

**UNIVERSIDAD POLITÉCNICA SALESIANA  
SEDE CUENCA**

**CARRERA DE INGENIERÍA MECATRÓNICA**

*Trabajo de titulación previo  
a la obtención del título de  
Ingeniero Mecatrónico*

**PROYECTO TÉCNICO:**

**DETECCIÓN Y DIAGNÓSTICO DE FALLOS DE  
CAJA DE ENGRANAJES RECTOS UTILIZANDO  
UN ALGORITMO DE CLASIFICACIÓN BASADO  
EN SIMILARIDAD DIFUSA APLICADO EN  
SEÑALES DE VIBRACIÓN**

**AUTOR:**

WILLY FABIÁN LOAIZA SÁNCHEZ

**TUTORA:**

ING. MARIELA CERRADA LOZADA, Ph.D

**CUENCA – ECUADOR**

**2021**

# CESIÓN DE DERECHOS DE AUTOR

Yo, Willy Fabián Loaiza Sánchez con documento de identificación N° 0705878874, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del trabajo de titulación: **“DETECCIÓN Y DIAGNÓSTICO DE FALLOS DE CAJA DE ENGRANAJES RECTOS UTILIZANDO UN ALGORITMO DE CLASIFICACIÓN BASADO EN SIMILARIDAD DIFUSA APLICADO EN SEÑALES DE VIBRACIÓN”**, mismo que ha sido desarrollado para optar por el título de: *Ingeniero Mecatrónico*, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en mi condición de autor me reservo los derechos morales de la obra antes citada. En concordancia, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, agosto del 2021.



Willy Fabián Loaiza Sánchez  
C.I. 0705878874

# CERTIFICACIÓN

Yo declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **“DETECCIÓN Y DIAGNÓSTICO DE FALLOS DE CAJA DE ENGRANAJES RECTOS UTILIZANDO UN ALGORITMO DE CLASIFICACIÓN BASADO EN SIMILARIDAD DIFUSA APLICADO EN SEÑALES DE VIBRACIÓN”**, realizado por Willy Fabián Loaiza Sánchez, obteniendo el *Proyecto Técnico*, que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana.

Cuenca, agosto del 2021.

A handwritten signature in blue ink, appearing to read 'M. Cerrada', with a long horizontal stroke extending to the right.

Ing. Mariela Cerrada Lozada, Ph.D  
C.I. 0151771813

# DECLARATORIA DE RESPONSABILIDAD

Yo, Willy Fabián Loaiza Sánchez con documento de identificación N° 0705878874, autor del trabajo de titulación: **“DETECCIÓN Y DIAGNÓSTICO DE FALLOS DE CAJA DE ENGRANAJES RECTOS UTILIZANDO UN ALGORITMO DE CLASIFICACIÓN BASADO EN SIMILARIDAD DIFUSA APLICADO EN SEÑALES DE VIBRACIÓN”**, certifico que el total contenido del *Proyecto Técnico* es de mi exclusiva responsabilidad y autoría.

Cuenca, agosto del 2021.



Willy Fabián Loaiza Sánchez  
C.I 0705878874

# Dedicatoria

A Dios y a mi familia.

# Agradecimientos

Agradezco a Dios por ser siempre mi mayor fortaleza para alcanzar la culminación de mis estudios de Pre-grado. A mis padres por el apoyo incondicional y porque han estado en todo momento de mi vida motivándome para que dé lo mejor de mí. Estaré muy agradecido con los docentes de la Universidad porque aportaron para mi crecimiento como profesional, además siempre recordaré a la Doctora Mariela Cerrada por la confianza que depositó en mí y por su guía durante el trabajo de titulación. Y a los amigos que estuvieron durante este proceso de estudio.

# Resumen

El proyecto técnico con enfoque general titulado “Detección y diagnóstico de fallos de caja de engranajes rectos utilizando un algoritmo de clasificación basado en similaridad difusa aplicado en señales de vibración”, presenta la aplicación de un marco metodológico para la detección y diagnóstico de fallos usando señales de vibración de la caja de engranajes disponible en el Laboratorio de Vibraciones de la Universidad Politécnica Salesiana, Sede Cuenca, Ecuador. El contenido del proyecto desarrolla un enfoque de investigación cuantitativa y de experimentación, ya que se analizan datos numéricos de las series temporales de señales de vibración, que va desde su procesamiento hasta su uso específico por un modelo de clasificación de fallos.

La introducción se presenta en la sección 1, en donde se aborda de forma general la temática que se trata en el trabajo. El problema es tratado en la sección 2, se presentan los antecedentes del problema de estudio, la importancia y sus alcances, además de la delimitación geográfica. En la sección 3, se presentan los objetivos planteados, tanto general como específico, para la realización del proyecto técnico.

La fundamentación teórica se presenta en sección 4. Se presentan los aspectos generales del mantenimiento basado en la condición considerando el monitoreo de las señales de vibración, y se mencionan los modos de fallas en las cajas de engranajes como diente roto, agrietamiento, picadura y rayadura, que serán estudiados en este proyecto. Luego se detallan los aspectos teóricos sobre los diagramas de Poincaré, y la extracción de características de tales diagramas, que se asocian con indicadores de condición de falla. Finalmente, se abordan los aspectos fundamentales sobre la selección de las mejores características usando ANOVA y métricas de validación de grupos de datos, se se presenta el algoritmo LAMDA HAD (Learning Algorithm for Multivariate Data with High Adequacy Degree).

En el capítulo 5 se desarrolla todo el marco metodológico usado en este proyecto. Primero se describe el proceso de adquisición de señales de vibración de la caja de engranajes rectos, con todos los modos de fallas estudiados en este proyecto. Posteriormente, se detalla el proceso de expansión de muestras de señales de vibración, para poder mejorar desde el punto de vista estadístico, la información disponible en las señales originales. Posteriormente se explica el proceso de extracción y selección de características a partir de diagramas de Poincaré, acompañado con análisis exploratorio de datos, usando la técnica de visualización T-SNE.

Finalmente, se explica la aplicación del algoritmo LAMDA HAD en este caso de estudio, tanto en la fase de entrenamiento como en la de prueba. Se explicará acerca de la necesidad de obtener diferentes clasificadores por grupos de fallos y condiciones de operación de velocidad, los cuales mostrarán buenos resultados a nivel de precisión en la clasificación del fallo.

En el capítulo 6 se presentan los resultados obtenidos, usando el marco metodológico previamente explicado. La implementación de los algoritmos tanto de procesamiento de señales usando diagramas de Poincaré, extracción y selección de atributos, y el ajuste y pruebas de los modelos de clasificación usando LAMDA HAD se desarrollaron en la herramienta Matlab. Como un análisis adicional de este proyecto, se presentará un estudio exploratorio para determinar cómo se podrían usar simultáneamente los diferentes grupos de clasificadores para estimar la clase de falla de una muestra desconocida.

Finalmente en el capítulo 7 se presenta las conclusiones y recomendaciones finales del proyecto de titulación. En esta sección se expondrán las problemáticas presentadas en el desarrollo de este trabajo, también se hará mención de la importancia del mantenimiento predictivo para una detección y diagnóstico temprano de fallos en un proceso industrial que tenga cajas de engranajes rectos como parte de la maquinaria que lo soporta.



# Índice

<b>Cesión de Derechos de Autor</b>	<b>I</b>
<b>Certificación</b>	<b>II</b>
<b>Declaratoria de Responsabilidad</b>	<b>III</b>
<b>Dedicatoria</b>	<b>IV</b>
<b>Agradecimientos</b>	<b>V</b>
<b>Resumen</b>	<b>VI</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Problema</b>	<b>3</b>
2.1. Antecedentes del problema de estudio . . . . .	3
2.2. Importancia y alcances . . . . .	4
2.3. Delimitación . . . . .	5
<b>3. Objetivos</b>	<b>7</b>
3.1. Objetivo general . . . . .	7
3.2. Objetivos específicos . . . . .	7
<b>4. Fundamentación teórica</b>	<b>8</b>
4.1. Mantenimiento basado en la condición . . . . .	8
4.2. Monitoreo de las señales de vibración . . . . .	10
4.3. Fallas en sistemas de cajas de engranajes . . . . .	10
4.4. Clasificación de fallos basado en datos . . . . .	17
4.5. Diagramas de Poincaré . . . . .	20
4.5.1. Método de Poincaré . . . . .	20
4.5.2. Características de Poincaré . . . . .	21
4.6. Selección de características utilizando ANOVA e índices de validación de agrupaciones . . . . .	26
4.6.1. Análisis de varianza (ANOVA) y prueba de Tukey . . . . .	26
4.6.2. Evaluación de la validez de la agrupación . . . . .	27

4.6.3.	Algoritmo de selección de características utilizando ANOVA y el índice $CDbw$ . . . . .	30
4.7.	Clasificador basado en LAMDA . . . . .	32
4.7.1.	LAMDA . . . . .	32
4.7.2.	LAMDA HAD . . . . .	36
4.7.3.	Entrenamiento y funcionamiento del clasificador LAMDA HAD . . . . .	39
<b>5.</b>	<b>Marco Metodológico</b>	<b>40</b>
5.1.	Adquisición de señales . . . . .	40
5.1.1.	Plan experimental . . . . .	40
5.1.2.	Modos de fallos y niveles de severidad . . . . .	40
5.2.	Preprocesamiento de las señales . . . . .	45
5.2.1.	Expansión con las ventanas . . . . .	45
5.2.2.	Determinación de los retardos . . . . .	46
5.2.3.	Obtención de los gráficos de Poincaré . . . . .	47
5.3.	Extracción de los atributos de Poincaré . . . . .	48
5.4.	Selección de atributos con ANOVA + $CDbw$ . . . . .	50
5.4.1.	Datos Train-Test y normalización . . . . .	50
5.4.2.	Selección de atributos y análisis exploratorio usando t-SNE . . . . .	51
5.5.	Clasificación de fallos con LAMDA-HAD . . . . .	54
5.5.1.	Determinación de las mejores modelos de clasificadores con los datos de entrenamiento . . . . .	54
5.5.2.	Comprobación del clasificador con los datos de Test . . . . .	55
<b>6.</b>	<b>Validación y pruebas del diseño</b>	<b>59</b>
6.1.	Resultados con los datos de entrenamiento . . . . .	59
6.2.	Análisis de resultados con los datos de prueba . . . . .	62
6.3.	Clasificación de una muestra con el uso simultáneo de todos los clasificadores	64
<b>7.</b>	<b>Conclusiones y recomendaciones</b>	<b>69</b>
	<b>ANEXOS</b>	<b>71</b>
<b>A.</b>	<b>Clasificación de una muestra con el uso simultáneo de todos los clasificadores</b>	<b>72</b>
<b>B.</b>	<b>Códigos de MatLab utilizados para el desarrollo del algoritmo para la cla-</b>	



## Lista de Tablas

1.	Especificación del equipo mecánico utilizado en el experimento. . . . .	43
2.	Niveles de Severidad para el modo de fallo de diente roto . . . . .	44
3.	Ejemplo de Selección de los mejores atributos con ANOVA + CDbw con el acelerómetro 1, modo de fallo Picadura, Grupo 1, Velocidad 1 . . . . .	51
4.	Ejemplo de Selección de los mejores atributos con ANOVA + CDbw con el acelerómetro 1, modo de fallo Picadura, Grupo 2, Velocidad 1 . . . . .	52
5.	Ejemplo de Selección de los mejores atributos con ANOVA + CDbw con el acelerómetro 1, modo de fallo Picadura, Grupo 3, Velocidad 2 . . . . .	53
6.	Mejor clasificador LAMDA-HAD para los grupos con con el acelerómetro 1, modo de fallo picadura, con datos de entrenamiento . . . . .	55
7.	Tabla de precisión de los datos de prueba de Picadura . . . . .	56
8.	Resultados de los clasificadores en el modo de fallo de Picadura con los datos de entrenamiento . . . . .	59
9.	Resultados de los clasificadores en el modo de fallo de Diente Roto con los datos de entrenamiento . . . . .	60
10.	Resultados de los clasificadores en el modo de fallo por agrietamiento con los datos de entrenamiento . . . . .	61
11.	Resultados de los clasificadores en el modo de fallo por rayadura con los datos de entrenamiento . . . . .	61
12.	Tabla de precisión de los datos de prueba de Diente Roto . . . . .	62
13.	Tabla de precisión de los datos de Test del modo de fallo por agrietamiento .	63
14.	Tabla de precisión de los datos de Test del modo de fallo por picadura . . . .	63
15.	Tabla de precisión de los datos de Test del modo de fallo por rayadura . . .	64

## Lista de Figuras

2.1. Ubicación de la Universidad Politécnica Salesiana . . . . .	5
4.1. Enfoque de diagnóstico de fallos basado en modelos (Sánchez 2018) . . . . .	9
4.2. Esquema del proceso de diagnóstico de fallos basado en datos para el Mantenimiento Basado en la condición (Niu 2017; Vachtsevanos y Vachtsevanos 2006). . . . .	9
4.3. Grieta en el flanco del diente (Fuente: documentos internos del grupo GIDTEC-UPS) . . . . .	11
4.4. Diente roto en un 50 % (Fuente: documentos internos del grupo GIDTEC-UPS)	12
4.5. Dientes de engranaje con 2 agujeros por picadura, de Diámetro: 1,5 mm y Profundidad: 1,5 mm (Fuente: documentos internos del grupo GIDTEC-UPS)	14
4.6. Dientes de engranaje con rayadura donde existe una línea de Profundidad 0,5 mm, Ancho 1 mm y Longitud 16 mm. (Fuente: documentos internos del grupo GIDTEC-UPS) . . . . .	16
4.7. Esquema del proceso para clasificación de fallos basado en datos . . . . .	17
4.8. Mapa de Poincaré para el atractor de Lorenz 3D (Rubén Medina y col. 2020)	20
4.9. Gráfica de Poincaré generada en la extracción de atributos. . . . .	21
4.10. Gráfica que indica SD1 y SD2 (Brennan y col. 2001). . . . .	22
4.11. Casco convexo y área del casco convexo (Rubén Medina y col. 2017). . . . .	23
4.12. T-norma and T-conorma (Mizumoto 1989). . . . .	35
4.13. Procedimiento para el entrenamiento y funcionamiento del clasificador LAMDA-HAD . . . . .	39
5.1. Diagrama de flujo del proceso . . . . .	41
5.2. Banco de pruebas del laboratorio de vibraciones de la UPS . . . . .	42
5.3. Montaje experimental esquemático (Fuente: documentos internos del grupo GIDTEC-UPS) . . . . .	42
5.4. Imágenes del diente roto con los niveles de severidad . . . . .	44
5.5. Señal de vibración y ventanas deslizantes que indica el solapamiento y duración de un sub-señal de vibración . . . . .	45
5.6. Autocorrelación promedio con el Acelerometro 1 para el modo de fallo picadura, con la velocidad 1 . . . . .	47
5.7. Señal de vibración vs Señal de vibración retrasada . . . . .	48
5.8. Comparación del gráfico de Poincaré clase Normal vs clase con severidad alta	49
5.9. Comparación del Área convexa clase Normal vs clase con severidad alta () .	50

5.10. Gráfico de los 6 atributos con TSNE para el Grupo 1 con modo de fallo picadura	52
5.11. Gráfico de los 4 atributos con TSNE para el Grupo 2 con modo de fallo picadura	53
5.12. Gráfico de los 7 atributos con TSNE para el Grupo 3 con modo de fallo picadura	54
5.13. Resultados de la clasificación de los datos de Test con Picadura . . . . .	57
5.14. Diagrama de caja de Picadura con los datos de prueba para todos los grupos	57
6.1. Esquema del proceso para el análisis complementario . . . . .	65
6.2. Representación de las letras en los diagramas de caja . . . . .	65
6.3. Diagrama de caja de P1 con todos los grupos . . . . .	66
6.4. Datos clasificados en P1 en los grupos de Picadura . . . . .	66
6.5. Diagrama de caja de P2 con todos los grupos . . . . .	67
6.6. Datos clasificados en P2 en los grupos de Picadura . . . . .	67
6.7. Diagrama de caja de P3 con todos los grupos . . . . .	67
6.8. Datos clasificados en P3 en los grupos de Picadura . . . . .	68
A.1. Diagrama de caja de P4 con todos los grupos . . . . .	72
A.2. Datos clasificados en P4 en los grupos de Picadura . . . . .	72
A.3. Datos clasificados en P5 en los grupos de Picadura . . . . .	73
A.4. Diagrama de caja de P5 con todos los grupos . . . . .	73
A.5. Diagrama de caja de P6 con todos los grupos . . . . .	74
A.6. Datos clasificados en P6 en los grupos de Picadura . . . . .	74
A.7. Datos clasificados en P7 en los grupos de Picadura . . . . .	75
A.8. Diagrama de caja de P7 con todos los grupos . . . . .	75
A.9. Diagrama de caja de P8 con todos los grupos . . . . .	76
A.10. Datos clasificados en P8 en los grupos de Picadura . . . . .	76
A.11. Diagrama de caja de P9 con todos los grupos . . . . .	77
A.12. Datos clasificados en P9 en los grupos de Picadura . . . . .	77

# 1. Introducción

Una de las medidas que se tiene en consideración para disminuir el costo de la producción en la industria es el costo de mantenimiento de los equipos, considerando como el segundo más alto o incluso, en algunas empresas, el más alto de los costos de operación (Al-Najjar y col. 2016; Kim y col. 2016). Particularmente, la maquinaria rotativa como las cajas de engranajes, tiene gran presencia en la mayoría de los procesos industriales por ser elementos transformadores de energía mecánica y las encargadas de transmitir potencia entre ejes (AlShorman y col. 2020; Z. Liu y L. Zhang 2020; Raghav y R. B. Sharma 2021). Esta es la razón principal para implementar un sistema para la detección de fallos para esta maquinaria, y de este modo facilitar la implementación de un programa de mantenimiento basado en la condición, importante para reducir los costes de producción y mejorar la calidad de los productos o servicios.

Para el monitoreo de la condición se consideran tres pasos fundamentales: (i) Adquisición de datos, (ii) Procesamiento de datos, (iii) Toma de decisiones. La adquisición de datos es el primer paso para el monitoreo de la maquinaria, en donde se recoge y almacenan señales relacionada con la condición (vibración, corriente, emisión acústica, entre otras), siendo las señales de vibración las más usadas para el monitoreo de la condición en cajas de engranajes (V. Sharma 2021). En el procesamiento de datos, con base a la información obtenida en el paso anterior, se analiza con detenimiento las características esenciales que aportan información para la interpretación de lo que ocurre en las máquinas, normalmente relacionadas con indicadores estadísticos tales como valor medio, desviación estándar, kurtosis, entre otros calculados sobre la señal monitoreada (Sánchez y col. 2018); y en la toma de decisiones se determina las acciones pertinentes que se deben realizar para solucionar los fallos basándose en los resultados de diagnóstico y/o pronóstico (Jardine y col. 2006).

Con el auge de la inteligencia artificial y del aprendizaje automático, el paso de procesamiento de datos puede incluir el uso de modelos computacionales capaces de identificar o clasificar las posibles fallas en las que se encuentra la máquina. Estos modelos usan como información de entrada los indicadores que se extraen de las señales, y mediante algún algoritmo aprendido se genera la información relevante sobre la condición de la máquina (R. Liu y col. 2018).

En este proyecto se propone el uso de señales de vibración como señal informativa de la condición, se procesa la señal con el método de generación de gráficos de Poincaré para luego

extraer un conjunto de indicadores que caracterizan la forma del gráfico obtenido para cada condición de fallos, y finalmente se usará un algoritmo de clasificación denominado LAMDA (Learning Algorithm for Multivariable Data Analysis) para estimar la clase de falla que puede tener el equipo.

Con el desarrollo de la modalidad de investigación bibliográfica, se obtendrá la información necesaria para la construcción de los gráficos de Poincaré, los diferentes indicadores que se van a extraer de estos gráficos, así como todo lo referente al algoritmo LAMDA, i.e. sus actualizaciones, referencias de aplicación, y demás aspectos teóricos, con lo que se estructurará y desarrollará el proyecto de una manera plausible, segura y enfocada en el cumplimiento de la detección de fallos en un caja de engranajes rectos. El algoritmo LAMDA permite manipular información que tiene alto grado de incertidumbre; es así, que se establecerían clasificadores automáticos con base en grados de adecuación difusos; se determinarán las fallas en el proceso industrial y se realizará un diagnóstico.

Finalmente, se tendrá un alcance explicativo en el cual se deben analizar e interpretar los resultados generados por el algoritmo de aprendizaje usado, para así evaluar el estado de la máquina y determinar si se encuentra en condiciones normal o de fallos, con el uso simultáneo de todos los clasificadores disponibles, por grupos de fallas.



## 2. Problema

### 2.1. Antecedentes del problema de estudio

Debido al avance de la ciencia y la tecnología a una gran velocidad, las empresas e industrias se ven en la necesidad de mantener a sus equipos en estados óptimos de operación para evitar pérdidas en las producciones, razón por la cual muchas empresas buscan mejorar sus procesos para la detección de fallos en las diferentes áreas de la ingeniería que involucran máquinas; por lo cual el mantenimiento predictivo es una alternativa de solución debido a sus nuevas técnicas, ya que nos permite evaluar el estado de las máquinas de manera que se pueda diagnosticar temprano los fallos.

En la industria moderna se requiere que la maquinaria esté en uso constantemente, por lo cual es beneficioso conocer el estado en el que se encuentran los equipos debido a que una parada de la máquina genera un gasto muy significativo a nivel industrial, que afecta principalmente en la productividad de la empresa, y por ende tiene efectos económicos. Actualmente se dispone mayormente de equipos automatizados y mecatrónicos, los cuales pueden ser motores, compresores, bombas, cajas de engranajes, etc., en donde se debería aplicar técnicas de monitoreo usando señales de vibración, emisiones acústicas, corriente y voltaje, entre otras. (Cerrada, Sánchez y col. 2018).

Algunas investigaciones realizadas muestran que las fallas en las máquinas pueden ocurrir por varias razones que incluyen imperfección del diseño, corrosión, erosión, mala lubricación y deformación plástica, razón por la cual la detección y el diagnóstico de fallas es un desafío y generalmente requiere la adquisición y análisis de señales de vibración (Rubén Medina y col. 2017). De modo que ampliar la investigación sobre el monitoreo de la condición para el diagnóstico de fallos basado en un conjunto de datos recolectados es recomendado para un diagnóstico temprano, además del beneficio de contar con máquinas en buen estado y con un consumo adecuado de energía.

Una de las técnicas mayormente utilizadas enfocado en el estudio de monitoreo en las cajas de engranajes es el análisis de las señales de vibración debido al avance de la tecnología que permite disponer de acelerómetros de alta precisión facilitando la tarea de detectar y/o diagnosticar los posibles fallos (Sánchez 2018). En este contexto, la detección de fallos nos permite determinar el estado actual de los equipos antes de producirse una falla que colapsaría con el proceso de la línea de producción, de tal modo de antelarse en cuál es el posible fallo y

dónde se encuentra para realizar el mantenimiento de manera oportuna.

Una vez establecidas las necesidades, el problema que se pretende solucionar es el de disponer de modelos de clasificación de fallos, ajustados mediante algoritmos de aprendizaje automático, que sean efectivos para monitorear la condición de cajas de engranajes, para un diagnóstico temprano de los fallos. Considerando estos antecedentes, surge la siguiente interrogante sobre la cual se basa el planteamiento de este trabajo de titulación: ¿Es posible y viable la detección de fallos basado en datos en las cajas de engranajes rectos a partir de señales de vibración usando un procesamiento de señales basado en mapas de Poincaré y aplicando un algoritmo de clasificación basado en lógica difusa como lo es el algoritmo LAMDA?

## **2.2. Importancia y alcances**

Tomando en consideración que las industrias en la actualidad debido a la globalización y al avance tecnológico se ven la obligación de mejorar los procesos para el mantenimiento de sus equipos, se ha optado por emplear el mantenimiento predictivo, de modo que se antelan a los posibles fallos en sus máquinas evitando que se ocasionen daños más graves. Sin embargo, este mantenimiento aún sigue siendo un campo en donde hay alternativas para explorar con técnicas nuevas de monitoreo.

Actualmente en la industria el mantenimiento predictivo es una técnica que se utiliza para la evaluación regular de las condiciones operativas reales de equipos dentro de una planta, especialmente en los sistemas de producción y demás para optimizar la operación total, es así que uno de los elementos más empleados en la industria son las cajas de engranajes, en el cual se centrará el proyecto técnico. Además la mayoría de las personas definen a este mantenimiento como un medio para prevenir fallas catastróficas de máquinas rotativas críticas. Por tal motivo es muy recomendado establecer un programa de mantenimiento predictivo integral debe incluir técnicas de monitoreo y diagnóstico como: análisis de vibraciones, termografía, tribología, parámetros de proceso, inspección visual y otras técnicas de prueba no destructiva. Los recientes avances en la tecnología de microprocesadores, junto con la experiencia de las empresas que se especializan en la tecnología de análisis y diagnóstico de maquinaria, han evolucionado los medios para proporcionar un mantenimiento predictivo basado en la vibración que se puede utilizar de manera rentable en la mayoría de las aplicaciones de fabricación y procesos (Mobley y col. 2014).

Este proyecto de titulación sirve para proporcionar una herramienta alternativa para la detección de fallos en cajas de engranajes, y esto constituye un aporte del cual podría beneficiarse la industria local que utilice maquinarias rotativas en sus procesos de producción, debido a que si se utiliza correctamente, el análisis de vibración puede identificar componentes específicos de la máquina que están degradados o el modo de falla de la maquinaria de la planta antes de que ocurra un daño grave.

## 2.3. Delimitación

- a.- Delimitación Espacial: El proyecto va a tener lugar en la Universidad Politécnica Salesiana, sede Cuenca, que se encuentra ubicada en la Calle Vieja 12-30 y Elia Liut. Específicamente, proyecto se refiere al caso de estudio de detección y diagnóstico de fallos en cajas de engranajes rectos bajo diferentes modos de fallo, cuyas señales de vibración están disponibles en el laboratorio de vibraciones de esta universidad.



**Figura 2.1:** Ubicación de la Universidad Politécnica Salesiana

- b.- Delimitación Temporal: El periodo seleccionado para la investigación y desarrollo del proyecto es de 8 meses.
- c.- Delimitación Sectorial e Institucional: La Universidad Politécnica Salesiana sede Cuenca es de carácter privada y se localiza en el barrio El Vecino, parroquia El Vecino.

d.- Área de conocimiento: El proyecto se enfoca de manera principal en el área de Ciencia y Tecnología orientada al monitoreo de los procesos industriales, el cual es una línea de investigación del grupo GIDTEC de la UPS-Cuenca.

## **3. Objetivos**

### **3.1. Objetivo general**

- Aplicar un algoritmo basado en lógica difusa para detectar y diagnosticar fallos en cajas de engranajes rectos a partir del procesamiento de las señales de vibración.

### **3.2. Objetivos específicos**

- Extraer características indicadoras de fallos de las señales de vibración de las cajas de engranajes rectos usando diagramas de Poincaré.
- Determinar las mejores características indicadoras de fallos extraídas de diagramas de Poincaré, a través de un análisis de correlación basado en ANOVA e índices de validación de agrupamiento de datos.
- Implementar un algoritmo de detección de fallos basado en el algoritmo LAMDA.
- Determinar la eficacia del algoritmo para realizar el diagnóstico al problema de la máquina.

## 4. Fundamentación teórica

En este capítulo se describen los conceptos básicos necesarios para entender el problema de detección y diagnóstico de fallas basados en datos en una caja de engranajes, y el proceso para obtener los indicadores que permitan evaluar las condiciones de cada modo de fallo y así realizar el posible diagnóstico.

