 33*Pruebas Fase 3 con DT y MI*

Para la prueba de pesos de MI con DT se observa en la Figura 33, se muestra que se despliega un banner con los valores del mejor resultado de porcentaje de acierto, alcanza un 94.7% con un peso de 0.19, y esto se logra mediante la consideración de un conjunto total de 10 características.

Resultados

Resultado de pruebas

Los resultados obtenidos en las pruebas de variación de parámetros, permiten el poder definir cada cuales son los mejores de la interfaz gráfica, los resultados de esta son detallados a continuación:

Fase 1

En esta fase permiten obtener con que parámetro es el mejor resultado de clasificación entre los cuales se tiene los siguientes valores:

Tabla 34

Mejores resultados de la Fase 1

Selector	Clasificador	Valor de parámetro del clasificador	Porcentaje de clasificación
SFS	Autoencoder (neuronas)	10	94.1
SFFS		20	93.8
Ninguno		50	94.6
SFS	KNN (vecinos)	8	93.24
SFFS		30	92.96
Ninguno		8	93.8
SFS	DT (divisiones)	8	93.24
SFFS		3	94.65
Ninguno		3	94.08

En la Tabla 34, se tiene como selectores: Ninguno, SFS y SFFS, para estos se toma el mejor resultado del selector Ninguno, ya que con este parámetro obtenido es con todas las características, y nos permite que, al aplicar otros selectores, se varía el número de características determinar las mejores. El valor obtenido de cada clasificador es: Autoencoder con 50 neuronas, KNN con 8 vecinos y DT con 3 divisiones.

Fase 2

Esta fase se varió el valor de los niveles de cuantización Q de cada selector los mejores resultados son los obtenidos a continuación:

Tabla 35*Mejores resultados de Fase 2*

Selector	Clasificador	Valor de parámetro del clasificador	Valor de Q	Porcentaje de clasificación
SD	Autoencoder	50	30	93.8
MI			9	94.37
SD	KNN	8	12	93.24
MI			12	93.24
SD	DT	3	12	94.08
MI			12	93.52

En la Tabla 35 se puede observar que en el único que presenta variación al momento de variar los niveles de cuantización es el Autoencoder con 30 niveles para SD y 9 niveles para MI. Al no encontrar variación en los niveles de cuantización de KNN y DT, se toma el valor por defecto que es 12.

Fase 3

En esta fase se varía el número de características se tiene los siguientes mejores resultados:

Tabla 36*Mejores resultados Fase 3*

Selector	Clasificador	Valor de parámetro del clasificador	Valor de Q	Pesos	Número de características	Porcentaje de clasificación
SD	Autoencoder (neuronas)	50	30	0.48	4	94.10
MI			9	0.2	4	94.64
SD	KNN (vecinos)	8	12	0.082	1352	95.21
MI			12	0.7	769	95.77
SD	DT (divisiones)	3	12	0.33	11	94.64
MI			12	0.19	10	94.64

Para la Fase 3 los mejores resultados mostrados en la Tabla 36, por lo que se analiza el número de características, se tiene que para Autoencoder con SD y MI se toman 4 características, para KNN con SD se toman 1352 y con MI 769; y para DT con SD se toman 11 y MI 10. KNN, los mejores resultados de clasificación son de SD un porcentaje de acierto de 95.21%, y con MI 95.8%.

Resultados de los clasificadores

A continuación, se muestra la ejecución de la interfaz gráfica de los mejores resultados obtenidos de cada clasificador, donde se puede observar las métricas de cada uno y las características seleccionadas por el selector

Autoencoder

Figura 34

Resultado con los mejores parámetros obtenidos de Autoencoder



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Clasificador de señales del volcán Cotopaxi

Steve Alexander Morales Zapata

Ejecutable
Informativo

Selectores:

MI
 SD
 SFS
 SFFS
 Ninguno

Clasificadores:

