Instituto Tecnológico de Ciudad Madero

DIVISION DE STUDIOS DE POSGRADO E INVESTIGACION

"POR MI PATRIA Y POR MI BIEN"

Nombre de Tesis:

Análisis Computacional para la Clasificación de Fluidos Mediante Vibraciones y Herramientas de Aprendizaje Automatizado

Que para obtener el grado de

Maestro en Ciencias de la Computación

presenta

Ing. Carlos Andrés Delgado Cedillo

G16070826

No. De CVU de CONAHCYT

1204511

Director de tesis

Dr. Nelson Rangel Valdez

Codirector de tesis

Dr. Juan Javier González Barbosa

Cd. Madero Tamaulipas Diciembre 2024

EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA
TECNOLÓGICO NACIONAL DE MÉXICO

Instituto Tecnológico de Ciudad Madero
Subdirección Académica
División de Estudios de Posgrado e Investigación

Ciudad Madero, Tamaulipas, 13/junio/2024

OFICIO No.: U.087/2024
ASUNTO: Autorización de impresión de tesis

C. CARLOS ANDRÉS DELGADO CEDILLO
No. DE CONTROL G16070826
P R E S E N T E

Me es grato comunicarle que después de la revisión realizada por el Jurado designado para su Examen de Grado de Maestría en Ciencias de la Computación, se acordó autorizar la impresión de su tesis titulada:

"ANÁLISIS COMPUTACIONAL PARA LA CLASIFICACIÓN DE FLUIDOS MEDIANTE VIBRACIONES Y HERRAMIENTAS DE APRENDIZAJE AUTOMATIZADO"

El Jurado está integrado por los siguientes catedráticos:

PRESIDENTA:	DRA.	MARÍA LUCILA MORALES RODRÍGUEZ
SECRETARIA:	DRA.	CLAUDIA GUADALUPE GÓMEZ SANTILLÁN
VOCAL:	DR.	NELSON RANGEL VALDEZ
SUPLENTE:	DR.	JUAN JAVIER GONZÁLEZ BARBOSA
DIRECTOR DE TESIS:	DR.	NELSON RANGEL VALDEZ
CO-DIRECTOR:	DR.	JUAN JAVIER GONZÁLEZ BARBOSA

Es muy satisfactorio para la División de Estudios de Posgrado e Investigación compartir con usted el logro de esta meta. Espero que continúe con éxito su desarrollo profesional y dedique su experiencia e inteligencia en beneficio de México.

A T E N T A M E N T E
Excelencia en Educación Tecnológica®
"Por mi patria y por mi bien"®

EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA
TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE CIUDAD MADERO
DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

MARCO ANTONIO CORONEL GARCÍA
JEFE DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

c.c.p.- Archivo

MACG/MLMR

Av. 1° de Mayo y Sor Juana I. de la Cruz S/N Col. Los Mangos C.P. 89440 Cd. Madero, Tam.
Tel. 01 (833) 357 48 20, ext. 3110, e-mail: depi_cdmadero@tecnm.mx
tecnm.mx | cdmadero.tecnm.mx

2024 Felipe Carrillo PUERTO

Dedicatoria

Quiero dedicar el presente trabajo a mi madre y padre:

Juan Carlos y Cristina Turrubiates

Agradecimientos

Quiero agradecer primeramente a la vida, al destino o la casualidad de haberme permitido estar en el momento adecuado para mejorar profesionalmente y personalmente.

A mis padres, los cuales agradezco la comprensión en todo mi proceso académico los cuales forman parte de mis éxitos.

Al Dr. Nelson Rangel Valdez, siendo este un maestro que me ha permitido mejorar tanto a académicamente como personalmente, además de apoyarme en todo mi proceso de formación.

Al Dr. Juan Javier Barbosa por sus enseñanzas y consejos para mejorar puntos importantes de mi formación

A la Dra. Claudia Guadalupe Gómez Santillán por su paciencia y sus enseñanzas que han salido valiosas a lo largo de mi proceso académico

A la Dra. Dra. María Lucila Morales Rodríguez por todas las observaciones que me han permitido en cada una de ellas mejorar en mi perspectiva académica.

Agradezco a mis mejores amigos Alejandro Paulin y Carlos Robledo por estar conmigo en las buenas y en las malas, así como todos los consejos que nos hemos dado mutuamente.

Resumen

El presente documento explora la efectividad de las Redes Neuronales Artificiales en la clasificación de fluidos utilizando datos de vibraciones provenientes de una tubería, dando como resultado una metodología que tiene como finalidad reducir el número de mediciones sin afectar estadísticamente el desempeño del algoritmo. La recopilación de datos se llevó a cabo en un laboratorio experimental diseñado específicamente para este estudio para lograr obtener información de un ambiente controlado, de este laboratorio experimental se recolectaron muestras de dos fluidos de referencia agua y agua con jabón, además de dos escenarios secundarios que corresponden a periodos de inactividad tubería vacía con la bomba encendida y tubería vacía con la bomba apagada. Los hallazgos indican que, la reducción de mediciones conlleva una disminución en la precisión de clasificación, sin embargo, los modelos pueden no tener una diferencia significativa lo que se traduce como una mejora tanto espacial como temporal. Los resultados obtenidos fueron una precisión mayor al 97% y realizando una comparativa entre conjuntos que representan la reducción en el número de medición se concluyó que de 2048 mediciones a 1500 mediciones no existió diferencia significativa.

Palabras clave: Aprendizaje Supervisado, Redes Neuronales Artificiales, Clasificación, Fluidos, Vibraciones, Aceleración

Abstract

This document explores the effectiveness of Artificial Neural Networks in the classification of fluids using vibration data from a pipe, resulting in a methodology that aims to reduce the number of measurements without statistically affecting the performance of the algorithm. Data collection was carried out in an experimental laboratory designed specifically for this study to obtain information from a controlled environment. From this experimental laboratory, samples of two reference fluids, water and water with soap, were collected, in addition to two secondary scenarios that correspond to periods of empty inactivity with the pump on and empty pipe with the pump off. The findings indicate that the reduction of measurements leads to a decrease in classification accuracy; however, the models may not have a significant difference, which translates into both a spatial and temporal improvement. The results obtained were a precision greater than 97% and by making a comparison between sets that represent the reduction in the number of measurements, it was concluded that from 2048 measurements to 1500 measurements there was no significant difference.

Keywords: Supervised Learning, Neural Networks, Classification, Fluids, Vibrations, Acceleration

Índice General

Dedicatoria ..... *IV*

Agradecimientos ..... *V*

Resumen VI

Abstract 7

Índice General *8*

Índice de Tablas *10*

Índice de Figuras *11*

Glosario de Acrónimos *12*

Índice de Nomenclaturas *13*

1. Introducción *14*

1.2. Justificación y beneficios **16**

1.3. Hipótesis **17**

1.4. Objetivos de la Investigación **17**

1.4.1 Objetivo General 18

1.4.2 Objetivos Específicos 18

1.5. Alcances y limitaciones **18**

1.6. Organización de la Tesis **19**

2. Estado del Arte *20*

3. Fundamento Teórico *25*

3.1 Keras **26**

3.2 Aprendizaje Supervisado **26**

3.3 Redes Neuronales Artificiales **27**

3.4 Información Mutua **30**

3.5 Transformada de Rápida de Fourier **30**

3.6 Vibraciones **31**

3.7 Acelerómetro **34**

3.8 Mecánica de Fluidos **35**

3.8.1 Densidad 35

3.8.2 Viscosidad 36

4. Metodología *37*

4.1 Laboratorio Experimental **39**

4.1.1 Aparatos y materiales 39

4.1.2 Instalación 40

4.1.3 Escenarios 41

4.2 Toma de muestras **42**
4.2 Conjunto de Mediciones **43**
4.3 Preprocesamiento para la caracterización mediante la FFT **44**
4.4 Entrenamiento con Diferentes conjuntos **44**
4.5 Comparación de diversos Conjuntos mediante Friedman **45**
5. Propuesta de Solución ***47***
5.1 Preprocesamiento mediante la Transformada de Rápida de Fourier **48**
5.2 Instancias de Entrenamiento, Validación y Prueba **50**
5.3 Arquitectura de Red Neuronal **51**
6. Resultados ***54***
6.1 Presentación de Resultados Experimentales **55**
6.2 Presentación de Resultados Comparativos entre conjuntos **63**
6.3 Resultados Adicionales **65**
6.3.1 Análisis de la Precisión en Conjuntos Reducidos 65
6.3.2 Prueba con la Base de Datos Unificada 67
6.3.4 Comparativa Contra la Reducción de Componentes 69
6.4 Discusión de Resultados **71**
7. Conclusiones y Trabajos Futuros ***74***
7.1 Aportaciones **75**
7.2 Conclusiones **75**
7.3 Trabajo Futuro **78**
Bibliografía ***80***
Anexos ***82***

Índice de Tablas

Tabla 1 Tabla del estado del arte *23*
Tabla 2. Diagrama de la tabla MYSQL de muestras de aceleración *43*
Tabla 3. Parámetros fijos de la Red Neuronal Artificial *51*
Tabla 4. Resultados de Experimentación variando Neuronas por Capa y Números de Capas en la RNA ... *52*
Tabla 5. Resultados de precisión y tiempo del conjunto de 2048 mediciones *56*
Tabla 6. Resultados de precisión y tiempo del conjunto de 2000 mediciones *57*
Tabla 7. Resultados de precisión y tiempo del conjunto de 1708 mediciones *58*
Tabla 8. Resultados de precisión y tiempo del conjunto de 1500 mediciones *59*
Tabla 9. Resultados de precisión y tiempo del conjunto de 1400 mediciones *60*
Tabla 10. Resultados de precisión y tiempo del conjunto de 1366 mediciones *61*
Tabla 11. Resultados de precisión y tiempo del conjunto de 1024 mediciones *62*
Tabla 12. Resultados de la comparación de los diferentes conjuntos con la prueba de Friedman *64*

Índice de Figuras

Figura 1 Construcción de Redes Neuronales *28*

Figura 2. Representación de las características de la longitud de onda. *33*

Figura 3. Representación de las características de la frecuencia *33*

Figura 4. Representación de las características de la Amplitud *34*

Figura 5. Bomba Sumergible *39*

Figura 6. Diagrama del laboratorio (No está a escala). A: Recipiente, B: Bomba, C1: Manguera 1, C2: Manguera 2, E: Acelerómetro, D: Tubería. *40*

Figura 7. Transmisor Instalado en la tubería de cobre *41*

Figura 8. Muestras correspondientes a cada escenario *48*

Figura 9. Datos preprocesando mediante la fórmula 1 ..... *49*

Figura 10. Gráfica de muestras en el dominio de la frecuencia. *50*

Figura 11. Arquitectura de RNA seleccionada *53*

Figura 12. Gráfica de entrenamiento de la RNA utilizando únicamente los datos de aceleración *53*

Figura 13. Comparativa de aprendizaje de los diferentes conjuntos *63*

Figura 14. Medias de precisión desde el conjunto 2048 hasta el conjunto 1400 ..... *65*

Figura 15. Diagrama de Flujo del experimento de comparativo en función del tamaño del conjunto *66*

Figura 16. Gráfica de prueba de Wilcoxon de grupos de conjuntos desde 2046 hasta 1400 en contra del conjunto 2048 y 2047 ..... *67*

Figura 17. Distribución de conjuntos de entrenamiento, validación y prueba *67*

Figura 18. Aprendizaje del conjunto de 2048 juntando todas las muestras de derivado de los 3 experimentos *68*

Figura 19. Medias a lo largo del conjunto 2049 hasta el conjunto 1400 utilizando la base de datos unificada *68*

Glosario de Acrónimos

Símbolo	Descripción
RNA	Redes Neuronales Artificiales
MFCC	Coeficientes Cepstrales en las Frecuencias de Mel
SVM	Máquinas de Soporte Vectorial
MI	Información Mutua
CNN	Redes Neuronales Convolucionales
FFT	Transformada Rápida de Fourier
DFT	Transformada Discreta de Fourier
AA	Aprendizaje Automático

Índice de Nomenclaturas

Símbolo	Descripción
x	Variable de entrada en la función ReLU
z_i	Entrada i-ésima en la función SoftMax
n_n	Número de neuronas en una capa densa
n_i	Número de entradas a una capa densa
I(X;Y)	Información mutua entre las variables X y Y
u(x,y)	Función conjunta de x e y
p(x)	Función de probabilidad marginal de x
u(y)	Función de probabilidad marginal de y
c_k	Coeficiente de Fourier
j	Unidad imaginaria
k	Índice entero
N	Número total de puntos en la serie de Fourier
d	Desplazamiento
D	Amplitud máxima del desplazamiento
ω	Frecuencia angular
t	Tiempo
v	Velocidad
p	Momentum (cantidad de movimiento)
m	Masa
Re	Número de Reynolds
ρ	Densidad del fluido
U	Velocidad característica del fluido
L	Longitud característica
μ	Viscosidad dinámica del fluido
xi	Muestra
Xi′	Muestra con media en 0
μ	Media de la muestra

CAPÍTULO 1

1. Introducción

En diferentes industrias la habilidad para distinguir y clasificar diferentes fluidos con rapidez y precisión son cruciales para la correcta operación. Desde la industria del petróleo y gas hasta el tratamiento de aguas residuales, la identificación del tipo de fluido es vital para el control de calidad, optimización de proceso, seguridad operativa, sostenibilidad ambiental y la realización de pronósticos. Tradicionalmente, la clasificación de fluidos se realiza con técnicas que pueden ser invasivas estructuralmente, lentas y sin automatizar, lo que puede llevar a retrasos o errores, por ello, se propone realizar la clasificación mediante vibraciones debido a que este tipo de sensores no modifican las estructuras al no requerir perforación para tomar mediciones, además, es posible monitorear con mayor tasa de frecuencia.

Por ejemplo un caso de estudio en donde existen estas necesidades es el tratamiento de aguas residuales [1] ya que describe necesidades de nuevas técnicas de monitoreo de la calidad del agua como nuevos equipos que no conlleven personal especializado, laboratorios costosos, rapidez, que sean automatizados y que tengan un mínimo tiempo de muestreo, esto debido a que técnicas como análisis de laboratorios o equipos especializados se vuelven complejos de implementar e inviables para la automatización.

Otro caso de estudio que puede beneficiarse es la industria del petróleo y gas, debido a que las propiedades del flujo son factores que influyen directamente en la eficiencia de la producción de pozos petroleros [2] y los sensores tradicionales limitan la capacidad de monitoreo a de una o dos veces por día de acuerdo con [3] limitando la capacidad de toma de decisiones en tiempo real.

Para lograr lo anterior, se enfrenta el reto de la eficiencia de los algoritmos desde la perspectiva del tiempo de muestreo. Una estrategia para mejorar lo anterior, implica la reducción del número de mediciones por muestra debido a que cada muestra puede contener múltiples mediciones, y al disminuir el número de estas, se reduce también el tiempo necesario para recolectar cada muestra. Esta reducción puede resultar una mayor eficiencia operativa, mejoras en el preprocesamiento, eficientar los tiempos de entrenamiento del modelo, en los tiempos de clasificación, mejoras en el almacenamiento de la información. Por lo tanto, es esencial desarrollar metodologías que puedan disminuir el número de muestras, permitiendo que mantengan la integridad en la precisión de la clasificación del modelo sin comprometer estadísticamente el desempeño.

Por ello, el propósito principal de esta investigación es la minimización del intervalo de muestreo sin perder la eficiencia en términos de precisión en la clasificación de fluidos de manera significativa. Este documento detalla la experimentación que incluye la recolección de muestras de vibraciones emitidas por una tubería cuando hay tránsito de fluido (agua y agua con jabón), así como en situaciones de ausencia de fluido (Con la bomba apagada y con

la bomba en operación). Se describe el preprocesamiento de los datos y los diversos conjuntos de datos que representan la reducción del intervalo de muestreo siendo estos los datos de entrada de una RNA para la obtención de valores estadísticos que permitan evaluar el desempeño y la comparación entre conjuntos de datos para determinar el intervalo más pequeño en dónde no tenga una diferencia significativa en su desempeño.