### 4.1. Mantenimiento basado en la condición

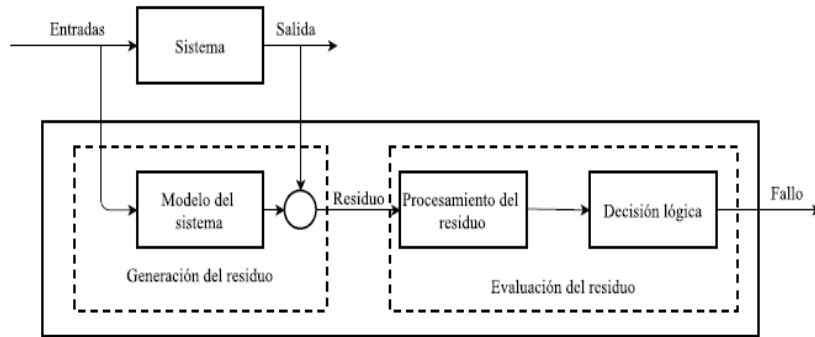
El enfoque para un mantenimiento basado en la condición radica en que tan confiables es su diagnóstico de fallos. Los algoritmos para diagnóstico de fallos están desarrollados con la intención explorar el rendimiento de un sistema, monitorear su estado e identificar fallos y/o averías con fenómenos detectables que permitan observar variaciones en alguna propiedad física del sistema (Vachtsevanos y Vachtsevanos 2006). El diagnóstico de fallos incluye tareas como la detección, el aislamiento y la identificación del fallo para tratar de relacionar los indicadores de condición con los fallos (Bayar y col. 2015; Niu 2017; Vachtsevanos y Vachtsevanos 2006). En el mantenimiento basado en la condición, el diagnóstico de las cajas de engranajes se realiza mediante el análisis de señales de monitoreo de la condición; éste es el método no destructivo más comúnmente utilizado para la detección de fallos en etapas tempranas (Gros 1996; Niu 2017; Stadhouders y col. 2011).

Los enfoques de diagnóstico de fallos se pueden dividir en dos: basados en modelos (model-based) y basados en datos (data-based or data-driven methods) (Kan y col. 2015; Xu y col. 2017; Yin y col. 2014).

- **El diagnóstico basado en modelos**

Consiste en disponer de un modelo matemático del sistema a explorar, generando estimaciones de las salidas del sistema que se comparan con valores reales medidos sobre el proceso, para obtener una señal de residuo y así evaluar sus propiedades e identificar las posibles condiciones de fallo (Isermann 2005; Li y Limmer 2000; Park y col. 2016). El esquema general para implementar este tipo de monitoreo se muestra en la Figura 4.1. Es importante conocer el sistema físico para de este modo describir su comportamiento a través de un modelo matemático, debido a que se requiere incorporar características físicas que detallen el fallo para su diagnóstico. En la industria se torna difícil establecer estas características por las condiciones de trabajo de las cajas de engranajes porque en

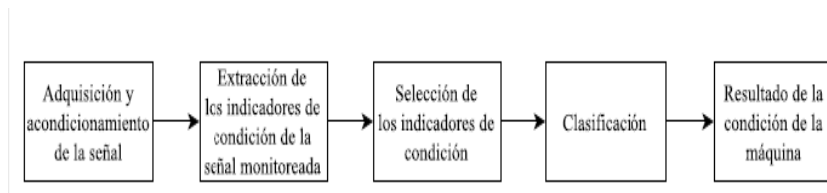
su mayoría no son estacionarias, además de que existen variables como la lubricación, material del engranaje y temperatura, lo cual impide obtener un modelo muy cercano a la realidad (Khalastchi y col. 2015; Xu y col. 2017; Yin y col. 2014).



**Figura 4.1:** Enfoque de diagnóstico de fallos basado en modelos (Sánchez 2018)

- **El diagnóstico de fallos basado en datos**

Se caracteriza por la utilización de métodos que permitan reconocer patrones en las señales monitoreadas, y mapear indicadores (características) extraídos de la señal en condiciones normales y en condiciones de fallo. Por lo tanto estos sistemas de diagnóstico deben ser entrenados con patrones en condiciones de trabajo normales y con fallo en el engranaje, y después los resultados podrán reconocer o inferir si está en condiciones normales o de fallo. Este es un enfoque alternativo y que ha sido eficiente en diagnóstico de fallos en cajas de engranajes (Khalastchi y col. 2015; Vachtsevanos y Vachtsevanos 2006; Yin y col. 2014). Las etapas de este proceso de diagnóstico basado en datos se muestran en la Figura 4.2, las cuáles se detallarán en la sección 4.4.



**Figura 4.2:** Esquema del proceso de diagnóstico de fallos basado en datos para el Mantenimiento Basado en la condición (Niu 2017; Vachtsevanos y Vachtsevanos 2006).

## 4.2. Monitoreo de las señales de vibración

Actualmente la técnica del monitoreo de señales de vibración es el más utilizado para el diagnóstico de fallos en cajas de engranajes porque ha demostrado ser eficiente ya que muestra la fluctuación de carga y fallos en los engranajes (Randall y Monitoring 2011; V. Sharma y Parey 2017). Las razones principales para el monitoreo de las vibraciones en cajas de engranajes son (Jayaswal y col. 2008; Kar y Mohanty 2006; Randall y Monitoring 2011):

- Las cajas de engranajes presentan patrones de la señal de vibración que caracterizan el estado o condición de las mismas. Estos patrones se denominan firma, por lo que una modificación en la misma puede reflejar síntomas de fallo.
- Los mecanismos de vibración de los engranajes son teóricamente entendidos haciendo posible predecir las características de las respuestas de vibración debido a los fallos.
- El desarrollo de sistemas comerciales confiables como analizadores y medidores de vibración portátiles.
- Las aplicaciones de computación han facilitado el desarrollo de métodos avanzados de procesamiento de señales y técnicas de reconocimiento de patrones.

Actualmente, la técnica más utilizada en la industria es la del monitoreo de vibraciones con alrededor del 90 % del número total de fallos en máquinas identificados por el cambio que se presentan en las señales de vibración; es así que con este mantenimiento basado en la condición se puede llevar a cabo para evitar fallos en las máquinas y alargar su vida útil (Devendiran y Manivannan 2016). Sin embargo, en este monitoreo los fallos incipientes no presentan la misma sensibilidad, esto es por algunas razones entre ellas por la presencia simultánea de otros ruidos, vibraciones por otras fuentes, fenómenos dinámicos no estacionarios, variación en carga y velocidad, presentando una dificultad en los sensores debido al ambiente agresivo que perjudica la sensibilidad para detectar estos fallos pequeños (Dalpiaz y Rivola 1997; J. Zhang y col. 2012).

## 4.3. Fallas en sistemas de cajas de engranajes

Se conoce que la industria en sus procesos de producción utiliza cajas de engranajes y estas representan un costo de mantenimiento debido a los fallos que pueden llegar a mostrar. Por lo tanto, a continuación se presentarán los fallos más comunes y los que se tratarán en



este proyecto técnico. Los modos de fallos mas comunes son por picadura (pitting, en inglés), rayadura (scuffing, en inglés), diente roto (break tooth, en inglés) y agrietamiento (crack, en inglés), que serán descritas a continuación.

## 1. Agrietamiento

Básicamente es la formación de grietas en los flancos de los dientes, en la zona cercana a la circunferencia exterior, por ende ocurren desmoronamientos locales progresivos, ver Figura 4.3. Esto puede producirse por variaciones abruptas en la temperatura, una mala lubricación, dureza excesiva del material, o variación de las condiciones del esfuerzo transmitido.



**Figura 4.3:** Grieta en el flanco del diente (Fuente: documentos internos del grupo GIDTEC-UPS)

El origen de las grietas puede ser atribuido a varias causas, en donde la dirección y la posición de ellas nos darán a notar la posible causa (material, forja, tratamiento térmico, etc.). Frecuentemente, estas grietas o fisuras solamente pueden ponerse en evidencia mediante el uso de ensayos no destructivos por líquidos penetrantes y por partículas magnéticas (Aznar 1996).

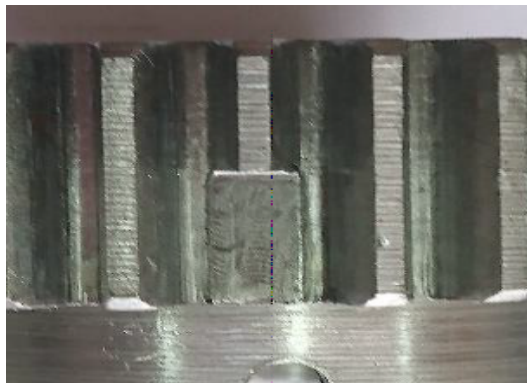
Esta falla se puede presentar de la siguiente manera (Aznar 1996):

- Falla por agrietamiento en el tratamiento térmico.
- Falla por agrietamiento en el esmerilado.
- Falla por agrietamiento en la rueda

- Falla por agrietamiento debido a la fatiga.

## 2. Diente Roto

Esta falla ocurre debido a las sobrecargas o por ciclos de esfuerzo más allá del límite permisible del material, ver Figura 4.4. Es ahí cuando un diente se rompe por fatiga desde un punto focal donde comienza la fractura, que suele ser una raya o una entalla en la raíz, en ocasiones, inclusive el diente se puede romper por alguna incrustación o grieta que ocurrió en un tratamiento térmico.



**Figura 4.4:** Diente roto en un 50% (Fuente: documentos internos del grupo GIDTEC-UPS)

Cuando hay una rotura de diente, ya sea por un choque repentino o una sobrecarga, esta fractura generalmente tiene apariencia fibrosa. En el caso de romperse dientes consecutivos se debe a que uno o dos se rompen por fatiga, pero como el engranaje sigue girando bajo un par torsional específico, el golpe de la rueda que engrana al saltar en el boquete dejado por el diente fatigado, romperá dientes adicionales. Podemos determinar si los dientes fracturados son por fatiga cuando presentan una apariencia lisa, y si son por sobrecarga mostrarán una apariencia fibrosa (Hernández 1985).

Una fractura generalmente inicia desde la raíz, debido a que estos dientes son más débiles en su punto base. Además, puede llegar a ser muy severo cuando se produce un picado en la línea de paso ya que está puede causar una fractura. A continuación se presentan formas que puede presentar un diente roto.

- **Rotura por fatiga.** Se puede notar que una rotura de diente se da por fatiga cuando esta presenta curvas semielípticas con marcas perpendiculares justo en

la superficie que irradian desde el foco o núcleo de la fractura. Esta fatiga se produce debido a cargas periódicas lo suficientemente altas para ocasionar un agrandamiento en una grieta existente, aunque para que se produzca una fractura de diente deberían existir más grietas debido a que una sola difícilmente provocaría una fractura (Hernández 1985).

- **Rotura por sobrecarga.** La rotura por sobrecarga muestra una apariencia fibrosa dejando como evidencia que el material fue arrancado o rasgado en rapidez. Esto sucede por una sobrecarga que excede la resistencia a la tensión del material, comúnmente esta rotura inicia sobre el lado de tensión de la raíz del filete. Se conoce que hay casos donde esta sobrecarga viene acompañada de una desalineación de los engranajes, por lo que en aquellos casos la fractura comienza en un extremo del diente y se prolonga en línea diagonal.
- **Rotura por desalineación.** La desalineación es una causa común de dientes rotos en engranajes rectos, helicoidales y cónicos. Los cojinetes son los responsables de ocasionar esta desalineación en los engranajes puesto que por estar flojos o dañados causan una deflexión del eje, que conlleva a la fractura del diente por las cargas a los extremos (Hernández 1985).

### 3. Picadura

Las picaduras en general son un modo de falla en los dientes de engranajes causadas por una fatiga hertziana que está relacionada con la lubricación. Esto se debe a que los dientes de engranajes están sometidos a altos esfuerzos de contacto hertzianos y a muchos ciclos de tensión. Por ejemplo, los engranajes endurecidos a fondo están diseñados típicamente para soportar tensiones de aproximadamente 700 MPa (100 Ksi), mientras que las tensiones de contacto en engranajes pueden alcanzar 2100 MPa (300 Ksi). Además, un diente en un piñón que gira a 3600 r.p.m acumula más de  $5 \times 10^6$  ciclos de esfuerzo cada 24h (Totten 2017). La Figura 4.5 muestra un ejemplo de picadura.



**Figura 4.5:** Dientes de engranaje con 2 agujeros por picadura, de Diámetro: 1,5 mm y Profundidad: 1,5 mm (Fuente: documentos internos del grupo GIDTEC-UPS)

Básicamente, las picaduras tienen como fenómeno de origen a la fatiga puesto que inicia como una fisura en la superficie del diente del engranaje o bien sea debajo de la superficie con una pequeña profundidad. Por lo general, la grieta se extiende a una corta distancia paralela a la superficie del diente antes de ramificarse en distintas direcciones de la superficie. Con el tiempo las grietas crecen hasta el punto de separar material de la superficie formándose así un hoyo. Dado que si varios hoyos crecen cerca, el pozo más grande se considera como cascajo. Para reducir las picaduras, y así aumentar la vida del conjunto de engranajes, se debe mantener (Campbell 2012):

- **Baja tensión de contacto.** Se consigue con factores geométricos en el diseño como: diámetro, ancho de la cara, el número de dientes, ángulo de presión.
- **Apropiada resistencia del material.** Para tener dientes duros se utiliza aleaciones de materiales y el tratamiento térmico, para superficies duras con dientes cementados, además de la carburación para disminuir tensiones de carga aunque su costo es elevado.
- **Correcto espesor del lubricante.** Se recomienda utilizar lubricante frío, limpio y seco con una viscosidad adecuada.

Hay una creencia de que la contaminación del agua en el lubricante hace fragilización del metal por el hidrógeno generando así picaduras, por lo que las partículas abrasivas presentes en el lubricante, al hacer indentaciones en las superficies del diente, causan tensión y alteran la película del lubricante generando picaduras; es por eso que existen

aditivos de lubricantes pero aun así no está resuelto. A continuación, se presentan algunas recomendaciones para prevenir la aparición de picaduras en conjuntos de engranajes (Committee 2004):

- Reduzca las tensiones de contacto al reducir cargas u optimizar la geometría del engranaje.
- Use acero limpio, tratado térmicamente de manera adecuada para una alta dureza, preferiblemente mediante cementación.
- Use superficies dentadas lisas producidas por un esmerilado o rectificado cuidadoso.
- Use una cantidad adecuada de lubricante fresco, limpio, seco y de viscosidad adecuada.

#### 4. Rayadura

El raspado (o rayadura) es causado por la unión de dos superficies sólidas deslizantes, teniendo como síntomas superficies ásperas en donde existe la transferencia de metal de una superficie a otra debido a la unión. La Figura 4.6 muestra un ejemplo de diente de engranaje con ralladura. Esto se da cuando hay deslizamiento y contacto rodante debido a que la capa de la película del aceite no tiene el espesor adecuado para separar las superficies, incluso pueden quedar marcas de los dientes de los engranajes cuando no hay la suficiente lubricación. Es así que las capas de óxido que protegen los dientes de engranaje pueden romperse y dejar libres las superficies del metal en donde se pueden unir los engranajes. La prueba de raspado consiste en ver la existencia de transferencia de metal de una superficie a otra, porque el deslizamiento que se dio lugar en los dientes de engranaje desgarran las uniones, generando transferencia de metal y dañando el engranaje (Romero Lozano y Giraldo Ramos 2018).

El raspado puede aparecer después de poner en uso al engranaje, a diferencia de otras fallas por fatiga. Se dice también que, cuando son nuevos, son propensos a esta falla porque no han sido suavizadas las superficies. Por tal motivo, se aconseja colocar los engranajes a mitad de carga/marcha al menos por 10 h para reducir la rugosidad de la superficie de los dientes antes de aplicar la carga/marcha completa. Los dientes del engranaje se pueden recubrir con hierro manganeso fosfato o chapado con cobre o plata

para protegerlos de raspaduras (Wulpi 2013).



**Figura 4.6:** Dientes de engranaje con rayadura donde existe una línea de Profundidad 0,5 mm, Ancho 1 mm y Longitud 16 mm. (Fuente: documentos internos del grupo GIDTEC-UPS)

Se considera que el rallado es causado por un alto calentamiento por fricción generado por la combinación de una alta velocidad de deslizamiento y una fuerte presión superficial. La teoría de la temperatura crítica de Block propone que este es el mejor criterio para predecir la aparición de raspados o gripados. La teoría establece que se puede producir raspaduras en los dientes del engranaje, que se deslizan bajo las condiciones de lubricación limitada, cuando la temperatura máxima de contacto de los dientes del engranaje alcanza una magnitud crítica. Por tal razón se debe escoger cuidadosamente los lubricantes que cumplan con las condiciones de funcionamiento de los engranajes, ya que un aumento de la viscosidad también ayuda a que reduzca el riesgo de raspado al aumentar el espesor de la película de lubricante, y reducir la temperatura de contacto generada por el contacto de metal con metal (Totten 2017).

Se deben observar las siguientes pautas para evitar raspaduras en el conjunto de engranajes (Totten 2017):

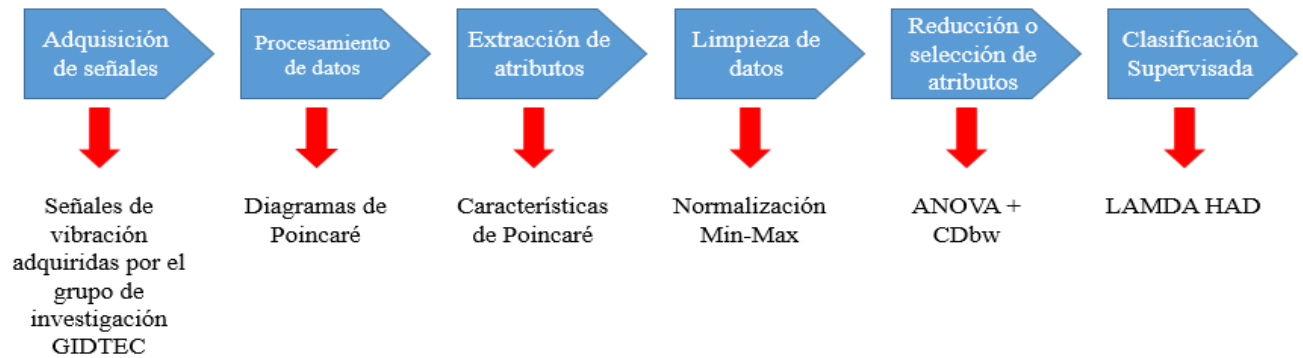
- Use lubricantes de alta viscosidad con aditivos anti suciedad, como azufre, fósforo o borato.
- Enfríe los dientes del engranaje suministrando una cantidad adecuada de lubricante frío.
- Para sistemas de circulación de aceite, use un intercambiador de calor para enfriar

el lubricante.

- Optimizar la geometría de los dientes del engranaje mediante el uso de dientes pequeños, y la modificación del perfil.
- Use dientes de engranaje precisos, montajes de engranajes rígidos y buena alineación helicoidal.
- Use aceros nitrurados para una máxima resistencia al rayado.

#### 4.4. Clasificación de fallos basado en datos

La Figura 4.7 muestra un mejor detalle genérico de la etapas de la aplicación del aprendizaje de máquinas, que fue inicialmente presentado en la Figura 4.1, las cuáles serán explicadas a continuación para el caso de aplicación en la detección y diagnóstico de fallos.



**Figura 4.7:** Esquema del proceso para clasificación de fallos basado en datos

En la **etapa de adquisición** de la señal se puede considerar diversas variable físicas como temperatura, vibración, corriente, sonido, emisión acústica, voltaje o señales asociadas a fenómenos físicos que se pueden adquirir y almacenar mediante diferentes instrumentos (Henriquez y col. 2013). Cuando nos disponemos a la adquisición de estas señales con los sensores debemos tener presente que las señales son cambiantes, por lo tanto se requiere de una planificación en el que se conozca las condiciones normales y de fallo de la máquina, seleccionar adecuadamente el sensor y también su localización, y luego la validación de la señal

adquirida (Chaari y col. 2012; Greñ y col. 2012; Kępski y Barszcz 2012). En este proyecto se usó la base de datos de señales de vibración disponible en el grupo GIDTEC, la cual esta organizada según las condiciones de funcionamiento de una caja de engranajes rectos emplazada en un banco de pruebas. El plan experimental que permitió adquirir estas señales, se detallará en la sección 5.1.

En la **etapa de procesamiento** de la señal, una vez obtenida la señal es necesario representarla en el dominio adecuado. La señal originalmente medida en la etapa de adquisición de señales es una serie temporal, y usualmente es necesario representarla en otro dominio, sea de frecuencia, usando la Transformada Rápida de Fourier, o tiempo-frecuencia, a través de la Transformada Wavelets, la Descomposición de Modos Empíricos, o cualquier otra representación que se quiera estudiar (Feng y col. 2013; Lei, Lin, He y Zuo 2013; Yan y col. 2014). En este trabajo se usará el procesamiento de la serie temporal aplicando el método de Poincaré para generar un conjunto de gráficos de la señal de vibración  $x(t)$  contra la señal desplazada  $x(t + \tau)$ . Este método ha demostrado ser versátil como forma alternativa de representación de la señal monitoreada, y que provee de atributos (indicadores) apropiados (Rubén Medina y col. 2020). Los detalles sobre este método se presentarán en la sección 4.5.1.

En la **etapa de extracción de atributos** se extraen los indicadores de condición pertinentes de la señal en el dominio de representación elegido. En el dominio de tiempo y de frecuencia, usualmente se extraen indicadores estadísticos tales como la media, la desviación estándar, la kurtosis (Sánchez 2018). En las aplicaciones de aprendizaje basado en datos, estos indicadores reciben el nombre de atributos o características (features, en inglés), por lo que en lo sucesivos en este trabajo podrán usarse invariablemente estas denominaciones. En este trabajo no se extraen indicadores estadísticos propiamente dichos, sino que se extraen indicadores que miden la dispersión de los datos en el gráfico de Poincaré. Estos indicadores se describirán en la sección 4.5.2. Usualmente, luego de la etapa de extracción de atributos se realiza una **etapa de limpieza de datos** donde se hace un análisis exploratorio para detectar valores atípicos, los cuales deben ser descartados para las etapas posteriores. Estos valores pueden deberse a situaciones no controladas en el proceso de adquisición de señales. También se requiere hacer un proceso de normalización o estandarización del dominio de atributos, para unificar las magnitudes de los indicadores (Sánchez 2018).

En la **etapa de reducción o selección de atributos** (o indicadores de la condición) es importante eliminar aquellos atributos redundantes o bien que no son relevantes para la



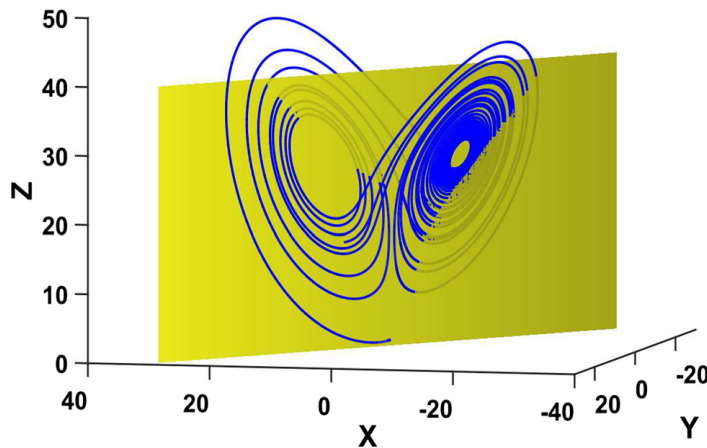
discriminación entre condición normal y de fallo (J. Tang y col. 2014). También es posible hacer una reducción de la dimensión del vector de indicadores con técnicas de mapeo de un espacio de atributos a otro espacio reducido (N. Sharma y Saroha 2015). Sin embargo, esto produce atributos artificiales no interpretables físicamente. La minería de datos puede aportar herramientas para reducir y clasificar la información basadas en datos, y ha demostrado ser muy eficaz en el manejo de información compleja (Baqqar y col. 2011; Perzyk y col. 2014; Raheja y col. 2006; Z. Zhang 2014). En este trabajo se propone realizar esta etapa usando la medida de correlación para la eliminación de los atributos redundantes, y luego se aplicará un algoritmo de selección basado en el análisis de varianza (ANOVA) y la aplicación de un índice de validación de grupos (clústers) que permite seleccionar atributos que producen grupos de datos asociados a una clase (tipo de fallos) que son compactos y apropiadamente separados del otras clases. Esto se detallará en la sección 4.6.

En la **etapa de clasificación**, es donde se reconocen los patrones y se hace un mapeo de los datos con los diferentes estados de fallo, o de severidad de fallo. En esta etapa, la inteligencia artificial y el aprendizaje automático proveen de una variedad de modelos y algoritmos para el desarrollo de sistemas de diagnóstico de fallos de alta precisión y de alta complejidad, tanto como se requiera, según el caso de estudio. Como se ha mencionado, las cajas de engranajes son equipos con dinámica altamente no lineal y es posible que se requieran apropiada dimensión de atributos y modelos de clasificación complejos, como los modelos de aprendizaje profundo (Duan y col. 2018; S. Tang y col. 2019). En esta etapa también es posible utilizar la técnica de fusión de datos cuando se tiene información de múltiples sensores de la misma señal o de diferentes señales puesto que podrías ayudar a mejorar la fiabilidad del sistema de diagnóstico (Banerjee y Das 2012; Hang y Min 2004; Khaleghi y col. 2013; Lei, Lin, He y Kong 2012). La fusión de datos propone usar un vector de atributos que combina indicadores de condición provenientes de diferentes fuentes, como atributos de entrada a un modelo de clasificación de fallos. En este trabajo, se desea probar la efectividad de un modelo de clasificación basado en la ideas de la lógica difusa. La característica de este modelo es que es simple de ajustar/aprender, se basa en una medida de pertenencia a un grupo de datos de una clase, y tiene la posibilidad de asignar una muestra de señal monitoreada a una clase desconocida, si es el caso. Este modelo se explicará en detalle en la sección 4.7.

## 4.5. Diagramas de Poincaré

### 4.5.1. Método de Poincaré

El método de Poincaré consiste en una herramienta metodológica que puede utilizarse para analizar la dinámica no lineal de sistemas que tienen un comportamiento caótico (Alligood y col. 1997). Este tipo de sistemas dinámicos se puede representar mediante un conjunto de ecuaciones diferenciales de primer orden y la solución para este conjunto de ecuaciones describe una órbita o trayectoria en el espacio de fase. Esto se ilustra en la Fig 4.8 donde se muestran las trayectorias de espacio de fase para el atractor de Lorenz 3D interceptando una superficie correspondiente a un plano. La interceptación de tales trayectorias con el plano corresponde al mapa de Poincaré (Cerrada, Macancela y col. 2020).

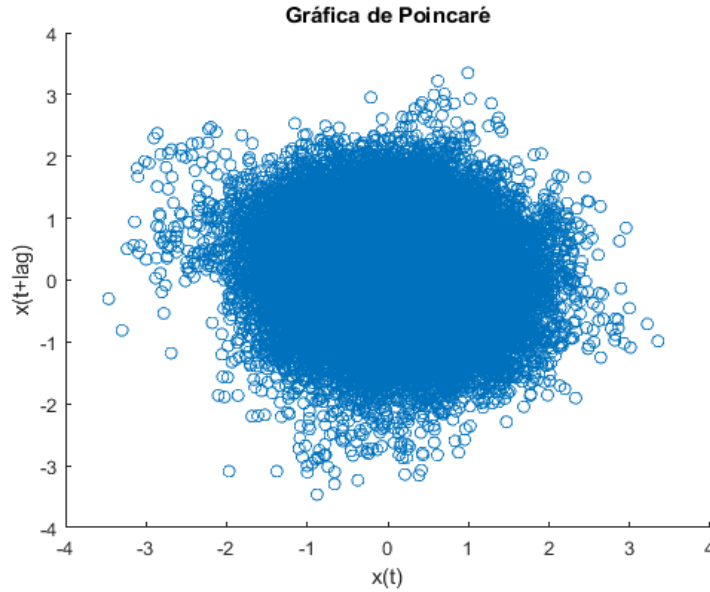


**Figura 4.8:** Mapa de Poincaré para el atractor de Lorenz 3D (Rubén Medina y col. 2020)

En aplicaciones prácticas, se desconocen las ecuaciones diferenciales que modelan el sistema y no se puede obtener el mapa de Poincaré. En consecuencia, es necesario utilizar soluciones aproximadas conocidas como la gráfica de Poincaré (Poincaré Plot, en inglés), a partir del conocimiento parcial del sistema. Este gráfico se obtiene de manera sencilla trazando la serie de tiempo  $x(t)$  de alguna variable del sistema con respecto a su representación desplazada  $x(t + \tau)$  (Alligood y col. 1997; Brennan y col. 2001; S.-T. Wu y col. 1998).