Autoencoder
 KNN
 DT

Tipo	Exactitud	Presición	Sensibilidad	Especificidad	Ber
VT	97.1831	90.0000	79.4118	97.8462	0.1137
LP	95.4930	98.4026	96.5517	73.8095	0.1482
Otros	96.6197	8.3333	50.0000	99.7085	0.2515
Overall	94.6479	94.4909	93.4422	76.7720	0.1489

Ingrese el número de neuronas

Ingrese el valor de Q de MI

Ingrese el peso de las características de MI

Off On

Train y Test Aleatorio

No	#Caracterisitica	Caracteristica
1	72	w_t_peak2peak_D2
2	37	w_f_maxval_D2
3	24	f_peak_1020_pos
4	30	f_PeaksAboveRMSDensity_fun

En la Figura 34 se muestra el mejor resultado de Autoencoder, en el cual con el selector MI, tiene un porcentaje de clasificación del 94.64%, una precisión del 94.49%, sensibilidad del 93.44%, especificidad del 76.77% y Ber de 0.14. En la Figura 35 se muestra la matriz de confusión del clasificador.

Figura 35

Matriz de confusión de clasificación del Autoencoder

Output Class	VT	27 7.6%	5 1.4%	2 0.6%	79.4% 20.6%
	LP	2 0.6%	308 86.8%	9 2.5%	96.6% 3.4%
	Otros	1 0.3%	0 0.0%	1 0.3%	50.0% 50.0%
		90.0% 10.0%	98.4% 1.6%	8.3% 91.7%	94.6% 5.4%
	VT	LP	Otros		Target Class

KNN

Figura 36

Resultado con los mejores parámetros obtenidos de KNN



Clasificador de señales del volcán Cotopaxi

Steve Alexander Morales Zapata

Ejecutable
Informativo

Selectores:

MI
 SD
 SFS
 SFFS
 Ninguno

Clasificadores:

Autoencoder
 KNN
 DT

Ingrese el número de vecinos

Ingrese el valor de Q de MI

Ingrese el peso de las características de MI

Off On

Train y Test Aleatorio

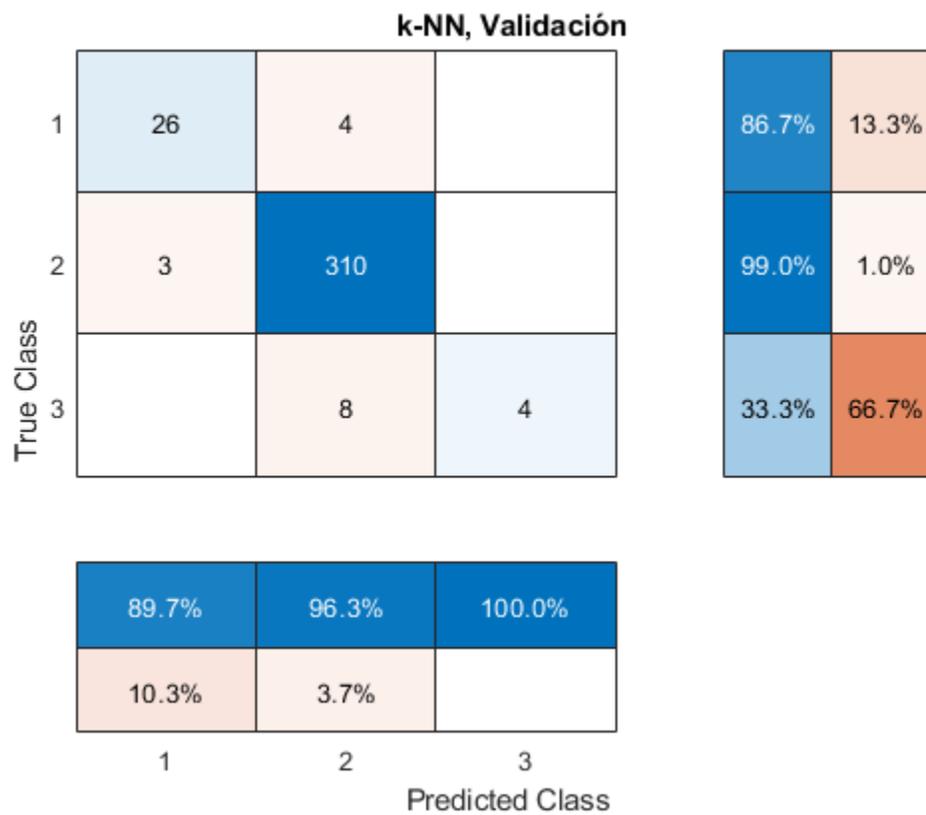
Tipo	Exactitud	Presición	Sensibilidad	Especificidad	Ber
VT	98.0282	89.6552	86.6667	98.7730	0.0728
LP	95.7746	96.2733	99.0415	90.9091	0.0502
Otros	97.7465	100.0000	33.3333	97.7208	0.3447
Overall	95.7746	95.8430	95.6570	91.8193	0.0626