1.2. Justificación y beneficios

En el desarrollo de algoritmos de clasificación en tiempo real de vibraciones o señales acústicas la clave de un algoritmo eficiente es el tiempo de respuesta y la calidad del resultado, esto se convierte en una problemática debido al volumen de datos que se pueden requerir para dicha tarea, por ejemplo, cada segundo puede contener miles de mediciones y estás deben ser preprocesadas para obtener sus atributos, además del procesamiento de entrenamiento y clasificación.

En el campo de clasificación con modelos de aprendizaje supervisado utilizando señales acústicas o vibraciones se destaca una ausencia de enfoques o metodologías que permitan reducir el intervalo de muestreo sin afectar dicha clasificación, esto cobra relevancia en la clasificación de fluidos debido a que el flujo puede ir a más de 15 litros por segundo, lo que vuelve crucial el tiempo de distinción de los fluidos ya que puede resultar en la monitorización de eventos pasados o la falta de rapidez para eventos críticos como la filtración de agua en pozos petroleros.

En este campo en especial cobra relevancia debido a que, el régimen del flujo se convierte en una tarea importante a monitorear debido a estas propiedades influyen para la maximización y optimización de la producción de petróleo [4]. Otra de las áreas dónde se puede ver beneficiada es el monitoreo de la calidad del agua, los autores [1] describen la necesidad de nuevos equipos que no conlleven personal especializado o laboratorios costosos, además añade que los dispositivos sean rápidos, automáticos y con un mínimo tiempo de muestreo.

Además, ya que los datos de entrada son proporcionados por un sensor basado en vibraciones este es una excelente opción para la industria debido a que no modifica la estructura actual de la instalación [5] lo que conlleva como beneficio una fácil instalación y por ende reducir costos al no necesitar maquinarias o personal especializado.

Otro de los beneficios de reducir el intervalo de muestreo, se puede determinar una mejora en el almacenamiento ya que como se mencionó anteriormente cada muestra puede tener miles mediciones lo cual puede resultar costoso en ser almacenadas en especial si se realiza un monitoreo constante de por ejemplo cada segundo, finalmente puede existir una mejora significativa en el procesamiento, esto cobra relevancia ya que, estos algoritmos pueden ser incluidos en microcontroladores en dónde la capacidad de hardware es limitado por lo tanto un coste alto en la complejidad computacional puede resultar inviable la integración de los mismos.

1.3. Hipótesis

H0: Un algoritmo de aprendizaje automático que clasifica fluidos mediante vibraciones es menos eficiente si reduce el tiempo de muestreo de datos requerido para la clasificación sin afectar estadísticamente su desempeño.

H1: Un algoritmo de aprendizaje automático que clasifica fluidos mediante vibraciones es más eficiente si reduce el tiempo de muestreo de datos requerido para la clasificación sin afectar estadísticamente su desempeño

1.4. Objetivos de la Investigación

A continuación, se presentan los objetivos de esta investigación divididos en objetivos general y objetivos específicos.

1.4.1 Objetivo General

Analizar la eficiencia de un algoritmo de aprendizaje automático aplicado a la clasificación de fluidos a partir de vibraciones.

1.4.2 Objetivos Específicos

- Analizar estrategias de clasificación de fluidos a partir de vibraciones.
- Definir caso de estudio.
- Caracterizar clasificación de fluidos mediante la transformada rápida de Fourier
- Desarrollar modelo de clasificación.
- Validar el modelo de clasificación propuesto.

1.5. Alcances y limitaciones

- En seguida se enlistan algunas limitaciones:
- Se uso únicamente un algoritmo de Redes Neuronales para realizar la clasificación.
- Se recolectaron únicamente 2 fluidos (Agua y Agua con jabón).
- El estudio no considera tamaño de longitud o diámetro de tuberías distintos
- No hubo sensores adicionales para medir viscosidad, densidad u otros elementos.

1.6. Organización de la Tesis

La estructura del presente trabajo se describe de la siguiente manera: en el capítulo 2, titulado Estado del Arte, se presenta un resumen de trabajos seleccionados relacionados con la clasificación mediante aprendizaje automático utilizando señales de vibración. En el capítulo 3 se desarrolla el Fundamento Teórico, donde se proporciona información detallada sobre los tópicos clave de la investigación, como redes neuronales, transformadas y mecánica de fluidos. El capítulo 4 describe la Metodología, incluyendo el diseño del laboratorio experimental, los aparatos utilizados y los escenarios evaluados. El capítulo 5 presenta la Propuesta de Solución, donde se detalla el preprocesamiento de datos y la arquitectura de la red neuronal utilizada. En el capítulo 6 se exponen los Resultados, incluyendo tanto resultados experimentales como comparativos. Finalmente, el capítulo 7 aborda las Conclusiones y Trabajos Futuros, discutiendo las principales aportaciones de este proyecto y planteando posibles líneas de investigación a futuro.

CAPÍTULO 2

2. Estado del Arte

En seguida se presenta de forma resumida algunos trabajos relacionados con la clasificación de vibraciones o señales acústicas con diferentes aplicaciones como el reconocimiento de embarques, detección de fugas, clasificación de sonidos, etc. En particular se analizan los diferentes atributos y diferentes técnicas de clasificación debido a que se encontraron pocas investigaciones que se relacionen directamente, sin embargo, las investigaciones mencionadas en la Tabla 1 comparten características debido a que los datos de entrada son ondas acústicas que pueden ser preprocesadas con las mismas técnicas encontradas en la literatura para obtener sus atributos.

En la investigación [8] aborda técnicas para el reconocimiento de huellas acústicas utilizando análisis espectral. Este implementa Análisis de Componentes Principales para reducir la dimensionalidad de las características espectrales extraídas de las señales acústicas. Posteriormente se realiza una comparativa de clasificación utilizando RNA y KNN obteniendo como resultado 95.9% y 86.73% respectivamente, además incluye un análisis de reducción de frecuencias en dónde se utiliza técnicas de reducción de componentes para eliminar frecuencias menos relevantes.

En la investigación realizada por [6] realizaron una clasificación de señales acústicas con la finalidad de detectar fugas en tuberías y su gravedad. Se realizó una comparativa de atributos

de los datos arrojados por la transformada de wavelet entre ello estacan la media absoluta, desviación estándar y el factor cresta. Utilizando la técnica de clasificación de Maquinas de Soportes Vectoriales lograron un éxito de 99.4% y para clasificar la gravedad de la fuja un 95.6%.

La investigación [7] realizaron una clasificación de diferentes audios proporcionados por una base de datos extraída de la plataforma Kaggle, entre algunas de las diferentes clases se incluyen: Partes de baterías musical, tijeras, líquido, pasos, entre otras. Utilizaron la técnica de Redes Neuronales Convulsiónales con atributos proporcionados por los coeficientes cepstrales de Mel, lograron un porcentaje de acierto de 88.9% con un total de 80 clases.

En la investigación [4] clasificaron petróleo, agua y gas utilizando redes neuronales convulsiónales logrando un 99.3% de acierto, utilizaron datos proporcionado por sensores de efecto Doppler dónde su principal función es medir flujo basado en la velocidad, estos datos fueron procesados para transformarlos al dominio de la frecuencia utilizando la transformada de Fourier 2D.

En el artículo [10] describe el desarrollo de un algoritmo robusto para la búsqueda de canciones las características principales se transforman en un hash esto con la finalidad de poder realizar una comparación directa con la base de datos de las canciones registradas para realizar la búsqueda, este algoritmo además es optimizado para tener tolerancia al ruido y a las distorsiones que puedan existir en el ambiente, obteniendo como resultados para la búsqueda un tiempo requerido de 5 a 500 milisegundos para encontrar la canción.

En la referencia número [11] expone una metodología para realizar clasificaciones de sonidos ambientales, entre ellos se incluyen sonidos de aire acondicionados, sirenas, disparos, niños jugando entre otras 6 clases más. Se utilizó como técnica de preprocesamiento para obtener las características más importantes los MFCC y una Red Neuronal tipo Convolucional obteniendo un 100% de acierto en el clasificador.

La investigación [12] tiene como propósito el uso de acelerómetros para monitorear el flujo de una tubería para lograrlo se utilizó la frecuencia media, temperatura y retraso como datos de entrada y una RNA para realizar una regresión obteniendo como resultado una pérdida de validación de 0.0084 tras 976 épocas, obteniendo como resultado que se logró generalizar razonablemente la medición de flujo.

En el artículo [13] se describe una experimentación que consiste detectar tempranamente el *slugging* severo. Se colocaron 6 acelerómetros instalados a lo largo de una tubería vertical y registraron los datos respectivos de aceleración, posterior a la recopilación de datos se preprocesaron para la obtención de características que permitan la clasificación entre ellos la media, varianza, Skewness, Curtosis, media cuadrática, coeficientes lineales predictivos. Logrando un 99% utilizando la técnica de SVM y 100% con la técnica de RNA para la clasificación.

En el último artículo [13] mostrado en la Tabla 1, describe una metodología utilizando acelerómetros para el monitoreo de flujo. La idea central fue demostrar que existe una relación lineal entre la el flujo y las amplitudes emitidas por la vibración de la tubería. Utilizaron las frecuencias para realizar el análisis de las muestras logrando visualizar la relación lineal inclusive cuando se cambiaba el material de la tubería y los diámetros.

Tabla 1. Análisis del estado del arte.

Referencia	**Aplicación**	**Instrumento**	**Atributos**	**Técnicas**
[8]	Reconocimiento de embarques	Muestras de audio	Frecuencia	KNN/RNA
[6]	Detección de fugas de gas	Sensor acústico	Desviación estándar, media absoluta, frecuencia media (Transformada de wavelet)	SVM
[9]	Análisis de señales acústicas	hidrófono	Frecuencia	-
[10]	Reconocimiento de canciones	Micrófono	Máximos locales (Frecuencia y tiempo)	Algoritmo Propio
[11]	Clasificación de sonidos	Muestras de Audio	MFCC	RNA
[12]	Medición de Flujo	Acelerómetro	Frecuencia media, retraso y temperatura	RNA
[7]	Clasificación de sonidos	Micrófono	MFCC	CNN
[4]	Clasificación de Aceite, Agua y Gas	Sensor Doppler	Transformada de Fourier 2D	CNN/RNA
[13]	Monitoreo de “Slugging” (Tuberías)	Acelerómetro	Media, varianza, Skewness, Curtosis, media cuadrátibca, coeficientes lineales predictivos	SVM/RNA
[14]	Medición de Flujo	Acelerómetro	Frecuencia	-

Como primera instancia de acuerdo a la Tabla número 1 correspondiente al estado del arte se puede concluir que pocas investigaciones se centran en la clasificación de fluidos mediante señales acústicas, hasta el momento fue encontrado solo una [4], con ello se observa un interés para la realizar del tipo de proyecto que cubre la presente investigación. Solo la investigación

[6] cubre el aspecto de la reducción de componentes, sin embargo, esta se centra en reducir únicamente las frecuencias y no en reducir el intervalo de muestreo. Finalmente se concluye la importancia del preprocesamiento en señales acústicas con una tendencia en las investigaciones [8],[12],[9],[12] y [13] de preprocesar las señales acústicas mediante a FFT obteniendo las frecuencias de las señales y 4 de ellas fueron aplicadas a tuberías cuando existía el tránsito de un fluido.

Con ello podemos concluir que si comparamos el estado del arte con la presente investigación observamos que solo una se centra en la clasificación de fluidos mediante señales acústicas, esta utiliza sensores de efecto Doppler lo cual hay una diferenciación al utilizar vibraciones como señales de entrada, además de que se utilizó una CNN para obtener un mejor porcentaje de acierto lo cual se traduce en un mayor coste computacional tanto de preprocesamiento como de entrenamiento y clasificación si lo comparamos con una RNA densa o convencional. [15] [16]

CAPÍTULO 3

3. Fundamento Teórico

La presente investigación se sustenta en un marco teórico presentado a continuación, en donde se integra varios dominios de diferentes áreas para la comprensión de la metodología desarrollada. En primer lugar, se expone los conceptos relacionados con el área de ciencias de la computación, seguido por el área y finalmente por el área física.

3.1 Keras

Keras es un *framework* de aprendizaje profundo echo para Python el cual provee casi cualquier tipo de modelo de aprendizaje profundo. Fue desarrollado principalmente para investigadores para realizar experimentación más rápido esto debido a que funciona como interfaz con la biblioteca de tensorflow [15].

Además de la facilidad para la implementación de modelos de Aprendizaje Profundo, incluye varios tipos de RNA, como los son las Redes Neuronales Convolucionales, Redes Neuronales Recurrentes y combinaciones de ambos. Incluye varios tipos de entrada y de salida de los modelos y las capas son intercambiables, esto hace a Keras apropiado para montar esencialmente cualquier modelo de Aprendizaje Profundo desde modelos generativos hasta máquinas de Turing.

Es distribuido bajo la licencia MIT, lo cual significa que puede ser distribuido para proyectos comerciales de manera gratuita. Por estas razones Keras es usado por más de 200,000 profesionales que van desde investigadores, ingenieros, estudiantes y personas aficionados. Además, es usado por las grandes empresas como Google, Netflix, Uber entre otras, lo cual lo convierte en una herramienta bastante poderosa al momento de el desarrollo de modelos de Inteligencia Artificial.

3.2 Aprendizaje Supervisado

De acuerdo con el autor [17] el Aprendizaje Supervisado se define como la ciencia de programar computadoras para que puedan aprender de los datos. El autor nos ilustra mediante un ejemplo práctico en dónde el mecanismo atrás del correo no deseo existe un algoritmo basado en Aprendizaje Supervisado en donde cada a la máquina se le proporcionó ejemplos de correos no deseos y correos electrónicos regulares, en dónde a cada ejemplo de entrenamiento se denomina como instancia de entrenamiento.

El beneficio más importante por lo cual el Aprendizaje Supervisado es la capacidad de modelar patrones esto debido a que, realizar reglas de manera manual que nos permitan por ejemplo, clasificar correos regulares o no deseados sería una tarea difícil de realizar, además que, el Aprendizaje Supervisado es capaz de detectar de manera autónoma variaciones de palabras que tengan el mismo significado y por lo tanto tengan el mismo potencial de ser consideradas dentro de un correo no deseado por ejemplo las palabras: “Four U” o “4U”.

Existen múltiples aplicaciones del Aprendizaje Supervisado por ejemplo en la clasificación automática de noticias, chatbots, detección de fraudes, reconcomiendo de voz, segmentación de clientes etc. Por ello se ha demostrado el gran valor que cuenta esta práctica para la industria obteniendo múltiples beneficios como la generación de automatizaciones.

Existen dos grandes categorías del Aprendizaje Supervisado que se dividen si están capacitados o no con supervisión humana (algoritmos no supervisados y supervisados). El primero de ellos se definen cuando los datos de entrenamiento no están etiquetados y el sistema intenta aprender por su cuenta, algunos algoritmos no supervisados son: K-means, SVM de una clase, kernel PCA, a priori, entre otros. Por otra parte, los algoritmos supervisados son aquellos que, si contienen una etiqueta, es decir, fueron validados generalmente por un especialista, alguna de las tareas típicas es de clasificación o regresión. En el siguiente capitulo describe un algoritmo de AA (Aprendizaje Automático).

3.3 Redes Neuronales Artificiales

Una de las técnicas más populares del AA son las RNA, estas son una técnica de la inteligencia artificial inspirada en la biología del cerebro humano, sin embargo, las RNA no tienen ni una fracción del poder del cerebro humano [18] Como se mencionó anteriormente al ser algoritmos del área del AA tienen la capacidad de aprender, generalizar y abstraer [19]

por lo cual resulta demasiado útiles en diversas áreas por ejemplo en la medicina, agricultura, marketing, etc.

El Backpropagation es un método de entrenamiento en el cual las RNA aprenden, su potencia reside en la capacidad de entrenar las capas ocultas [19]. En la Figura 1 se muestra una arquitectura básica de 3 capas en donde la primera capa es la capa de entrada y su objetivo es recibir los valores desde el exterior. La siguiente capa es una capa oculta y esta se interconecta con las demás capas. Por último, la capa de salida representa la salida o respuesta de la RNA.