La situación más común en los sistemas mecánicos consiste en la detección de variables como la vibración, que es una serie de tiempo  $x(t)$ , y la gráfica de Poincaré se obtiene seleccionando

correctamente del desplazamiento  $\tau$ , el cual generalmente se selecciona identificando el valor para el cual la autocorrelación de  $x(t + \tau)$  es cero o cerca de cero. Más detalles sobre el método de Poincaré y la gráfica de Poincaré se presentan en (Cerrada, Macancela y col. 2020). A partir de esta representación gráfica, es posible extraer características útiles para la clasificación de fallas. El gráfico de Poincaré genera, entonces una agrupación de datos en el espacio bidimensional  $x(t)$  vs  $x(t + \tau)$ , como se muestra en la Figura 4.9.



**Figura 4.9:** Gráfica de Poincaré generada en la extracción de atributos.

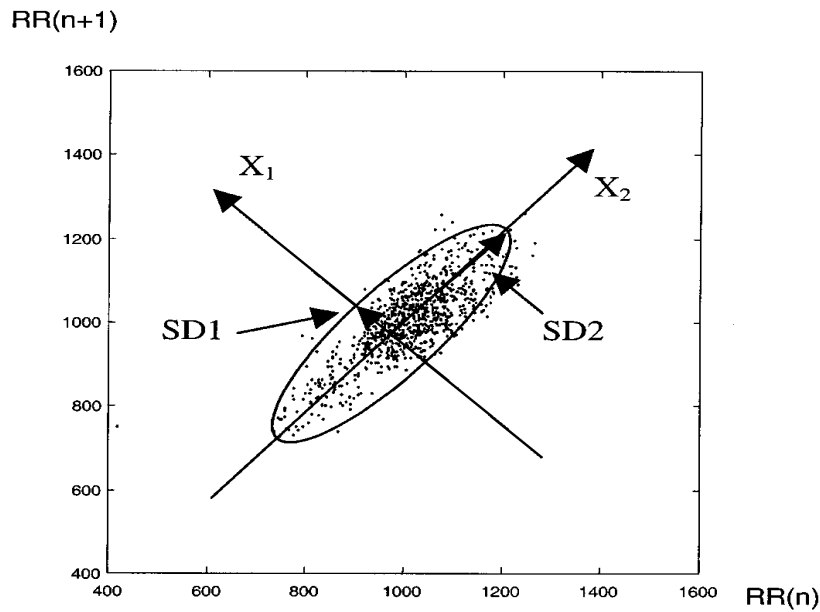
#### 4.5.2. Características de Poincaré

Algunos atributos, también denominados características, se pueden extraer del diagrama de Poincaré, algunas de ellas se han analizado en (Brennan y col. 2001) y (Piskorski y Guzik 2007) para resaltar la asimetría significativa en el mapa de Poincaré y su relación con las aceleraciones y desaceleraciones de la señal de variabilidad de la frecuencia cardíaca (VFC) en los intervalos entre latidos, también conocida como variabilidad RR. En el campo del diagnóstico de fallas de maquinaria rotativa, algunas de esas características han sido extraídas de señales de vibración en (Rubén Medina y col. 2017), y se estudió su capacidad de producir agrupaciones bien separadas para diagnosticar el estado saludable actual de una maquinaria rotativa. En particular, (Rubén Medina y col. 2017) se centra en tres características extraídas del Poincare Plot bidimensional.

A continuación se analizan 11 características de Poincaré extraídas del gráfico de Poincaré de la serie temporal de señales de vibración, denotadas como  $x(t)$ , y que serán obtenidas en este proyecto:

### 1. Dispersión de puntos SD1

$SD1$  mide la dispersión de puntos (desviación estándar) que son perpendiculares al eje de la línea de identidad  $y = x$ , ver Figura 4.10. Entonces,  $SD1$  se calcula mediante la Ec. (1):



**Figura 4.10:** Gráfica que indica SD1 y SD2 (Brennan y col. 2001).

$$SD1^2 = \frac{1}{2} \text{var}\{x(t) - x(t + \tau)\} = \frac{1}{2} SDSD^2 \quad (1)$$

### 2. Dispersión de puntos SD2

$SD2$  es la medida de la dispersión de puntos (desviación estándar) con respecto al eje  $y = -x$ , ver Figura 4.10, luego  $SD2$  se calcula mediante la Ec. (3):

$$SDRR2 = \text{promedio}\{x(t)^2\} - \text{promedio}\{x(t)\}^2 \quad (2)$$

$$SD2^2 = 2 \cdot SDRR2 - SD1^2 \quad (3)$$

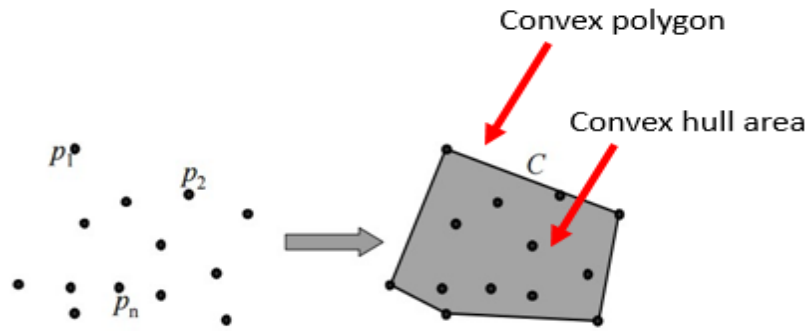
### 3. Relación SD

Esta característica mide la relación cuantitativa entre  $SD1$  y  $SD2$ , como se muestra en la Ec. (4):

$$SD = \frac{SD1}{SD2} \quad (4)$$

### 4. Zona de casco convexo

Dados los puntos de ajuste  $P = P_i$ , el casco convexo (convex hull, en inglés) es el conjunto convexo más pequeño  $C$ , tal que  $P \subset C$ , ver en la Figura. 4.11. Hay varios algoritmos para determinar los vértices convexos del casco. En este trabajo, el algoritmo propuesto en (Barber y col. 1996) se utiliza para calcular el casco convexo del diagrama de Poincaré. El casco convexo  $C$  define un polígono convexo cuya área se puede calcular, y esta área se usa como característica.



**Figura 4.11:** Casco convexo y área del casco convexo (Rubén Medina y col. 2017).

### 5. Eje centrado SDC

Esta característica es calculada por Ec. (5):

$$SDC = \frac{\sqrt{\frac{1}{2} \sum_{i=1}^n (d_i^1)^2}}{\sqrt{\frac{1}{2} \sum_{i=1}^n (d_i^2)^2}} \quad (5)$$

Donde  $d_i^1$  y  $d_i^2$  se calculan mediante las ecuaciones (6) y (7) siendo  $\bar{x}_i$  y  $\bar{y}_i$  valores medios sobre los datos disponibles:

$$d_i^1 = \frac{\|(x_i - \bar{x}_i) - (y_i - \bar{y}_i)\|}{\sqrt{2}} \quad (6)$$

$$d_i^2 = \frac{\|(x_i - \bar{x}_i) + (y_i - \bar{y}_i)\|}{\sqrt{2}} \quad (7)$$

## 6. Dispersión de puntos SD1C

Esta característica es propuesta por (Piskorski y Guzik 2007). También es una medida de la dispersión puntual del diagrama de Poincaré, pero difiere de la métrica propuesta por (Brennan y col. 2001) en la Ec. (1), calculando la varianza de la distancia entre los puntos y la línea  $y = x$  centrada alrededor del centroide del gráfico, de la siguiente manera:

$$SD1C = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i^1)^2} \quad (8)$$

## 7. Dispersión de puntos SD2C

Esta característica es propuesta (Piskorski y Guzik 2007) y es similar a las de la Ec. (8). Es una versión modificada de la métrica propuesta por (Brennan y col. 2001) en la Ec. (3). Esta característica se define de la siguiente manera según la Ec. (9):

$$SD2C = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i^2)^2} \quad (9)$$

## 8. Asimetría superior $SD1_{up}$

Esta característica está relacionada con la distancia  $D_{up}^i$  entre los puntos del diagrama de Poincaré ubicados sobre la línea de identidad  $y = x$  y esta línea de identidad. Esto se calcula mediante la Ec. (10):

$$SD1_{up} = \sqrt{\frac{1}{n_{up}} \sum_{i=1}^{n_{up}} (D_{up}^i)^2} \quad (10)$$

Donde  $n_{up}$  es el número total de puntos sobre la línea  $y = x$ .

### 9. Asimetría inferior SD1down

Esta característica está relacionada con la distancia  $D_{down}^i$  entre los puntos del diagrama de Poincaré ubicados debajo de la línea de identidad  $y = x$  y esta línea de identidad. Esto se calcula mediante la Ec. (11):

$$SD1_{down} = \sqrt{\frac{1}{n_{down}} \sum_{i=1}^{n_{down}} (D_{down}^i)^2} \quad (11)$$

donde  $n_{down}$  es el número total de puntos sobre la línea  $y = x$ .

### 10. Relación de asimetría superior

Esta característica es la medida de la relación entre la asimetría superior con respecto a la distancia media  $SD1_A$  entre todos los puntos y la recta  $y = x$ . Esto se calcula como se muestra en la Ec. (12):

$$C_{up} = \frac{SD1_{up}^2}{SD1_A^2} \quad (12)$$

$$SD1_A = \sqrt{\frac{1}{n} \left( \sum_{i=1}^{n_{up}} (D_{up}^i)^2 + \sum_{i=1}^{n_{down}} (D_{down}^i)^2 \right)} \quad (13)$$

### 11. Relación de asimetría inferior

Esta característica es análoga a la característica de relación de asimetría superior, y se calcula mediante la Ec. (14):

$$C_{down} = \frac{SD1_{down}^2}{SD1_A^2} \quad (14)$$

## 4.6. Selección de características utilizando ANOVA e índices de validación de agrupaciones

El análisis de varianza (ANOVA) se puede usar para medir las características relevantes con respecto a un conjunto de etiquetas de clase. Son muy utilizados para la selección de características en minería de textos, reconocimiento de emociones, bioinformática y medicina. Su uso en aplicaciones de ingeniería también se ha reportado, específicamente en el diagnóstico de fallas. En (Peña y col. 2018) propone un marco metodológico supervisado para la selección de características utilizando ANOVA y una métrica de distancia entre agrupaciones (clúster, en inglés).

En el primer paso, se utiliza ANOVA para clasificar individualmente la relevancia estadística de las características respecto a las clases identificadas; en el segundo paso, se utiliza un índice de evaluación de la validez de las agrupaciones de los datos por clase disponibles, para identificar los subconjuntos de características que producen mejores estructuras de los grupos relacionadas con los modos de falla, es decir, las clases que se están estudiando. Como resultado, se identifica un subconjunto de características, que refleja la información discriminativa con respecto a las diferentes clases. ANOVA es una prueba estadística utilizada para la clasificación supervisada, puesto que es capaz de identificar características que discriminan un conjunto de datos con respecto al valor medio, lo cual es importante para seleccionar características que produzcan estructuras de agrupamiento adecuadas.

### 4.6.1. Análisis de varianza (ANOVA) y prueba de Tukey

ANOVA es una poderosa prueba estadística para rechazar la hipótesis nula  $H_0$  expresada como  $H_0 : \mu_1 = \mu_2 = \dots = \mu_K$ , donde  $\mu_i, i = 1, \dots, K$  son las medias de la población de datos diferentes de  $K$ , bajo el supuesto de independencia, normalidad y homocedasticidad de la población (Ott y Longnecker 2015). Sin embargo, ANOVA puede funcionar bastante bien incluso bajo ligeras violaciones de supuestos.

La hipótesis alternativa  $H_a$  establece que  $H_a : \mu_i \neq \mu_j$  para algunos  $i, j$  t.q.  $i \neq j, j \leq i$ . Cada población, es decir, los valores continuos de la variable de respuesta  $Y$ , se caracteriza por uno o más factores categóricos, luego ANOVA compara las medias de la variable de respuesta en los diferentes niveles de factores. El ANOVA de una dirección solo considera un factor con dos o más niveles (o grupos). Para aceptar  $H_0$  en un ANOVA, la prueba F se usa calculando las estadísticas dadas por la Ec. (15):



$$F = \frac{\sum_{i=1}^K n_i (\bar{Y}_i - \bar{Y})^2}{K-1} \quad (15)$$

$$\frac{\sum_{i=1}^K \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{N-K}$$

donde  $\bar{Y}_i$  denota la media muestral en el  $i$ -ésimo grupo,  $n_i$  es el número de muestras en el  $i$ -ésimo grupo,  $\bar{Y}$  denota la media general de las muestras,  $Y_{ij}$  es la  $j$ -ésima muestra en el  $i$ -ésimo grupo y  $K$  denota el número de grupos.

Se espera que el valor  $F$  sea aproximadamente 1 por debajo de  $H_0$ , sin embargo, para rechazar  $H_0$ , se necesita un valor  $F$  alto. Un solo valor  $F$  alto es difícil de interpretar por sí solo. Entonces, el valor de referencia para rechazar  $H_0$  se basa en el valor  $p$ , que es la probabilidad de observar un valor  $F$  que es al menos tan alto como el valor que obtuvo nuestro estudio, bajo el supuesto de que  $H_0$  es verdadero. Comúnmente,  $H_0$  se rechaza si  $p \leq 0,05$ . Una vez que la prueba ANOVA rechaza la hipótesis nula, se requiere una comparación múltiple usando la prueba de Tukey para identificar cualquier diferencia entre dos medias  $\mu_i - \mu_j$ , calculando la razón en la Ec. (16):

$$qs = \frac{\mu_i - \mu_j}{SE_{\bar{x}}}, \quad \mu_i - \mu_j > 0 \quad (16)$$

donde  $SE_{\bar{x}}$  es el error estándar esperado de la media. El valor de  $qs$  se compara con el valor crítico  $q$  de la distribución del rango estudentizado. Si  $qs$  es mayor que el valor crítico  $q$  obtenido de la distribución, las dos medias son significativamente diferentes. La distribución de  $q$  y sus valores críticos se han tabulado y están disponibles para la comparación por pares.

#### 4.6.2. Evaluación de la validez de la agrupación

En algunas aplicaciones de clasificación supervisada, la estructura de grupos (o conglomerados) de un conjunto de datos podría ser una información interesante para seleccionar las características adecuadas que permitan mejores clasificadores. En este trabajo, se utiliza un índice de validez de agrupamiento global, llamado Densidad Compuesta entre y con agrupaciones (*CDbw*, Composed Density between and within clusters, en inglés), para medir la estructura de agrupaciones (Halkidi y Vazirgiannis 2002; S. Wu y Chow 2004). Este índice

pone énfasis en las características geométricas de los agrupamientos, como la compacidad y la separación. Según (Halkidi y Vazirgiannis 2002), un clúster  $i$  está compuesto por  $n_i$  muestras  $x \in R^n$  y, además del centroide (punto medio del clúster), hay  $r_i$  puntos representativos fijos  $v_{ir_i}$  que se generan en un procedimiento iterativo. En la primera iteración, el punto más alejado del centroide se elige como primer punto representativo. El siguiente punto representativo se elige de modo que esté más alejado del punto elegido previamente, y así sucesivamente.

Dada una partición de datos, en lugar de usar solo información de distancia, la información de densidad intra-grupo e inter-grupo también se considera mediante el índice  $CDbw$ , como se indica en la Ec. (17):

$$CDbw(c) = Intra\_D(c)Sep(c) \quad (17)$$

donde  $c$  es el número de conglomerados en la partición,  $Sep(c)$  mide la separación entre los conglomerados e  $Intra\_D(c)$  mide la densidad dentro del conglomerado.

Por un lado,  $Sep(c)$  en (17) considera tanto las distancias entre conglomerados como la densidad entre conglomerados, como se propone en la Ec. (18):

$$Sep(c) = \sum_{i=1}^c \sum_{\substack{j=1 \\ j \neq i}}^c \frac{\|C\_R(i) - C\_R(j)\|}{1 + Inter\_D(c)}, \quad c > 1 \quad (18)$$

donde  $C\_R(i)$  y  $C\_R(j)$  son el par más cercano de representaciones de dos conglomerados vecinos  $i$  y  $j$ , e  $Inter\_D(c)$  mide la densidad entre conglomerados definida como la densidad en las áreas entre conglomerados.  $Inter\_D(c)$  se define en la Ec. (19):

$$Inter\_D(c) = \sum_{i=1}^c \sum_{\substack{j=1 \\ j \neq i}}^c \frac{\|C\_R(i) - C\_R(j)\|}{\|sd(i) - sd(j)\|} d(\mu_{ij}), \quad c > 1 \quad (19)$$

donde  $\mu_{ij}$  es el punto medio entre el par  $(C\_R(i), C\_R(j))$ ,  $sd(\cdot)$  es el vector de desviación estándar de un grupo, y

$$d(\mu_{ij}) = \sum_{i=1}^{n_i+n_j} f(x_k, \mu_{ij}) \quad (20)$$

es una densidad, donde  $n_i$  y  $n_j$  son el número de muestras que pertenecen a los conglomerados  $i$  y  $j$ , respectivamente, y  $x_k$  es un punto de los conglomerados. La función  $f(x_k, \mu_{ij})$  está definida por la Ec. (21):

$$f(x_k, \mu_{ij}) = \begin{cases} 1 & \text{if } \|x_k, \mu_{ij}\| \leq \frac{\|sd(i)-sd(j)\|}{c} \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

Por otro lado,  $Intra\_D(c)$  en (17) se define como el número de puntos pertenecientes a la vecindad de puntos representativos de los conglomerados, como se propone en la Ecuación (22):

$$Intra\_D(c) = \frac{1}{c} \sum_{i=1}^c \sum_{j=1}^{r_i} Den(v_{ij}), \quad c > 1 \quad (22)$$

donde  $r_i$  es el número de puntos representantes del  $i$ -ésimo grupo, y

$$Den(v_{ij}) = \sum_{l=1}^{n_i} f(x_l, v_{ij}) \quad (23)$$

es una densidad, donde  $x_l$  es una muestra perteneciente al  $i$ -ésimo grupo,  $v_{ij}$  es el  $j$ -ésimo punto representativo del  $i$ -ésimo grupo;  $f(x_l, v_{ij})$  viene dada por la Ec (24):

$$f(x_l, v_{ij}) = \begin{cases} 1 & \text{if } \|x_l, v_{ij}\| \leq sd_a \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

donde  $sd_a$  es el promedio de la desviación estándar de todos los vectores de desviación estándar  $sd$  asociados con cada grupo.

La densidad y separación intra-grupo será significativamente alta para grupos bien sepa-

rados, luego usamos esta métrica para evaluar la estructura de grupo obtenida cuando se selecciona un subconjunto de características significativas. Se pueden encontrar más detalles sobre esta métrica en (S. Wu y Chow 2004).

#### 4.6.3. Algoritmo de selección de características utilizando ANOVA y el índice *CDbw*

En la selección de características, se ha reconocido que la combinación de características individuales buenas no conduce necesariamente a una buena clasificación (Peng y col. 2005). En este sentido, la selección de características por su aporte informativo individual, no se considera como la mejor manera de elegir las características más significativas. Al usar el valor  $p$  en el proceso de selección a través de ANOVA, varias características pueden tener un valor muy cercano, o incluso algunos subconjuntos de las características relevantes discriminan mejor las clases de fallas que otros subconjuntos con el mismo mejor valor  $p$ . Luego, un análisis adicional después de aplicar ANOVA podría orientarse a una nueva clasificación de acuerdo con la estructura de los grupos asociada con las características relevantes.

El análisis de características de (Peña y col. 2018) incluye un análisis adicional de las características relevantes clasificadas por ANOVA, mediante el uso de la métrica de distancia de grupos, aplicada a los conglomerados que se han creado a partir de las características relevantes. Esto es con el objetivo de identificar las características que mejor discriminan ciertos modos de falla. Se pueden seleccionar varios subconjuntos de características como entrada a un clasificador dedicado a modos de fallas particulares. Esto es especialmente útil para diseñar clasificadores múltiples en aplicaciones industriales en tiempo real donde el ruido y las perturbaciones hacen de la clasificación una tarea delicada que no podría resolverse fácilmente con un clasificador centralizado. Por supuesto, se debe desarrollar una búsqueda exhaustiva para encontrar el subconjunto final de características; luego, la selección de características podría formalizarse como un problema de optimización combinatoria para encontrar un conjunto de características para maximizar la calidad de la hipótesis aprendida de estas características. El trabajo en (Peña y col. 2018) no se centra en el algoritmo de búsqueda sino en el análisis de características.

Sean  $F$ ,  $P$  y  $S$  los conjuntos de características, clases de fallas y conjunto de datos, respectivamente, de manera que  $F = f_j$ ,  $j = 1, \dots, n$ ,  $P = P_i$ ,  $i = 1, \dots, c$ ,  $S = (x^k, y^k)$ ,  $k = 1, \dots, m$ ,  $x^k \in R^n$  es un vector de características y  $y^k$  es una variable categórica,  $y_k \in P$ . En la primera etapa, la metodología ejecuta la prueba ANOVA clásica seguida de la prueba

de Tukey para una comparación múltiple por pares, para clasificar las características más significativas. Debido a que varias características pueden tener un valor  $p < 0,05$  para rechazar la hipótesis nula, se establece un umbral de significancia  $Th_s$  para seleccionar una característica  $f_j$  tal que  $p_j < Th_s < 0,05$ . Para cada característica significativa  $f_j$ , se aplica la prueba de Tukey y se establece un umbral de significancia por pares  $Th_{PW_s}$  para seleccionar finalmente una característica significativa  $\hat{f}_j$  tal que  $\hat{p}_j < Th_{PW_s} < 0,05$ . Los umbrales de significancia se eligen como una proporción sobre el valor p máximo, para evitar el conjunto vacío de las características seleccionadas. Como resultado, se obtienen varios conjuntos  $F_{P_i^*} \subset F$  de características significativas para cada par  $P_i^* = (P_r, P_s)$  en estudio. Este primer análisis se repite varias veces sobre conjuntos  $S^* \subset S$  compuestos por  $q$  muestras seleccionadas al azar del conjunto de datos  $S, q < m$ , con el fin de tener una significación estadística. En nuestro caso, aquellas características  $\hat{f}_j \in F_{P_i^*}$  que aparecen todas las veces, para todos los pares, se guardan para su análisis en la siguiente etapa. Entonces, se obtiene el conjunto  $F_P$  de características significativas finales. A los efectos de un análisis más detallado, las características del conjunto  $F_P$  se clasifican en orden ascendente de acuerdo con la Ec. (25):

$$S_{f_j} = \sum_i \hat{P}_j^i \quad (25)$$

Donde  $S_{f_j}$  denota el índice de clasificación de características, y  $\hat{P}_j^i$  denota el valor  $\hat{P}_j$  de la característica  $j$  en el conjunto  $F_{P_i^*}$ .

En la segunda etapa, se define un conjunto de clases  $F_{P_i^*} \subseteq P$  para su posterior análisis. Sea  $\tilde{F}$  un conjunto de todas las n-tuplas posibles de características en  $F_P$ ,  $\tilde{F} = \{\tilde{f}_z\}$ . La estructura de conglomerados para cada par  $\hat{P}_i^* = \{P_r, P_s\}$ ,  $P_r, P_s \in \hat{P}$  con respecto a la n-tupla  $\tilde{f}_z$  se mide mediante el índice  $CDbw_i^z$ . La métrica de clúster final  $CDbwFz$ , con respecto a  $\hat{P}$  y  $\tilde{f}_z$  se calcula con la Ec. (26):

$$CDbwFz = \frac{\prod_i CDbw_i^z}{\sum_i CDbw_i^z} \quad (26)$$

Finalmente, la n-tupla de las características más significativas que  $f_{best}$  crea una estructura de conglomerado adecuada para el conjunto  $\hat{P}$  se selecciona de acuerdo con la condición de la

Ec. (27):

$$f_{best} = \tilde{f}z \quad t.q \quad CDbwFz \quad es \quad máxima \quad (27)$$

Los algoritmos 1 y 2 muestran el pseudocódigo para ejecutar el análisis de características (Peña y col. 2018).

---

**Algoritmo 1** Anova y algoritmos de prueba de Tukey

---

**Entrada:** Conjunto de datos  $S, S = \{(x^k, y^k)\}, k = 1, \dots, m$

Establecer  $F = \{f_j\}, i = 1, \dots, n$

Establecer  $P = \{P_i\}, j = 1, \dots, c$

Valor de la significancia del umbral  $Th_s$

Valor del umbral Importancia por pares  $Th_PW_s$

Número de ejemplos aleatorios  $q, q < m$

Número de repeticiones  $rep$

**Salida:** Subconjunto de características importantes  $F_P$

**para**  $p \leftarrow 1$  a  $rep$  **hacer**

Crear  $F_{P_i}^p = 0$

Elija aleatoriamente el conjunto  $S^* = \{(x^k, y^k)\}, S^* \subset S,$

**para**  $j \leftarrow 1$  a  $n$  **hacer**

Calcule  $p_j$  - valor de la función  $j$  th a través del ANOVA unidireccional aplicado en  $S^*$

**si**  $p_j$  - valor  $\leq Th_s$  **entonces**

Cree el conjunto  $P^* = \{P_i^* = \{P_r, P_s\} | P_r, P_s \in P, P_r \neq P_s\}$

**para**  $i \leftarrow 1$  a  $|P^*|$  **hacer**

Calcule el valor  $\hat{P}_j$  mediante la prueba de Tukey para la comparación por pares de  $P_i^*, F_{P_i^*}^p = F_{P_i^*}^p \cup \{f_j \mid \hat{p}_j\text{-valor} \leq Th_{PW_s}\}$

**fin para**

**fin si**

**fin para**

**fin para**

Obtenga el conjunto  $F_P$  tal que  $f_j \in F_{P_i^*}^p \forall i, p$

---

## 4.7. Clasificador basado en LAMDA

### 4.7.1. LAMDA

Es un algoritmo de aprendizaje para análisis de datos multivariados, basado en el cálculo del Grado de Adecuación Global (GAD) de un individuo a una clase, a través de las aportaciones de todos sus descriptores. Además utilizarse tanto para el aprendizaje supervisado y no

---

**Algoritmo 2** Algoritmo de evaluación de validez de conglomerados

---

**Entrada:** Conjunto de datos  $S, S = \{(x^k, y^k)\}, k = 1, \dots, m$

Establecer  $\hat{P} \in P$

Establecer  $F_P$

$n$  características a analizar,  $n \leq |F_P|$

**Salida:** Subconjunto de  $n$  características significativas  $f_{best}$  para el ser  $\hat{P}$  en estudio

Cree  $\tilde{F} = \{\tilde{f}_z = (n - \text{tupla}) \text{ sobre } F_P\}$

Cree el conjunto  $\hat{P}^* = \hat{P}_i^* = \{P_r, P_s\} | P_r, P_s \in \hat{P}, P_r \neq P_s\}$

**para**  $z \leftarrow 1$  a  $|\tilde{F}|$  **hacer**

**para**  $z \leftarrow 1$  a  $|\tilde{F}|$  **hacer**

    Cree el conjunto  $\tilde{S} = \{(x^k, y^k)\}, x_k$  es un vector de características  $\tilde{f}_z, y_k \in \hat{P}_i^*$

    Calcule  $CDbw_z^{\tilde{z}}$  sobre el conjunto de datos  $\tilde{S} \subset S$ , de acuerdo con la Ecuación (17)

**fin para**

  Calcule  $CDbwF_z$  de acuerdo con la ecuación (26)

**fin para**

  Seleccione  $f_{best}$  de acuerdo con la ecuación (27)

---

supervisado. En el contexto de este trabajo, un individuo es una muestra de señal de vibración, y un descriptor es un atributo o característica (indicador de condición).

LAMDA tiene la capacidad de crear nuevas clases después de la etapa de ajuste (training). Si un individuo no tiene suficiente similitud con las clases preexistentes, se evalúa con respecto a un umbral denominado Clase No Informativa (NIC). Básicamente, aquí es donde radica el problema del algoritmo LAMDA original, puesto que el NIC es constante para todas las clases, o porque el cálculo de GAD no es confiable. Estas contribuciones se calculan mediante funciones de probabilidad difusa, obteniendo como resultado el Grado de Adecuación Marginal (MAD). Al mezclar los MAD para una muestra, los GAD se calculan utilizando operadores de agregación difusa, que se asignan a cada dato como un grado de pertenencia a cada clase. Con esta información es posible determinar el estado actual, i.e. la clase, de un sistema (Morales y col. 2020).