No	#Caracterisitica	Caracteristica
1	72	w_t_peak2peak_D2
2	37	w_f_maxval_D2
3	65	w_t_rms_D2
4	58	w_t_PEC_D2
5	30	f_PeaksAboveRMSDensity_fun
6	24	f_peak_1020_pos
7	1814	spectralDecrease

En la Figura 36 se muestra el mejor resultado de KNN, en el cual con el selector SD, tiene un porcentaje de clasificación del 95.77%, una precisión del 95.84%, sensibilidad del 95.65%, especificidad del 91.81% y Ber de 0.06. En la Figura 37 se muestra la matriz de confusión del clasificador.

Figura 37

Matriz de confusión de clasificación de KNN



DT

Figura 38

Resultado con los mejores parámetros obtenidos de DT


ESPE
 UNIVERSIDAD DE LAS FUERZAS ARMADAS
 INNOVACIÓN PARA LA EXCELENCIA

Clasificador de señales del volcán Cotopaxi

Steve Alexander Morales Zapata

Ejecutable Informativo

Selectores: Clasificadores:
 MI Autoencoder
 SD KNN
 SFS DT
 SFFS Ninguno
 Ninguno

Ingrese el numero de splits

Ingrese el valor de Q de MI

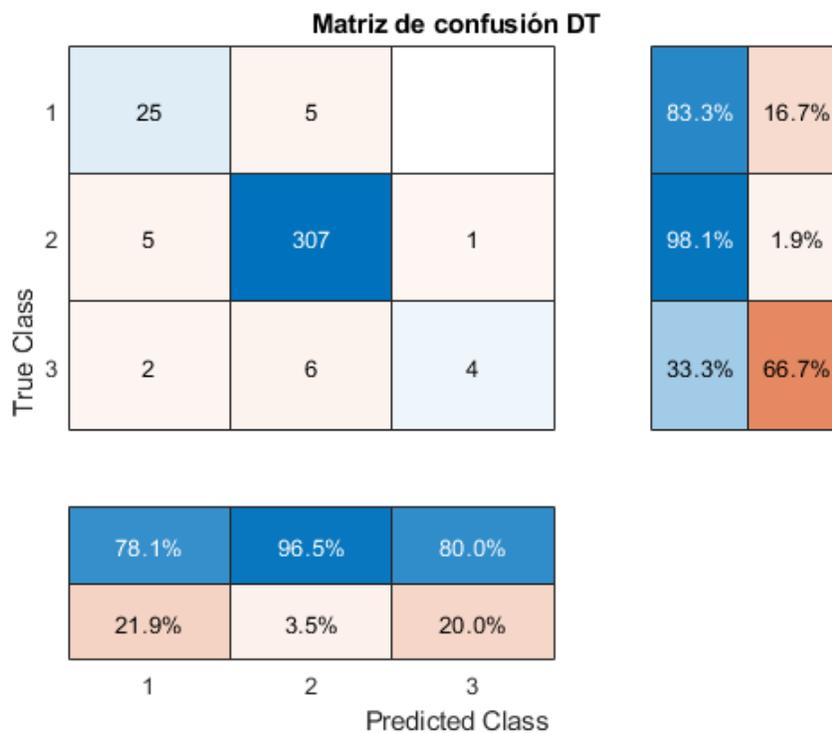
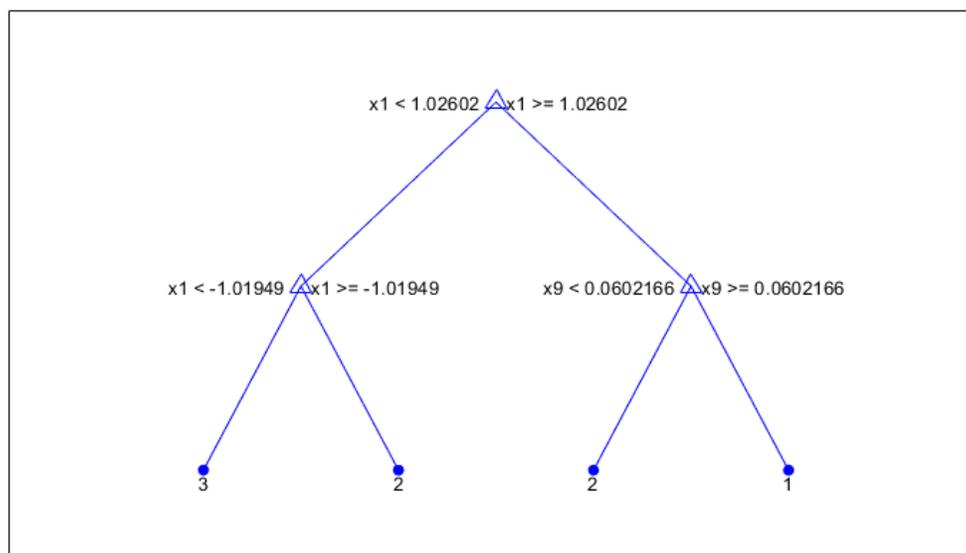
Ingrese el valor de peso de las características de MI