Una de las partes más importantes de la RNA es la función de activación debido a que, introduce no linealidades en la red neuronal, lo que permite a la red aprender y modelar relaciones complejas en los datos con la finalidad de activa una neurona o ponderar cuál será su nivel de activación. Sin una función de activación no lineal, la red neuronal se reduciría a una combinación lineal de las entradas y los pesos, lo cual limitaría su capacidad para aprender patrones no lineales en los datos.

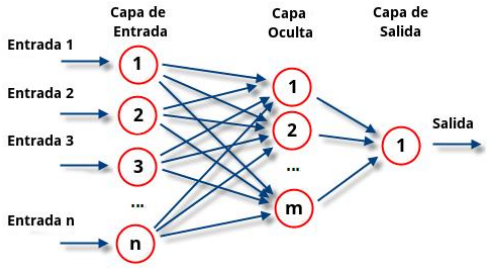

Figura 1. Construcción de Redes Neuronales

Existen diferentes tipos de funciones de activación utilizadas en las redes neuronales. Algunos ejemplos comunes incluyen:

Función ReLU (Rectified Linear Unit): La función ReLU es una función no lineal que mapea los valores de entrada negativos a cero y mantiene los valores positivos sin cambios. Es simple y computacionalmente eficiente, y se utiliza en muchas arquitecturas de redes neuronales, especialmente en las capas ocultas, la fórmula que describe esta función de activación esta dada por la fórmula número 1.

$$ReLU(x) = \max(0, x) \tag{1}$$

Función Softmax: La función softmax es utilizada comúnmente en la capa de salida de una red neuronal para problemas de clasificación multiclase. Toma un vector de entrada y normaliza sus valores para que sumen 1, representando las probabilidades de pertenencia a cada clase, es descrita por la fórmula número 2.

$$SoftMax(zi) = \frac{e^{zi}}{\sum_{j}^{n}{}_{=1} e^{zj}} \tag{2}$$

Para el cálculo de la complejidad computacional casi siempre se atribuye a una notación Big-O, sin embargo, de acuerdo a [20] la notación de Big-O puede ser a menudo una medida simplificada que no se puede traducir inmediatamente en los recursos de hardware necesarios para la realización un algoritmo de RNA, debido a esto se han realizado algunas métricas de evaluación. En la fórmula 3 se describe la métrica de multiplicaciones reales realizadas por cada capa.

$$\text{RMDense} = n_n n_i \tag{3}$$

3.4 Información Mutua

De acuerdo con [21] la información mutua (MI) es una estadística perfecta para medir el grado de relación entre conjuntos, además es sencilla de interpretar como la cantidad de información compartida esto comparado con otras estadísticas como el orden de rango resulta más fácil la interpretación. En la fórmula número 4 se realiza su descripción.

$$I(X;Y)\sum_x \int log\frac{u(x,y)}{p(x)u(y)}dy \qquad (4)$$

Por ello se vuelve una herramienta poderosa para evaluar la relación entre dos variables, proporcionando la medida de la cantidad de información que una variable tiene sobre otra. En el ámbito de ciencia de datos, se puede utilizar para la selección de características principales, es decir, aquellos componentes que muestren una mayor información mutua son seleccionados como los más significativos en este contexto.

3.5 Transformada de Rápida de Fourier

Los siguientes conceptos relacionados a la transformada de Fourier son extraídos del libro tratamiento digitales de señales [22] en él se describe que la representación de las series de Fourier como la descomposición de las señales en función de componentes sinusoidales, cuando ocurre dicha descomposición se dice que está representada en el dominio de la frecuencia.

Para las señales periódicas discretas en el tiempo, supongamos que existe una secuencia periódica $x(n)$ es decir, $x(n) = x(n + N)$. La representación en serie de Fourier consta de N funciones exponenciales. En la fórmula número 5 se describe la serie de Fourier discreta en el tiempo (DFT), esto proporcionará un cambio al dominio de la frecuencia. En dónde {ck} son los coeficientes de la representación en serie.

$$ck = \sum_{n=0}^{N-1} x(n) e^{-\frac{j(2\pi)}{N}(kn)} = \quad (5)$$

Sin embargo, este método tiene un coste computacional alto de O(N^2) por lo cual se desarrolló un método más eficiente para el cálculo de la DFT. El algoritmo más utilizado es el Cooley-Turkey [23] este cuenta con una notación *O(Nlog2N)* lo que indica una mejora significativa. En dónde el principio básico se establece bajo el método de divide y vencerás.

Por último, se destaca el teorema de muestreo esta es importante tenerla en cuenta al momento de la digitalización de señales, esta establece que, la frecuencia de muestreo debe ser al menos el doble de la frecuencia presente máxima. Esto implica que, a menor tiempo de muestreo obtendremos menos frecuencia lo que se puede traducir como disminución a la resolución lo que puede afectar a la aplicación final en la cual vaya ser usada.

3.6 Vibraciones

Una definición básica de vibración es descrita por [24] como: la oscilación o el movimiento repetitivo de un objeto alrededor de una posición de equilibrio. Se puede describir que la posición de equilibrio es cuando la fuerza que actúa sobre él es cero.

Esta definición corresponde a una vibración de cuerpo entero la cual se explica como una combinación de 6 movimientos diferentes. Estos pueden moverse en tres direcciones correspondientes a las direcciones (*x, y, z)*, además de las rotaciones correspondientes a los ejes (*5*).

Al total de estos movimientos se conoce como los seis grados de libertad, un ejemplo sencillo es que un cubo puede desplazarse hacia adelante y hacia atrás (eje *x*), desplazarse hacia los lados (eje *y*) y desplazarse hacia arriba y hacia abajo (eje *z*), además de que es posible la

rotación en los 3 ejes (*x*, *y*, *z*).

Si se ejemplifica un desplazamiento de un objeto sencillo y tomamos en cuenta únicamente una dirección la curva resultante resultará en una onda senoidal esta es descrita por la formula número 6:

$$d = Dsin(\omega\ t) \qquad (6)$$

Además, con esta información es posible obtener la velocidad instantánea o en otras palabras que tan rápido cambia su posición. Por lo cual si tomamos como referencia la formula número 1 podemos obtener dicha velocidad instantánea aplicando la derivada descrita en la formula número 7:

$$v = \frac{dd}{dt} = \ \omega Dcos(\omega\ t) \qquad (7)$$

Finalmente, para obtener la aceleración se describe como la proporción de cambio de la velocidad o que tan rápido cambia la velocidad en cualquier punto. Es descrita por la derivada en la fórmula número 8:

$$v = \frac{d}{dt}(\ \omega Dcos(\omega\ t)) \qquad (8)$$

Como se mostró anteriormente las vibraciones son representadas por una función senoidal que es una característica común de las ondas. A continuación, se muestran algunas propiedades descritas por [25] en el libro Ingeniería de sonido Conceptos, fundamentos y casos prácticos

Longitud de Onda (λ): Es un periodo espacial de la misma, es decir, la distancia a la que se repite la forma de la onda. Se puede traducir como el valor máximo tanto positivo como negativo, esto se observa en la Figura 2.

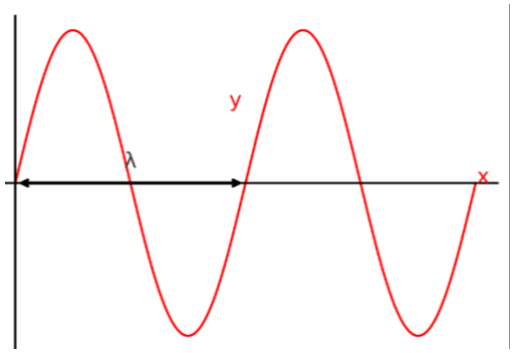

Figura 2. Representación de las características de la longitud de onda.

Frecuencia (f): Es el número de repeticiones por unidad de tiempo de cualquier evento periódico (ver Figura 3). Se traduce como la cantidad de ciclos vibratorios completos por una cantidad de tiempo dada, está relacionada el tiempo que tarda algo en suceder se denomina como periodo (cantidad en tiempo) y la frecuencia es la cantidad de velocidad (ciclos por segundo) su unidad de medida son los Hertz (Hz).

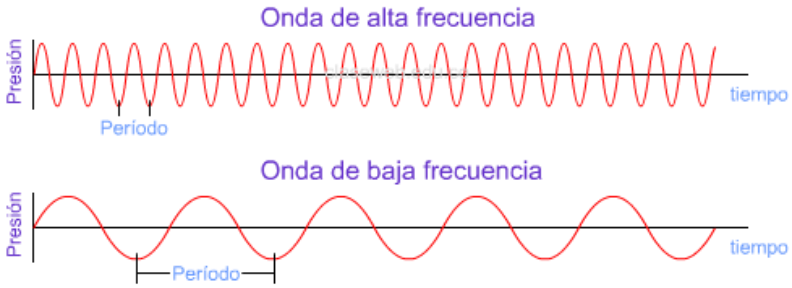

Figura 3. Representación de las características de la frecuencia

Amplitud: Indica la cantidad de cambios en a la presión de aire. Puede medirse como a la distancia vertical máxima desde una presión de aire cero hasta el valor más alto. En la figura 4 se observa el rango de valores que toma la amplitud desde 0 hasta su límite superior.

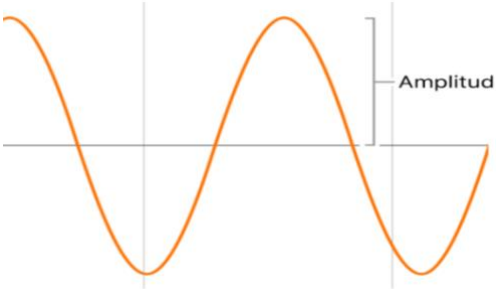

Figura 4. Representación de las características de la Amplitud

3.7 Acelerómetro

Los sensores imitan la capacidad de percepción de los seres humanos, esto hace poder obtener información sin que un humano este presente. Podemos encontrar sensores relacionado con los sentidos de los humanos como: vista, oído, tacto, etc. Estos no se limitan a únicamente la percepción humana si no van más allá donde existen sensores que reaccionan a componentes químicos o a cuestiones físicas [26].

Un acelerómetro es un dispositivo que es capaz de medir y analizar la aceleración linear y angular. Utilizan mediciones de aceleración gravitacional estática lo que permite determinar el ángulo de desviación del objeto medido de la vertical, así como en mediciones de aceleración dinámica debido a golpes movimiento o vibración.

Lo cual estos dispositivos son útiles en aplicaciones de estabilidad, por ejemplo, los celulares tienen un acelerómetro integrado que es usado para realizar estabilidad al momento de tomar fotos, tomar videos o en aplicaciones de realidad aumentada para poder fijar objetos y poder observarlos en un plano 3D.

3.8 Mecánica de Fluidos

La mecánica de fluidos es una rama de la física que se define como estudio de leyes del movimiento de los fluidos y sus procesos de interacción con los cuerpos sólidos [16]. Un fluido se define como una sustancia que puede fluir y adaptarse a la forma del recipiente, estos pueden ser líquidos o gaseosos por ejemplo el agua o el aire.

La característica principal de los fluidos se denomina fluidez. Puede cambiar de manera continua cuando está sometido a un esfuerzo constante, en otras palabras, el fluido no puede soportar un esfuerzo. Lo que determina que un líquido se mueva más rápido que otros fluidos es denominado viscosidad y está relacionada con las acciones de fueras de rozamiento, esto se puede observar en la figura 4 indicando que las flechas que los fluidos requieren de menos fuerza para ser movidos.

3.8.1 Densidad

Esta característica de los fluidos mide compactibilidad del material, es decir, la densidad (p). se define como la masa por unidad de volumen esto es escrito en la fórmula número 9. Las unidades en el sistema internacional son kg/m3. Para un fluido homogéneo (la densidad es la misma en todas sus partes) se define mediante la siguiente formula.

$$\underline{p = v/m} \tag{9}$$

En los líquidos al tener baja compresibilidad (disminución de tamaño al someter una presión) la densidad depende de la temperatura, pero depende poco de la presión y para los fluidos compresibles, la densidad depende en general tanto de la presión como la temperatura [27].

3.8.2 Viscosidad

La característica de la viscosidad es la resistencia al movimiento del fluido, esto ocasiona que el fluido pase más o menos rápido en un medio, por ejemplo, en una tubería. En los fluidos ideales no se toma en cuenta el proceso de disipación de la energía por fricción y conductividad térmica por lo tanto las fuerzas internas es únicamente la presión.

Los fluidos se aproximan a lo antes mencionada por lo que en la mayoría de los casos se pueden analizar como fluidos ideales. El parámetro primario que determina el comportamiento de los fluidos newtonianos es el número adimensional de Reynolds descrito en la fórmula número 10, donde U y L representan la velocidad y longitud características del flujo, como u y p entran como cociente en el parámetro v, el cociente tiene significado propio y se denomina viscosidad cinemática.

$$Re = \frac{\rho UL}{\mu} = \frac{UL}{v} \tag{10}$$

El número de Reynolds (*Re*) es importante ya que valores pequeños nos indicarán un movimiento más lento del fluido por lo tanto más viscosos. En este caso los valores de inercia son despreciables. Para valores moderados de Re corresponden al flujo laminar, caracterizados por variaciones suaves y valores altos de Re suelen estar asociados con el flujo turbulento [27].

CAPÍTULO 4

4. Metodología

El objetivo de la presente metodología es implementar un sistema de clasificación de fluidos en tuberías mediante redes neuronales artificiales (RNA) basado en la información de vibraciones obtenida en un laboratorio experimental descrito en este capítulo, optimizando la cantidad de mediciones sin comprometer la precisión del modelo de clasificación.

En la Figura 8 se muestra el flujo de trabajo seguido en la presente metodología, El proceso comienza con la obtención de un conjunto de aceleración de 2048 mediciones, provenientes del laboratorio experimental. Posteriormente, estas mediciones se reducen a subconjuntos de hasta 1024 mediciones para evaluar el impacto en el rendimiento del modelo. A continuación, se aplica la Transformada Rápida de Fourier (FFT) para convertir las señales al dominio de la frecuencia y extraer características clave. Con los diferentes conjuntos de datos preprocesados, se entrenan varios modelos de clasificación basados en redes neuronales artificiales (RNA). Los datos de precisión obtenidos a partir del entrenamiento son recopilados y, finalmente, se realiza una comparación estadística mediante el test de Friedman para determinar si la reducción de mediciones afecta significativamente la precisión del modelo.

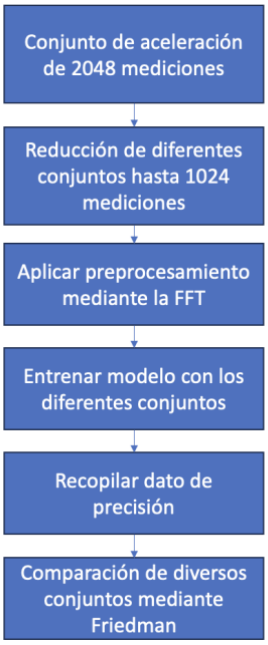

Figura 8. Metodología de Comparación de Diferentes Conjuntos de Mediciones

4.1 Laboratorio Experimental

Para abordar la primera etapa de la metodología (conjunto de aceleración de 2048 mediciones) se ha diseñado e implementado un laboratorio experimental de circulación de fluidos líquidos con el propósito de estudiar un ambiente controlado en donde las variaciones de los datos sean producto de factores estudiados y no de externos, de este modo se facilita la recolección de datos en tiempo real creando un entorno que permita validar el modelo de RNA para la clasificación.

4.1.1 Aparatos y materiales

El laboratorio consta de una bomba sumergible con unas características de 14 centímetros de largo y 15 centímetros de altura con la capacidad de un caudal de 3500L/H con una altura máxima de elevación de 3.5 metros, esta es mostrada en la Figura 5.

Figura 5. Bomba Sumergible

La tubería es de material de cobre y tendrán un diámetro de ¾ de pulgada con una longitud de 1.5 metros, así mismo se usarán 2 mangueras de ¾ de pulgada y 1.5 metros de largo.

Un contenedor de plástico con una capacidad de 72 litros y de dimensiones 38x44x62 centímetros.