### Algoritmo de aprendizaje para análisis de datos multivariantes

LAMDA es un método difuso, basado en el concepto de grado de adecuación. A diferencia de los algoritmos convencionales, no es un método basado en distancia, que realiza un análisis de similitud entre descriptores (características) de los individuos, con el fin de establecer una relación entre cada uno y su respectiva clase. El grado de pertenencia de un individuo  $X$  a una clase  $C = \{C_1; C_2; \dots; C_k, C_m\}$  se estima en un proceso no iterativo. El individuo  $X$  está

formado por un vector de  $n$  descriptores (Kempowsky-Hamon 2004):

$$X = [x_1; x_2; \dots; x_j; \dots; x_n] \quad (28)$$

donde  $x_j$  es el descriptor  $j$  del individuo  $X$

Los descriptores deben normalizarse en un rango entre  $[0,1]$ . Esta normalización se realiza con respecto a los valores máximo y mínimo de cada descriptor, como se muestra en (29).

$$\bar{x}_j = \frac{(x_j - x_{jmin})}{(x_{jmax} - x_{jmin})} \quad (29)$$

donde  $X_{jmin}$  es valor mínimo de  $x_j$ ,  $X_{jmax}$  es valor máximo  $x_j$  y  $\bar{x}_j$  es el descriptor normalizado.

### **Grado de adecuación marginal (MAD)**

El MAD establece qué tan similar es un descriptor con respecto al mismo descriptor en una clase determinada. Para un individuo definido, se determina el vector MAD para caracterizar su situación con respecto a cada clase. Para el cálculo de MAD se utilizan funciones de densidad de probabilidad, las más comunes son: Función Binomial Difusa y Función Gaussiana (Ruiz y col. 2017).

### **Función binomial difusa**

Esta es una extensión difusa de la función binomial que permite calcular el grado de pertenencia, como una probabilidad bayesiana (Doncescu y col. 2007), y es determinada en la Ec. (30).

$$MAD_{k,j}(\bar{x}_j | \rho_{k,j}) = (\rho_{k,j})^{(\bar{x}_j)} \cdot (1 - \rho_{k,j})^{(1-\bar{x}_j)} \quad (30)$$

donde  $\rho_{k,j}$  es el valor medio del descriptor  $j$  que pertenece a la clase  $k$ , y se calcula para el caso de aprendizaje supervisado mediante la Ec. (31).



$$\rho_{k,j} = \frac{1}{n_{k,j}} \sum_{t=1}^{n_{k,j}} \bar{x}_j(t) \quad (31)$$

donde  $n_{k,j}$  es el número de datos del descriptor  $j$  perteneciente a la clase  $k$ .

### Grado de adecuación global (GAD)

El GAD establece la adecuación de un individuo a cada clase, la cual se calcula mezclando MADs usando funciones de agregación de manera recurrente (cualquiera de las tuplas mostradas en la tabla de la Figura 4.12), con el nivel de exigencia  $\alpha \in [0, 1]$ , que permite una estricta o permisible clasificación (Ruiz y col. 2017). Si  $\alpha$  aumenta, la clasificación es más estricta, por lo tanto, no se reconocerán más individuos (se enviarán a la NIC), lo que hará que el algoritmo sea más selectivo (Kempowsky y col. 2003). El GAD de un individuo  $\bar{X}$  en la clase  $k$  se calcula mediante la Ec. (32)

$$GAD_{k,\bar{X}}(MAD_{k,1}, \dots, MAD_{k,n}) = \alpha T(MAD_{k,1}, \dots, MAD_{k,n}) + (1 - \alpha) S(MAD_{k,1}, \dots, MAD_{k,n}) \quad (32)$$

donde  $\bar{X}$  es el individuo  $X$  normalizado,  $T$  es una función t-norma (intersección), y  $S$  es una función t-conorm (unión).

Type	T-norms	T-conorms
Min-Max	$T(a, b) = \min(a, b)$	$S(a, b) = \max(a, b)$
Product-Sum	$T(a, b) = ab$	$S(a, b) = a + b - ab$
Hammacher	$T(a, b) = \frac{ab}{p + (1-p)(a+b-ab)}$	$S(a, b) = \frac{a+b-ab-(1-p)ab}{1-(1-p)ab}$

**Figura 4.12:** T-norma and T-conorma (Mizumoto 1989).

El individuo  $\bar{X}$  se asigna a la clase con el GAD máximo, de acuerdo con la Ec. (33), donde el índice corresponde al número de la clase. El Algoritmo 3 establece los pasos para la clasificación supervisada.

$$\text{índice} = \max(GAD_{1,\bar{x}}, GAD_{k,\bar{x}}, \dots, GAD_{m,\bar{x}}, GAD_{NIC,\bar{x}}) \quad (33)$$

---

**Algoritmo 3** LAMDA para clasificación supervisada
 

---

**Entrada:** Individual  $X$

**Salida:** índice de clase

- 1: Normalice los descriptores del objeto usando la Ec. (29).
  - 2: Calcule la DMA para cada descriptor y cada clase con la Ec. (30).
  - 3: Calcule la MAD del NIC, considerando  $\rho_{NIC} = 0.5$ , para Función Binomial y  $\rho_{NIC} = 0.5$ ;  $\sigma_{NIC} = 0,25$  para la función gaussiana.
  - 4: Calcule el GAD de cada clase usando la t-norma y t-conorm (4.12) y la Ec. (32).
  - 5: Para identificar la clase a la que pertenece  $\bar{X}$ , elija el GAD máximo con la ecuación. (33).
- 

#### 4.7.2. LAMDA HAD

En (Morales y col. 2020) se mejora la eficiencia de LAMDA original mediante dos estrategias. La primera, por el cálculo de NICs adaptables para cada clase, lo que impide que los individuos correctamente clasificados creen nuevas clases; y el segundo, mediante el cálculo del Grado Superior de Adecuación (HAD), que otorga mayor robustez al algoritmo. Se propone atacar los dos problemas anteriores utilizando dos estrategias, que extienden los alcances originales de LAMDA (Morales y col. 2020):

- Primero, se propone calcular tantos NIC como el número de clases existentes. En este caso, cada NIC (y su GAD) se calcula en base a las características intrínsecas de cada clase, lo que impide enviar individuos bien clasificados a la NIC, como ocurre en el algoritmo LAMDA original, donde solo se calcula una NIC para todas las clases existentes. A esto se le denomina  $GAD_{NIC}$  adaptable.
- La segunda estrategia propuesta se basa en calcular el Grado Superior de Adecuación (HAD), que consiste en medir el grado de similitud del GAD de un individuo en relación al promedio de los GADs de las clases existentes a través de operadores probabilísticos, permitiendo conocer con mayor certeza la clase a la que pertenece un individuo. Permite mejorar la robustez del algoritmo y evitar que se clasifique erróneamente a individuos con características similares entre clases.

A continuación se presentan las definiciones que soportan las extensiones propuestas al algoritmo LAMDA.

**Definición 1.**  $MGAD_{k,p}$  es el valor medio de los GAD de la clase  $p$  en una clase  $k$ . Considerando que  $n_k$  es el número de individuos pertenecientes a la clase  $k$  y  $p = \{1, \dots, m\}$ , podemos calcular este parámetro mediante la Ec. (34).

$$MGAD_{k,p} = \frac{1}{n_k} \sum_{t=1}^{n_k} GAD_{p,t} \quad (34)$$

donde  $GAD_{p,t}$  es el GAD del individuo  $t$  para la clase  $p$ , en la clase  $k$ .

**Definición 2.** Sea  $GAD_{NIC_k}$  el GAD del NIC para la clase  $k$ , y se calcula como el valor promedio de todos los  $MGAD_{k,p}$  en cada clase. El cálculo de  $GAD_{NIC_k}$  se realiza mediante la Ec. (35).

$$GAD_{NIC_k} = \frac{1}{m} \sum_{p=1}^m MGAD_{k,p} \quad (35)$$

Como se muestra en la Ec. (35),  $GAD_{NIC_k}$  depende de las contribuciones de todos los MGAD, por lo tanto, también depende de la distribución de todos los GAD en cada clase.

**Definición 3.**  $AD_{GAD_{k,p},\bar{X}}$  es un parámetro que permite obtener una medida de similitud entre el GAD de un individuo  $\bar{X}$  y cada  $MGAD_{k,p}$ . Se calcula mediante el uso de la función de densidad de probabilidad como se muestra en la Ec. (36).

$$AD_{GAD_{k,p},\bar{X}} = (MGAD_{k,p})^{GAD_{p,\bar{X}}} \cdot (1 - MGAD_{k,p})^{1-GAD_{p,\bar{X}}} \quad (36)$$

para  $\bar{X}$  evaluado en cada clase.

Esta expresión, al ser una extensión difusa de la función binomial, presenta los mejores resultados cuando el objetivo es medir la similitud entre el individuo y las clases.

**Definición 4.** Sea  $AD_{GAD_{k,p},\bar{X}}$  el nuevo grado de adecuación del GAD, el cálculo del Grado de Adecuación Superior (HAD) de un individuo a una clase se realiza sumando todas las  $AD_{GAD_{k,p},\bar{X}}$  de la clase  $k$  según la Ec. (37).

$$HAD_{k,\bar{X}} = \sum_{p=1}^m AD_{GAD_{k,p,\bar{X}}} \quad (37)$$

Con base en este cálculo, se determina que el valor más alto del  $HAD_{k,\bar{X}}$ , y la clase estimada para  $\bar{X}$  será el que tenga mayor grado de pertenencia. Así, el índice estimado  $EI$  de la Ec. (38) es la clase a la que el individuo tiene mayor probabilidad de pertenecer.

$$EI = \max(HAD_{1,\bar{x}}, \dots, HAD_{k,\bar{x}}, \dots, HAD_{m,\bar{x}}) \quad (38)$$

**Definición 5.** Sea  $EI$ , el índice de la clase con mayor grado de adecuación, luego es necesario verificar si el valor máximo de  $GAD_{EI,\bar{X}}$  es mayor que el  $GAD_{NIC_{EI}}$ . Si se cumple esto, entonces el individuo  $\bar{X}$  pertenece a la clase  $EI$ ; de lo contrario, se enviará a la clase  $NIC$ , de acuerdo con lo expresado en la Ec. (39).

$$index = \max(GAD_{EI,\bar{X}}, GAD_{NIC_{EI}}) \quad (39)$$

El Algoritmo 4 presenta los pasos para realizar la clasificación supervisada, según LAMDA HAD.

---

**Algoritmo 4** LAMDA-HAD para clasificación supervisada

---

**Entrada:** Individual  $X$

**Salida:** índice de clase

- 1: Normalice los descriptores del individuo usando la Ec. (29).
  - 2: Calcule la DMA para cada descriptor y cada clase con base en la función de densidad de probabilidad seleccionada descrita por la Ec. (30)
  - 3: Calcule el GAD de cada clase usando la t-norma y t-conorm con cualquiera de las ecuaciones que se muestran en la Tabla 1, y la Ec. (32).
  - 4: Calcule el MGAD en cada clase mediante la ecuación. (34) de la Definición 1.
  - 5: Utilice la ecuación. (35) en la Definición 2, para calcular de  $GAD_{NIC}$  para cada clase  $k$ .
  - 6: Calcule el grado de adecuación  $AD_{GAD}$  con la ecuación. (36) en la Definición 3, y el HAD como la suma de ellos con la Ec. (37) en la Definición 4.
  - 7: Encuentre el índice de estimación  $EI$  de la clase a la que podría pertenecer el individuo con la ecuación. (38).
  - 8: En la clase estimada, determine si el GAD máximo es mayor que el  $GAD_{NIC}$  correspondiente con la Definición 5.
-

### 4.7.3. Entrenamiento y funcionamiento del clasificador LAMDA HAD

En las aplicaciones de aprendizaje automático, los modelos se ajustan en una fase de entrenamiento utilizando una porción de las señales disponibles, las cuales en los casos ideales deben estar balanceadas respecto a las clases que se desean clasificar. Usualmente se seleccionan aleatoriamente entre el 70 % y 80 % de los datos disponibles para el entrenamiento, y el restante se deja para la fase de prueba. En este trabajo se usará el Algoritmo 4 en la fase de entrenamiento (ajuste del modelo) aplicando el algoritmo bajo diferentes valores de los parámetros y funciones ajustables, esto es  $\alpha$ ,  $S$  y  $T$ . El mejor modelo se seleccionará con base en la precisión en la clasificación de fallos. Una vez seleccionado el mejor modelo, se fijan los valores como: Alfa ( $\alpha$ ),  $P_{kj}$ ,  $T$  (Intersección),  $S$  (Unión), GADNIC, MGAD, Número de atributos. Con estos valores se crea una instancia del modelo de clasificación y se prueba el desempeño con los datos de prueba, calculando de nuevo la precisión en la clasificación. La precisión es uno de los indicadores que más se usan para medir el desempeño de clasificadores, y se puede calcular a partir de la matriz de confusión (Tharwat 2021). Una representación de este procedimiento se ilustra en la figura 4.13



**Figura 4.13:** Procedimiento para el entrenamiento y funcionamiento del clasificador LAMDA-HAD

## 5. Marco Metodológico

En este capítulo se detalla el procedimiento para la detección y diagnóstico de fallos de caja de engranajes rectos utilizando el algoritmo LAMDA aplicado en señales de vibración. Todo el estudio de este proyecto se encuentra dividido en varias secciones, en donde se explica cada paso de la metodología descrita en la sección 4, y específicamente en la Figura 4.7 y diagrama de flujo de la Figura 5.1, este último detallando el marco metodológico que se siguió para aplicar la minería de datos y el aprendizaje automático requerido en este proyecto.

### 5.1. Adquisición de señales

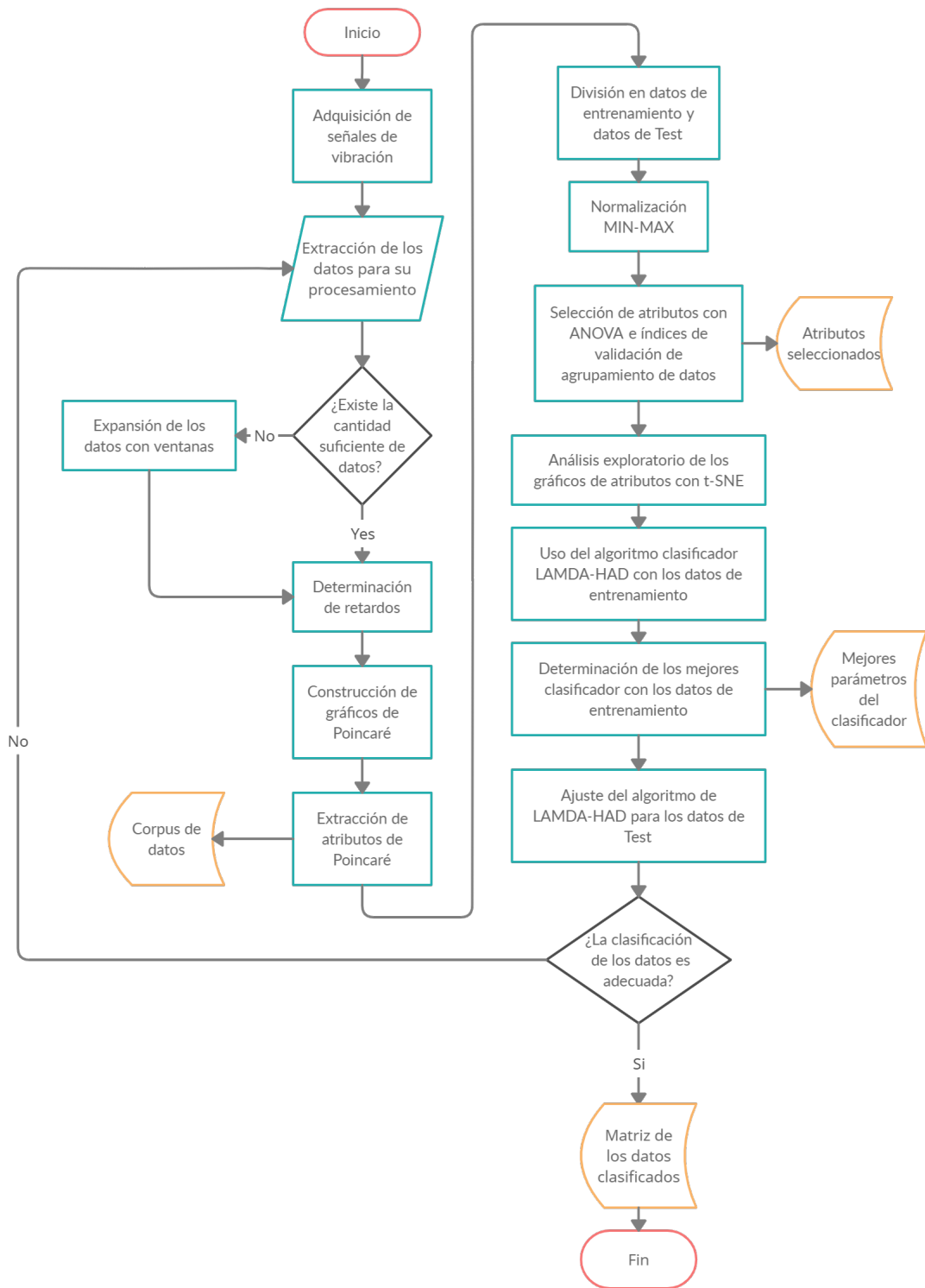
#### 5.1.1. Plan experimental

El estudio se aplica en las señales de vibración medidas en una caja de engranajes rectos que está localizada en el Laboratorio de Vibraciones de la Universidad Politécnica Salesiana, Cuenca, Azuay, Ecuador. Dependiendo del tipo de fallo y la severidad, se presentan diferentes patrones en dichas señales que permiten identificar y así diferenciar el estado en el que se encuentra la máquina. La caja de engranajes esta acoplada en un banco de pruebas como el mostrado en a Figura 5.2

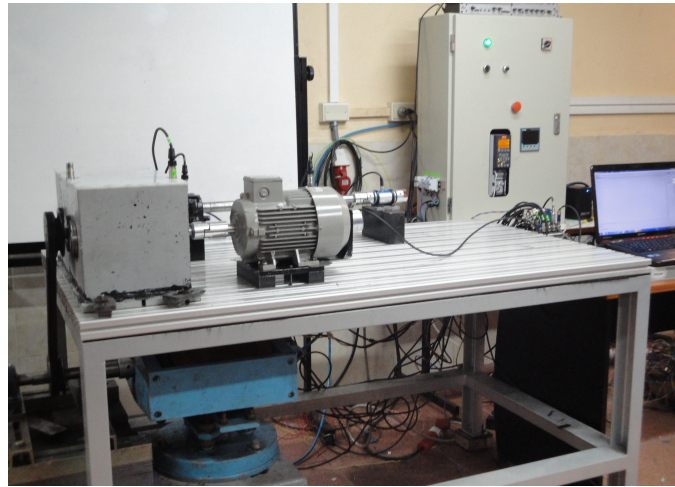
El montaje experimental esquemático se observa en la Figura 5.3. El plan experimental aplicado para obtener las señales de vibración se denomina "Severidad del engranaje de fallas múltiples" y esta codificado como *EP\_007V0*. Se pueden observar en el esquemático varios sensores que captan las diferentes señales (corriente, Voltaje, Aceleración, Velocidad) que han sido configuradas para representar la mayor cantidad de información extraíble; nuestro estudio sólo usa las señales de vibración del acelerómetro A1 emplazado de forma vertical sobre el engranaje de entrada. El plan experimental considera diferentes modos de fallos con diferentes niveles de severidad, aportando con una base de datos completa para fallos comunes en cajas reductoras.

#### 5.1.2. Modos de fallos y niveles de severidad

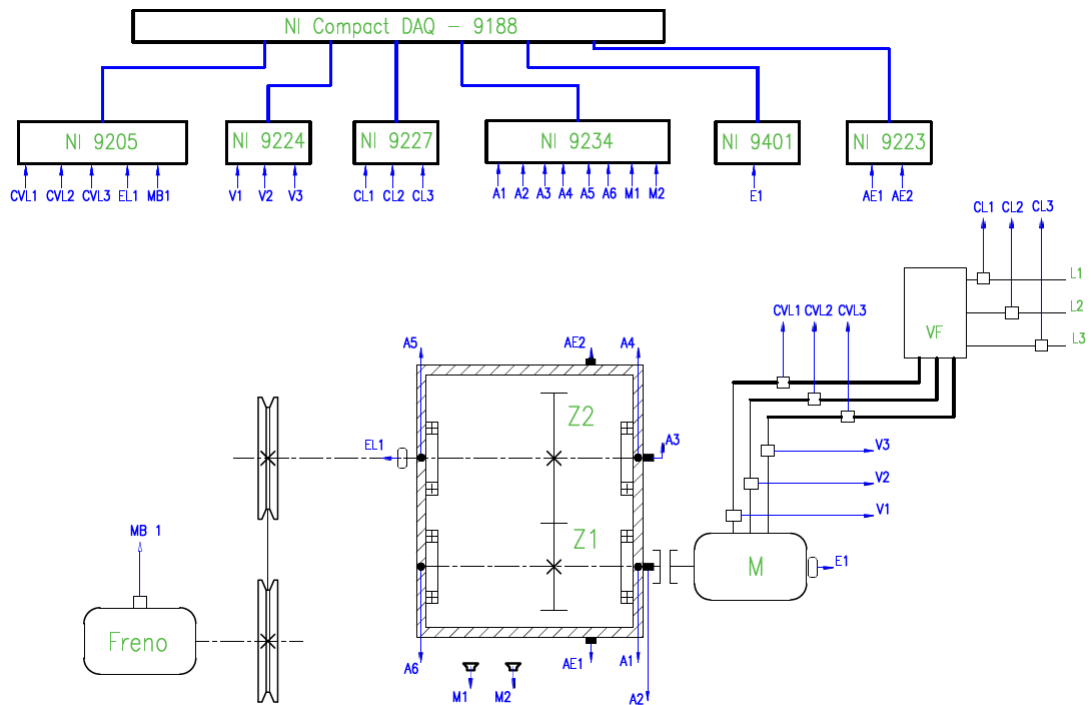
En la base de datos *EP\_007V0* se establecieron fallos por diente roto, rayadura, picadura y agrietamiento, cada uno con 9 niveles de severidad, en el cual el primero es con el comportamiento normal y otros ocho niveles con daños parciales hasta totales del dientes. La adquisición de datos se realizó con las siguientes consideraciones:



**Figura 5.1:** Diagrama de flujo del proceso



**Figura 5.2:** Banco de pruebas del laboratorio de vibraciones de la UPS



**Figura 5.3:** Montaje experimental esquemático (Fuente: documentos internos del grupo GIDTEC-UPS)

- Duración de cada señal: 10 s a 50 KHz de frecuencia de muestreo.
- Repetición de las pruebas: R1, R2, R3, . . . , R15, con el fin de preservar validez estadística en el plan experimental.



- Frecuencia de rotación del motor (velocidad constante): 6Hz (F1), 12Hz (F2), 16Hz (F3).
- Frecuencia de rotación del motor (velocidad variable 8-13Hz): onda sinusoidal 0.5Hz (F4), onda triangular 0.5Hz (F5), onda cuadrada 0.5Hz (F6).
- Carga: sin carga (L1), 5 V (L2), 10 V (L3), que representan los voltajes de alimentación del freno magnético que induce un par de carga constante.
- Sistema de adquisición de señales: NI cDAQ 9234 y un programa de adquisición de señales con LabView.

En este trabajo solo se ha utilizado velocidad constante que nos permiten generar mejores resultados de clasificación, se usaron todas las repeticiones y los niveles de carga (L1, L2, L3). En la tabla 1 se presentan las características del equipo mecánico utilizado para el experimento. En la tabla 2 se muestra un ejemplo de los niveles de severidad para el modo de fallo del diente roto y las clases representadas, así mismo en la Figura 5.4 se muestra al diente roto en sus ocho niveles de severidad y el porcentaje que fue retirado del diente de engranaje.

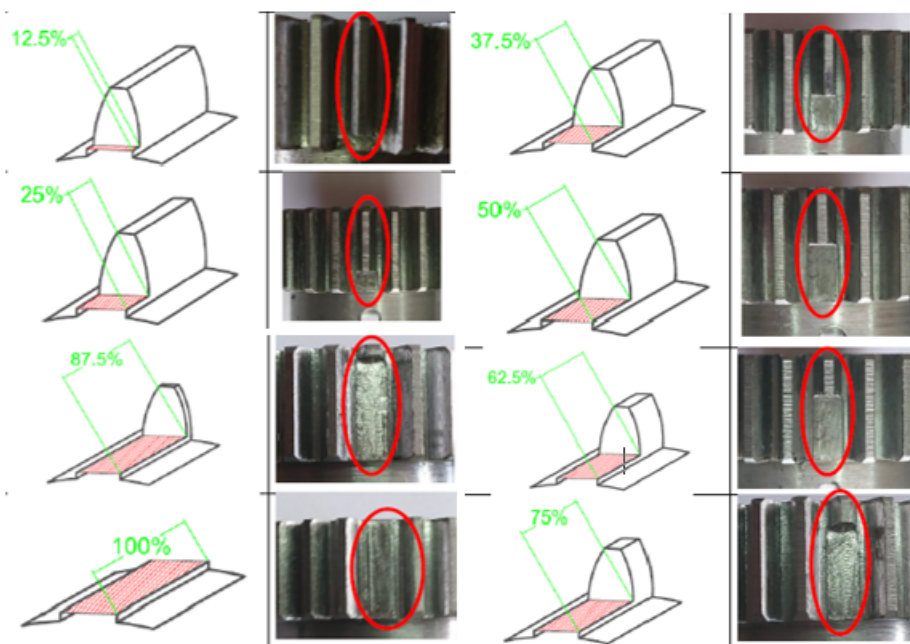
<b>Componente</b>	<b>Especificación</b>
Caja de cambios	Una fase con engranajes rectos
Motor	Siemens 1LA7 096-6YA60 2Hp
Acoplamiento	Lovejoy
Roldana	Optibelt SPA 100-2 1610
Tipo de carga	Electromagnético a través de una transmisión por correa.

**Tabla 1:** Especificación del equipo mecánico utilizado en el experimento.

Cabe mencionar que para desarrollar los clasificadores para cada modo de fallo, los niveles de severidad se distribuirán por grupos, estos son: Grupo 1 (P1,P2,P3), Grupo 2 (P1,P4,P5,P6), Grupo 3 (P1,P7,P8,P9), con severidad baja, media y alta respectivamente. También se consideró la separación por velocidades, de tal manera que se tendrán 3 clasificadores, uno por cada velocidad, para cada grupo de severidades, lo que da un total de 9 clasificadores para cada modo de falla (36 grupos de clasificadores). La restricción de separar por velocidades es realista ya que, en general, la velocidad es un parámetro que se conoce como parte de la condición de operación de la máquina, en tanto que la carga puede deberse a factores de perturbación sobre el proceso. El nivel P1 (diente sin fallo, o clase Normal) se considera en todos los grupos con el fin de potenciar la capacidad de detección con el uso simultáneo de

Modo de fallo diente roto		
Clases	Severidad	Porcentaje [%] del diente roto
P1	Normal	0
P2	1	12,5
P3	2	25
P4	3	37.5
P5	4	50
P6	5	62.5
P7	6	75
P8	7	87.5
P9	8	100

**Tabla 2:** Niveles de Severidad para el modo de fallo de diente roto



**Figura 5.4:** Imágenes del diente roto con los niveles de severidad

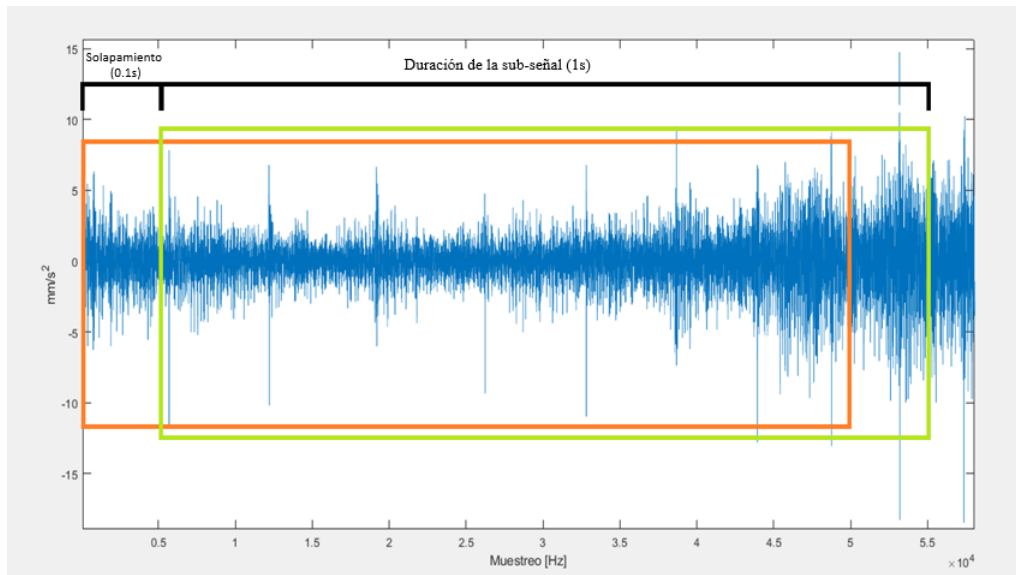
los 3 clasificadores que se van a desarrollar, por velocidad y por modo de fallo.