 Off On
 Train y Test Aleatorio

Tipo	Exactitud	Presición	Sensibilidad	Especificidad	Ber
VT	96.6197	78.1250	83.3333	98.4520	0.0911
LP	95.2113	96.5409	98.0831	83.7838	0.0907
Otros	97.4648	80.0000	33.3333	97.7143	0.3448
Overall	94.6479	94.3883	94.5307	85.5251	0.0997

No	#Caracterisitica	Caracteristica
1	72	w_t_peak2peak_D2
2	37	w_f_maxval_D2
3	65	w_t_rms_D2
4	58	w_t_PEC_D2
5	30	f_PeaksAboveRMSDensity_fun
6	24	f_peak_1020_pos
7	1814	spectralDecrease

En la Figura 38 se muestra el mejor resultado de DT, en el cual con el selector MI, tiene un porcentaje de clasificación del 94.64%, una precisión del 94.38%, sensibilidad del 94.53%, especificidad del 85.52% y Ber de 0.09. En la Figura 39 se muestra la matriz de confusión del clasificador y la Figura 40 son las divisiones realizadas por el clasificador.

Figura 39*Matriz de confusión de clasificación de DT***Figura 40***Árbol de decisión en la clasificación de DT*

Resultados de las mejores características

En análisis de las mejores características se obtiene de los mejores resultados de clasificación, para los cuales son las siguientes:

Autoencoder

Este clasificador tiene un resultado del 94.64% de acierto, para esto, toma 4 características mostradas en la Tabla 37.

Tabla 37

Características con mejor resultado de Autoencoder

Número	Característica número	Característica
1	72	w_t_peak2peak_D2
2	37	w_f_maxval_D2
3	24	f_peak_1020_pos
4	30	f_PeaksAboveRMSDensity_fun

KNN

Este clasificador tiene un resultado del 95.77% de acierto, para esto, toma 769 características, se muestran a continuación una parte mostrándose en la Tabla 38 las 30 iniciales:

Tabla 38*Características con mejor resultado de KNN*

Número	Característica número	Característica
1	72	w_t_peak2peak_D2
2	37	w_f_maxval_D2
3	65	w_t_rms_D2
4	58	w_t_PEC_D2
5	30	f_PeaksAboveRMSDensity_fun
6	24	f_peak_1020_pos
7	1814	spectralDecrease
8	1812	spectralDecrease
9	1813	spectralDecrease
10	1483	spectralSkewness
11	1824	spectralDecrease
12	1815	spectralDecrease
13	1816	spectralDecrease
14	2807	spectralCentroid
15	1490	spectralSkewness
16	1821	spectralDecrease
17	1825	spectralDecrease
18	1822	spectralDecrease
19	1481	spectralSkewness
20	1495	spectralSkewness
21	2814	spectralCentroid
22	1482	spectralSkewness
23	1484	spectralSkewness
24	2805	spectralCentroid
25	1493	spectralSkewness
26	1826	spectralDecrease
27	2808	spectralCentroid
28	2815	spectralCentroid
29	1811	spectralDecrease
30	1494	spectralSkewness

DT

Este clasificador tiene un resultado del 94.64% de acierto, para esto toma 10 características:

Tabla 39*Características con mejor resultado de DT*

Número	Característica número	Característica
1	72	w_t_peak2peak_D2
2	37	w_f_maxval_D2
3	65	w_t_rms_D2
4	58	w_t_PEC_D2
5	30	f_PeaksAboveRMSDensity_fun
6	24	f_peak_1020_pos
7	1814	spectralDecrease
8	1812	spectralDecrease
9	1813	spectralDecrease
10	1483	spectralSkewness

Las mejores características que se muestran repetitivamente en los mejores resultados de los clasificadores son:

- D2 Peak-to-peak in time-domain
- D2 Max. peak in freq.-domain
- D2 RMS in time-domain
- Percentage of energy for D2
- Density of peaks above RMS
- Freq. of max. peak in 10–20 Hz Band
- Spectral Decrease

Comparación con trabajo previo.

En el trabajo de (Minango, 2022), se establece el uso de dos clases LP y VT, además de solo trabajar con 3 características, las cuales son: Atenuación, Planitud y Disminución Espectral; con una escala ERB 256 bandas obtuvo los valores mostrados en Tabla 40:

Tabla 40

Resultado de ERB 256 bandas de (Minango, 2022)

Clasificador	Exactitud%	Precisión%	Especificidad%	Sensibilidad%	Ber
Autoencoder	92.06	88.09	94.05	88.09	0.089
KNN	97.35	96.03	98.02	96.03	0.03
DT	92.06	88.09	94.05	88.09	0.089

Para este trabajo se establece el uso de 3 clases LP, VT y Otros. Con un análisis de 98 características, se muestran los mejores resultados en la Tabla 41:

Tabla 41

Mejores resultados obtenidos

Clasificador	Exactitud%	Precisión%	Especificidad%	Sensibilidad%	Ber
Autoencoder	94.64	94.49	76.77	93.44	0.14
KNN	95.77	95.84	91.81	95.65	0.06
DT	94.64	94.38	85.52	94.53	0.09

Se compara los resultados de Tabla 40 y 41, se observa una mejora en Autoencoder y DT, pero cabe recalcar que existe una mejora referente a que se utilizaron tres clases.

Capítulo V

Conclusiones y Recomendaciones

Conclusiones

Se concluye que se tiene una mejora del clasificador mediante la selección de características de las señales de micro terremotos del volcán Cotopaxi, con esto al reducir las características logrando redes menos complejas, además que se presenta una interfaz gráfica la cual permite observar de manera detallada los resultados.

El mejor resultado de este clasificador es obtenido con el clasificador KNN y el selector MI, el cual tiene un porcentaje de clasificación de 95.77%, con una precisión del 95.84%, Especificidad de 91.81%, Sensibilidad de 95.65% y un Ber de 0.06.

Las características que presentan mejores resultados para las señales de microsismos del volcán Cotopaxi, en las cuales se tiene como mejores seis lineales y una psicoacústica, las cuales son: D2 Peak-to-peak in time-domain, D2 Max. peak in freq.-domain, D2 RMS in time-domain, Percentage of energy for D2, Density of peaks above RMS, Freq. of max. peak in 10–20 Hz Band y spectralDecrease; estas son las que se muestran en los selectores.