El transmisor que se utilizará se compone de un microcontrolador ESP32 con una velocidad de reloj de 240MHz una memoria flash de 4MB y memoria RAM de 520KB. Las conectividades usadas fueron WIFI (2,5GHz) e I2C, además se compone de un módulo MPU-

6050 este cuenta con acelerómetro y giroscopio, finalmente la capacidad de medición utilizada fue de ±2g.

El servidor en dónde recibieron los datos y en dónde se entrenaron los algoritmos fue una MacBook Air con procesador M1 con 8 GB de RAM y un SSD 240GB.

4.1.2 Instalación

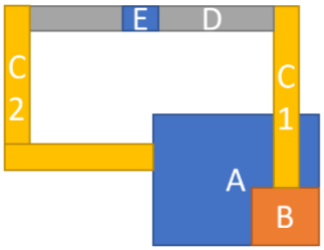

Figura 6. Diagrama del laboratorio (No está a escala). A: Recipiente, B: Bomba, C1: Manguera 1, C2: Manguera 2, E: Acelerómetro, D: Tubería.

El diagrama del laboratorio se muestra en la Figura 6 en dónde la bomba (B) se coloca dentro del recipiente (A) y en la salida de descarga de la bomba se introducirá la manguera (C1) y en el otro extremo de la misma manguera (C1) va colocado un extremo de la tubería (D) mediante una abrazadera para unir ambas piezas, la otra manguera (C2) va colocado igualmente con una abrazadera en el otro extremo de la tubería (D) y el otro extremo de la manguera va instalado en un orifico del recipiente (A). La tubería deberá ser colocada paralela al piso asegurándose que tenga una cierta altura y no este pegada al suelo. Finalmente, la carcasa del transmisor (E) es instalada mediante una abrazadera y el acelerómetro es colocado a la mitad de la tubería mediante una cinta doble cara como se muestra en la Figura 7.

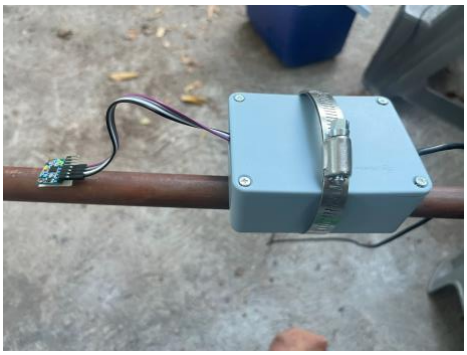

Figura 7. Transmisor Instalado en la tubería de cobre

4.1.3 Escenarios

A continuación, se describe a detalle cada escenario que corresponderá a cada clase de la RNA.

1. Tubería vacía con la bomba apagada:

En este escenario, la tubería no tiene contenido y la bomba permanece inactiva. Sirve como una línea base o referencia, permitiendo observar las mediciones en las condiciones más estáticas y sin intervención de fluidos o activación de la bomba. En este escenario se pretende identificar el ruido y cuando no este circulando ningún fluido.

2. Tubería vacía con la bomba encendida:

La tubería permanece vacía, sin embargo, en este escenario la bomba se activa, esto permitirá identificar y medir aceleraciones que provenga directamente de la operación de la bomba sin la influencia de fluidos.

3. Tubería con agua en tránsito:

El recipiente se llenará con 2 galones de agua con la bomba activa, y tras un minuto de espera para estabilizar el sistema, se comenzará la recolección de datos. Este escenario refleja las condiciones de una operación normal con un fluido puro (agua) circulando a través del sistema.

4. Tubería con mezcla de agua y jabón:

Al recipiente con agua, se le añaden 200 gramos de jabón en polvo y después de esperar 20 minutos para que el jabón se mezcle adecuadamente con el agua, se comenzará a recolectar datos. Esta etapa permitirá observar cómo un aditivo (en este caso el jabón) altera las mediciones, realizando la comparación con el agua pura. Es importante monitorizar la formación de espuma y retirar el exceso con cuidado, ya que puede afectar las mediciones.

4.2 Toma de muestras

En el contexto de este estudio, una muestra se define como un conjunto de mediciones de aceleración relacionadas por la vibración de la tubería inducidas por los fluidos o en la ausencia de estos en un periodo determinado.

Considerando lo anterior se llevará a cabo el desarrollo de los 4 diferentes escenarios descritos en la sección anterior, cada escenario tuvo una duración de 15 minutos con la finalidad de obtener suficientes muestras para realizar la experimentación. A continuación, se detalla la lectura de las muestras, Cada muestra contiene 2048 mediciones de aceleración, cada medición en promedio es registrada en un tiempo de 0.28ms por el microcontrolador lo que indica que cada muestra requiere en promedio 577.1ms, sin embargo, se realiza un proceso de serialización para enviar las muestras al servidor, este proceso toma un promedio de 277ms lo que daría un total de 854ms en promedio en enviar una muestra al servidor.

Las muestras fueron recibidas por el servidor mediante el protocolo MQTT y fueron almacenadas en formato de arreglo en una base de datos MYSQL con el esquema descrito en la Tabla 2.

Tabla 2. Diagrama de la tabla MYSQL de muestras de aceleración

aceleration_samples
aceleration_samples_id int (11) samples JSON (longtext) date (datatime)

4.2 Conjunto de Mediciones

En este estudio, se han diseñado siete conjuntos distintos para representar diferentes niveles de reducción en el número de mediciones, con el objetivo de evaluar cómo la cantidad de mediciones afecta la eficiencia del algoritmo de clasificación. Esta descripción corresponde a las etapas 1 y 2 de la metodología. La etapa 1, "Conjunto de aceleración de 2048 Mediciones", se refiere a las mediciones obtenidas a partir de los experimentos descritos en el capítulo 4, "Laboratorio Experimental". Por otro lado, la etapa 2, "Reducción de diferentes conjuntos de mediciones hasta 1024 mediciones", es la elección de distintos conjuntos que tiene como finalidad analizar el impacto de reducir el número de mediciones en el rendimiento del algoritmo de clasificación, estos se describen a continuación:

- **Conjunto 2048**: Este conjunto es el conjunto original, contiene 2048 mediciones aceleración y representa el conjunto con la mayor cantidad de mediciones.
- **Conjunto 2000**: Este conjunto está compuesto de 2000 mediciones, representa poca reducción con respecto al conjunto original.
- **Conjunto 1708:** Este conjunto contiene 1708 mediciones, representa una reducción intermedia con respecto a los demás conjuntos.
- **Conjunto 1500:** Este conjunto contiene 1500 mediciones representa una reducción intermedia con tendencia a menos datos con respecto a los demás conjuntos.
- **Conjunto 1400:** Este conjunto está compuesto por 1400 mediciones, representa una reducción más cercana a la mitad de los datos con respecto al conjunto original.
- **Conjunto 1366**: Este conjunto contiene 1366 mediciones, representa una reducción

significativa con respecto al conjunto original.

- **Conjunto 1024:** Este conjunto está compuesto por 1024 mediciones, es el conjunto con menor cantidad de mediciones.

Cada uno de estos conjuntos constará de 1000 muestras por cada escenario o clase con la finalidad de obtener una equivalencia en cada escenario y asegurarse que el modelo entrene de manera equitativa.

4.3 Preprocesamiento para la caracterización mediante la FFT

La precisión en la clasificación de fluidos mediante vibraciones depende de la calidad del preprocesamiento para la extracción de características de las muestras, dado lo anterior es esencial implementar técnicas que permitan la representación adecuada de la información, además de considerar el coste computacional para asegurar la eficiencia temporal, de acuerdo al estado del arte se analizaron algunas de estas técnicas. Se descartaron los MFCC y la FFT 2D debido a que esta estructura de datos requiere una mayor resolución espacial. Por ello la comparación fue más directamente con la transformada de Wavelet, sin embargo, debido a la naturaleza de los datos de vibraciones al ser señales que provienen de un flujo constante es más cercano a una señal estacionaria por lo cual se utilizó la FFT para realizar el preprocesamiento, además de ser computacionalmente más eficiente.

4.4 Entrenamiento con Diferentes conjuntos

El entrenamiento de las RNA con diferentes conjuntos de datos preprocesados es un paso crucial en la clasificación de fluidos mediante señales de vibración debido a que, afecta directamente la precisión del modelo. Para ello, se utilizaron varias configuraciones de RNA con el objetivo de determinar la arquitectura óptima que maximice la precisión del modelo. Esta configuración fue obtenida mediante una búsqueda en cuadrícula (grid search), lo que

permitió seleccionar tanto el número adecuado de neuronas por capa como la cantidad de capas ocultas. El tipo de RNA empleado en este proceso fue una **red neuronal densa**, debido a que los datos de frecuencia, derivados del preprocesamiento con la FFT, son representados en vectores, lo que optimiza tanto el entrenamiento como la clasificación.

Esto debido a que, a diferencia de redes convolucionales, que requieren estructuras de datos más complejas y multidimensionales, las RNA densas resultan más eficientes en términos computacionales para este tipo de datos vectoriales. Además, una vez completado el entrenamiento, el modelo se comporta como un **algoritmo reactivo**, lo que significa que puede realizar clasificaciones de manera eficiente en tiempo real. Esta característica es especialmente importante en este caso de estudio, donde la capacidad de procesar información de forma rápida y precisa es un requisito esencial. Cabe destacar que los diferentes conjuntos fueron utilizados para entrenar los modelos y posteriormente evaluar cómo la reducción de mediciones impacta en el rendimiento del sistema.

4.5 Comparación de diversos Conjuntos mediante Friedman

Una vez entrenados los modelos con los diferentes conjuntos de datos, se realizó un análisis estadístico utilizando el **test de Friedman**. Este test es particularmente útil cuando se comparan múltiples métodos o conjuntos bajo las mismas condiciones, ya que permite determinar si existen diferencias significativas entre ellos. En este caso, el test de Friedman se utilizó para evaluar si la reducción en el número de mediciones tuvo un impacto considerable en la precisión de clasificación de las RNA.

El análisis involucró los resultados obtenidos de la precisión del modelo para cada conjunto de datos y permitió identificar si la reducción en el tamaño de los conjuntos afectaba de manera estadísticamente significativa el rendimiento de las RNA. Esta prueba no paramétrica es ideal para este tipo de comparación debido a que los datos no requieren seguir una

distribución normal, y es útil para analizar las clasificaciones en función de los diferentes conjuntos evaluados. Los resultados del test se detallarán en la sección de resultados para interpretar el impacto de la reducción de datos sobre la eficacia del modelo.

CAPÍTULO 5

5. Propuesta de Solución

La solución propuesta para la clasificación de fluidos mediante vibraciones se basa en el uso de la FFT para convertir las señales de aceleración al dominio de la frecuencia, permitiendo extraer las características más relevantes para el proceso de clasificación. Posteriormente, estas características fueron utilizadas para entrenar modelos de RNA.

El diseño experimental se llevó a cabo en un entorno controlado descrito en la sección de laboratorio experimental, donde se capturaron datos de cuatro escenarios distintos: tubería vacía con bomba apagada, tubería vacía con bomba encendida, agua en tránsito y agua con jabón en tránsito. Estos datos se utilizaron para la búsqueda de configuración adecuada, seleccionando el número eficiente de capas ocultas y neuronas por capa mediante una búsqueda en cuadrícula (*grid search*).

5.1 Preprocesamiento mediante la Transformada de Rápida de Fourier

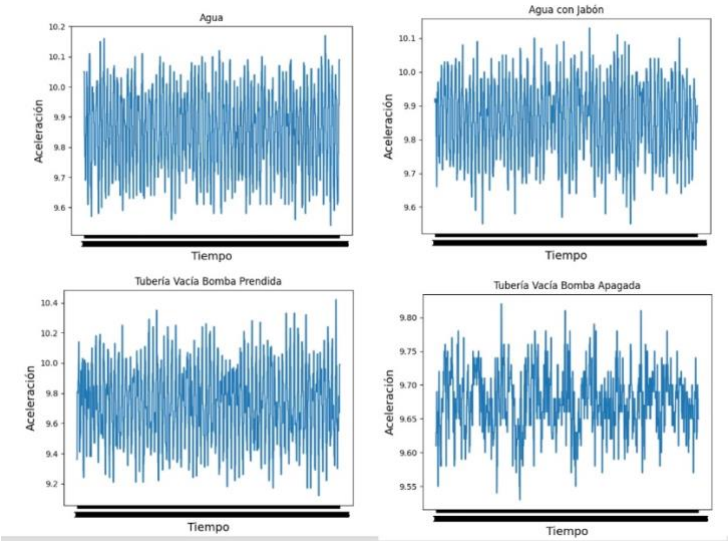

Figura 9. Muestras correspondientes a cada escenario

Derivado de los escenarios descritos con anterioridad y de la descripción en la sección de toma de muestras, en la Figura 9 se observa una muestra de cada escenario del conjunto original, se reveló que las muestras no tienen un valor medio de cero, este valor correspondiente al componente DC o 0 Hz esto interfiere en el análisis de frecuencias debido a que no se realiza correctamente la transformación al espectro de frecuencias por lo cual es crucial eliminar este componente antes de aplicar la transformación de frecuencia. Este preprocesamiento corresponde a la fórmula 11.

$$X'_i = x_i - \mu \tag{11}$$

Una vez que las muestras fueron preprocesadas mediante la fórmula 11 se puede observar en la Figura 9 que la forma de las muestras no cambia con respecto a la Figura 10 y la media de los datos es igual a 0 por lo cual, se procede a realizar el cambio al dominio de la frecuencia mediante la FFT.

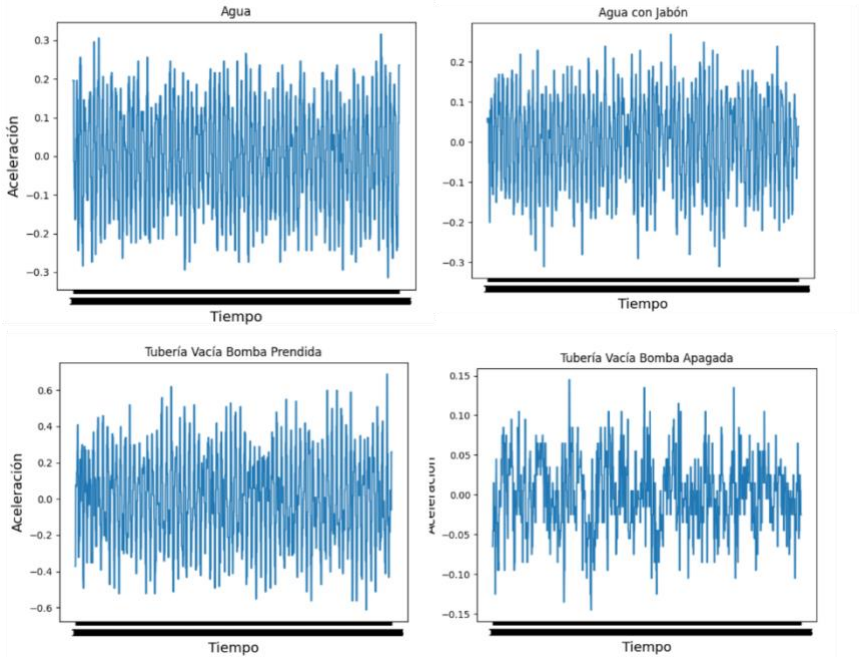

Figura 10. Datos preprocesando mediante la fórmula 1

En la Figura 11 se puede observar las muestras de la Figura 10 transformadas al domino de la frecuencia, se puede observar una diferencia significativa en las frecuencias de la tubería vacía con la bomba apagada, por otro lado cuando se activa la bomba sin el tránsito de fluido existe un pico de amplitud cerca de los 80hz que no existe en las demás muestras, para los fluidos (agua y agua con jabón) la forma de las gráficas es parecida, sin embargo, se aprecia una diferencia de amplitudes y diferencia en las frecuencias más bajas.