## 5.2. Preprocesamiento de las señales

A continuación se mencionarán con detalles las adecuaciones a las señales de vibración, así como la determinación del retardo con base a la autocorrelación y los gráficos de Poincaré.

### 5.2.1. Expansión con las ventanas

El ajuste de los clasificadores en este trabajo se hizo para cada velocidad constante y todas las cargas, entonces se procedió a separar los datos en diferentes conjuntos de acuerdo a la velocidad. Dado que al momento de utilizar las señales de vibración y separarlas por velocidades se obtenían pocas muestras, para ser exactos alrededor de 45 señales, esto no suele considerarse suficiente para alcanzar resultados estadísticamente significativos que permitan un buen entrenamiento. Es así que para realizar un aumento de datos (data augmentation, en inglés) se aplicó una ventana deslizante sobre la señal de vibración; esto consiste tomar de la señal original, de duración 10 s, señales de corta duración y con un cierto solapamiento entre ellas. En nuestro caso, se determinó un intervalo de solapamiento de 0,1 s (equivalente a 5000 muestras) y de duración 1 s (equivalente a 50000 muestras) como el tamaño de la nueva muestra como se indica en la Figura 5.5. Aplicando este procedimiento se generaron 90 sub-señales por cada señal de vibración original y 4050 como el total de señales de una clase para cada velocidad.



**Figura 5.5:** Señal de vibración y ventanas deslizantes que indica el solapamiento y duración de un sub-señal de vibración

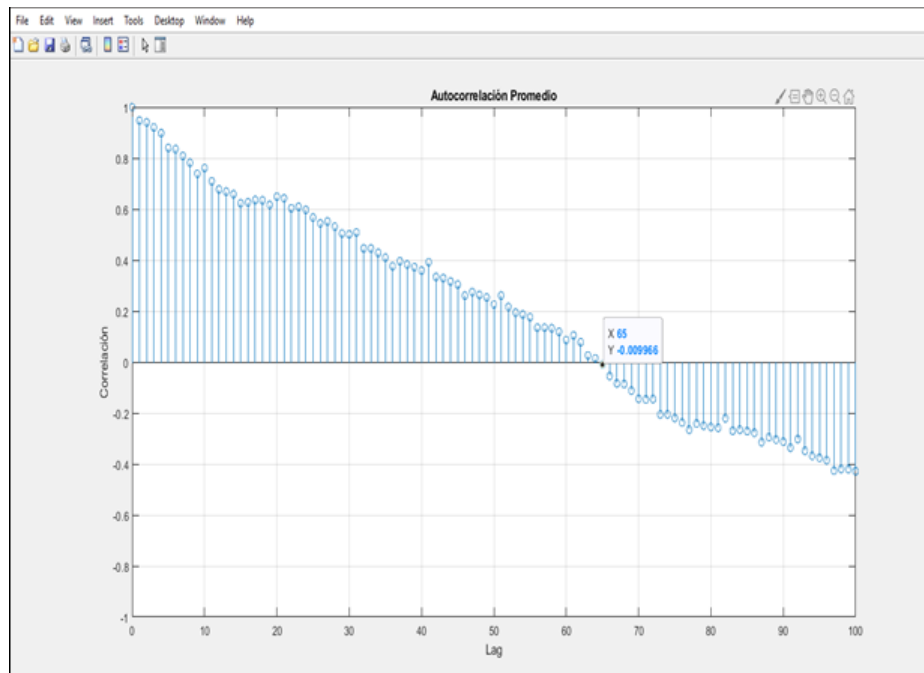
### 5.2.2. Determinación de los retardos

Para determinar el retardo óptimo ( $\tau$ ) que se requiere para la generación de los gráficos de Poincaré, hay que tener como pautas la información mutua y la secuencia de autocorrelación promedio. Comúnmente en los sistemas mecánicos la detección de variable a utilizar suele ser la vibración, que es una serie de tiempo  $x(t)$ , y la gráfica de Poincaré se obtiene seleccionando correctamente el retraso  $\tau$ , que generalmente se selecciona identificando el valor para el cual la autocorrelación de  $x(t + \tau)$  es cero o cerca de cero (Cerrada, Macancela y col. 2020).

Para la reconstrucción de la dinámica espacio-fase se logra mediante el uso de la incrustación de retardo de tiempo. Esto implica la construcción del vector de estado  $X(t) = [x(t), x(t + \tau), \dots, x(t + D\tau)]$ , donde  $x(t)$  es un señal de serie temporal. La dimensión del estado de fase se denota como  $D$  (también conocida como dimensión fractal) y el retardo de tiempo (o *lag*, en inglés) se denota como  $\tau$ . La dinámica del espacio-fase se reconstruye trazando cada uno de los puntos  $X(t)$  en el espacio dimensional  $D$  ( $D = 2$  en nuestra aplicación). Cuando la señal medida se considera como una cantidad infinita de datos sin ruido, es factible una selección arbitraria del retardo  $\tau$  (Takens 1981), sin embargo, cuando los datos disponibles son finitos y ruidosos, el retardo podría seleccionarse de manera que las muestras en  $X(t)$  son independientes o no están correlacionados. Esto se logra seleccionando el retardo de dos maneras, de la serie de tiempo disponible: el valor para el cual la gráfica Autocorrelación vs Retraso tiene cruce por cero (normalmente el primer cruce por cero), o el valor para el cual la gráfica Información mutua vs Retraso tiene un cruce por cero mínimo local (Fraser y Swinney 1986; Rubén Medina y col. 2020).

En la Figura 5.6 se muestra el ejemplo de una gráfica Autocorrelación vs Retraso, construida con las sub-señales promedio, que tiene cruce por cero; en este caso el cruce se da en  $\tau=65$  por lo tanto ese es el *lag* que se ocupa. Para los diferentes modos de fallo, el cruce por cero en las diferentes clases no siempre era el mismo variaba sobre todo para las velocidades 2 y 3, es así que en busca de obtener los mejores resultados se utilizó un *lag* por grupo con el objetivo de que las características de Poincaré sean las mejores.

Existen 2 formas para determinar la autocorrelación promedio: (i) la primera consiste en extraer el cruce por cero de cada señal y al final promediar esos valores, (ii) la otra opción es promediar todas las señales de autocorrelación y obtener un único cruce por cero de la señal promedio; haciendo comparaciones esto variaba muy poco, además esta segunda opción se ejecutaba más rápido, lo cual ahorra tiempo de máquina.



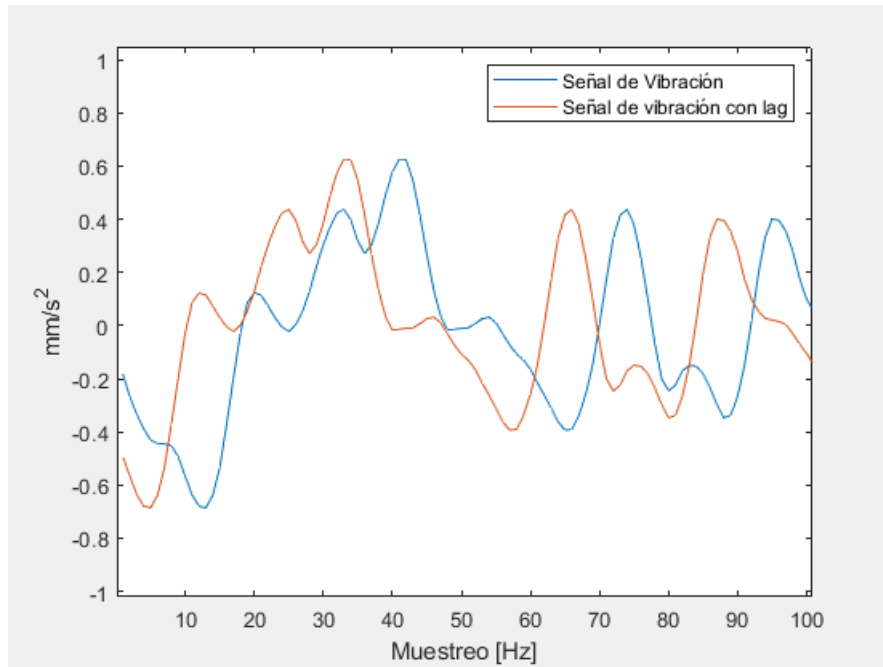
**Figura 5.6:** Autocorrelación promedio con el Acelerómetro 1 para el modo de fallo picadura, con la velocidad 1

Para determinar el *lag* por grupos se podía hacer de 2 alternativas: (i) o bien el *lag* de cada clase se promedia con las demás clases que pertenecen al grupo, además que se facilita el uso de la consideración por inspección visual al disponer de las gráficas, lo cual consiste en observar en donde es el cruce por cero y donde las clases coinciden mayormente por ejemplo puede ser por el primer, segundo, o tercer cruce por cero (cabe mencionar que aquello queda a criterio propio); o (ii) también promediar todo el grupo y obtener una señal promedio, lo cual se hizo para algún caso y es igual de bueno que la primera alternativa pero conlleva mucho tiempo de ejecución por lo cual tarda más. En este trabajo se optó por la primera opción.

La Figura 5.7 muestra una señal de vibración y la señal con el correspondiente retraso. Estas dos señales se representan en un gráfico 2D, generando así el diagrama de Poincaré.

### 5.2.3. Obtención de los gráficos de Poincaré

Los gráficos de Poincaré, como se explicó en la sección 4.5.1, se contruyen una vez determinado el retardo. La forma del gráfico de Poincaré a partir de las señales de vibración del conjunto de datos de la caja de cambios muestra que su forma varía según el nivel de severidad del fallo. Lo novedoso de este enfoque es que luego nos permite extraer las características



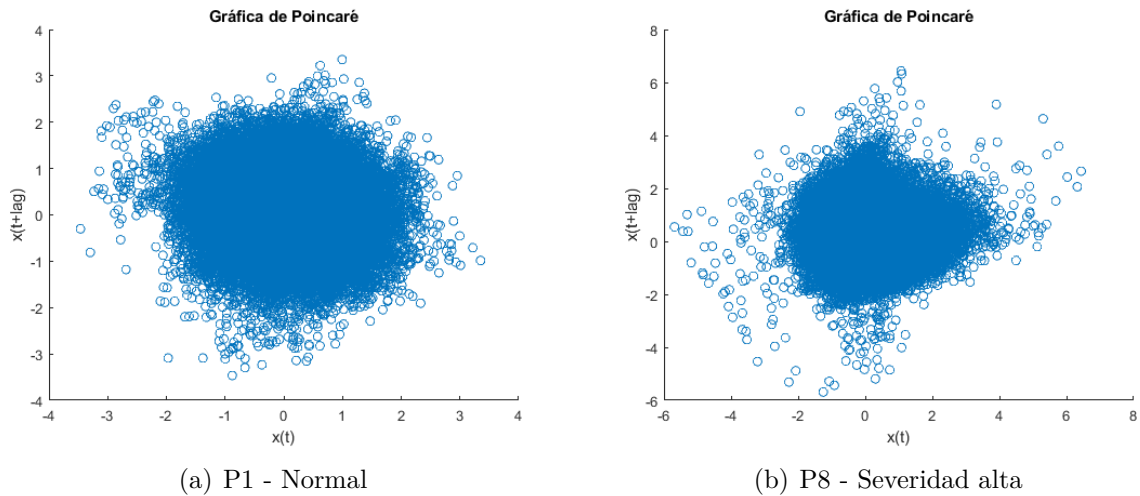
**Figura 5.7:** Señal de vibración vs Señal de vibración retrasada

cuantitativas del gráfico de Poincaré para lograr una clasificación de fallas precisa en lugar de utilizar este gráfico como una simple herramienta gráfica cualitativa.

En los gráficos de Poincaré de la Figura 5.8 se muestra la comparación de la clase P1 (Normal) con la clase P8 (Severidad alta) del modo de fallo de ralladura, con el acelerómetro 1. Para el caso (a) se puede observar que tiene una forma circular con unas pequeñas dispersiones fuera de la circunferencia, lo que viene a ser el caso idóneo. Para el caso (b), el cual tiene una severidad alta, se esperaría que tenga características que lo diferencien de la normal; tal caso es así, y se aprecia que éste gráfico tiene una forma densa diferente con unas cuantas dispersiones fuera de la concentración de datos, siendo este análisis un punto de partida para la posterior extracción de las característica o atributos de Poincaré.

### 5.3. Extracción de los atributos de Poincaré

Para la extracción de atributos de Poincaré es necesario construirlas a partir de las señales de vibración de la caja de engranajes rectos, de modo que estas características serán las que nos permitirán realizar las clasificaciones en los diferentes modos de fallos con su severidad permitiendo así una herramienta de apoyo al mantenimiento predictivo de la máquina, antes de producirse una falla en la misma. Este proceso de extracción de características forma parte



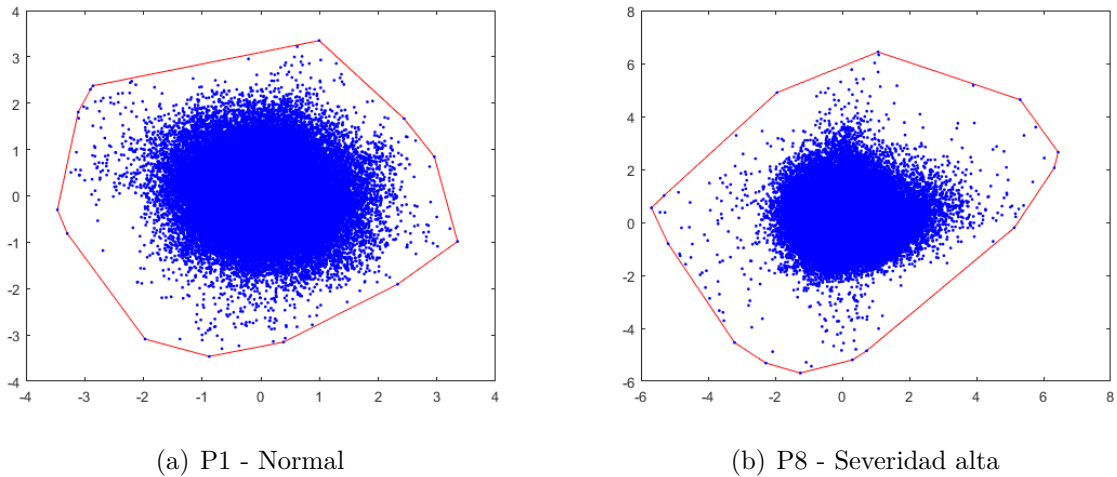
**Figura 5.8:** Comparación del gráfico de Poincaré clase Normal vs clase con severidad alta

del análisis de señales de vibración, lo cual es fundamental para la detección y exploración de las condiciones de las cajas de engranajes.

Las características extraídas en este proyecto se corresponden a las descritas en la sección 4.5.2, siendo las más comunes la *SD1*, la *SD2* y en casco convexo (*convex hull*, en inglés), el cual define un polígono cuya área se puede calcular, y se utiliza como una característica. En la figura 5.9 tenemos al área convexa obtenida de las señales de vibración del modo de fallo rayadura con el acelerómetro 1, en donde (a) representa el caso Normal en P1 y el caso (b) indica el modo de fallo con una severidad alta en P8; en ambos se observa un cúmulo de puntos, con puntos dispersos en diferentes direcciones. La diferencia principal radica en la forma en la que se encuentra estos puntos dispersos encerrados en el área convexa, sumado a esto que para el caso (a) la escala es casi la mitad del caso (b).

Por lo tanto, una vez realizado este análisis y verificando que en efecto las gráficas de Poincaré pueden producir indicadores informativos para la detección y el diagnóstico de fallos, se procede a extraer todas las 11 características de Poincaré con las ecuaciones que fueron descritas en la sección 4.5.2. Esto produce una matriz de datos (también llamado corpus) de 4050x11 para cada clase. Considerando los grupos de clases definidos, entonces se obtuvieron 9 corpus (3 grupos para cada velocidad), para cada modo de fallas:

- Para el grupo 1, una matriz de 16200x11



**Figura 5.9:** Comparación del Área convexa clase Normal vs clase con severidad alta ()

- Para los grupo 2 y 3, una matriz de 12150x11

## 5.4. Selección de atributos con ANOVA + CDbw

### 5.4.1. Datos Train-Test y normalización

Para la realización de la clasificación supervisada de los modos de fallo con respecto a su severidad es necesario dividir las muestras totales en dos grupos:

- **Datos de entrenamiento:** Se corresponde a datos para el ajuste del modelo de clasificación y representa el 80 % de los datos totales de cada corpus de datos, por modo de fallo y para cada grupo como lo son  $G1(P1,P2,P3)$ ,  $G2(P1,P4,P5,P6)$ ,  $G3(P1,P7,P8,P9)$ .
- **Datos de Test:** El 20 % de los datos restantes de cada corpus de datos se usarán para probar el desempeño de los clasificadores obtenidos en la con los datos de entrenamiento. Esto permitirá evaluar las mejores características obtenidas en la fase en entrenamiento y además de la información relevante que LAMDA-HAD nos proporcione el mejor modelo.

La técnica de normalización empleada para nuestro estudio es el min-max que usualmente ha sido utilizado para el procesamiento de imágenes; esta técnica se caracteriza por re-escalar las magnitudes de cada característica en el corpus de datos en valores que están en el rango  $[0 - 1]$ , manteniendo así valores positivos. Para realizar la normalización se debe conocer el valor mínimo y máximo del dominio de cada característica en el corpus respectivo, con la



siguiente ecuación (40).

$$\bar{x} = \frac{x - xmin(x)}{xmax(x) - xmin(x)} \quad (40)$$

#### 5.4.2. Selección de atributos y análisis exploratorio usando t-SNE

La aplicación de algoritmo de la sección 4.6.3, se generó una jerarquización de grupos de atributos, de los 11 analizados, que producen las mejores estructuras de clúster para cada grupo de fallas analizada, por velocidad y por modo de fallo. La selección de atributos es un desafío debido que aquí radica el éxito posterior en la clasificación de los fallos. En la tabla 3 se muestra un ejemplo de los resultados obtenidos de la aplicación de los algoritmos de selección de características, respecto a las dos métricas propuestas, i.e. para Max\_F1\_Extendido y Max\_CDbw\_Global. El número identificador se corresponde con el atributo listado en el orden de la sección 4.5.2

Nro. de atributos	Max_F1_Extendido	Max_CDbw_global
3	[9,3,1]	[6,3,1]
4	[9,6,3,1]	[7,5,3,1]
5	[9,6,5,3,1]	[9,6,5,3,1]
6	[11,9,8,6,5,1]	[9,8,6,5,3,1]
7	[10,9,6,5,3,2,1]	[9,8,7,6,5,3,1]
8	[11,10,9,8,7,6,5,1]	[9,8,7,6,5,3,2,1]
9	[11,10,9,8,7,6,5,3,1]	[11,10,9,8,7,6,5,3,1]
10	[11,10,9,8,7,6,5,3,2,1]	[11,10,9,8,7,6,5,3,2,1]
11	[11,10,9,8,7,6,5,4,3,2,1]	[11,10,9,8,7,6,5,4,3,2,1]

**Tabla 3:** Ejemplo de Selección de los mejores atributos con ANOVA + CDbw con el acelerómetro 1, modo de fallo Picadura, Grupo 1, Velocidad 1

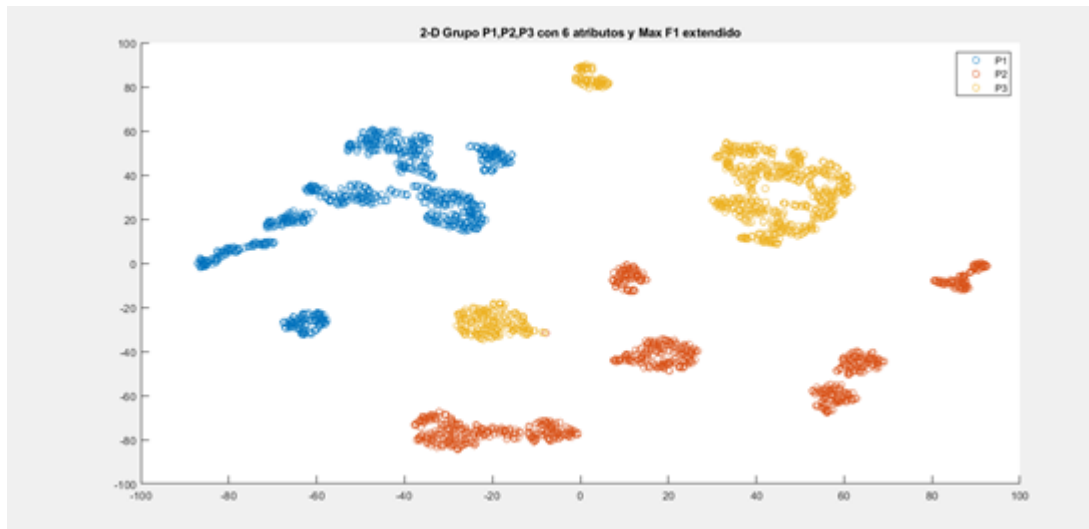
Estas métricas determinan los mejores grupos de atributos que luego se usarán en la aplicación de algoritmo de clasificación LAMDA-HAD. Siendo LAMDA-HAD un clasificador que depende fuertemente de la estructura de *clúster* de las clases (*clústers* densos y separados), resulta indispensable aplicar una técnica de selección de atributos que refleje este requerimiento de la mejor manera. En esta tabla vemos que hasta el grupo de 8 atributos, hay diferencias en los atributos que se selecciona, según la métrica usada.

Con el fin de hacer un análisis exploratorio de las estructuras de *clústers* obtenidas con las características seleccionadas, se usó la técnica de visualización t-SNE (Van der Maaten y

Nro. de atributos	Max_F1_Extendido	Max_CDbw_global
3	[8,3,9]	[8,6,4]
4	[8,3,1,9]	[8,3,1,9]
5	[8,6,3,1,9]	[8,6,3,1,9]
6	[8,6,3,1,9,7]	[11,10,8,6,3,9]
7	[11,8,6,5,3,1,9]	[11,10,8,6,3,1,9]
8	[11,10,8,6,5,3,1,9]	[11,8,6,5,3,1,9,4]
9	[11,10,8,6,5,3,1,9,7]	[11,8,6,5,3,1,9,4,2]
10	[11,10,8,6,5,3,1,9,7,2]	[11,8,6,5,3,1,9,4,7,2]
11	[11,10,8,6,5,3,1,9,4,7,2]	[11,10,8,6,5,3,1,9,4,7,2]

**Tabla 4:** Ejemplo de Selección de los mejores atributos con ANOVA + CDbw con el acelerómetro 1, modo de fallo Picadura, Grupo 2, Velocidad 1

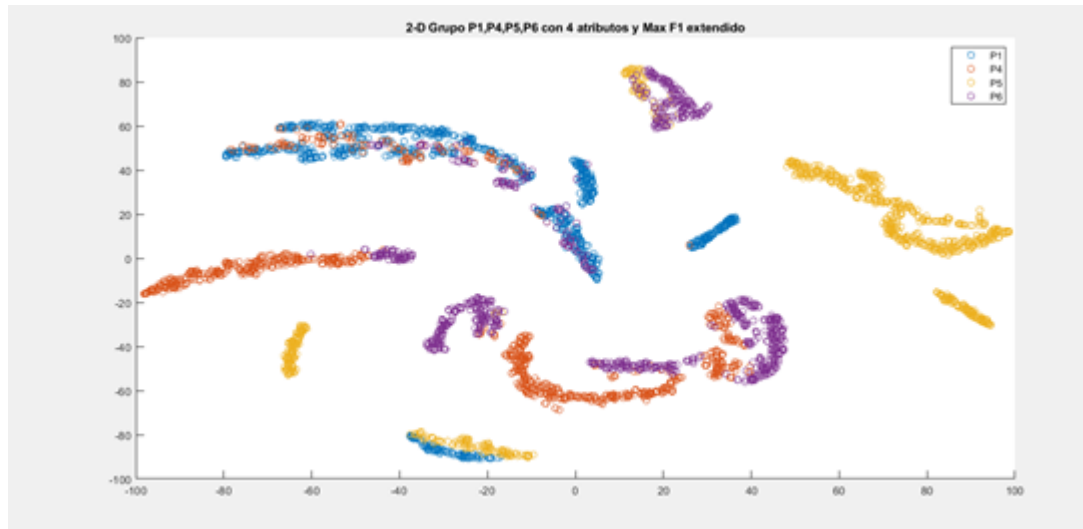
Hinton 2008); esta técnica sirve para reducir la dimensión del vector de atributos y así permite visualizar los datos de alta dimensión. Con fines demostrativos, hemos tomado los resultados con 6 atributos de la Tabla 3. La Figura 5.10 muestra la visualización 2D y con estas gráficas se puede constatar que existen separaciones apropiadas entre las clases que harán probable una buena clasificación supervisada con LAMDA-HAD; esto lo comprobaremos más adelante. En la figura 5.10 se observa que se logra separar de muy buena manera la clase normal P1 de las clases P2 y P3 que representan un nivel de severidad bajo que aun así guardan una distancia de separación considerablemente buena.



**Figura 5.10:** Gráfico de los 6 atributos con TSNE para el Grupo 1 con modo de fallo picadura

En la figura 5.11 se observan los resultados, escogidos de la Tabla 4 con 4 atributos seleccionados para el grupo 2, velocidad 1. Se observa que no hay separación adecuada de

la clase Normal con las clases de severidad media, es así que esto puede conllevar a que en la siguiente etapa exista un cierto margen de error, bajando así su precisión de clasificación porque el algoritmo va confundirse y ubicará a ciertas clases donde no correspondan pero esto se determinará con mayor detalle en la siguiente fase.