Se implementó un sistema en el software MATLAB® en el cual se engloban todas las fases del proyecto, el cual permite el análisis de características y la evaluación del clasificador se visualiza los resultados de una manera detallada y gráfica.

Recomendaciones.

En la etapa de preprocesamiento, se recomienda automatizarla para que de esa manera solo incluir la base de datos con las señales y evitar la manipulación de las mismas.

Se sugiere que investigaciones futuras profundicen en la clasificación, con la inclusión de las cinco clases totales presentes en la base de datos. Este enfoque permitiría una

clasificación integral y precisa de todas las señales, con una mayor comprensión y efectividad del clasificador.

Se recomienda incluir un análisis más profundo de los selectores SFS y SFFS, los cuales pueden permitir una selección de diferentes características.

Trabajos Futuros

Crear diferentes tipos de señales sintéticas que representen una variedad de condiciones y características que se pueden encontrar en las señales de microsismos reales del volcán Cotopaxi. Con esto lograr garantizar que el conjunto de datos utilizado para evaluar esté equilibrado. Esto implica que cada clase debe ser representada de manera uniforme para evitar sesgos en la evaluación del clasificador.

Adaptar nuevos clasificadores y selectores al sistema, con esto tener un análisis más profundo de las características y la clasificación de las señales.

Referencias

- Abu Alfeilat, H. A., Hassanat, A. B. A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., Eyal Salman, H. S., & Prasath, V. B. S. (2019). Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *https://home.liebertpub.com/big*, 7(4), 221–248. <https://doi.org/10.1089/BIG.2018.0175>
- Bank, D., Koenigstein, N., & Giryas, R. (2021). *Autoencoders*.
- Bolívar Cáceres, Jair Ramírez, Bernard Francou, Jean-Philippe Eissen, Jean-Denis Taupin, Ekkebard Jordan, Lars ngerechts, Luis Maisincho, Diego Barba, Eric Cadier, Rodolphe Bucher, Arturo Peñafiel, Pablo Samaniego, & Patricia Mothes. (2008). *DETERMINACIÓN DEL VOLUMEN DEL CASQUETE DE HIELO DEL VOLCAN COTOPAXI*.
- Brown, S. (2018). *¿A cuánta gente matan realmente los volcanes? - BBC News Mundo*. Bristol University. <https://www.bbc.com/mundo/noticias-44274988>
- Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, 81(3–4), 429–450. <https://doi.org/10.1007/S10472-017-9564-8/METRICS>
- Carniel, R., & Raquel Guzmán, S. (2020). Chapter Machine Learning in Volcanology: A Review. *Updates in Volcanology - Transdisciplinary Nature of Volcano Science*. <https://doi.org/10.5772/INTECHOPEN.94217>
- Divine, G., Norton, H. J., Hunt, R., & Dienemann, J. (2013). A review of analysis and sample size calculation considerations for wilcoxon tests. *Anesthesia and Analgesia*, 117(3), 699–710. <https://doi.org/10.1213/ANE.0B013E31827F53D7>
- IGEPN. (2024). *Descarga de Señales de Infrasonido - Instituto Geofísico - EPN*. <https://www.igepn.edu.ec/senales-sismicas/fomulario-infrasonido>

Instituto Geofísico - EPN. (2014, junio 24). *COTOPAXI - Instituto Geofísico - EPN*.

<https://www.igeppn.edu.ec/cotopaxi>

Lara-Cueva, R. A., Benítez, D. S., Carrera, E. V., Ruiz, M., & Rojo-Álvarez, J. L. (2016). Feature selection of seismic waveforms for long period event detection at Cotopaxi Volcano.

Journal of Volcanology and Geothermal Research, 316, 34–49.

<https://doi.org/10.1016/J.JVOLGEORES.2016.02.022>

Lara-Cueva, R. A., Moreno, A. S., Larco, J. C., & Benítez, D. S. (2016). Real-Time Seismic Event Detection Using Voice Activity Detection Techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(12), 5533–5542.