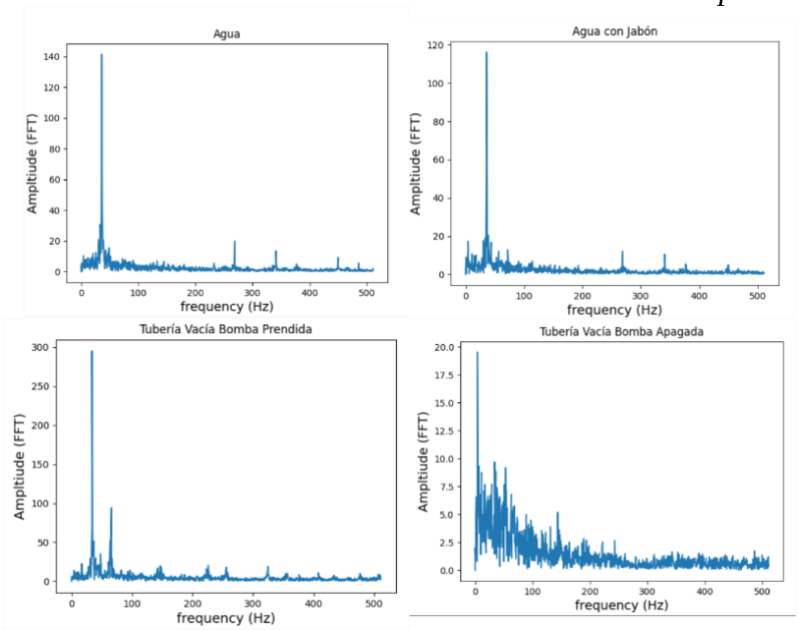

Figura 11. Gráfica de muestras en el dominio de la frecuencia.

5.2 Instancias de Entrenamiento, Validación y Prueba

Una vez que los datos fueron preprocesados mediante la FFT, los datos de cada conjunto se dividieron en 3 diferentes instancias que serán entrenamiento, validación y prueba. Cada uno de estas instancias contiene el mismo número de muestras por cada escenario o clases para garantizar una distribución equitativa, esto para prevenir posibles sesgos en el proceso de entrenamiento y validación. A continuación, se describen cada uno de las instancias.

Instancia de entrenamiento: Esta instancia de datos corresponde al 70% total de los datos (700 por cada escenario y 2800 en total) este será el conjunto principal el cual será encargado de alimentar la red neuronal para aprender e identificar patrones.

Instancia de validación: Esta instancia de datos corresponde al 15% total de los datos (150 por escenario y 600 en total) que trabaja en conjunto con la instancia validación para ajustar los parámetros de la RNA en el proceso de entrenamiento.

Instancia de prueba: Finalmente, una vez que la RNA terminó el proceso de entrenamiento se evaluará con esta instancia correspondiente al 15% del total de los datos (150 por escenario y 600 en total). Con esta instancia se evalúa el rendimiento de la RNA con datos no vistos en el proceso de entrenamiento para evitar sesgos de la RNA.

Una vez concluido la división de instancia estos serán mezclados aleatoriamente para evitar todos lo sesgos posibles usando la librería sklearn con la función shuffle con el parámetro random state = 42.

5.3 Arquitectura de Red Neuronal

Dentro del proceso de la realización de la búsqueda grid para la elección de la configuración de una arquitectura adecuada existirán parámetros de la RNA que se mantendrán fijos a lo largo del experimento, estos parámetros son descritos en la Tabla 3, en ellas en la primera columna se describen los nombres de los parámetros y en la segunda columna el nombre del parámetro seleccionado.

Tabla 3. Parámetros fijos de la Red Neuronal Artificial

Nombre del Parámetros	**Parámetro**
Función de Activación	ReLU
Optimizador	Adam
Pérdida	Categorical Cross Entropy
Métrica	Accuracy
Tasa de aprendizaje	0.001
Tamaño del Bache	5
Épocas	35

En la Tabla 4 se presentan los resultados de la experimentación de la búsqueda grid las columnas muestran los hiperparámetros elegidos en cada corrida (número de capas y neuronas por cada capa) si la celda no tiene ningún número significa que no se consideró esa capa

oculta, seguido de esas columnas se muestras métricas de evaluación correspondientes a la precisión de cada instancia de entrenamiento, validación y entrenamiento, finalmente se muestras los tiempos de entrenamiento y clasificación

Tabla 4. Resultados de Experimentación variando Neuronas por Capa y Números de Capas en la RNA

No.	Neuronas Capas de Entrada	Neuronas Capa Oculta 2	Neuronas Capa Oculta 3	Neuronas Capa Oculta 4	Neuronas Capa Oculta 5	Precisión Entrenam.	Precisión Validación	Precisión Prueba	Tiempo Entrenam.	Tiempo Clasificación
1	2	2	-	-	-	43%	44%	44%	12.16s	0.14s
2	4	4	-	-	-	63%	63%	63%	12.76s	0.14s
3	6	6	6	6	-	84%	81%	81%	13.31s	0.15s
4	12	12	12	12	-	94%	92%	91%	14.07s	0.16s
5	12	12	12	12	-	97%	94%	93%	14.82s	0.16s
6	16	16	16	16	-	99%	97%	96%	13.97s	0.15s
7	20	20	20	20	-	100%	97%	96%	16.27s	0.16s|
8	12	12	12	12	12	99%	94%	94%	14.83s	0.17s
9	20	20	20	20	20	100%	96%	96%	16.18s	0.16s
10	30	30	30	30	30	100%	97%	96%	16.63s	0.16s

Después de analizar la Tabla 4, se observa que la configuración número 7 que consiste en 4 capas ocultas (incluyendo la capa de entrada) con 20 neuronas por cada capa destaca debido a que se observa una mejora en la precisión en la instancia de precisión con respecto a configuraciones con menor número de neuronas y menor número de capas ocultas, además las configuraciones con mayor número de neuronas y mayor número de capas se observa una disminución en la instancia de precisión, esto puede deberse a problemas de sobre ajuste. Por consecuencia la configuración número 7 fue la elegida como arquitectura para la obtención de resultados, esta es representada en la Figura 11.

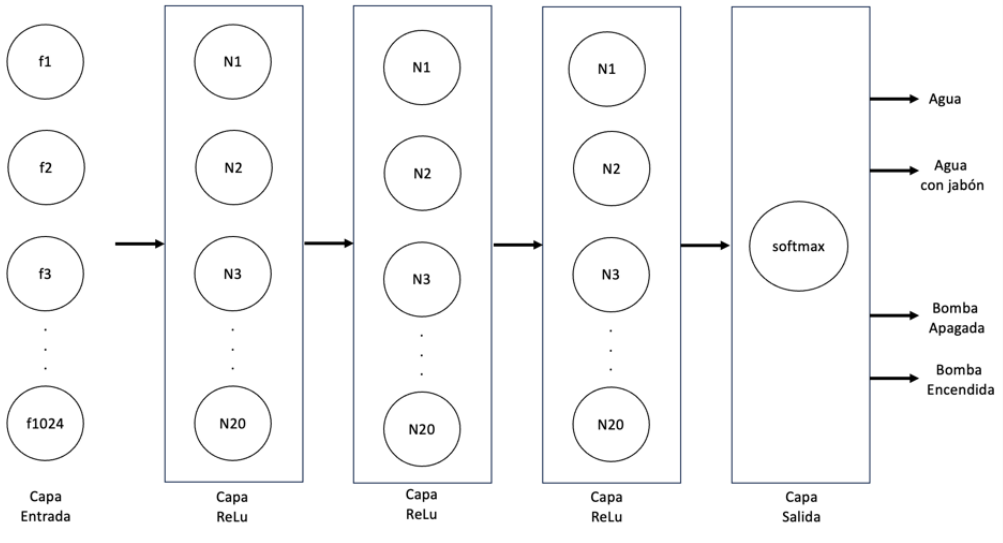

Figura 11. Arquitectura de RNA seleccionada

Finalmente se realizó una prueba adicional con la finalidad de observar como la arquitectura seleccionada se comportaba utilizando los datos de aceleración sin preprocesar como datos de entrada, en la Figura 12 se muestra los resultados en ella se relevan que la precisión de los datos de entrenamiento fue menor a 25% indicando que los datos sin preprocesar no son suficientes para realizar una caracterización, dando mayor sustento al uso de la FFT como parte del proceso para clasificar fluidos.

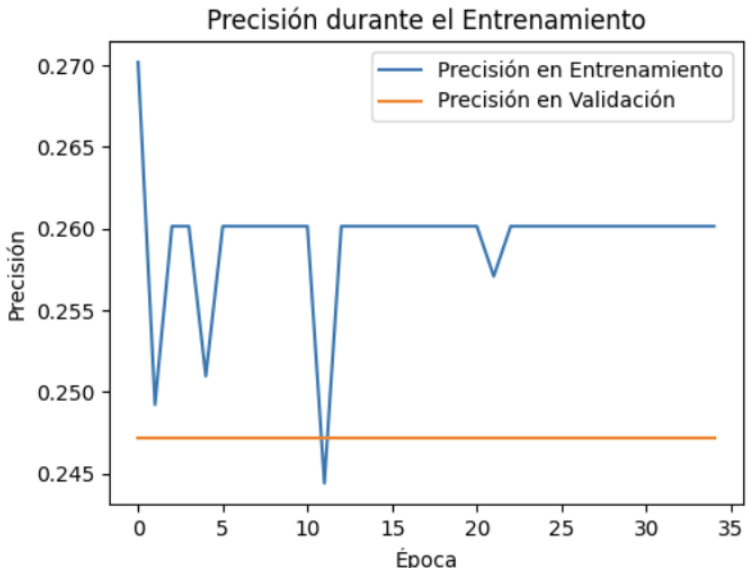

Figura 12. Gráfica de entrenamiento de la RNA utilizando únicamente los datos de aceleración

CAPÍTULO 6

6. Resultados

Este capítulo presenta los resultados obtenidos tras correr el algoritmo de RNA en los diversos conjuntos mostrados en la sección de conjuntos de mediciones, la finalidad es la obtención de métricas estadísticas que permitan comparan los conjuntos entre sí y posteriormente observar si existe una diferencia significativa al reducir el número de mediciones.

6.1 Presentación de Resultados Experimentales

Siguiendo el procedimiento descrito en la sección 4.1, “Laboratorio Experimental”, el laboratorio fue montado en tres ocasiones. En cada una de estas tres ocasiones, se llevó a cabo el procedimiento de toma de muestras descrito en la sección 4.2, “Toma de Muestras”, lo que resultó en tres bases de datos distintas de aceleración. Cada una de estas bases de datos fue dividida en los siete conjuntos de datos descritos en la sección 4.3, “Conjunto de Mediciones”, y posteriormente fueron preprocesadas según lo indicado en la sección 4.4, “Preprocesamiento para la caracterización mediante la FFT”. Finalmente, para cada una de las bases de datos, el algoritmo de redes neuronales artificiales (RNA) fue ejecutado con 11 corridas, lo que suma un total de 33 corridas. Este enfoque permite evaluar de manera directa y comparable cómo la variación en la cantidad de datos influye en el desempeño del modelo. A continuación, se presentan las Tablas 5 a 11 en dónde cada una de ellas corresponde a los diferentes conjuntos, el nombre de la Tabla corresponde al conjunto seleccionado, las 3 primeras columnas corresponden a la precisión de la instancia de entrenamiento, validación y prueba y las últimas dos columnas corresponden al tiempo de entrenamiento y de clasificación respectivamente, las filas corresponden a las 33 corridas realizadas por el algoritmo de RNA.

Tabla 5. Resultados de precisión y tiempo del conjunto de 2048 mediciones

Conjunto 2048				
Entrenamiento	Validación	Prueba	Tiempo Entrenamiento (s)	Tiempo prueba (s)
1.000	0.970	0.973	15.377	0.161
1.000	0.959	0.962	14.515	0.152
1.000	0.974	0.966	14.485	0.152
1.000	0.974	0.972	14.931	0.153
1.000	0.967	0.974	14.665	0.153
1.000	0.960	0.967	15.271	0.157
1.000	0.970	0.969	14.951	0.163
1.000	0.972	0.973	14.730	0.156
1.000	0.965	0.973	14.657	0.152
1.000	0.986	0.984	14.496	0.153
1.000	0.982	0.977	14.541	0.152
1.000	0.991	0.992	14.920	0.153
1.000	0.994	0.988	14.442	0.150
1.000	0.995	0.992	14.471	0.150
1.000	0.997	0.994	14.525	0.454
1.000	0.994	0.983	15.093	0.152
1.000	0.996	0.992	14.544	0.151
1.000	0.994	0.992	14.585	0.152
1.000	0.996	0.992	14.554	0.153
1.000	0.994	0.990	14.899	0.158
1.000	0.996	0.991	14.730	0.151
1.000	0.995	0.993	14.632	0.160
1.000	0.984	0.981	14.503	0.157
1.000	0.984	0.989	14.440	0.151
1.000	0.977	0.979	14.559	0.152
1.000	0.977	0.975	14.474	0.150
1.000	0.980	0.984	14.469	0.158
1.000	0.978	0.984	14.455	0.152
1.000	0.979	0.973	14.503	0.155
1.000	0.980	0.979	14.488	0.150
1.000	0.985	0.979	14.515	0.157
1.000	0.986	0.984	14.496	0.153
1.000	0.982	0.977	14.541	0.152

Tabla 6. Resultados de precisión y tiempo del conjunto de 2000 mediciones

Conjunto 2000				
Entrenamiento	Validación	Prueba	Tiempo Entrenamiento (s)	Tiempo prueba (s)
1.000	0.970	0.963	14.540	0.153
1.000	0.974	0.970	14.469	0.148
1.000	0.974	0.962	17.635	0.160
1.000	0.974	0.961	15.056	0.152
1.000	0.973	0.967	15.097	0.153
1.000	0.975	0.966	14.614	0.150
1.000	0.968	0.963	14.666	0.152
1.000	0.976	0.968	14.617	0.157
1.000	0.973	0.957	14.911	0.162
1.000	0.970	0.964	14.696	0.149
1.000	0.974	0.963	14.828	0.152
1.000	0.995	0.990	14.504	0.155
1.000	0.996	0.993	14.814	0.151
1.000	0.993	0.990	14.746	0.153
1.000	0.996	0.991	14.740	0.151
1.000	0.996	0.991	14.829	0.153
1.000	0.995	0.992	14.847	0.150
1.000	0.996	0.989	14.630	0.157
1.000	0.990	0.986	14.780	0.537
1.000	0.995	0.994	14.769	0.150
1.000	0.996	0.991	14.780	0.155
1.000	0.998	0.993	14.560	0.150
1.000	0.985	0.979	14.586	0.155
1.000	0.981	0.984	14.467	0.152
1.000	0.982	0.975	14.512	0.151
1.000	0.984	0.983	14.475	0.153
1.000	0.979	0.987	14.451	0.153
1.000	0.978	0.981	14.516	0.151
1.000	0.975	0.981	14.476	0.152
1.000	0.982	0.979	14.500	0.151

Tabla 7. Resultados de precisión y tiempo del conjunto de 1708 mediciones

Conjunto 1708				
Entrenamiento	Validación	Prueba	Tiempo Entrenamiento (s)	Tiempo prueba (s)
1.000	0.971	0.965	14.661	0.153
1.000	0.973	0.959	14.773	0.150
1.000	0.968	0.966	14.569	0.150
1.000	0.968	0.963	14.511	0.151
1.000	0.969	0.965	14.616	0.157
1.000	0.952	0.957	14.591	0.151
1.000	0.965	0.962	14.644	0.152
1.000	0.960	0.963	14.495	0.152
1.000	0.953	0.950	14.460	0.151
1.000	0.964	0.964	18.828	0.195
1.000	0.967	0.965	14.714	0.151
1.000	0.994	0.990	14.480	0.151
1.000	0.995	0.994	14.375	0.152
1.000	0.989	0.989	14.642	0.151
1.000	0.996	0.991	14.454	0.153
1.000	0.993	0.991	14.363	0.150
1.000	0.994	0.992	14.389	0.151
1.000	0.994	0.994	14.443	0.151
1.000	0.996	0.992	14.362	0.150
1.000	0.993	0.990	14.354	0.149
1.000	0.995	0.989	14.362	0.151
1.000	0.994	0.990	14.371	0.152
1.000	0.985	0.980	15.035	0.154
1.000	0.984	0.980	14.434	0.150
1.000	0.982	0.983	14.391	0.149
1.000	0.983	0.979	14.413	0.151
1.000	0.983	0.983	14.330	0.152
1.000	0.986	0.979	14.336	0.150
0.491	0.499	0.499	14.393	0.160
1.000	0.989	0.985	14.334	0.152
1.000	0.981	0.978	14.297	0.152
1.000	0.977	0.969	14.328	0.151
1.000	0.982	0.979	14.325	0.149

Tabla 8. Resultados de precisión y tiempo del conjunto de 1500 mediciones

Conjunto 1500				
Entrenamiento	Validación	Prueba	Tiempo Entrenamiento (s)	Tiempo prueba (s)
1.000	0.956	0.961	14.537	0.154
1.000	0.956	0.955	14.402	0.158
1.000	0.961	0.949	14.451	0.149
1.000	0.947	0.942	14.387	0.147
1.000	0.951	0.955	14.394	0.149
1.000	0.956	0.949	14.406	0.149
1.000	0.952	0.959	14.422	0.148
1.000	0.952	0.949	14.357	0.148
1.000	0.952	0.951	14.493	0.149
1.000	0.952	0.950	14.364	0.150
1.000	0.947	0.954	14.321	0.148
1.000	0.989	0.992	14.621	0.150
1.000	0.991	0.992	14.722	0.149
1.000	0.990	0.994	18.570	0.168
1.000	0.993	0.993	15.644	0.156
1.000	0.991	0.989	14.982	0.150
1.000	0.993	0.994	14.322	0.148
1.000	0.991	0.995	16.524	0.197
1.000	0.988	0.994	16.168	0.160
1.000	0.991	0.996	15.777	0.167
1.000	0.990	0.994	17.497	0.200
1.000	0.991	0.993	16.152	0.163
1.000	0.980	0.983	15.762	0.156
1.000	0.981	0.974	14.503	0.149
1.000	0.977	0.980	14.803	0.175
1.000	0.981	0.976	18.254	0.231
1.000	0.978	0.982	16.520	0.176
1.000	0.973	0.979	15.009	0.153
1.000	0.982	0.979	15.442	0.161
1.000	0.975	0.977	17.969	0.182
1.000	0.979	0.979	14.810	0.151
1.000	0.976	0.984	15.730	0.160
1.000	0.977	0.977	14.416	0.149

Tabla 9. Resultados de precisión y tiempo del conjunto de 1400 mediciones.