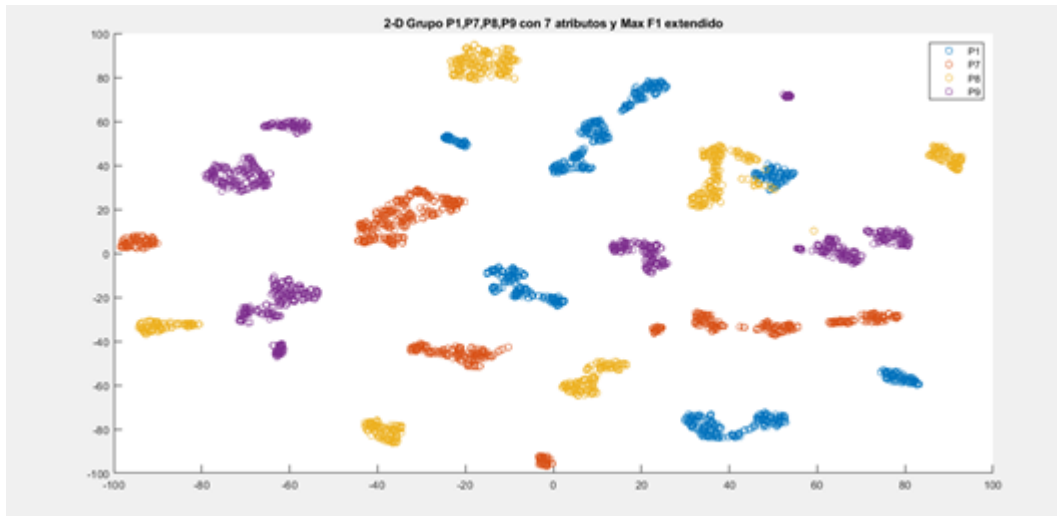


**Figura 5.11:** Gráfico de los 4 atributos con TSNE para el Grupo 2 con modo de fallo picadura

Nro. de atributos	Max_F1_Extendido	Max_CDbw_global
3	[11,3,5]	[9,1,6]
4	[11,10,3,5]	[8,9,1,6]
5	[7,4,8,9,6]	[4,8,9,1,6]
6	[7,4,8,9,1,6]	[7,2,8,9,1,6]
7	[7,10,4,8,9,1,6]	[7,2,4,8,9,1,6]
8	[7,2,10,4,8,9,1,6]	[7,11,10,3,5,8,1,6]
9	[11,10,3,5,4,8,9,1,6]	[7,11,10,3,5,8,9,1,6]
10	[7,2,11,10,5,4,8,9,1,6]	[7,2,11,10,3,5,8,9,1,6]
11	[7,2,11,10,3,5,4,8,9,1,6]	[7,2,11,10,3,5,4,8,9,1,6]

**Tabla 5:** Ejemplo de Selección de los mejores atributos con ANOVA + CDbw con el acelerómetro 1, modo de fallo Picadura, Grupo 3, Velocidad 2

En la figura 5.12, la gráfica de visualización 2D, utilizando los 7 atributos del grupo 3, velocidad 2, de la Tabla 5. Se aprecia con claridad una separación entre las clases de severidad alta con respecto al estado normal; con esta gráfica ya es un aporte que nos permite visualizar que nuestro algoritmo de clasificación muy probablemente tenga un buen desempeño en los resultados de clasificación.



**Figura 5.12:** Gráfico de los 7 atributos con TSNE para el Grupo 3 con modo de fallo picadura

## 5.5. Clasificación de fallos con LAMDA-HAD

### 5.5.1. Determinación de las mejores modelos de clasificadores con los datos de entrenamiento

Una vez obtenidos los grupos de atributos para cada corpus de datos en estudio, se procedió a obtener el mejor clasificador basado en el algoritmo LAMDA-HAD, con los datos de entrenamiento. La búsqueda del mejor modelo consistió en una búsqueda exhaustiva sobre la mejor combinación de los siguientes hiperparámetros:

1. **El conjunto de atributos.** Se consideraron los grupos de atributos establecidos en grupos desde 3 hasta los 11, determinados por la selección con ANOVA y CD<sub>bw</sub>.
2. **El valor de  $\alpha$ .** Se consideraron diferentes valores de  $\alpha$  sobre el intervalo de [0.9 a 0.5], con pasos de 0.1.
3. **Las funciones S y T.** Estas funciones se usan en el algoritmo LAMDA-HAD para el cálculo del grado de adecuación global, en donde intervienen las funciones de t-norma (intersección) y t-conorma (unión), ver la tabla de la Figura 4.12.

Para medir el rendimiento, se usó la precisión en la clasificación, la cual establece la tasa de cantidad de clasificaciones acertadas respecto a todas las clasificaciones disponibles, para cada clase. Esto es, el porcentaje de los positivos ciertos (*true positive*, en inglés). Por otro

lado, en este trabajo no se estimó la clase predicha por el modelo respecto a las clasificaciones absolutas para cada sub-señal, sino que la clasificación de una muestra de señal de vibración de longitud 10 s se determinó luego de evaluar la clase mayoritaria en frecuencia para las 90 subseñales que se obtienen de la muestra original.

A continuación, tomado como ejemplo el modo de fallo picadura con el acelerómetro 1, en la Tabla 6 se muestran los mejores parámetros que produjeron la mejor precisión en cada grupo G1, G2, G3. Se observa que para los grupos G1 y G2 bajo condición de operación en velocidad V1 y todas las cargas, se obtuvo una buena clasificación, siendo la mejor clasificación para el grupo G1 con una precisión del 100%. Este resultado se logró con el conjunto de 6 atributos, un valor  $\alpha = 0,6$ , y las funciones Min-Max para el cálculo de la adecuación global. En general, las funciones Min-Max para el cálculo de la norma y la conorma fueron las que permitieron mejores desempeños. Para el el grupo G2 y G3 se observa, por ejemplo, que los resultados se lograron con otro conjunto de atributos.

<b>Grupo/velocidad</b>	<b>Parámetros</b>	
<b>Grupo 1 Velocidad 1</b>	Alpha( $\alpha$ )	0.6
	Número de atributos	6
	Min-Max	X
	Producto-suma	
	Precisión	100 %
<b>Grupo 2 Velocidad 1</b>	Alpha( $\alpha$ )	0.5
	Número de atributos	4
	Min-Max	X
	Producto-suma	
	Precisión	89.58 %
<b>Grupo 3 Velocidad 2</b>	Alpha( $\alpha$ )	0.7
	Número de atributos	7
	Min-Max	X
	Producto-suma	
	Precisión	85.42 %

**Tabla 6:** Mejor clasificador LAMDA-HAD para los grupos con con el acelerómetro 1, modo de fallo picadura, con datos de entrenamiento

### 5.5.2. Comprobación del clasificador con los datos de Test

Una vez obtenidos los mejores clasificadores en la fase de entrenamiento, se almacenan también los otros parámetros obtenidos como consecuencia de la distribución de los datos de entrenamiento considerados, tales como  $\rho_{k,j}$ ,  $MGAD_{k,p}$  y  $GAD_{NIC_k}$ , ver sección 4.7. En lo

que sigue, con cada muestra en los datos de prueba se determinará a que clase pertenecen, usando el clasificador LAMDA HAD respectivo.

Con el fin de argumentar con los datos de prueba, la forma de estimar la clase, se analizan los diagrama de caja para cada grupo G-V, respecto al valor del GAD obtenido para cada sub-muestra, y las precisiones obtenidas. Con fines demostrativos se lo hará con grupo de clases G1-G2-G3, únicamente para el modo de fallo picadura. El diagrama de caja sirve para ubicar el posicionamiento del GAD en la escala [0-1], aportando la advertencia sobre las dispersiones de los datos. Los resultados nos exponen la cantidad de datos que fueron clasificados en esas clases proporcionando más información y evidenciando al clasificador como adecuado.

<b>Prueba de Picadura</b>				
<b>Grupos</b>		<b>Grupo 1</b>	<b>Grupo 2</b>	<b>Grupo 3</b>
<b>Velocidad Escogida</b>		<b>Velocidad 1</b>	<b>Velocidad 1</b>	<b>Velocidad 2</b>
<b>Parámetros</b>	$\alpha$	0.6	0.5	0.7
	<b># Atributos</b>	6	4	7
	<b>Min-Max</b>	X	X	X
	<b>Precisión</b>	99.95 %	81.02 %	69.6 %

**Tabla 7:** Tabla de precisión de los datos de prueba de Picadura

En la Figura 5.14 se observa a P1 por encima de 0.54, a P2 se lo identifica con valor por debajo de 0.53, y P3 tiene un valor por encima del 0.58, por lo tanto estos valores de GAD se diferencian y sugiere que se logrará una correcta clasificación. Además en la figura 5.13 se muestran las clases y sus clasificaciones como adicional para tener en cuenta en donde se alojan la mayor cantidad de datos, el grupo 1 muestra una precisión de 99.95 como se indica en la tabla 7. En el caso de P1 con una 100 % de precisión, P2 es la de menor precisión alcanza 99.87 %, esto considerando el total de las 810 señales, porque con el cálculo considerando las 90 sub-señales y clasificándola como clase adecuada a la que es mayoría en su conjunto y se obtiene mejores resultados y P3 también tiene una precisión del 100 %, algunas clases de P2 se clasificaron erróneamente en P3 pero es una cantidad de minoría.

En la figura 5.14 se observa que todas las clases tienen GADs muy bien separados, por lo cual la inspección mediante diagramas de caja para este grupo 2 sería muy favorable, también surgen clases NIC que de acuerdo a la figura 5.13 en (b) Grupo 2 son pocos datos. El grupo 2 muestra una precisión de 81.02 % como se indica en la tabla 7, por lo cual obteniendo la precisión para P1 es del 92.46 %, la clase P4 es la que posee la precisión más baja alrededor

del 56 %, P5 una precisión del 100 %, y P6 dispone del 74.44 %.

	1	2
1	'P1'	810x1 double
2	'P2'	809x1 double
3	'P3'	811x1 double

a) Grupo 1

	1	2
1	'P1'	984x1 double
2	'P4'	564x1 double
3	'P6'	813x1 double
4	'P5'	848x1 double
5	'NIC5'	[0.5743;0.57...
6	'NIC6'	24x1 double

b) Grupo 2

	1	2
1	'P1'	746x1 double
2	'P8'	803x1 double
3	'NIC8'	164x1 double
4	'P7'	714x1 double
5	'NIC9'	239x1 double
6	'P9'	543x1 double
7	'NIC7'	31x1 double

c) Grupo 3

Figura 5.13: Resultados de la clasificación de los datos de Test con Picadura

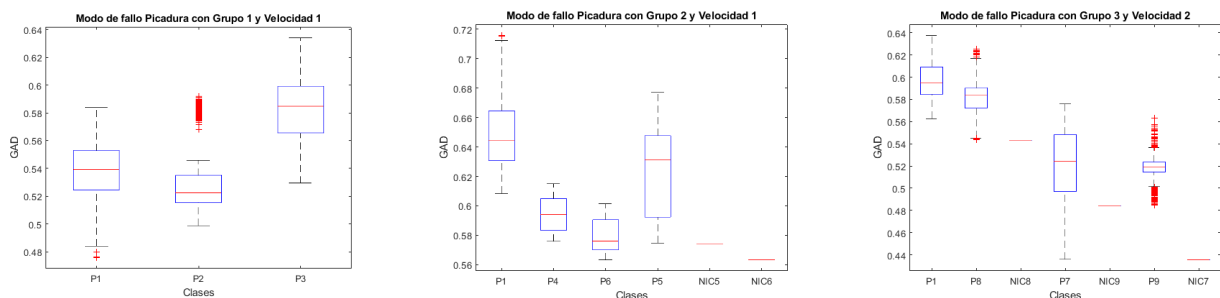


Figura 5.14: Diagrama de caja de Picadura con los datos de prueba para todos los grupos

En la figura 5.14 se muestran las clases P1 siendo el estado normal, y P7, P8, P9 que son clases con una severidad de fallo alta. Se puede verificar que los GADs están muy bien separados entre las clases siendo este medio visual mediante diagramas de cajas un punto referencial importante para el análisis, en donde según la tabla 7 el grupo 3 posee de la precisión menor con 69.59 %, donde específicamente en P1 dispone del 72.54 % de datos bien clasificados; para P7 es 74.56 %, P8 es 65.92 % y P9 es 65.31 % la precisión correspondiente. Esta es una de las razones por las que se reformuló el cálculo de la clase estimada, no por cada sub-muestra sino con base en la clasificación que más se repite en las sub-muestras de una muestra original. Además en la Tabla de la Figura 5.13 en c) Grupo 3 se puede observar que se crearon clases NIC en donde también se clasificaron datos y es la razón por la que baja la precisión, aunque esto nos indica que ahí es probable que se alojen datos con se podrían clasificar a la clase P correspondiente.

Con el análisis de los diagramas de cajas y los resultados de las clasificaciones nos comprueban que fueron escogidos los mejores clasificadores, con los datos de prueba obtuvimos precisiones exitosas, considerando que existen sub-muestras de todas las señales, razón por la cual se recomienda obtener una base de datos aleatoria con las señales de vibración con su paquete de 90 sub-señales que se las utilizó para agrandar la base de datos y que ahora podemos verificar la cantidad de clases en ese rango de señales y determinar la más adecuada siendo está la que más veces se repite, esto será expuesto en el análisis de resultados.



## 6. Validación y pruebas del diseño

En este capítulo presentan todos los resultados obtenidos con la utilización del algoritmo de clasificación basado en similaridad difusa aplicado en señales de vibración, para los diferentes niveles de severidad de los modos de fallos estudiados. Se presentará una sección en la cual se mostrará los resultados obtenidos con los datos de entrenamiento en los modos de fallo, además se presentará una sección que explique el análisis de resultados con los datos de prueba para la detección y diagnóstico de fallos de caja de engranajes rectos.

### 6.1. Resultados con los datos de entrenamiento

En la tabla 8 se presentan los mejores resultados de la fase de entrenamiento, para todos los grupos de fallos y velocidades para el modo de fallo picadura. Los mejores clasificadores para el grupo 1 y grupo 2 se obtuvieron con la velocidad 1 y Min-Max con una precisión de 100 % y 89.58 % respectivamente. Para el grupo 3 el mejor clasificador está dado por la velocidad 2 y Min-Max con una precisión de 85.42 %, siendo así estos los mejores clasificadores para el modo de fallo de picadura.

MODO DE FALLO PICADURA CON ACELERÓMETRO 1							
	Parámetros	Velocidad 1		Velocidad 2		Velocidad 3	
<b>G1</b>	Alfa	0,6	0,6	0,9	0,5	0,9	0,6
	Atributos	6	9	10	10	8	8
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	<b>100</b>	100	83,33	77,78	75	67,59
<b>G2</b>	Alfa	0,5	0,6	0,5	0,8	0,5	0,7
	Atributos	4	9	6	5	10	10
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	<b>89,58</b>	79.17	79.17	79,86	68,75	61,80
<b>G3</b>	Alfa	0.6	0.5	0.7	0.6	0.7	0.9
	Atributos	3	3	7	9	8	6
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	76.39	73.61	<b>85.42</b>	59.72	52.08	60.42

**Tabla 8:** Resultados de los clasificadores en el modo de fallo de Picadura con los datos de entrenamiento

En la tabla 9 se presentan los mejores resultados de la fase de entrenamiento para el modo

de fallo diente roto, en donde los mejores clasificadores para el grupo 1 y grupo 3 se obtuvieron con la velocidad 1 y Min-Max con una precisión de 92.59 % y 99.31 % respectivamente. Para el grupo 2 el mejor clasificador está dado por la velocidad 2 y Min-Max con una precisión de 78.47 %, siendo así estos los mejores clasificadores para el modo de fallo con diente roto.

<b>FALLO DIENTE ROTO CON ACELERÓMETRO 1</b>							
	<b>Parámetros</b>	<b>Velocidad 1</b>		<b>Velocidad 2</b>		<b>Velocidad 3</b>	
<b>G1</b>	Alfa	0.8	0.6	0.9	0.5	0.9	0.9
	Atributos	3	5	3	4	5	4
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	<b>92.59</b>	91.67	79.63	69.44	40.74	53.70
<b>G2</b>	Alfa	0.8	0.5	0.9	0.5	0.8	0.5
	Atributos	3	5	5	7	3	8
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	64.58	64.58	<b>78.47</b>	70.83	55.55	58.33
<b>G3</b>	Alfa	0.5	0.8	0.7	0.5	0.9	0.8
	Atributos	5	6	3	3	3	10
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	<b>99.31</b>	95.83	79.86	76.39	41.67	54.86

**Tabla 9:** Resultados de los clasificadores en el modo de fallo de Diente Roto con los datos de entrenamiento

En la tabla 10 se muestran los mejores clasificadores para el modo de falla agrietamiento. En donde los mejores clasificadores para el grupo 1 y grupo 2 se obtuvieron con la velocidad 3 y Min-Max con una precisión de 100 % y 74.31 % respectivamente. Para el grupo 3 el mejor clasificador está dado por la velocidad 1 y Producto-Suma con una precisión de 95.14 %, siendo así estos los mejores clasificadores para el modo de fallo por agrietamiento.

En la tabla 11 se muestran los mejores clasificadores en la fase de entrenamiento para el modo de fallo rayadura. Los mejores clasificadores para el grupo 1 y grupo 3 se obtuvieron con la velocidad 3 y Min-Max para el grupo 1 y Producto-Suma para grupo 3 con una precisión de 100 % y 91.67 % respectivamente. Para el grupo 2 el mejor clasificador está dado por la velocidad 1 y Min-Max con una precisión de 80.56 %, siendo así estos los mejores clasificadores para el modo de fallo por rayadura.

FALLO POR AGRIETAMIENTO CON ACELERÓMETRO 1							
	Parámetros	Velocidad 1		Velocidad 2		Velocidad 3	
<b>G1</b>	Alfa	0.5	0.8	0.5	0.5	0.7	0.9
	Atributos	9	10	3	6	5	5
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	66.67	83.33	63.89	64.81	<b>100</b>	99.07
<b>G2</b>	Alfa	0.6	0.9	0.5	0.5	0.6	0.9
	Atributos	6	11	3	7	11	10
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	61.11	68.75	61.81	59.72	<b>74.31</b>	62.5
<b>G3</b>	Alfa	0.5	0.5	0.5	0.7	0.6	0.9
	Atributos	3	4	3	7	9	5
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	89.58	<b>95.14</b>	84.03	88.89	79.17	79.17

**Tabla 10:** Resultados de los clasificadores en el modo de fallo por agrietamiento con los datos de entrenamiento

FALLO POR RAYADURA CON ACELERÓMETRO 1							
	Parámetros	Velocidad 1		Velocidad 2		Velocidad 3	
<b>G1</b>	Alfa	0.6	0.7	0.6	0.9	0.5	0.9
	Atributos	8	10	9	5	10	9
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	90.74	92.59	73.15	63.89	<b>100</b>	93.52
<b>G2</b>	Alfa	0.5	0.9	0.5	0.9	0.5	0.9
	Atributos	3	3	3	8	11	4
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	<b>80.56</b>	79.17	63.89	75	65.28	63.19
<b>G3</b>	Alfa	0.6	0.6	0.8	0.8	0.8	0.9
	Atributos	9	3	5	11	4	5
	Min-Max	X		X		X	
	Producto-Suma		X		X		X
	Precisión [%]	79.17	78.47	49.31	47.92	90.97	<b>91.67</b>

**Tabla 11:** Resultados de los clasificadores en el modo de fallo por rayadura con los datos de entrenamiento

## 6.2. Análisis de resultados con los datos de prueba

Con la obtención de los mejores clasificadores con los datos de entrenamiento es importante corroborar estos resultados con los datos de prueba. Se utilizan los parámetros de los clasificadores con la que se obtuvo la mejor precisión para los modos de falla por diente roto, agrietamiento, picadura y rayadura. A continuación se mostrará las precisiones alcanzadas que son calculadas para todas señales de vibración que pertenecen al grupo, es decir de las 90 sub-señales por cada señal se escogió como correcta la que más veces se repetía, de este modo se llega a la cantidad original de señales y se reduce la clasificación errónea.

### Análisis del clasificador para el modo de fallo por diente roto

Utilizamos los mejores parámetros de clasificación de la tabla 9 del diente roto con los datos de entrenamiento y los utilizamos para comprobar nuestros clasificadores con los datos de Test. En la tabla 12 están los parámetros escogidos de los mejores clasificadores por grupos, así como su velocidad y la precisión lograda. En la tabla 12 se muestra que alcanza en el grupo 1 una precisión del 85.19 % y velocidad 1, en el grupo 2 una precisión del 72.22 % y velocidad 2, en el grupo 3 una precisión del 94.44 % y velocidad 1.

Prueba del Diente roto				
Grupos		Grupo 1	Grupo 2	Grupo 3
Velocidad Escogida		Velocidad 1	Velocidad 2	Velocidad 1
Parámetros	$\alpha$	0.8	0.9	0.5
	# Atributos	3	5	5
	Min-Max	X	X	X
	Precisión	85.19 %	72.22 %	94.44 %

**Tabla 12:** Tabla de precisión de los datos de prueba de Diente Roto

### Análisis del clasificador para el modo de fallo por agrietamiento

Utilizamos los mejores parámetros de clasificación de la Tabla 10 del modo de fallo por agrietamiento con los datos de entrenamiento y los utilizamos para comprobar nuestros clasificadores con los datos de prueba. En la Tabla 13 están los parámetros escogidos de los mejores clasificadores por grupos, así como su velocidad y la precisión lograda. En la Tabla 13 se obtiene para el grupo 1 un 100 % de clasificación adecuada y con velocidad 3, para el grupo 2 un 69.44 % de clasificación y con velocidad 3, para el grupo 3 un 91.66 % de datos clasificados y con velocidad 3.

Prueba del modo de fallo por Agrietamiento				
Grupos		Grupo 1	Grupo 2	Grupo 3
Velocidad Escogida		Velocidad 3	Velocidad 3	Velocidad 1
Parámetro	$\alpha$	0.7	0.6	0.5
	# Atributos	5	11	4
	Min-Max	X	X	
	Producto-suma			X
	Precisión	100 %	69.44 %	91.66 %

**Tabla 13:** Tabla de precisión de los datos de Test del modo de fallo por agrietamiento

### Análisis del clasificador para el modo de fallo por picadura

Utilizamos los mejores parámetros de clasificación de la tabla 8 del modo de fallo por picadura con los datos de entrenamiento y los utilizamos para comprobar nuestros clasificadores con los datos de Test. En la tabla 14 están los parámetros escogidos de los mejores clasificadores por grupos, así como su velocidad y la precisión lograda. Esto es favorable para una buena clasificación de los datos y así se lo verifica en la tabla 14 puesto que se obtiene para el grupo 1 un 100 % de clasificación adecuado con velocidad 1, el grupo 2 un 83.33 % de clasificación con velocidad 1, el grupo 3 un 77.77 % de precisión de clasificación con velocidad 2.

Prueba del modo de fallo por Picadura				
Grupos		Grupo 1	Grupo 2	Grupo 3
Velocidad Escogida		Velocidad 1	Velocidad 1	Velocidad 2
Parámetro	$\alpha$	0.6	0.5	0.7
	# Atributos	6	4	7
	Min-Max	X	X	X
	Precisión	100 %	83.33 %	77.77 %

**Tabla 14:** Tabla de precisión de los datos de Test del modo de fallo por picadura

### Análisis del clasificador para el modo de fallo por rayadura

Utilizamos los mejores parámetros de clasificación de la tabla 11 del modo de fallo por rayadura con los datos de entrenamiento y los utilizamos para comprobar nuestros clasificadores con los datos de Test. En la tabla 15 están los parámetros escogidos de los mejores clasificadores por grupos, así como su velocidad y la precisión lograda. Según la tabla 15, el grupo 1 tiene una precisión del 100 % en el clasificador con velocidad 3, el grupo 2 tiene una precisión del 75 % en el clasificador con velocidad 1, el grupo 3 tiene una precisión del 88 % con velocidad 3.

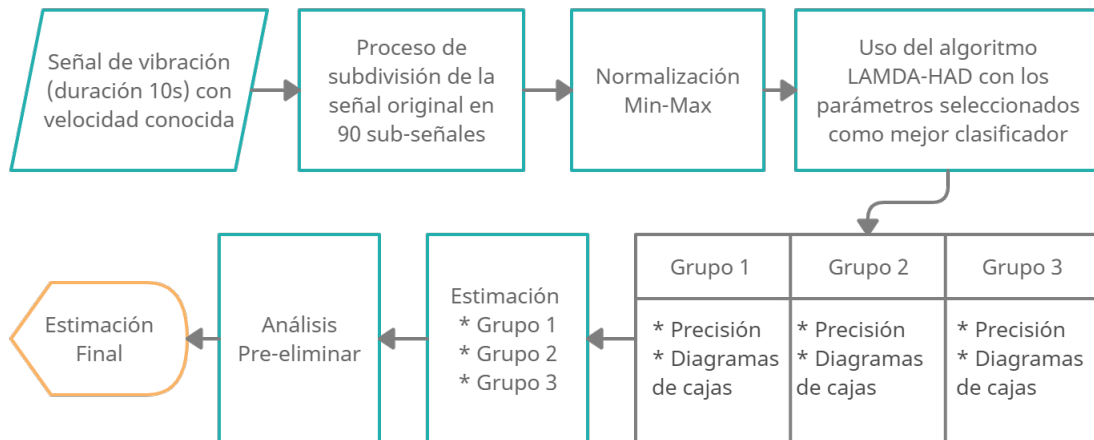
Prueba del modo de fallo por Rayadura				
Grupos		Grupo 1	Grupo 2	Grupo 3
Velocidad Escogida		Velocidad 3	Velocidad 1	Velocidad 3
Parámetro	$\alpha$	0.5	0.5	0.8
	# Atributos	10	3	4
	Min-Max	X	X	X
	Precisión	100 %	75 %	88 %

**Tabla 15:** Tabla de precisión de los datos de Test del modo de fallo por rayadura

Con los resultados de precisión en la clasificación con los datos de prueba, mostramos que si es posible tener un clasificador simple, como el construido con el algoritmo LAMDA-HAD, para la detección y diagnóstico de fallos en diferentes niveles de severidad de cuatro modos de fallo en engranajes rectos. Cabe mencionar que para la mayoría de casos se obtuvo una precisión considerablemente buena, únicamente para el modo de fallo por Agrietamiento en la fase de prueba se obtuvo una precisión menos al 70 % dejando entre ver la debilidad del clasificador para ciertos modos de fallo.

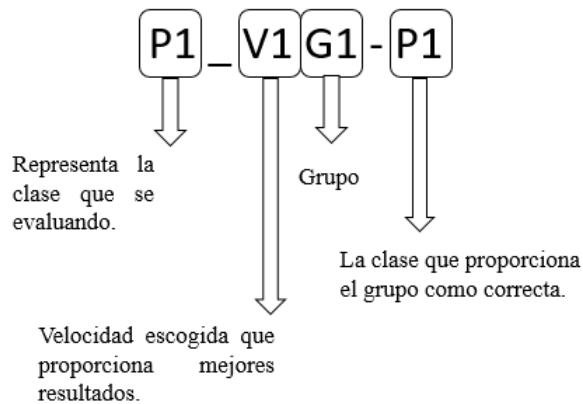
### 6.3. Clasificación de una muestra con el uso simultáneo de todos los clasificadores

Un análisis complementario que se realizó en este proyecto, es la estimación de la clase de una muestra de los datos de prueba, con el uso simultáneo de los 3 grupos de clasificadores. Esto es, dada una muestra de clase desconocida, a velocidad conocida de la máquina, procesarla con los 3 clasificadores por grupo y determinar cuál de los 3 clasificadores podría dar la estimación correcta, como se muestra en la figura 6.1. Para ello, se hace uso de los diagramas de caja de los resultados de la fase de prueba.



**Figura 6.1:** Esquema del proceso para el análisis complementario

Para comprender la notación usada en cada diagrama de caja que se presentará a continuación, se muestra la Figura 6.2 que indica lo que describe los próximos diagramas de caja.

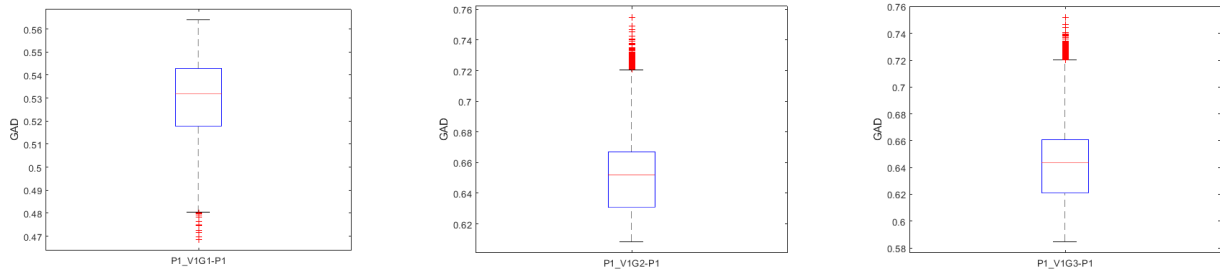


**Figura 6.2:** Representación de las letras en los diagramas de caja

### Modo de fallo por Picadura: análisis de la clase P1 con los 3 grupos

Se observa en la Figuras 6.3, que para el P1\_V1G1-P1 es donde el GAD promedio está

más cercano al 0.5, además según la tabla de la Figura 6.4 en el Grupo 1 se obtiene 4049 valores clasificados correctamente con una precisión del 100 %, por lo tanto es la que mejor proporciona información de estado normal de nuestra máquina; sin embargo los resultados arrojados por Grupo 2 y 3 equivalen a más del 92.5 % de precisión aproximadamente. Se concluye entonces que las muestras de la clase P1 estarían siendo clasificadas con un buen nivel de precisión.