<https://doi.org/10.1109/JSTARS.2016.2605061>

Lara-Cueva, R., Benitez, D. S., Paillacho, V., Villalva, M., & Rojo-Alvarez, J. L. (2017). On the use of multi-class support vector machines for classification of seismic signals at Cotopaxi volcano. *2017 IEEE International Autumn Meeting on Power, Electronics and Computing, ROPEC 2017, 2018-January*, 1–6. <https://doi.org/10.1109/ROPEC.2017.8261613>

Malfante, M., Dalla Mura, M., Metaxian, J. P., Mars, J. I., Macedo, O., & Inza, A. (2018).

Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives. *IEEE Signal Processing Magazine*, 35(2), 20–30. <https://doi.org/10.1109/MSP.2017.2779166>

Martín Neira, & Lilian C. (2021). RESPONSABILIDAD CIVIL POR DAÑOS DERIVADOS DE FENÓMENOS NATURALES. UNA REVISIÓN JURISPRUDENCIAL. *Revista chilena de derecho privado*, 36(36), 141–186. <https://doi.org/10.4067/S0718-80722021000100141>

Michelucci, U. (2022). *An Introduction to Autoencoders*. <https://arxiv.org/abs/2201.03898v1>

- Minango, M. (2022). *Clasificación de eventos sismo volcánicos usando características psicoacústicas mediante técnicas de aprendizaje automático supervisado y no supervisado*. <http://repositorio.espe.edu.ec/jspui/handle/21000/31541>
- Nasteski, V. (2017). *An overview of the supervised machine learning methods*. <https://doi.org/10.20544/HORIZONS.B.04.1.17.P05>
- Peluffo-Ordóñez, D. H., Rodríguez-Sotelo, J. L., & Castellanos-Domínguez, G. (2009). Estudio Comparativo de Métodos de Selección de Características de Inferencia Supervisada y No Supervisada. *TecnoLógicas*, 23, 149–166. <https://doi.org/10.22430/22565337.239>
- Pérez, N., Benítez, D., Grijalva, F., Lara-Cueva, R., Ruiz, M., & Aguilar, J. (2020). ESeismic: Towards an Ecuadorian volcano seismic repository. *Journal of Volcanology and Geothermal Research*, 396, 106855. <https://doi.org/10.1016/J.JVOLGEORES.2020.106855>
- Pohjalainen, J., Räsänen, O., & Kadioglu, S. (2013a). *Feature Selection Methods and Their Combinations in High-Dimensional Classification of Speaker Likability, Intelligibility and Personality Traits*.
- Pohjalainen, J., Räsänen, O., & Kadioglu, S. (2013b). *Feature Selection Methods and Their Combinations in High-Dimensional Classification of Speaker Likability, Intelligibility and Personality Traits*.
- Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11), 1119–1125. [https://doi.org/10.1016/0167-8655\(94\)90127-9](https://doi.org/10.1016/0167-8655(94)90127-9)
- Rehman, R., Rehman Assistant Professor, R., Chandra Hazarika, G., Kardong, D., & Professor, A. (2015). Auditory Scale Analysis and Evaluation of Phonemes in MISING Language Handling imbalanced dataset. View project Auditory Scale Analysis and Evaluation of

Phonemes in MISING Language. *Article in International Journal of Computer Applications*, 113(15), 975–8887. <https://doi.org/10.5120/19899-2001>

Rostami, M., Berahmand, K., & Nasiri, E. (2021). *Review of swarm intelligence-based feature selection methods. Engineering Applications of Artificial Intelligence*. <https://sci-hub.se/10.1016/j.engappai.2021.104210>

Taha Jijo, B., & Mohsin Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. <https://doi.org/10.38094/jastt20165>

Venegas, P., Perez, N., Benitez, D. S., Lara-Cueva, R., & Ruiz, M. (2019). Building Machine Learning Models for Long-Period and Volcano-Tectonic Event Classification. *IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies, CHILECON 2019*. <https://doi.org/10.1109/CHILECON47746.2019.8987505>