Conjunto 1400				
Entrenamiento	Validación	Prueba	Tiempo Entrenamiento (s)	Tiempo prueba (s)
0.995	0.939	0.938	14.533	0.158
1.000	0.958	0.957	14.543	0.149
1.000	0.963	0.958	15.093	0.155
1.000	0.962	0.955	14.267	0.147
0.999	0.963	0.959	14.331	0.150
1.000	0.966	0.961	14.285	0.149
1.000	0.958	0.956	14.328	0.150
1.000	0.962	0.959	14.350	0.149
1.000	0.966	0.955	14.347	0.162
0.996	0.961	0.960	14.277	0.149
1.000	0.966	0.955	14.333	0.146
1.000	0.990	0.986	14.480	0.149
1.000	0.990	0.989	14.424	0.148
1.000	0.992	0.989	14.328	0.149
1.000	0.989	0.986	14.369	0.150
1.000	0.988	0.990	14.415	0.151
1.000	0.990	0.989	14.347	0.149
1.000	0.986	0.990	14.321	0.150
1.000	0.986	0.989	14.349	0.150
1.000	0.986	0.985	14.331	0.149
1.000	0.986	0.989	14.365	0.149
1.000	0.988	0.989	14.345	0.149
1.000	0.975	0.979	14.417	0.153
1.000	0.974	0.982	14.370	0.148
1.000	0.978	0.980	14.393	0.149
1.000	0.976	0.980	14.384	0.150
1.000	0.978	0.979	14.357	0.150
1.000	0.978	0.979	14.377	0.148
1.000	0.976	0.983	14.387	0.148
1.000	0.978	0.982	14.379	0.151
1.000	0.977	0.978	14.390	0.150
1.000	0.977	0.978	14.401	0.151
1.000	0.972	0.974	14.421	0.149

Tabla 10. Resultados de precisión y tiempo del conjunto de 1366 mediciones.

Conjunto 1366				
Entrenamiento	Validación	Prueba	Tiempo Entrenamiento (s)	Tiempo prueba (s)
1.000	0.951	0.955	18.978	0.165
0.990	0.914	0.911	14.388	0.151
1.000	0.955	0.957	14.716	0.150
0.985	0.941	0.942	14.571	0.151
1.000	0.955	0.949	14.424	0.149
1.000	0.951	0.952	14.528	0.149
0.988	0.948	0.947	14.232	0.148
1.000	0.944	0.957	14.278	0.149
1.000	0.955	0.957	14.281	0.147
1.000	0.951	0.952	14.311	0.151
0.992	0.960	0.954	14.648	0.150
1.000	0.991	0.987	14.292	0.148
1.000	0.988	0.983	14.578	0.149
1.000	0.988	0.986	14.263	0.149
1.000	0.990	0.983	14.341	0.149
1.000	0.987	0.987	14.333	0.150
1.000	0.988	0.982	14.306	0.150
1.000	0.989	0.984	14.283	0.150
1.000	0.988	0.985	14.313	0.149
1.000	0.990	0.984	14.343	0.149
1.000	0.988	0.982	14.322	0.148
1.000	0.988	0.983	14.367	0.149
1.000	0.977	0.976	14.219	0.153
1.000	0.980	0.980	14.264	0.150
1.000	0.980	0.981	14.313	0.149
1.000	0.984	0.981	14.295	0.149
1.000	0.973	0.976	14.303	0.153
1.000	0.977	0.977	14.322	0.151
1.000	0.980	0.978	14.300	0.148
1.000	0.977	0.983	14.260	0.148
1.000	0.980	0.983	14.310	0.150
1.000	0.980	0.979	14.276	0.148
1.000	0.981	0.979	14.223	0.148

Tabla 11. Resultados de precisión y tiempo del conjunto de 1024 mediciones

Conjunto 1024				
Entrenamiento	Validación	Prueba	Tiempo Entrenamiento (s)	Tiempo prueba (s)
1.000	0.941	0.936	14.169	0.151
0.998	0.932	0.925	14.555	0.149
0.989	0.942	0.927	14.288	0.149
0.994	0.941	0.928	14.361	0.150
0.998	0.939	0.930	14.402	0.149
0.995	0.942	0.928	14.299	0.149
0.995	0.938	0.927	14.271	0.147
0.994	0.939	0.943	14.594	0.161
0.992	0.939	0.924	19.946	0.281
0.993	0.941	0.936	15.310	0.160
0.991	0.936	0.934	14.795	0.149
1.000	0.974	0.973	14.237	0.149
1.000	0.982	0.975	14.231	0.148
1.000	0.975	0.974	14.204	0.148
1.000	0.980	0.983	14.190	0.149
1.000	0.977	0.975	14.169	0.148
1.000	0.975	0.978	14.337	0.154
1.000	0.977	0.981	14.145	0.150
1.000	0.976	0.973	14.247	0.149
1.000	0.976	0.978	14.625	0.146
1.000	0.978	0.971	14.167	0.149
1.000	0.978	0.976	14.285	0.148
1.000	0.972	0.968	14.155	0.154
1.000	0.962	0.969	14.463	0.148
1.000	0.963	0.970	15.521	0.168
1.000	0.974	0.970	14.291	0.149
1.000	0.969	0.972	14.297	0.151
1.000	0.962	0.963	14.300	0.150
1.000	0.967	0.971	14.263	0.148
1.000	0.968	0.969	14.163	0.149
1.000	0.966	0.968	14.169	0.149
0.999	0.960	0.961	14.174	0.149
1.000	0.968	0.970	14.141	0.149

Para observar el comportamiento de aprendizaje de los diferentes conjuntos se realizó una corrida extra para registrar el aprendizaje de los diferentes conjuntos a lo largo de las 35 épocas en dónde fue entrenado el modelo de RNA, el eje X corresponde a cada época y el eje Y a la precisión de la instancia de validación como se muestra en la Figura 13.

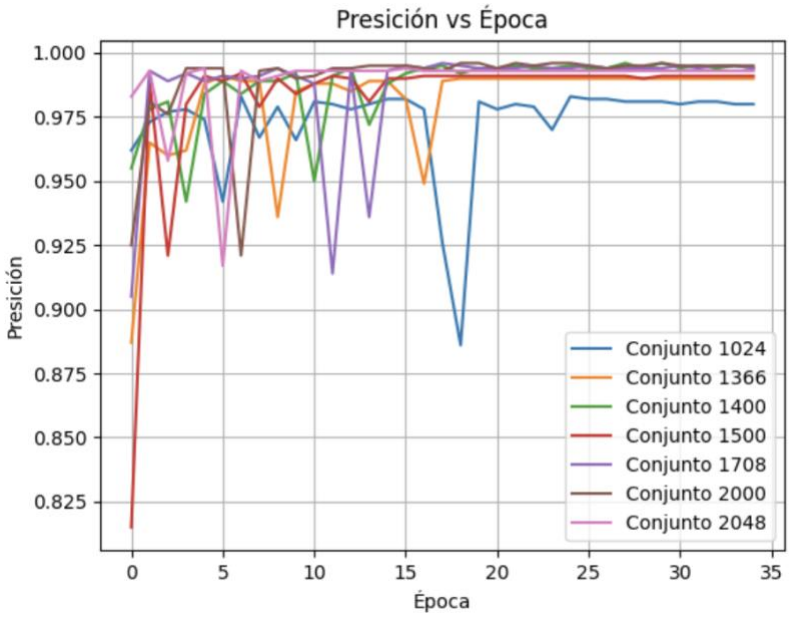

Figura 13. Comparativa de aprendizaje de los diferentes conjuntos

6.2 Presentación de Resultados Comparativos entre conjuntos

A continuación, en la Tabla 12 se muestras los resultados de la prueba de Friedman en dónde se utilizó un p-value = .5 y la hipótesis nula (H1) se define como la ausencia de diferencias estadísticamente significativas en el rendimiento de clasificación entre los modelos que utilizan diferentes tamaños de conjunto de datos, que varían desde 2048 hasta 1024 mediciones. El rechazo de esta hipótesis indicaría que las diferencias en el número de mediciones afectan significativamente la precisión del modelo, mientras que su aceptación sugeriría que la reducción de mediciones no impacta adversamente el rendimiento del modelo, permitiendo así optimizaciones en la captura y procesamiento de datos sin sacrificar la calidad de los resultados.

Comparación	**Estadístico**	**Ajuste p-value**	**Resultado**
1024 vs 2048	8.43308	0	H0 es rechazada
1024 vs 2000	8.0912	0	H0 es rechazada
1024 vs 1708	6.9516	0	H0 es rechazada
1024 vs 1500	6.29632	0	H0 es rechazada
1024 vs 1400	4.98578	0.00001	H0 es rechazada
2048 vs 1366	4.70087	0.00005	H0 es rechazada
1366 vs 2000	4.35899	0.00027	H0 es rechazada
1024 vs 1366	3.73221	0.00399	H0 es rechazada
2048 vs 1400	3.44731	0.01189	H0 es rechazada
1708 vs 1366	3.21939	0.02698	H0 es rechazada
2000 vs 1400	3.10543	0.0399	H0 es rechazada
1500 vs 1366	2.56411	0.21722	H0 es aceptada
1500 vs 2048	2.13676	0.68497	H0 es aceptada
1708 vs 1400	1.96582	1	H0 es aceptada
1500 vs 2000	1.79488	1	H0 es aceptada
1708 vs 2048	1.48149	1	H0 es aceptada
1500 vs 1400	1.31055	1	H0 es aceptada
1366 vs 1400	1.25357	1	H0 es aceptada
1708 vs 2000	1.13961	1	H0 es aceptada
1708 vs 1500	0.65527	1	H0 es aceptada
2048 vs 2000	0.34188	1	H0 es aceptada

Tabla 12. Resultados de la comparación de los diferentes conjuntos con la prueba de Friedman

6.3 Resultados Adicionales

En esta sección se propone presentar resultados adicionales con la finalidad de explorar más a fondo de cómo afecta reducir el número de mediciones en la precisión de clasificación de los fluidos propuestos, estas prueban se realizaron siguiendo el procedimiento de la sección pasada para generar las 3 bases de datos distintas, a continuación, se describe la primera prueba adicional.

6.3.1 Análisis de la Precisión en Conjuntos Reducidos

En la presente sección es descrito como la precisión de la RNA se ve influida directamente en la reducción de mediciones, abarcando desde el conjunto original de 2048 hasta 1400 mediciones realizando una base de datos por cada medición reducida. En la Figura 14 se muestra la gráfica derivada de este análisis en donde el eje x corresponde a número mediciones usado y el eje y el promedio de las 33 corridas derivado de realizar 11 corridas por cada experimento realizado.

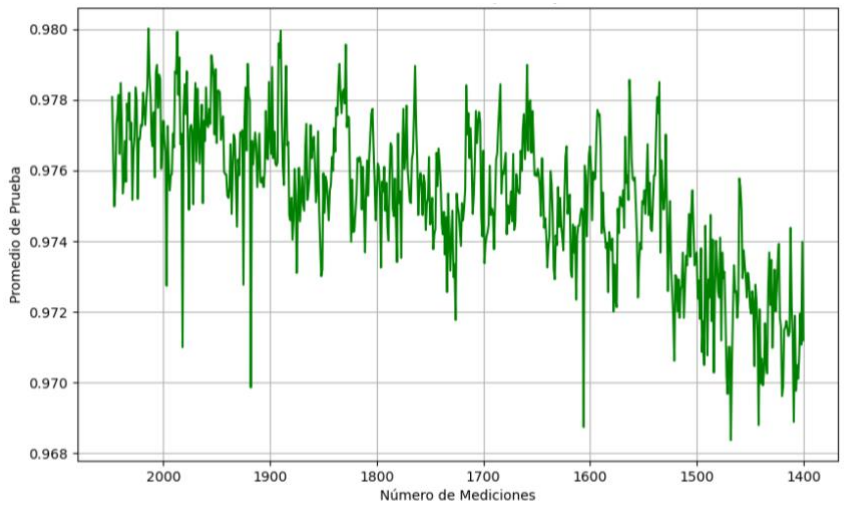

Figura 14. Medias de precisión desde el conjunto 2048 hasta el conjunto 1400

Tomando las bases de datos generadas, se realizó una prueba de Wilcox con la finalidad de evaluar de manera efectiva el impacto de la reducción del número de mediciones del modelo. Para garantizar la rigurosidad y la reproducibilidad del experimento se procedió a agrupar las bases de datos en pares utilizando los conjuntos 2048 y 2047 como el conjunto de referencia los cuales se asumen que tienen mayor información y por lo tanto más potencial a una mayor precisión. Este procedimiento es descrito en la Figura 15.

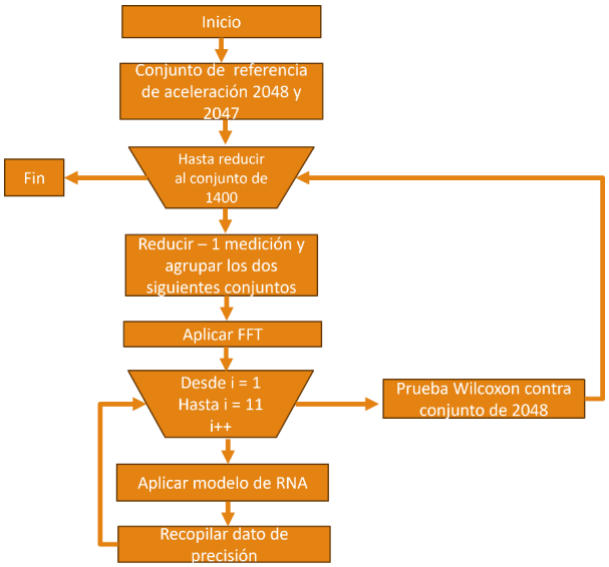

Figura 15. Diagrama de Flujo del experimento de comparativo en función del tamaño del conjunto

Derivado de esta experimentación en la Figura 16 se presenta el análisis del comportamiento de las diferencias significativas entre los conjuntos de datos a lo largo del experimento. El eje horizontal (eje *X*) muestra la distribución de los conjuntos de datos evaluados, que van desde el conjunto 2046 hasta el conjunto 1400, mientras que el eje vertical (eje *Y*) indica la presencia o ausencia de diferencias estadísticamente significativas, representada mediante un sistema binario (donde si indica diferencia significativa y no la ausencia de la misma).