**Figura 6.3:** Diagrama de caja de P1 con todos los grupos

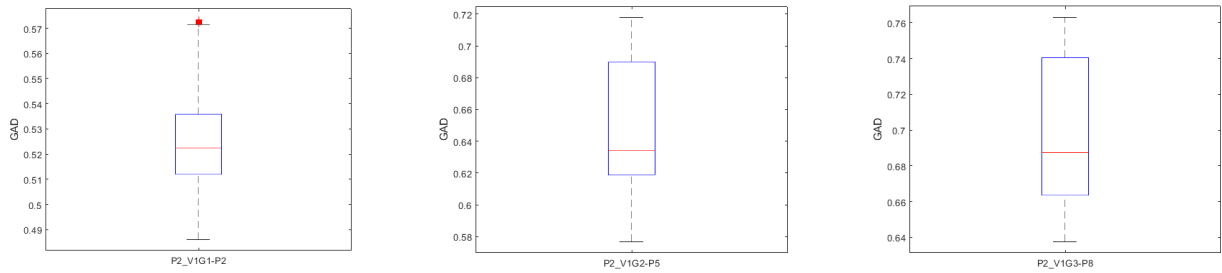
	1	2		1	2		1	2
1	'P1'	4049x1 dou...	1	'P1'	3750x1 dou...	1	'NIC9'	16x1 double
2	'NIC2'	0.4434	2	'NIC1'	18x1 double	2	'P1'	4005x1 dou...
			3	'P4'	278x1 double	3	'P9'	29x1 double
			4	'P5'	[0.6255;0.62...	4		

**Figura 6.4:** Datos clasificados en P1 en los grupos de Picadura

### Modo de fallo por Picadura: Análisis de la clase P2 con los 3 grupos

Como observa en la Figura 6.5 la clase P2 en el primer grupo la clasifica efectivamente en P2 con una precisión del 92.88 %, en el grupo 2 en P5 con el 99.9 % y el grupo 3 en P8 con el 57.93 %, el detalle que lleva a estos resultados se muestran en la tabla de la Figura 6.6. Ahora para determinar a qué clase pertenece, observamos que para fallas leves están relacionadas a GADs cercanos al 0.5, por ende nos inclinamos por la clasificación que indica el grupo 1.





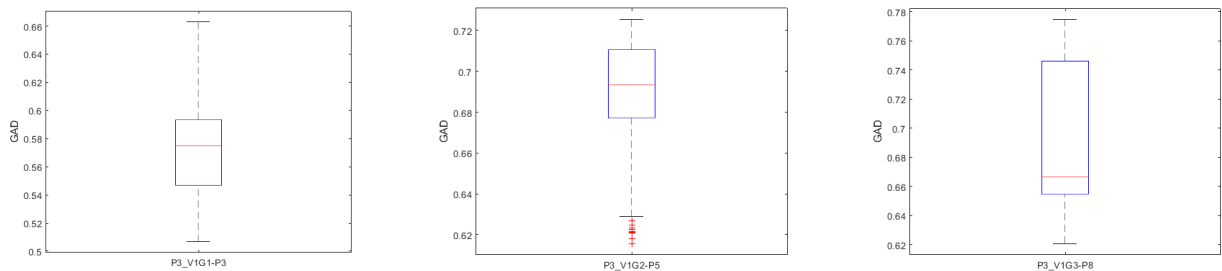
**Figura 6.5:** Diagrama de caja de P2 con todos los grupos

	1	2		1	2		1	2
1	'P2'	3762x1 dou...	1	'P5'	4046x1 dou...	1	'P7'	168x1 double
2	'P3'	288x1 double	2	'NIC5'	[0.5743;0.57...	2	'NIC9'	700x1 double
						3	'P9'	836x1 double
						4	'P8'	2346x1 dou...

**Figura 6.6:** Datos clasificados en P2 en los grupos de Picadura

### Modo de fallo por Picadura: Análisis de la clase P3 con los 3 grupos

Como se observa en la Figura 6.8 tanto P3 y P5 muestran una precisión del 100 % según la acumulación de datos en sus clases. Además, como se muestra en la figura 6.7 la diferencia es que P3\_V1G1-P3 tiene un GAD menor siendo así un fallo leve, por lo tanto el Grupo 1 se posiciona como nuestra mejor opción.



**Figura 6.7:** Diagrama de caja de P3 con todos los grupos

	1	2
1 'P3'		4049x1 dou...
2 'P2'		0.5407

	1	2
1 'P5'		4050x1 dou...

	1	2
1 'P8'		3971x1 dou...
2 'P9'		77x1 double
3 'NIC9'		[0.6039;0.60...

**Figura 6.8:** Datos clasificados en P3 en los grupos de Picadura

Con este análisis, es posible determinar el uso simultáneo de los grupos de clasificadores para estimar la clase de falla de una muestra desconocida, el cual debe ser formalizado apropiadamente en un trabajo futuro. Cabe destacar que este análisis no estaba previsto en los objetivos de este proyecto, pero se consideró conveniente proyectar la posibilidad de formalizar este enfoque de clasificación a partir de este estudio preliminar.

El análisis de las demás clases pueden revisarse en el Anexo A.

## 7. Conclusiones y recomendaciones

El problema de la detección y diagnóstico de fallos en cajas de engranajes rectos requiere de un análisis exhaustivo de señales de vibración, usando procesamiento de señales. Adicionalmente, las técnicas de aprendizaje automático proveen de algoritmos que pueden automatizar esta tarea de detección y diagnóstico de fallos. El uso combinado de técnicas de procesamiento de señales que generan indicadores (atributos) que pueden ser analizados con algoritmos de aprendizaje automático, es una excelente vía para producir clasificadores de fallas que apoyan a mantenimiento predictivo basado en el monitoreo de la condición.

En este trabajo se presentó la aplicación de los diagramas de Poincaré como técnica de procesamiento de señales para la extracción de atributos que son indicadores de la condición de fallo de las cajas de engranajes. También se presentó el uso del análisis de varianza (ANOVA) e índices de validación de agrupaciones CDbw, para la selección de atributos de Poincaré informativos. Luego se mostró la aplicación del algoritmo clasificador LAMDA-HAD para este problema.

Las señales analizadas en este proyecto se obtuvieron en un contexto de pruebas de una caja de engranajes rectos, en donde en un proyecto anterior se hizo la adquisición de las señales de vibración para el mantenimiento preventivo de la máquina. En este prototipo de laboratorio de la maquinaria rotativa, se simularon diferentes modos de fallo para su estudio.

El estudio realizado en este proyecto requirió determinar los mejores clasificadores para todos los modos de fallo y los diferentes grupos de estudio, en una fase de ajuste (entrenamiento) del modelo. Usando cada mejor modelos, el desempeño en la clasificación se basó en el análisis de los datos de prueba, para al ingresar nuevas muestras al modelo hacer la detección y el diagnóstico que indique que en efecto la máquina se encuentra en fallo y su diagnóstico permita ubicar su severidad y que necesita mantenimiento.

El desarrollo del proyecto tuvo su punto de partida en el aumento los datos a través de ventanas de tiempo con un solapamiento del 10 % aproximadamente del total de la duración de sub-señal. Esto se hizo para subsanar el problema de no tener suficientes muestras para un ajuste del modelo estadísticamente válido. Luego se seleccionó un retardo de tiempo adecuado de la serie temporal de vibración, identificando el valor en donde la autocorrelación es cero o cercano a ese valor.

En el proceso se extrajeron los atributos de Poincaré, éstas 11 características son las que brindan la información para la detección de fallos. Las magnitudes de estas características fueron normalizadas en el rango  $[0 - 1]$ , para posteriormente seleccionar las mejores agrupaciones de atributos con ANOVA e índices de validación de agrupaciones CDbw. Posteriormente se hizo una visualización 2D con el algoritmo t-SNE para observar el agrupamiento de datos que genera que tan dispersos se encuentran. Finalmente se aplicó el algoritmo LAMDA-HAD para obtener los mejores clasificadores. También se hizo un análisis de resultados con los diagramas de caja, y un análisis de sus GADs en las clases en sus respectivos grupos.

En este trabajo se verificó que para detección de anomalías en las cajas de engranajes funcionan para cierto modos de fallo con velocidad 1, para picadura y diente roto, y con velocidad 3 para agrietamiento y rayadura; además, el grupo G2 de severidad media fue el que menor precisión tuvo, aun así nos basamos en los demás clasificadores para un diagnóstico final y en los diagramas de caja. Para mejorar la precisión se dispuso evaluar la cantidad de veces en la que una clase se repetía, considerando las 90 sub-señales extraídas de la señal original, de este modo mejoró la precisión.

Como recomendación, se sugiere que para obtener modelos de clasificadores robustos se debe disponer de una base de datos de entrenamiento estadísticamente suficiente que permita el aprendizaje supervisado para su posterior uso con mayor precisión para las nuevas muestras ingresados al clasificador seleccionado.

Como trabajo futuro, se requiere formalizar la estimación de la clase de una muestra desconocida usando los diferentes grupos de clasificadores, con base en el valor de los GAD. Esto sugiere la formalización de una nueva propuesta referente al uso de clasificadores modulares y no el uso de un clasificador monolítico.

# ANEXOS

## A. Clasificación de una muestra con el uso simultáneo de todos los clasificadores

### Modo de fallo por Picadura: Análisis de la clase P4 con los 3 grupos

Como se observa en la figura A.1 de diagramas de caja y podemos descartar con facilidad a P4\_V1G1-P1 debido a que existe muestras dispersas y que su GAD es menor a 0.5. Nuestra mejor opción es la P4\_V1G2-P4 debido a que establece alrededor del 70 % de clasificación correcta según la tabla de la Figura A.2 donde se muestran los datos con las clasificaciones y además su GAD es mayor que todos siendo este un factor importante en la determinación de fallos medios y altos. Cabe mencionar que P4\_V1G3-P7 tiende a confundirse con la clase P1 de estado normal descartándose así esta opción. Por lo tanto nos decantaríamos por el grupo 2.

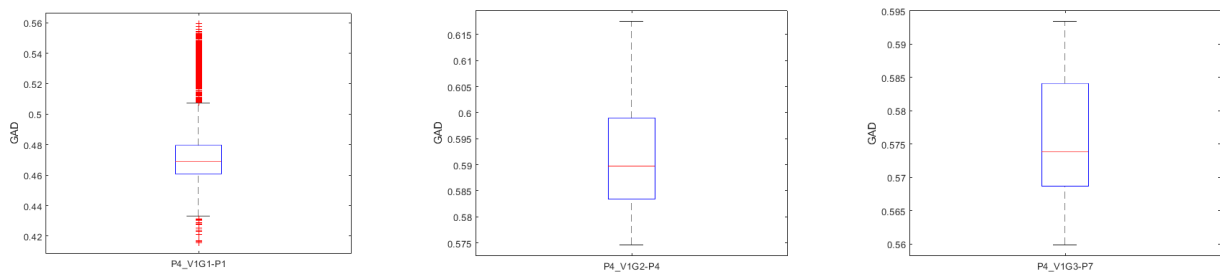


Figura A.1: Diagrama de caja de P4 con todos los grupos

	1	2		1	2		1	2
1	'P1'	3469x1 dou...	1	'P1'	393x1 double	1	'P1'	1069x1 dou...
2	'NIC2'	581x1 double	2	'P4'	2824x1 dou...	2	'P7'	2705x1 dou...
3			3	'P6'	799x1 double	3	'NIC1'	276x1 double
			4	'NIC1'	11x1 double			
			5	'P5'	14x1 double			
			6	'NIC4'	[0.5746;0.57...			

Figura A.2: Datos clasificados en P4 en los grupos de Picadura

## Modo de fallo por Picadura: Análisis de la clase P5 con los 3 grupos

En la figura A.4 claramente se puede observar que en P5\_V1G1-P3 existe mucha dispersión en las muestras, además en la figura A.3 se muestran los datos donde se crearon varias clases con lo que se identifica que existe una incertidumbre de modo que no es una clase correcta en ese grupo.

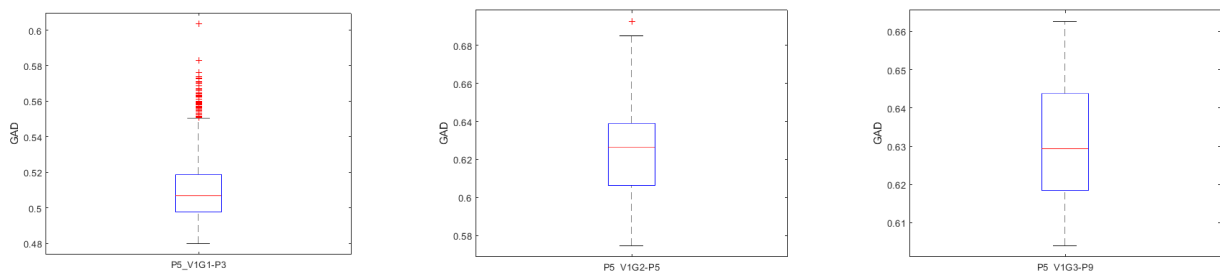
	1	2
1	'P1'	421x1 double
2	'NIC2'	476x1 double
3	'P3'	2417x1 dou...
4	'P2'	736x1 double

	1	2
1	'P1'	409x1 double
2	'NIC1'	17x1 double
3	'P5'	3516x1 dou...
4	'P6'	69x1 double
5	'P4'	11x1 double
6	'NIC5'	[0.5743;0.57...
7	'NIC6'	18x1 double

	1	2
1	'P1'	435x1 double
2	'P9'	2072x1 dou...
3	'P7'	662x1 double
4	'NIC9'	338x1 double
5	'NIC1'	0.5699
6	'NIC7'	27x1 double
7	'P8'	515x1 double

**Figura A.3:** Datos clasificados en P5 en los grupos de Picadura

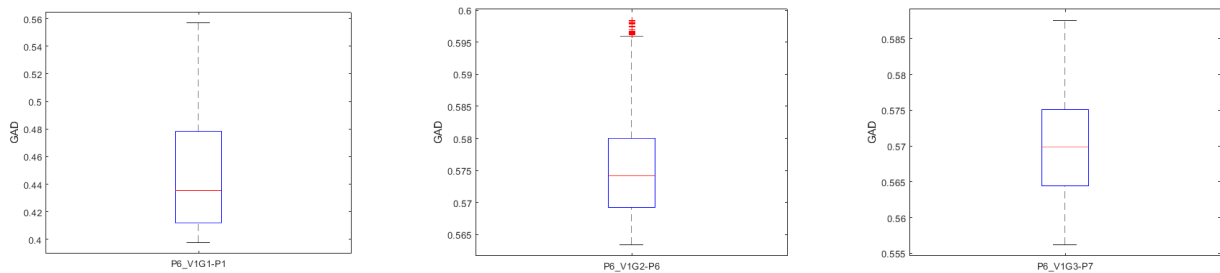
En la imagen A.4 en P5\_V1G3-P9 claramente dispone de un GAD promedio elevado lo que la calificaría como idónea, aquí recurrimos a la figura A.3 de los resultados en donde presenta varias clases y donde la clase P9 apenas alcanza el 51 % descartándose así. Finalmente la figura P5\_V1G1-P5 en la clasificación dispone de un GAD adecuado y la clase P5 proporciona una precisión de hasta el 85 % posesionándola como la mejor opción, siendo el grupo 2 el que mejor se acopla.



**Figura A.4:** Diagrama de caja de P5 con todos los grupos

### Modo de fallo por Picadura: Análisis de la clase P6 con los 3 grupos

Como observa en la figura A.5 en donde a la imagen P6\_V1G1-P1 la descartamos inmediatamente porque su GAD es menor a 0.5 además de tener varias clases como se muestra en la Figura A.6 con una cantidad menor al 50 % de los datos. Aquí surge la duda entre si es P6 o P7 debido a que existe cantidad similar en datos en ambas clases aunque P6 tiene un GAD mayor por lo tanto sería la opción más acertada con una precisión del 73 %, aun así podría confundirse con P7 que tiene una precisión del 83 % lo cual no estaría mal porque esto demuestra que existe un fallo con una severidad alta que haga que se realice el mantenimiento de inmediato, siendo así nuestra por el grupo 2 con inclinación por un fallo de severidad media con tendencia para alta.



**Figura A.5:** Diagrama de caja de P6 con todos los grupos

	1	2
1	'P1'	1981x1 dou...
2	'NIC2'	1842x1 dou...
3	'P3'	164x1 double
4	'NIC1'	58x1 double
5	'P2'	[0.4442;0.44...

	1	2
1	'P1'	439x1 double
2	'P4'	418x1 double
3	'P6'	2958x1 dou...
4	'P5'	161x1 double
5	'NIC5'	20x1 double
6	'NIC6'	54x1 double

	1	2
1	'P1'	641x1 double
2	'P7'	3375x1 dou...
3	'NIC1'	34x1 double

**Figura A.6:** Datos clasificados en P6 en los grupos de Picadura



### Modo de fallo por Picadura: Análisis de la clase P7 con los 3 grupos

En la Figura A.8 en P7\_V1G1-P2 observamos que el GAD es menor a 0.5 por lo tanto no es. Las siguientes dos imágenes P7\_V1G1-NIC6 y P7\_V1G1-P7 muestra algo interesante puesto que en el NIC 6 dispone de un buen GAD y está haciendo lo correcto el algoritmo lo pone en una clase NIC por lo tanto el grupo con la clasificación más adecuada está en el Grupo 3 con P7, el cual aparte tiene una efectividad del 100 % de datos bien clasificados como se obtiene de la tabla de la Figura A.7 siendo el grupo 3 el mayormente adecuado.

	1	2
1	'NIC1'	518x1 double
2	'P1'	48x1 double
3	'NIC2'	2312x1 dou...
4	'P2'	1172x1 dou...

	1	2
1	'NIC6'	3576x1 dou...
2	'P4'	15x1 double
3	'P6'	46x1 double
4	'NIC4'	0.5746
5	'NIC5'	412x1 double

	1	2
1	'P7'	4050x1 dou...

Figura A.7: Datos clasificados en P7 en los grupos de Picadura

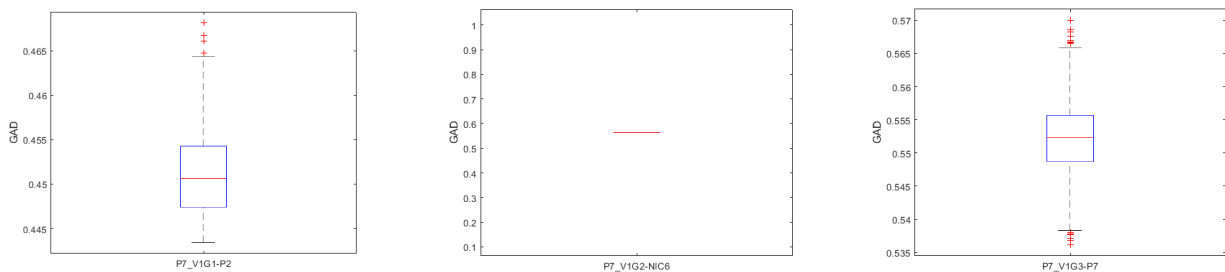
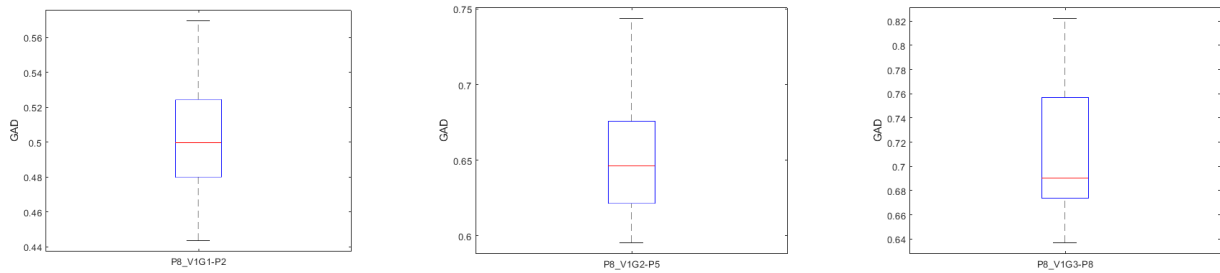


Figura A.8: Diagrama de caja de P7 con todos los grupos

### Modo de fallo por Picadura: Análisis de la clase P8 con los 3 grupos

Como observa la figura A.9 claramente en P8\_V1G3-P8 dispone del GAD más alto por lo tanto es la mejor opción, además la Figura A.10 confirma que dispone de las clases P8 y P9 dando a entre ver que existe un modo de fallo severo. En P8\_V1G3-P2 el GAD posee una precisión del 61.33% eso la descarta, y P8\_V1G3-P5 tiene un GAD alto aunque menor que P8\_V1G3-P8. Siendo el grupo 3 para este caso en particular el que más se acopla al análisis.



**Figura A.9:** Diagrama de caja de P8 con todos los grupos

	1	2
1	'P2'	2484x1 dou...
2	'NIC2'	106x1 double
3	'P3'	1460x1 dou...

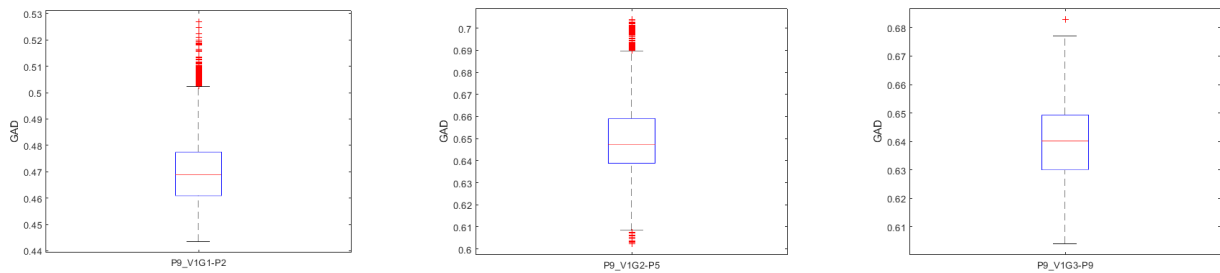
	1	2
1	'P5'	4050x1 dou...

	1	2
1	'P9'	1929x1 dou...
2	'P8'	1953x1 dou...
3	'NIC9'	168x1 double

**Figura A.10:** Datos clasificados en P8 en los grupos de Picadura

## Modo de fallo por Picadura: Análisis de la clase P9 con los 3 grupos

Como observa en la figura A.11 donde P9\_V1G1-P2 y P9\_V1G2-P5 muestran datos dispersos y además crearon todas las clases en sus respectivos grupos por ende no están bien definidas y se las descarta, entonces la clasificación adecuada se da en P9\_V1G3-P9 puesto dispone de datos muy bien organizados y GAD alto, hay que mencionar que P8 y P9 tienden a tener una cantidad de datos similares como se muestra en la Figura A.12 con una precisión cercana al 50% en ambas, son fallos severos lejos de ser un problema esto es visto como positivo debido a que nos da la certeza del fallo existente, siendo así el grupo 3 en donde se encuentra el posible fallo.



**Figura A.11:** Diagrama de caja de P9 con todos los grupos

	1	2
1	'NIC1'	0.3977
2	'P2'	2541x1 dou...
3	'P3'	652x1 double
4	'NIC2'	798x1 double
5	'P1'	58x1 double

	1	2
1	'NIC6'	27x1 double
2	'P6'	[0.5642;0.59...
3	'P4'	0.6007
4	'P1'	39x1 double
5	'P5'	3981x1 dou...

	1	2
1	'NIC7'	25x1 double
2	'P7'	[0.5194;0.54...
3	'NIC1'	0.5699
4	'P1'	[0.5840;0.60...
5	'NIC9'	48x1 double
6	'P9'	2146x1 dou...
7	'P8'	1825x1 dou...

**Figura A.12:** Datos clasificados en P9 en los grupos de Picadura

## B. Códigos de MatLab utilizados para el desarrollo del algoritmo para la clasificación de datos

```
% Este programa nos sirve para extraer los archivos .mat de una base de
% datos para utilizar solo los que requerimos ya sean filas o columnas
% Este programa los hicimos para P1 a P9
% A1_CRACK
tic
clear
clc
%% P1
a=1; %Representa la R(1:15)
b=1; %Representa la F(1:3)
c=1; %Representa la L(1:3)
i=1;
while a<=15
    while b<=3
        while c<=3
            j=['H:\DB_007 - Spur gear faults severity\5. Unpatched_Raw_Data_DB_007V0\Unpatched_Raw_Data_crack_DB_007V0\P1\R' num2str(a) 'F' num2str(b) 'L' num2str(c)
            %num2str (variable) se utiliza para convertir en caracter
            load(j); % para cargar la tabla
            P1(i,:)=data.Analog50k.TDMSAnalog50kAcelerometer0(:,1); %para almacenar
            c=c+1;
            i=i+1;
        end
        b=b+1;
        c=1;
    end
    a=a+1;
    b=1;
    c=1;
end
%% P2
a=1; %Representa la R(1:15)
b=1; %Representa la F(1:3)
c=1; %Representa la L(1:3)
i=1;
while a<=15
    while b<=3
        while c<=3
            j=['H:\DB_007 - Spur gear faults severity\5. Unpatched_Raw_Data_DB_007V0\Unpatched_Raw_Data_crack_DB_007V0\P2\R' num2str(a) 'F' num2str(b) 'L' num2str(c)
            %num2str (variable) se utiliza para convertir en caracter
            load(j); % para cargar la tabla
            P2(i,:)=data.Analog50k.TDMSAnalog50kAcelerometer0(:,1); %para almacenar
            c=c+1;
            i=i+1;
        end
        b=b+1;
        c=1;
    end
    a=a+1;
    b=1;
    c=1;
end
%% P3
a=1; %Representa la R(1:15)
b=1; %Representa la F(1:3)
c=1; %Representa la L(1:3)
i=1;
while a<=15
    while b<=3
        while c<=3
            j=['H:\DB_007 - Spur gear faults severity\5. Unpatched_Raw_Data_DB_007V0\Unpatched_Raw_Data_crack_DB_007V0\P3\R' num2str(a) 'F' num2str(b) 'L' num2str(c)
            %num2str (variable) se utiliza para convertir en caracter
            load(j); % para cargar la tabla
            P3(i,:)=data.Analog50k.TDMSAnalog50kAcelerometer0(:,1); %para almacenar
            c=c+1;
            i=i+1;
        end
    end
end
```

```

        end
        b=b+1;
        c=1;
    end
    a=a+1;
    b=1;
    c=1;
end
%% P4
a=1; %Representa la R(1:15)
b=1; %Representa la F(1:3)
c=1; %Representa la L(1:3)
i=1;
while a<=15
    while b<=3
        while c<=3
            j=['H:\DB_007 - Spur gear faults severity\5. Unpatched_Raw_Data_DB_007V0\Unpatched_Raw_Data_crack_DB_007V0\P4\R' num2str(a) 'F' num2str(b) 'L' num2str(c)
            %num2str (variable) se utiliza para convertir en caracter
            load(j); % para cargar la tabla
            P4(i,:)=data.Analog50k.TDMSAnalog50kAcelerometer0(:,1); %para almacenar
            c=c+1;
            i=i+1;
        end
        b=b+1;
        c=1;
    end
    a=a+1;
    b=1;
    c=1;
end
%% P5
a=1; %Representa la R(1:15)
b=1; %Representa la F(1:3)
c=1; %Representa la L(1:3)
i=1;
while a<=15
    while b<=3
        while c<=3
            j=['H:\DB_007 - Spur gear faults severity\5. Unpatched_Raw_Data_DB_007V0\Unpatched_Raw_Data_crack_DB_007V0\P5\R' num2str(a) 'F' num2str(b) 'L' num2str(c)
            %num2str (variable) se utiliza para convertir en caracter
            load(j); % para cargar la tabla
            P5(i,:)=data.Analog50k.TDMSAnalog50kAcelerometer0(:,1); %para almacenar
            c=c+1;
            i=i+1;
        end
        b=b+1;
        c=1;
    end
    a=a+1;
    b=1;
    c=1;
end
%% P6
a=1; %Representa la R(1:15)
b=1; %Representa la F(1:3)
c=1; %Representa la L(1:3)
i=1;
while a<=15
    while b<=3
        while c<=3
            j=['H:\DB_007 - Spur gear faults severity\5. Unpatched_Raw_Data_DB_007V0\Unpatched_Raw_Data_crack_DB_007V0\P6\R' num2str(a) 'F' num2str(b) 'L' num2str(c)
            %num2str (variable) se utiliza para convertir en caracter
            load(j); % para cargar la tabla
            P6(i,:)=data.Analog50k.TDMSAnalog50kAcelerometer0(:,1); %para almacenar
            c=c+1;
            i=i+1;
        end
        b=b+1;
        c=1;
    end
    a=a+1;
end