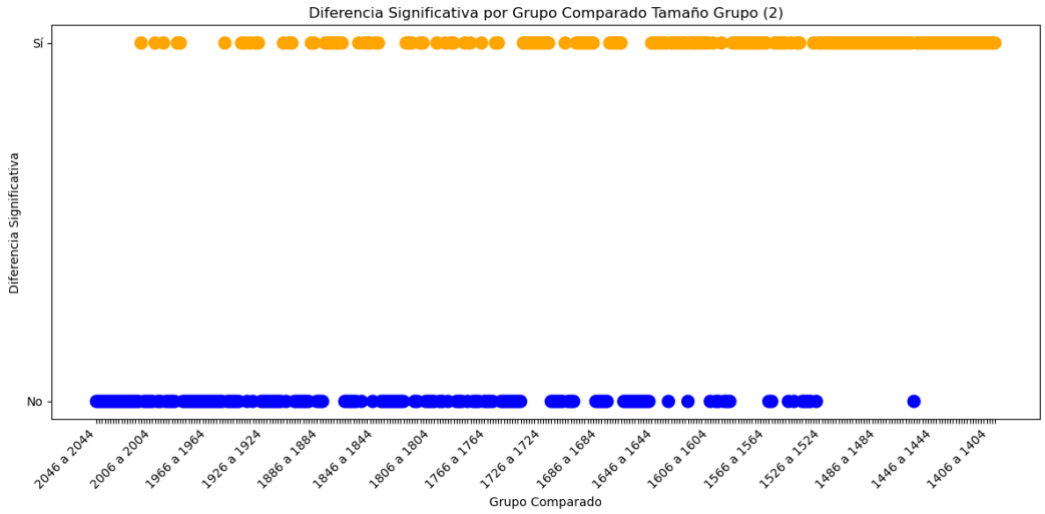

Figura 16. Gráfica de prueba de Wilcoxon de grupos de conjuntos desde 2046 hasta 1400 en contra del conjunto 2048 y 2047

6.3.2 Prueba con la Base de Datos Unificada

Figura 17. Distribución de conjuntos de entrenamiento, validación y prueba

En esta sección se busca robustez los resultados del modelo de RNA en un contexto más amplio, se llevó a cabo una prueba unificando las bases de datos de los tres experimentos previos. En lugar de realizar 11 corridas independientes para cada experimento, se combinaron las tres bases de datos en una sola base de datos, en la Figura 17 se muestra la distribución de instancias (Entrenamiento, validación y prueba) y se realizaron un total de 33 corridas con esta instancia unificada con el conjunto original o de 2048 mediciones para observar si el algoritmo aún era capaz de clasificar, estos resultados son mostrados en la Figura 18.

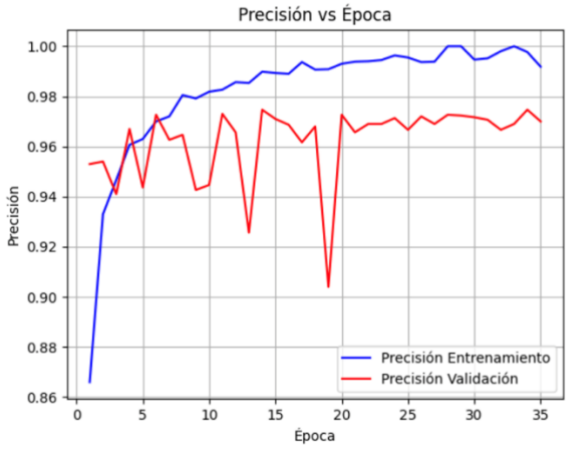

Figura 18. Aprendizaje del conjunto de 2048 juntando todas las muestras de derivado de los 3 experimentos

Finalmente, la base de datos unificada de igual manera que la sección anterior se obtuvieron bases de datos por cada conjunto reducido que varían desde 2048 hasta 1400 mediciones de aceleración. Este procedimiento aseguró una evaluación completa de cómo varían la precisión de clasificación del modelo. En la Figura 19 muestra este procedimiento en su totalidad en dónde el eje x muestra el número de mediciones del conjunto y el eje y promedio de la clasificación en la instancia de prueba.

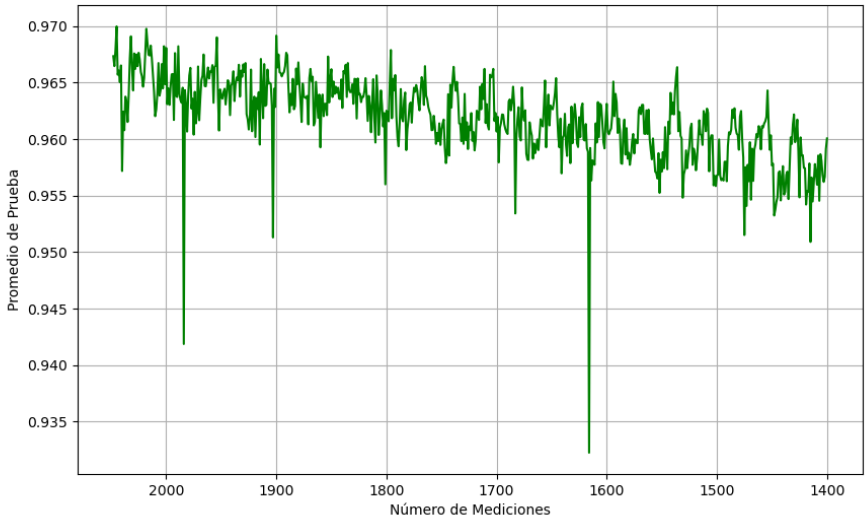

Figura 19. Medias a lo largo del conjunto 2049 hasta el conjunto 1400 utilizando la base de datos unificada

6.3.4 Comparativa Contra la Reducción de Componentes

En este capítulo se abordará una metodología adicional y se comparará con la metodología utilizada en el presente documento, para la clasificación de fluidos a partir de señales de vibración. La metodología expuesta en este documento se basa en la reducción de mediciones de aceleración antes de cualquier preprocesamiento, mientras que, la segunda metodología se basa en eliminar componentes menos significativos utilizando la información mutua identificando aquellos que tengan menos información para la realización de clasificación. Al comparar estos dos enfoques se busca tener una comparativa y determinar cuál de ellos ofrece una mejor compensación en la reducción de dimensionalidad de datos.

Para realizar lo anterior se utilizó la herramienta mutual_info_classif de scikit-learn, se identificaron las frecuencias según su relevancia informativa a partir de un espectro de frecuencias obtenido por FFT del conjunto original de 2048 mediciones. Se procedió a eliminar secuencialmente 24, 170, 274, 324, 341 y 512 de las frecuencias menos informativas. Estas cifras fueron seleccionadas para tener una comparativa con los tamaños reducidos de los conjuntos 2000, 1708, 1500, 1400, 1366 y 1024 mediciones, respectivamente. Esta metodología nos permitió emular la disminución de datos en escenarios donde se reduce el número de mediciones iniciales.

Conjunto Representado	Estadístico	p-valor	Diferencia significativa
2000	97.5	0.005	si
1708	37	0.000005	si
1500	112.5	0.0002	si
1400	130.05	0.0065	si
1366	181.0	0.076	si
1024	72	0.00007	si

Tabla 13. Prueba de Wilcoxon contra el conjunto original

Para evaluar estadísticamente las diferencias en la precisión de clasificación entre los conjuntos de datos con frecuencias reducidas y el conjunto original de 2048 mediciones, se aplicó la prueba de Wilcoxon. Esta prueba no paramétrica se utilizó para comparar dos muestras relacionadas, proporcionando una evaluación de si las medianas de dos conjuntos de datos difieren de manera significativa. Los resultados de esta prueba se resumen en la Tabla 13, donde cada fila corresponde a un conjunto de datos con un número específico de frecuencias eliminadas, y se comparan con el conjunto original.

Cada "Conjunto Representado" refleja las mediciones originales menos las frecuencias que fueron eliminadas basándose en su menor información mutua. Los valores del "Estadístico" y "p-valor" indican la magnitud de la diferencia y la probabilidad de que estas diferencias ocurran por casualidad, respectivamente. Los resultados muestran que todas las comparaciones resultaron en diferencias estadísticamente significativas, indicando que la eliminación de las frecuencias menos informativas tiene un impacto considerable en la precisión de la clasificación comparado con el uso del conjunto original.

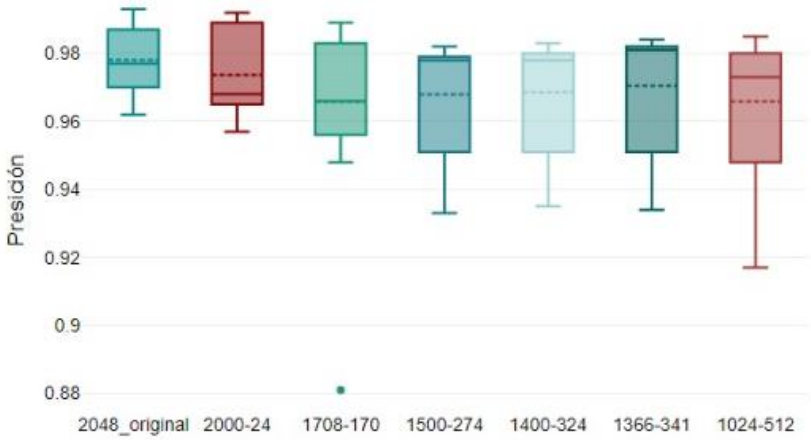

Figura 20 Comparación de la Precisión de Clasificación entre el conjunto y reducido.

Derivado de esta prueba en la Figura 20 se muestra como la distribución de la precisión de clasificación a través de la reducción de frecuencias, desde el conjunto original hasta el

conjunto donde se eliminaron 512 frecuencias. El conjunto original de 2048 mediciones, muestra una consistencia alta en la precisión. Los conjuntos subsiguientes, donde se eliminaron 24, 170, 274, 324, 341 y 512 frecuencias, respectivamente, exhiben una tendencia decreciente en la precisión a medida que disminuye el número de frecuencias. La variabilidad en la precisión también se incrementa con las reducciones más profundas, particularmente notable en los conjuntos 1366-341 y 1024-512, donde la dispersión y la presencia de valores atípicos se acentúan esto indicando que tiene afecta la precisión.

6.4 Discusión de Resultados

En este capítulo se aborda la discusión de los resultados presentados en la sección anterior obtenidos de la presente investigación titulada "Análisis Computacional para la Clasificación de Fluidos mediante Vibraciones y Aprendizaje Supervisado". Este estudio explora la eficiencia de las RNA en la clasificación de fluidos utilizando datos de frecuencia provenientes de datos de vibraciones, centrado en como la reducción de mediciones afecta la precisión del modelo.

Observando la Figura 13, se puede observar que a medida que a medida que los conjuntos tienen menos mediciones la precisión disminuye, sin embargo, se destaca que la precisión es mayor al 97%, indicando una precisión alta, además, se observa que, a partir de la época de la época 20 el algoritmo converge en los diversos conjuntos, por último, se observa que el conjunto 1024 se ve más alejado en el porcentaje de precisión con respecto a los demás conjuntos.

Esto último se puede observar de mejor manera en la Tabla 12 en dónde se muestran la comparativa entre conjuntos utilizando la prueba de Friedman con los diversos conjuntos, en dónde el conjunto de 1024 obtuvo una diferencia significativa con respecto a los demás conjuntos indicando que, existe perdida de información que afecta el desempeño del

algoritmo, esto debido a que a menor número de mediciones se obtienen menos frecuencias de acuerdo con el teorema de Nyquist-Shannon, lo que se traduce a menos características y menor resolución de la onda.

Por otro lado, el principal hallazgo de la investigación se destaca que el conjunto 2048 y el conjunto de 1500 no registraron diferencia significativa, esta reducción representa un decremento de 548 mediciones, lo que representa una reducción del 26.8%. Este hallazgo es crucial para las aplicaciones como en la industria del petróleo y gas o calidad del agua en donde puede premiar la eficiencia temporal y espacial.

Para observar la mejora temporal desde la perspectiva de la complejidad computacional, la FFT cuenta con una complejidad de O(Nlog2N) que para el conjunto de 2048 se obtiene una complejidad de 22528, y para el conjunto de 1500 se obtuvo una complejidad de aproximadamente 15826.12 obteniendo una mejora de aproximadamente de 6701.88 operaciones. Esto cobra relevancia en el proceso de entrenamiento en donde se preprocesaron 4000 muestras dando como resultado una mejora de 26,807,520 de operaciones indicando una mejora significativa.

Para el cálculo de la complejidad computacional de la RNA se desarrolló la formula número 8 en dónde se toma en cuenta la arquitectura realizada durante la experimentación, dando como resultado para el conjunto de 2084 21760 multiplicaciones y para el conjunto de 1500 16280 multiplicaciones realizando una mejora del 25.18%, de igual para se destaca una mejora tanto en el proceso de entrenamiento como de clasificación.

Por otro lado, para la mejora espacial parte que cada medición es de valor float esta utiliza 4 bytes. Por ende, cada muestra de 2,048 mediciones requiere 8192 bytes y para la muestra de 1500 mediciones se requieren 6000 bytes. Si realizamos la diferencia de memoria es de 2192

bytes lo que representa una mejora de 26.76%. Estos resultados representan una mejora considerable debido a que si se registra una muestra cada segundo en un día se pueden obtener una mejora de 189388800 bytes o 180.62MB.

La primera conclusión derivada de las pruebas adicional y observando la Figura 14, se muestra existe una tendencia a la baja indicando que efectivamente existe una pérdida de información que puede afectar al modelo, a pesar de esto se destaca una efectividad mayor al 97% a lo largo de los conjuntos indicando un rendimiento bueno del modelo inclusive reduciendo la cantidad de mediciones.

Observando la Figura 16 es notable que los conjuntos de 2046 y 2045 hasta los conjuntos de 1725 a 1724 predomina el resultado de no haber diferencia significativa y a partir de los conjuntos de 1526 a 1524 se observa que existe un cambio y comienza a observarse diferencia significativa. Este comportamiento destaca la importancia de analizar detalladamente este tipo de gráficos, en conjunto a gráficos de comportamiento de precisión ya que permiten identificar los umbrales críticos en los que la reducción del tamaño del conjunto de datos comienza a impactar significativamente en la precisión de la clasificación.

Finalmente observando la gráfica de las bases de datos unificadas, se concluye que el algoritmo de RNA fue capaz de realizar una clasificación con un porcentaje de acierto de mayor del 96% y en la Figura 19 que corresponde al comportamiento a lo largo de la reducción de mediciones se observó la misma tendencia a la baja de la precisión de clasificación, pero manteniendo el porcentaje de acierto alto.

CAPÍTULO 7

7. Conclusiones y Trabajos Futuros

A lo largo de esta investigación, se ha explorado el rendimiento de un modelo de RNA bajo condiciones controladas y experimentales, llegando a conclusiones significativas que no solo avanzan en el conocimiento científico, sino que también ofrecen orientación práctica para aplicaciones futuras. En este capítulo final, resumiremos los hallazgos clave obtenidos, Además, se sugieren direcciones futuras para investigaciones que puedan continuar donde este trabajo concluye.

7.1 Aportaciones

A diferencia de la mayoría de los estudios revisados que se enfocan en la clasificación de sonidos ambientales o en la detección de fugas, la presente investigación aborda específicamente la clasificación de fluidos (agua y agua con jabón) utilizando señales acústicas en específico de vibración. Esto es relevante ya que, solo una investigación anterior [4] se ha centrado en este enfoque, lo que subraya la necesidad y el potencial de tu trabajo en este ámbito.

Dado lo anterior, la aportación principal de esta investigación es una metodología que se basa en el uso de vibraciones generadas por el flujo de fluidos en una tubería, como se mencionó en la investigación [4] aborda esta problemática, sin embargo, se realiza mediante sensores del efecto Doppler y el mayor porcentaje de acierto se logró utilizando CNN y la transformada de Fourier 2D, lo que con lleva un mayor coste computacional a diferencia de la metodología presentada en esta tesis.

Finalmente, esta investigación no solo se limita a aplicaciones específicas, sino que también contribuye al conocimiento general sobre acústica y procesamiento de señales. Ofrece una metodología que permite observar si existen diferencias significativas al reducir el intervalo de muestreo. Esto facilita la toma de decisiones informadas en entornos industriales.

7.2 Conclusiones

El análisis de diversas estrategias proporcionó una base sólida para el desarrollo de una metodología eficiente para la clasificación de fluidos. Se optó por utilizar la FFT para el preprocesamiento de las vibraciones y una RNA de tipo densa como modelo de clasificación,

dado que estos algoritmos se ajustan mejor a la naturaleza de los datos vibracionales. Además, en comparación con otros enfoques como la CNN y la FFT 2D presentadas en el estado del arte, resultan ser más eficientes computacionalmente. Cumpliendo con esto el objetivo específico número 1.

Se definió un caso de estudio que involucró la clasificación de cuatro clases distintas: dos clases de fluidos (agua y agua con jabón) y dos clases correspondientes a periodos de inactividad (tubería vacía con la bomba encendida y tubería vacía con la bomba apagada). Para la evaluación de las mediciones, se utilizaron 2048 mediciones recolectadas en el Laboratorio Experimental , proporcionando una base sólida para la validación del modelo de clasificación. El cumplimiento de este segundo objetivo específico permitió una adecuada representación de las condiciones reales.

Se realizó la caracterización de las vibraciones mediante la FFT, lo que permitió descomponer las señales del dominio del tiempo al dominio de la frecuencia. Este paso fue clave para caracterizar los distintos escenarios correspondientes de cada clase de fluido y estado de inactividad, logrando una extracción de características significativa que alimentó el modelo de clasificación, este paso es relevante debido a que se trató de alimentar el modelo únicamente con los datos de aceleración y no logró superar el 25% de acierto. El cumplimiento de este tercer objetivo específico aseguró que la clasificación se basara en datos precisos y relevantes, además de reducir su dimensionalidad, mejorando la efectividad del modelo.

El modelo de clasificación basado en RNA fue desarrollado con éxito, alcanzando una precisión del 97%. Este modelo fue capaz de clasificar correctamente las cuatro clases definidas, demostrando un alto nivel de efectividad al utilizar las características extraídas mediante la FFT. Se realizaron ajustes iterativos para optimizar el rendimiento, cumpliendo satisfactoriamente con el cuarto y quinto objetivo específico planteado en el estudio.

El objetivo principal de esta investigación fue analizar la eficiencia de un algoritmo de AA aplicado a la clasificación de fluidos a partir de vibraciones, centrándose en cómo la reducción de mediciones afecta el rendimiento del modelo. Dado que los resultados muestran que se pudo reducir el conjunto original de 2048 a 1500 mediciones sin afectar estadísticamente el desempeño, se puede concluir que el objetivo fue alcanzado. Esta reducción no solo mejoró el rendimiento del algoritmo, sino que también proporcionó mejoras en la complejidad computacional y en el uso del espacio de almacenamiento.

A pesar de que el estudio [6] aborda la reducción de componentes a través de la eliminación de frecuencias menos relevantes, en el capítulo 7.3.4 "Comparativa Contra la Reducción de Componentes", se trató de abordar desde esta óptica, sin embargo, se detectó que cuando se utiliza esta metodología, en este contexto, no se logró reducir la misma o más cantidad de componentes de entrada de la RNA sin afectar estadísticamente el desempeño si se comparaba con el conjunto original de 2048 mediciones de aceleración.

En términos de precisión, se demostró que el modelo de RNA mantuvo un rendimiento superior al 97% en la mayoría de los conjuntos de datos, aunque la reducción a 1024 mediciones mostró una pérdida significativa de información, afectando el rendimiento. Esto refuerza la necesidad de encontrar un balance adecuado entre la cantidad de mediciones y la precisión del modelo. La prueba de Friedman indicó que, si bien la reducción del número de mediciones afecta el rendimiento, no se observó una diferencia significativa entre los conjuntos de 2048 y 1500 mediciones, lo que representa una reducción del 26.8% en el número de mediciones sin comprometer el desempeño.

Además, desde el punto de vista de la complejidad computacional, el uso de la FFT para procesar los datos de 1500 mediciones redujo la carga computacional en 6,701.88 operaciones en comparación con el conjunto de 2048 mediciones. Cuando se procesaron 4000 muestras, esto representó una mejora de 26,807,520 operaciones, optimizando el tiempo de preprocesamiento. Asimismo, la clasificación con el modelo de RNA presentó una reducción

del 25.18% en el número de multiplicaciones necesarias, mejorando la eficiencia tanto en el entrenamiento como en la clasificación.

Por otro lado, se observó una mejora significativa en el uso de espacio de almacenamiento. El conjunto de 2048 mediciones requería 8192 bytes, mientras que el de 1500 mediciones utilizaba 6000 bytes, lo que representa una mejora del 26.76% en el uso de memoria. Esta reducción cobra especial importancia en aplicaciones donde el monitoreo es continuo y la captura de datos es crítica, permitiendo ahorrar 180.62 MB al día al reducir el tamaño del conjunto de datos.

Finalmente, se resalta la importancia de elegir el intervalo de muestreo adecuado según las necesidades específicas del sector. Una reducción considerable en el intervalo puede resultar en una falta de información crítica para el algoritmo, lo que afectaría su eficiencia. Por otro lado, un intervalo de muestreo alto podría conllevar a tiempos más largos de clasificación y a mayores requerimientos de almacenamiento, lo que implicaría un aumento de costos.

7.3 Trabajo Futuro

A continuación, se describen los siguientes trabajados futuros derivados de la presente investigación:

Agregar fluido extra: Adicional a la clasificación de fluidos líquidos en una investigación futura podría considerar agregar un gas y observar si es existe una viabilidad en la clasificación utilizando vibraciones.

Explorar diversas arquitecturas de RNA: Una investigación futura interesante podría tratarse de explorar diversas arquitecturas de RNA centradas en la clasificación de fluidos a partir de vibraciones, en dónde un análisis más detallado podría traer mejoras en la eficiencia algorítmica.

Aplicación real en la industria: Observando que existe una potencial viabilidad en la industria para clasificar fluidos a partir de vibraciones se propone llevar la experimentación a un sector real y expandir el conocimiento derivado de esta investigación.

Monitoreo de propiedades del fluido: Dado los alcances y limitaciones de este estudio no contempla sensores como densidad o viscosidad, se propone el monitoreo de estas propiedades y observar cómo influyen a las vibraciones y a la clasificación de fluidos.

Diferentes materiales: De acuerdo al alcance y limitaciones de esta investigación se utilizó únicamente una medida de tubería, el uso de diferentes medidas de longitud o diámetro de tubería podría afectar al desempeño de un algoritmo de RNA basado en vibraciones para la clasificación de fluidos, estudiar este comportamiento a fondo sería de interés debido a que en la industria existen múltiples medidas.

Umbrales de Pérdida de Información: En este estudio se reveló que el conjunto de 1024 con respecto a los demás sufría de una perdida de información por lo cual, una investigación centrada en este fenómeno daría un mayor contexto de la existencia de posibles umbrales en donde se pierda información relevante para la realización de la clasificación de fluidos mediante vibraciones.

Bibliografía

[1] I. Yaroshenko, D. Kirsanov, M. Marjanovic, P. A. Lieberzeit, O. Korosynska, A. Mason, L. Frau y A. Legin, «Real-Time Water Quality Monitoring with Chemical Sensors,» 2020. https://doi.org/10.3390/s20123432.

[2] B. Guo, X. Liu y X. Tan, de *Petroleum Production Engineering*, 2017, p. 19. Gulf Professional Publishing. https://doi.org/10.1016/B978-0-12-809374-0.00001-5

[3] S. Elkatatny, Z. Tariq y M. Mahmoud, «Real Time Prediction of Drilling Fluid Rheological Propeties Using Artificial Neuronal Network Visible Mathematical Model,» *Journal of Petroleum Science and Engineering,* 2016. *Journal of Petroleum Science and Engineering*. https://doi.org/10.1016/j.petrol.2016.08.021

[4] N. Vahabi y D. R. Selviah, «Convolutional Neural Networks to Classify Oil, Water, Gas Wells Fluid Using Acoustic Signals,» *IEEE,* 2019. *Proceedings of the 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 1–6. https://doi.org/10.1109/ISSPIT47144.2019.9001845

[5] X. Xue, Y. Dong y X. Wu, «Motion Induced Eddy Current Sensor for Non-Intrusive Vibration Measurement,» *IEEE SENSORS JOURNAL,* 2020. . *IEEE Sensors Journal, 20*(2), 735-744. https://doi.org/10.1109/JSEN.2019.2943931

[6] R. Xiao, Q. Hu y J. Li, «Leak detection of gas pipelines using acoustic signals based on wavelet transform and Support Vector Machine,» 2019. https://doi.org/10.1016/j.measurement.2019.06.050

[7] B. Zhang y T. S. Leitner Jared, «Audio Recognition using Mel Spectrograms and Convolution Neural Networks,» 2019. *Proceedings of the IEEE International Conference on Audio and Signal Processing*

[8] E. E. Zurek, M. Gamarra R., J. Escorcia R., C. Gutierrez, H. Bayona, R. Pérez y X. García, «Análisis espectral para el reconocimiento de huellas acústicas,» *Revista De Investigaciones Universidad Del Quindío*, 2016.

[9] D. Wenbo, J. Leitne y T. Sam, «Time Frequency Analysis of Acoustic Signals for Condition Monitoring of Fluid Filled Pipelines,» 2007. *Proceedings of the 14th International Congress on Sound and Vibration.*

[10] C.-L. A. Wang, «An Industrial-Strength Audio Search Algorithm,» 2003. *Proceedings of the International Society for Music Information Retrieval*

[11] D. Calvo, «www.diegocalvo.es,» 26 Junio 2020. [En línea]. Available: https://www.diegocalvo.es/clasificacion-de-sonido-mediante-redes-neuronales-convolucionales/.

[12] O. Penttinen, M. Ulveström, K. Karlsson, V. Andersson, H. Andersson, J. Pettersson y O. Büker, «Towards flow measurement with passive accelerometers,» *Flow Measurement and Instrumentation, https://doi.org/10.1088/0957-0233/26/11/115306*

[13] J. Sunah, Y. Haesang, P. Kiheum, S. Yutaek y S. Woojae, «Monitoring Severe Slugging in Pipeline-Riser System Using Accelerometers for Application in Early Recognition,» 2019. https://doi.org/10.3390/s19183930

[14] M. M. Campagna, G. Dinardo, L. Fabbiano y G. Vacca, Fluid flow measurements by means of vibration monitoring, 2015.

[15] F. Chollet, Deep Learning with Python, 2017. ISBN: 9781617294433

[16] M. A. Domingo, Apuntes de Mecánica de Fluidos, 2011. Recuperado de http://creativecommons.org/licenses/by-sa/3.0/es

[17] A. Géron, Aprende Machine Learning con Scikit-Learn, Keras y Tensorflow, 2019. ISBN: 9781492071589

[18] T. M. Hagan, B. H. Demuth, H. M. Beale y O. De Jesús, Neuronal Network Design, 2014. ISBN: 978-0971732117

[19] X. Olabe, Redes Neuronales Artificiales y sus Aplicaciones, 2008.

[20] P. J. Freire, S. Srivallapondh, A. Napoli, E. J. Prilepsky y K. S. Turistsyn, Computational Complexity Evaluation of Neuronal Netowrk Applications in Signal Processing, 2022. *IEEE Journal on Selected Areas in Communications*. https://doi.org/10.1109/JSAC.2022.3159371

[21] C. B. Ross, Mutual Information between Discrete and Continuous Data Sets, 2014. https://doi.org/10.1371/journal.pone.0087357

[22] P. G. John y M. G. Dimitris, Tratamiento digital de señales, 2007.

[23] G. Mario, «A Survey on Pipelined FFT Hardware Architectures,» *Journal of Signal Processing System,* 2022. https://doi.org/10.1007/s11265-021-01655-1

[24] G. White, Análisis de Vibraciones, Azima DLI, 2019.

[25] J. Zafra, Ingeniería de sonido conceptos, fundamentos, y casos prácticos, Ra-Ma, 2018. . RA-MA Editorial. ISBN: 978-84-9964-743-2

[26] S. A. Ruiz, G. R. Francisco y R. J. C. Noguera, «Guía práctica de sensores,» 2010. . Siemens AG

[27] N. S. Yuri, Introducción a la dinámica de fluidos, Centro de Ciencias de la Atmosfera Univerisdad Nacional Autónoma de México, 2008. Universidad Nacional Autónoma de México. Recuperado de https://www.researchgate.net/publication/236164971

[28] R. T. Odgen, Essential Wavelets for Statistical Applications and Data Analysis, 1997. Springer Science+Business Media. https://doi.org/10.1007/978-1-4612-0709-2

Anexos

ANEXO A

1. Código Principal para la evaluación de la arquitectura propuesta para la clasificación de fluidos mediante vibraciones y AA

```
array_samples = [2048]
dbs = ['db']

for db in dbs:
    print(db)
    for samples in array_samples:
        print(samples)
        # samples = 2048
        prom_accuaray_train = []
        prom_accuaray_validation = []
        prom_accuaray_test = []
        prom_time_trainig = []
        prom_time_test = []

        size_fft = 750
        train = pd.read_csv('train.csv')
        validation = pd.read_csv('validation.csv')
        test = pd.read_csv('test.csv')

        # Separar características y etiquetas para el conjunto de entrenamiento
        X_train = train.drop(columns=[str(size_fft)])
        y_train = vectorizar(train[str(size_fft)])

        # Separar características y etiquetas para el conjunto de validación
```

```
        X_val = validation.drop(columns=[str(size_fft)])
        y_val = vectorizar(validation[str(size_fft)])

        # Separar características y etiquetas para el conjunto de prueba
        X_test = test.drop(columns=[str(size_fft)])
        y_test = vectorizar(test[str(size_fft)])

        for i in range(0, 11):
            print("corrida: " + str(i))
            size_x = X_train.shape[1]
            model = models.Sequential()
            model.add(layers.Dense(20, activation='relu', input_shape=(size_x,)))
            model.add(layers.Dense(20, activation='relu'))
            model.add(layers.Dense(20, activation='relu'))
            model.add(layers.Dense(20, activation='relu'))
            model.add(layers.Dense(4, activation='softmax'))

            model.compile(optimizer='adam',
                          loss='categorical_crossentropy',
                          metrics=['acc']
                          )
            start_time = time.time()
            history = model.fit(X_train, y_train,
                                batch_size=75, epochs=35, validation_data=(X_val,
y_val))
            end_time = time.time()
            trainig_time = end_time - start_time
            start_time_2 = time.time()
            test_loss, test_acc = model.evaluate(X_test, y_test)
            end_time_2 = time.time()
            testing_time = end_time_2 - start_time_2

            # estadísticos
            epocas = range(1, len(history.history['loss']) + 1)
            perdidas = history.history['loss']
            exactitud = history.history['acc']
            precision_entrenamiento = history.history['acc']
            precision_validacion = history.history['val_acc']

            prom_time_trainig.append(trainig_time)
            prom_accuaray_train.append(precision_entrenamiento[34])
            prom_accuaray_validation.append(precision_validacion[34])
            prom_accuaray_test.append(test_acc)
            prom_time_test.append(testing_time)

        df = pd.DataFrame(
            {'Entrenamiento': prom_accuaray_train, 'Validación':
prom_accuaray_validation, 'Prueba': prom_accuaray_test,
             'Tiempo Entrenamiento (s)': prom_time_trainig, 'Tiempo prueba (s)':
prom_time_test})

promedio = np.mean(prom_accuaray_train)
print(prom_accuaray_train)

promedio_validacion = np.mean(prom_accuaray_validation)
print(prom_accuaray_validation)

promedio_test = np.mean(prom_accuaray_test)
print(prom_accuaray_test)
```

```
print("Accuracy on test set:", test_acc)
```