```

```

        b=1;
        c=1;
    end
    %% P7
    a=1; %Representa la R(1:15)
    b=1; %Representa la F(1:3)
    c=1; %Representa la L(1:3)
    i=1;
    while a<=15
        while b<=3
            while c<=3
                j=['H:\DB_007 - Spur gear faults severity\5. Unpatched_Raw_Data_DB_007V0\Unpatched_Raw_Data_crack_DB_007V0\P7\R' num2str(a) 'F' num2str(b) 'L' num2str(c)
                %num2str (variable) se utiliza para convertir en caracter
                load(j); % para cargar la tabla
                P7(i,:)=data.Analog50k.TDMSAnalog50kAccelerometer0(:,1); %para almacenar
                c=c+1;
                i=i+1;
            end
            b=b+1;
            c=1;
        end
        a=a+1;
        b=1;
        c=1;
    end
    %% P8
    a=1; %Representa la R(1:15)
    b=1; %Representa la F(1:3)
    c=1; %Representa la L(1:3)
    i=1;
    while a<=15
        while b<=3
            while c<=3
                j=['H:\DB_007 - Spur gear faults severity\5. Unpatched_Raw_Data_DB_007V0\Unpatched_Raw_Data_crack_DB_007V0\P8\R' num2str(a) 'F' num2str(b) 'L' num2str(c)
                %num2str (variable) se utiliza para convertir en caracter
                load(j); % para cargar la tabla
                P8(i,:)=data.Analog50k.TDMSAnalog50kAccelerometer0(:,1); %para almacenar
                c=c+1;
                i=i+1;
            end
            b=b+1;
            c=1;
        end
        a=a+1;
        b=1;
        c=1;
    end
    %% P9
    a=1; %Representa la R(1:15)
    b=1; %Representa la F(1:3)
    c=1; %Representa la L(1:3)
    i=1;
    while a<=15
        while b<=3
            while c<=3
                j=['H:\DB_007 - Spur gear faults severity\5. Unpatched_Raw_Data_DB_007V0\Unpatched_Raw_Data_crack_DB_007V0\P9\R' num2str(a) 'F' num2str(b) 'L' num2str(c)
                %num2str (variable) se utiliza para convertir en caracter
                load(j); % para cargar la tabla
                P9(i,:)=data.Analog50k.TDMSAnalog50kAccelerometer0(:,1); %para almacenar
                c=c+1;
                i=i+1;
            end
            b=b+1;
            c=1;
        end
        a=a+1;
        b=1;
        c=1;
    end
end
toc
%%

```

```

save('A1_CRACK.mat','P1','P2','P3','P4','P5','P6','P7','P8','P9')

tic %ver el tiempo de ejecución del programa funciona con el toc al final
% 50000 Hz muestras por segundo [s]
% Obtuve un promedio de
% P1_V1=45
% P2_V1=64
% P3_V1=65
% P4_V1=64
% P9_V1=65

close all
clc
clear

load('A1PittingP1.mat','P1_V3');
P1=P1_V3;
[lagt,H]=obtener_lag(P1);
lag(1)=lagt;
C(1,:)=H;
toc
%%
tic
load('A1PittingP2.mat','P2_V3');
P1=P2_V3;
[lagt,H]=obtener_lag(P1);
lag(2)=lagt;
C(2,:)=H;
toc
%%
tic
load('A1PittingP3.mat','P3_V3');
P1=P3_V3;
[lagt,H]=obtener_lag(P1);
lag(3)=lagt;
C(3,:)=H;
toc
%%
tic
load('A1PittingP4.mat','P4_V3');
P1=P4_V3;
[lagt,H]=obtener_lag(P1);
lag(4)=lagt;
C(4,:)=H;
toc
%%
tic
load('A1PittingP5.mat','P5_V3');
P1=P5_V3;
[lagt,H]=obtener_lag(P1);
lag(5)=lagt;
C(5,:)=H;
toc
%%
tic
load('A1PittingP6.mat','P6_V3');
P1=P6_V3;
[lagt,H]=obtener_lag(P1);
lag(6)=lagt;
C(6,:)=H;
toc
%%
tic
load('A1PittingP7.mat','P7_V3');
P1=P7_V3;
[lagt,H]=obtener_lag(P1);
lag(7)=lagt;
C(7,:)=H;
toc
%%

```

```

tic
load('A1PittingP8.mat','P8_V3');
P1=P8_V3;
[lagt,H]=obtener_lag(P1);
lag(8)=lagt;
C(8,:)=H;
toc
%%
tic
load('A1PittingP9.mat','P9_V3');
P1=P9_V3;
[lagt,H]=obtener_lag(P1);
lag(9)=lagt;
C(9,:)=H;
toc

function [lagt,H]=obtener_lag(P1)
h=1; %local
%ht=1; %global
[m,n] = size(P1);
while h<=m
    x = P1(h,:); %Almacenar en el vector x
    [c,lags] = xcorr(x,'coeff'); % c es la correlación y lags el retraso
    X(h,:)=x; % Acumulación de las señales de vibración
    C(h,:)=c; % Acumulación de la correlación
    h=h+1;
    %ht=ht+1;
end
X(h,:)=mean(X,1); %promedio de las señales de vibración
prom=1;
H(prom,:)=mean(C,1); %esto para obtener en la última fila el promedio de las
% V_cero(h,:)=mean(V_cero,1); %correlaciones (se utiliza 1 y para las columnas el 2)
C(h,:)=mean(C,1);
figure
stem(lags,C(h,:))
axis([0 100 -1 1])
title('Autocorrelación Promedio')
xlabel('Lag')
ylabel('Correlación')
grid on
% fprintf('El promedio retraso con cruces por cero es: %1.2f \n', V_cero(h,2));
%Para determinar el retraso cuando la correlación es cero el promedio
q=1;
i=n;
while q==1
    if (sign(C(h,i))-sign(C(h,i+1)))&(lags(1,i)>0) %para determinar el cambio de signo
        q=0;
    else
        i=i+1;
    end
end
i=-(i-n);
fprintf('El retraso con la autocorrelación promedio es : %1.0f \n', i);
lagt(prom)=i;

%% cargas la base de datos con la clase correspondiente
%% P1
clc
clear
tic
load('A1_CRACK.mat','P1')
[P1V1,P1V2,P1V3] = dividirP_velocidades(P1);
[P1_V1,P1_V2,P1_V3] = obtener_ventanas(P1V1,P1V2,P1V3);
save('A1CrackP1.mat','P1_V1','P1_V2','P1_V3')
toc

function [P_V1,P_V2,P_V3] = obtener_ventanas(P1V1,P1V2,P1V3)
%% obtengo las 9 ventanas para las clases P con las 3 velocidades
r=1;
[m,n]=size(P1V1);

```



```

for i=1:m
    j=1;
    k=50000;%1s
    for q=1:90
        P_V1(r,:)=P1V1(i,j:k);
        P_V2(r,:)=P1V2(i,j:k);
        P_V3(r,:)=P1V3(i,j:k);
        j=j+5000;%0.1s
        k=k+5000;
        r=r+1;
    end
end

function [P1V1,P1V2,P1V3] = dividirP_velocidades(P1)
%% divide las clases P con sus 3 velocidades respectivas
i=1;
h=1;
[m,n]=size(P1);
while (i<=m)
    cont=1;
    while cont<=3
        P1V1(h,:)=P1(i,:);
        P1V2(h,:)=P1(i+3,:);
        P1V3(h,:)=P1(i+6,:);
        i=i+1;
        h=h+1;
        cont=cont+1;
    end
    i=i+6;
end

%% LAMDA-HAD para clasificación supervisada
% Algoritmo macro
% Entrada: Individual X
% Procedimiento:
% 1. Normalice los descriptores del individuo usando la Ec. (2).
% 2. Calcule la MAD para cada descriptor y cada clase con base en la función de densidad de probabilidad seleccionada descrita por las Ecs. (3) o (5).
% 3. Calcule el GAD de cada clase usando la t-norma y t-conorm con cualquiera de las ecuaciones que se muestran en la Tabla 1, y la Ec. (7).
% 4. Calcule el MGAD en cada clase mediante la ecuación. (9) de la Definición 1.
% 5. Utilice la ecuación. (10) en la Definición 2, para calcular de GADNIC para cada clase k.
% 6. Calcule el grado de adecuación ADGAD con la ecuación. (11) en la Definición 3, y el HAD como la suma de ellos con la Ec. (12) en la Definición 4.
% 7. Encuentre el índice de estimación EI de la clase a la que podría pertenecer el individuo con la ecuación. (13).
% 8. En la clase estimada, determine si el GAD máximo es mayor que el GADNIC correspondiente con la Definición 5.
% 9. Fin.
% Salida: índice de clase

clear
clc

%% 1. Normalizar los descriptores(atributos)
%Se caso cargamos la base de datos con los atributos normalizados [0 1]
%caso contrario se debe normalizar

tic
% load('DataV2_A1Break_tooth.mat')
% load('Selec_A1Break_tooth_V2.mat')

% load('Selec_A1Crack_V1.mat')
% load('DataV1_A1Crack.mat')

% load('DataV2_A1Pitting.mat')
% load('Selec_A1Pitting_V2.mat')

load('Selec_A1Scuffing_V3.mat')
load('DataV3_A1Scuffing.mat')

alfa=0.9;
w=1;
while (w<=5)
    j=1;
    while j<=3
        if (j==1)

```

```

        target_train=grpTrain_G1;
        pos=length(atr_escogidos_G1);
        data=dataTrain_G1;
    elseif (j==2)
        target_train=grpTrain_G2;
        pos=length(atr_escogidos_G2);
        data=dataTrain_G2;
    else
        target_train=grpTrain_G3;
        pos=length(atr_escogidos_G3);
        data=dataTrain_G3;
    end
end
fila=1;
col_clase=1;
col_gad=2;
while (fila<=pos)
    if (j==1)
        AUX=atr_escogidos_G1(fila).Max_F1_extendido;
    elseif (j==2)
        AUX=atr_escogidos_G2(fila).Max_F1_extendido;
    else
        AUX=atr_escogidos_G3(fila).Max_F1_extendido;
    end
    data_train = data(:,[AUX]);
    % 2. Grado de adecuación marginal (MAD)
    [Pkj] = valor_medio(data_train,target_train);
    [MAD] = grado_MAD(data_train,target_train,Pkj);
    % 3. Grado de adecuación global (GAD)
    fprintf('Ingrese el valor de alfa entre [0 1]: ') %Valor de alfa
    %alfa = input('');
    %alfa=0.6;
    [Tnorm,Sconorm,GAD] = grado_GAD(data_train,target_train,MAD,alfa);
    % 4. Calcule el MGAD es el valor medio de los GAD de la clase p en una clase k
    [MGAD] = valor_medio(GAD,target_train);
    % 5. Calcular de GADNIC para cada clase k
    c = unique(target_train,'stable');
    k = length(c); % Cantidad de clases
    for i = 1:k
        GADNIC(i,1)=mean(MGAD(i,:));
    end
    % 6. Calcule el grado de adecuación ADGAD
    [ADGAD] = grado_MAD(GAD,target_train,MGAD);
    h = size(ADGAD);
    for at=1:h(1,1)
        for i = 1:k
            mat1 = ADGAD{at,1}(i,:);
            HAD(at,i)=sum(mat1);
        end
    end
    % 7. Índice de estimación EI de la clase a la que podría pertenecer el individuo
    for at=1:h(1,1)
        mat1 = HAD(at,:);
        [EI(at,2),EI(at,1)]=max(mat1);
    end
    % 8. En la clase estimada, determine si el GAD máximo es mayor que el GADNIC correspondiente
    [Clases_K,Clases_NIC] = crear_clases(target_train);
    for at=1:h(1,1)
        index(at,1)=max(GAD(at,EI(at,1)),GADNIC(EI(at,1)));
        if (index(at,1)==GAD(at,EI(at,1)))
            if(j==1)
                matri_final_G1{at,col_clase}=Clases_K{EI(at,1)};
                matri_final_G1{at,col_gad}=GAD(at,EI(at,1));
            elseif (j==2)
                matri_final_G2{at,col_clase}=Clases_K{EI(at,1)};
                matri_final_G2{at,col_gad}=GAD(at,EI(at,1));
            else
                matri_final_G3{at,col_clase}=Clases_K{EI(at,1)};
                matri_final_G3{at,col_gad}=GAD(at,EI(at,1));
            end
        end
    end
    else
        if(j==1)

```

```

                matri_final_G1{at,col_clase}=Clases_NIC{(EI(at,1))};
                matri_final_G1{at,col_gad}=GADNIC(EI(at,1));
            elseif (j==2)
                matri_final_G2{at,col_clase}=Clases_NIC{(EI(at,1))};
                matri_final_G2{at,col_gad}=GADNIC(EI(at,1));
            else
                matri_final_G3{at,col_clase}=Clases_NIC{(EI(at,1))};
                matri_final_G3{at,col_gad}=GADNIC(EI(at,1));
            end
        end
    end
    end
    col_clase=col_clase+2;
    col_gad=col_gad+2;
    Ba_Pkj{fila}=Pkj;
    Ba_MGAD{fila}=MGAD;
    Ba_GADNIC{fila}=GADNIC;
    fila=fila+1;
    clear Pkj MGAD GADNIC
end
B_Pkj{w,j}=Ba_Pkj;
B_MGAD{w,j}=Ba_MGAD;
B_GADNIC{w,j}=Ba_GADNIC;
clear Ba_Pkj Ba_MGAD Ba_GADNIC
j=j+1;
clear c EI GAD GADNIC HAD index MAD MGAD Pkj Sconorm Tnorm target_train
end
mat_final{w,1}=matri_final_G1;
mat_final{w,2}=matri_final_G2;
mat_final{w,3}=matri_final_G3;
w=w+1;
alfa=alfa-0.1;
end
%% Guardar la matriz final

%save('Matriz_final_A1Crack_V1.mat','matri_final_G1','matri_final_G2','matri_final_G3')
%save('Matriz_final_A1Break_tooth_V2_minmax.mat','mat_final')
save('BASE_A1Scuffing_V3.mat','B_Pkj','B_MGAD','B_GADNIC')
toc

%% Boxplot del GAD
clear
close all
clc
tic

% Dibujar con boxplotr
load('Test_A1Break_tooth_V1G1_minmax.mat')
load('Test_A1Break_tooth_V2G2_minmax.mat')
load('Test_A1Break_tooth_V1G3_minmax.mat')

%[mat] = new_precision(matri_final_G2(:,1));

[c1,num_c1,num_m1,ind1]=clase_box(matri_final_G1);
[c2,num_c2,num_m2,ind2]=clase_box(matri_final_G2);
[c3,num_c3,num_m3,ind3]=clase_box(matri_final_G3);

gad1=cell2mat(matri_final_G1(:,2));
figure
boxplot(gad1,matri_final_G1(:,1))
title('Modo de fallo Diente Roto con Grupo 1 y Velocidad 1')
xlabel('Clases')
ylabel('GAD')

gad2=cell2mat(matri_final_G2(:,2));
figure
boxplot(gad2,matri_final_G2(:,1))
title('Modo de fallo Diente Roto con Grupo 2 y Velocidad 2')
xlabel('Clases')
ylabel('GAD')

gad3=cell2mat(matri_final_G3(:,2));

```

```

figure
boxplot(gad3,matri_final_G3(:,1))
title('Modo de fallo Diente Roto con Grupo 3 y Velocidad 1')
xlabel('Clases')
ylabel('GAD')

j1=[c1,ind1];
j2=[c2,ind2];
j3=[c3,ind3];

toc
clc
clear
tic

%load('MATRIZ_test_A1Pitting_V1G2_minmax.mat')
% load('Test_A1Pitting_V2G3_minmax.mat')
% load('Test_A1Break_tooth_V1G3_minmax.mat')
% load('Test_A1Crack_V1G3.mat')
load('Test_A1Scuffing_V3G3_minmax.mat')

K=9;

% target_G1=[linspace(1,1,K)';linspace(2,2,K)';linspace(3,3,K)'];
% grpTrain_G1 = {};
% for i = 1:length(target_G1)
%     grpTrain_G1(i,1) = cellstr(strcat('P',num2str(target_G1(i))));
% end
% [mat] = new_precision(matri_final_G1(:,1));
% C{1,1}=confusionmat(grpTrain_G1,mat,'Order',{'P1','P2','P3','NIC1','NIC2','NIC3'});

% target_G2=[linspace(1,1,K)';linspace(4,4,K)';linspace(5,5,K)';linspace(6,6,K)'];
% grpTrain_G2 = {};
% for i = 1:length(target_G2)
%     grpTrain_G2(i,1) = cellstr(strcat('P',num2str(target_G2(i))));
% end
% [mat] = new_precision(matri_final_G2(:,1));
% C{1,1}=confusionmat(grpTrain_G2,mat,'Order',{'P1','P4','P5','P6','NIC1','NIC4','NIC5','NIC6'});

target_G3=[linspace(1,1,K)';linspace(7,7,K)';linspace(8,8,K)';linspace(9,9,K)'];
grpTrain_G3 = {};
for i = 1:length(target_G3)
    grpTrain_G3(i,1) = cellstr(strcat('P',num2str(target_G3(i))));
end
[mat] = new_precision(matri_final_G3(:,1));
C{1,1}=confusionmat(grpTrain_G3,mat,'Order',{'P1','P7','P8','P9','NIC1','NIC7','NIC8','NIC9'});

%clear mat

[m,n]=size(C);
e=sum(C{1,1},2);
for j=1:n
    for i=1:m
        t=length(C{i,j})/2;
        for k=1:t
            Porcentaje{i,j}(k,1)=(C{i,j}(k,k)*100)/e(1,1);
        end
        Precision{i,j}=mean(Porcentaje{i,j});
    end
end

%save('Confu_Matriz_final_A1Pitting_V1_minmax.mat','Conf','Porc','Prec');
toc

```

## Referencias

- Alligood, K. T., Sauer, T. D. & Yorke, J. A. (1997). An introduction to dynamical systems. *Chaos*. Springer, Berlin.
- Al-Najjar, B., Ingwald, A. & Kans, M. (2016). Maintenance in Real Estate and Manufacturing Industries: Differences, Problems, Needs and Potentials-Four Case Studies. *Proceedings of the 10th World Congress on Engineering Asset Management (WCEAM 2015)*, 13-27.
- AlShorman, O., Masadeh, M., Alkahtani, F. & AlShorman, A. (2020). A Review of Condition Monitoring and Fault Diagnosis and Detection of Rotating Machinery Based on Image Aspects. *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, 1-5.
- Aznar, A. F. (1996). *Modelo de cálculo a flexión de engranajes cilíndricos de perfil de evolvente* (Tesis doctoral). UNED. Universidad Nacional de Educación a Distancia (España).
- Banerjee, T. P. & Das, S. (2012). Multi-sensor data fusion using support vector machine for motor fault detection. *Information Sciences*, 217, 96-107.
- Baqqar, M., Ahmed, M. & Gu, F. (2011). Data mining for gearbox condition monitoring. *The 17th International Conference on Automation and Computing*, 138-142.
- Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*, 22(4), 469-483.
- Bayar, N., Darmoul, S., Hajri-Gabouj, S. & Pierreval, H. (2015). Fault detection, diagnosis and recovery using Artificial Immune Systems: A review. *Engineering Applications of Artificial Intelligence*, 46, 43-57.
- Brennan, M., Palaniswami, M. & Kamen, P. (2001). Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE transactions on biomedical engineering*, 48(11), 1342-1347.
- Campbell, F. C. (2012). *Fatigue and fracture: understanding the basics*. ASM International.
- Cerrada, M., Macancela, J.-C., Cabrera, D., Estupiñan, E., Sánchez, R.-V. & Medina, R. (2020). Reciprocating compressor multi-fault classification using symbolic dynamics and complex correlation measure. *Applied Sciences*, 10(7), 2512.
- Cerrada, M., Sánchez, R.-V., Li, C., Pacheco, F., Cabrera, D., Valente de Oliveira, J. & Vásquez, R. E. (2018). A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, 99, 169-196.
- Chaari, F., Zimroz, R., Bartelmus, W., Fakhfakh, T. & Haddar, M. (2012). Model based investigation on a two stages gearbox dynamics under non-stationary operations. *Condition Monitoring of Machinery in Non-Stationary Operations* (pp. 133-142). Springer.

- Committee, A. I. H. (2004). *ASM Handbook. Volume 9, Metallography and Microstructures: George F. Vander Voort*. ASM International.
- Dalpiaz, G. & Rivola, A. (1997). Condition monitoring and diagnostics in automatic machines: comparison of vibration analysis techniques. *Mechanical Systems and Signal Processing*, 11(1), 53-73.
- Devendiran, S. & Manivannan, K. (2016). Vibration based condition monitoring and fault diagnosis technologies for bearing and gear components-a review. *International Journal of Applied Engineering Research*, 11(6), 3966-3975.
- Doncescu, A., Aguilar-Martin, J. & Atine, J.-C. (2007). Image color segmentation using the fuzzy tree algorithm T-LAMDA. *Fuzzy Sets and Systems*, 158(3), 230-238.
- Duan, L., Xie, M., Wang, J. & Bai, T. (2018). Deep learning enabled intelligent fault diagnosis: Overview and applications. *Journal of Intelligent & Fuzzy Systems*, 35(5), 5771-5784.
- Feng, Z., Liang, M. & Chu, F. (2013). Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mechanical Systems and Signal Processing*, 38(1), 165-205.
- Fraser, A. M. & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, 33, 1134-1140. <https://doi.org/10.1103/PhysRevA.33.1134>
- Greń, B., Kępski, P. & Barszcz, T. (2012). Challenges in maintenance of vibration monitoring systems dedicated to underground mining machinery. *Diagnostyka*, 13-18.
- Gros, X. (1996). *NDT data fusion*. Elsevier.
- Halkidi, M. & Vazirgiannis, M. (2002). Clustering validity assessment using multi representatives. *Proceedings of the Hellenic Conference on Artificial Intelligence, SETN*, 237-249.
- Hang, G. & Min, Y. (2004). Data fusion in distributed multi-sensor system. *Geo-spatial Information science*, 7(3), 214-217.
- Henriquez, P., Alonso, J. B., Ferrer, M. A. & Travieso, C. M. (2013). Review of automatic fault diagnosis systems using audio and vibration signals. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(5), 642-652.
- Hernández, L. E. B. (1985). Las fallas en los engranajes. *Ingeniería e investigación*, (12), 40-52.
- Isermann, R. (2005). *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Science & Business Media.

- Jardine, A. K., Lin, D. & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7), 1483-1510.
- Jayaswal, P., Wadhvani, A. & Mulchandani, K. (2008). Machine fault signature analysis. *International Journal of Rotating Machinery*, 2008.
- Kan, M. S., Tan, A. C. & Mathew, J. (2015). A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mechanical Systems and Signal Processing*, 62, 1-20.
- Kar, C. & Mohanty, A. (2006). Monitoring gear vibrations through motor current signature analysis and wavelet transform. *Mechanical systems and signal processing*, 20(1), 158-187.
- Kempowsky, T., Aguilar, J., Subias, A. & Le Lann, M.-V. (2003). Classification tool based on interactivity between expertise and self-learning techniques. *IFAC Proceedings Volumes*, 36(5), 675-680.
- Kempowsky-Hamon, T. (2004). *Surveillance de procédés à base de méthodes de classification: conception d'un outil d'aide pour la détection et le diagnostic des défaillances* (Tesis doctoral). INSA de Toulouse.
- Kępski, P. & Barszcz, T. (2012). Validation of vibration signals for diagnostics of mining machinery. *Diagnostyka*, 25-30.
- Khalastchi, E., Kalech, M., Kaminka, G. A. & Lin, R. (2015). Online data-driven anomaly detection in autonomous robots. *Knowledge and Information Systems*, 43(3), 657-688.
- Khaleghi, B., Khamis, A., Karray, F. O. & Razavi, S. N. (2013). Corrigendum: Corrigendum to 'Multisensor data fusion: A review of the state-of-the-art'[Information Fusion 14 (1)(2013) 28-44]. *Information Fusion*, 14(4), 562.
- Kim, J., Ahn, Y. & Yeo, H. (2016). A comparative study of time-based maintenance and condition-based maintenance for optimal choice of maintenance policy. *Structure and Infrastructure Engineering*, 12(12), 1525-1536.
- Lei, Y., Lin, J., He, Z. & Kong, D. (2012). A method based on multi-sensor data fusion for fault detection of planetary gearboxes. *Sensors*, 12(2), 2005-2017.
- Lei, Y., Lin, J., He, Z. & Zuo, M. J. (2013). A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical systems and signal processing*, 35(1-2), 108-126.
- Li, C. J. & Limmer, J. (2000). Model-based condition index for tracking gear wear and fatigue damage. *Wear*, 241(1), 26-32.

- Liu, R., Yang, B., Zio, E. & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 108, 33-47.
- Liu, Z. & Zhang, L. (2020). A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. *Measurement*, 149, 107002.
- Medina, R. [Rubén], Alvarez, X., Jadán, D., Cerrada, M., Sánchez, R.-V. & Macancela, J. C. (2017). Poincaré plot features from vibration signal for gearbox fault diagnosis. *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, 1-6.
- Medina, R. [Rubén], Macancela, J. C., Lucero, P., Cabrera, D., Sánchez, R.-V. & Cerrada, M. (2020). Gear and bearing fault classification under different load and speed by using Poincaré plot features and SVM. *Journal of Intelligent Manufacturing*, 1-25.
- Mizumoto, M. (1989). Pictorial Representations of fuzzy connectives, Part I: Cases of t-norms, t-conorms and averaging operators. *Fuzzy sets and systems*, 31(2), 217-242.
- Mobley, R. K., Higgins, L. R. & J., W. D. (2014). *Maintenance engineering handbook*. McGraw-Hill Education.
- Morales, L., Aguilar, J., Chávez, D. & Isaza, C. (2020). LAMDA-HAD, an Extension to the LAMDA Classifier in the Context of Supervised Learning. *International Journal of Information Technology & Decision Making*, 19(01), 283-316. <https://doi.org/10.1142/S0219622019500457>
- Niu, G. (2017). *Data-driven technology for engineering systems health management*. Springer.
- Ott, R. L. & Longnecker, M. T. (2015). *An introduction to statistical methods and data analysis*. Cengage Learning.
- Park, J., Ha, J. M., Oh, H., Youn, B. D., Choi, J.-H. & Kim, N. H. (2016). Model-based fault diagnosis of a planetary gear: A novel approach using transmission error. *IEEE Transactions on Reliability*, 65(4), 1830-1841.
- Peng, H., Long, F. & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8), 1226-1238.
- Peña, M., Cerrada, M., Alvarez, X., Jadán, D., Lucero, P., Milton, B., Guamán, R. & Sánchez, R.-V. (2018). Feature engineering based on anova, cluster validity assessment and knn for fault diagnosis in bearings. *Journal of Intelligent & Fuzzy Systems*, 34(6), 3451-3462.
- Perzyk, M., Kochanski, A., Kozłowski, J., Soroczynski, A. & Biernacki, R. (2014). Comparison of data mining tools for significance analysis of process parameters in applications to process fault diagnosis. *Information Sciences*, 259, 380-392.



- Piskorski, J. & Guzik, P. (2007). Geometry of the Poincaré plot of RR intervals and its asymmetry in healthy adults. *Physiological measurement*, 28(3), 287.
- Raghav, M. S. & Sharma, R. B. (2021). A review on fault diagnosis and condition monitoring of gearboxes by using AE technique. *Archives of Computational Methods in Engineering*, 28(4), 2845-2859.
- Raheja, D., Llinas, J., Nagi, R. & Romanowski, C. (2006). Data fusion/data mining-based architecture for condition-based maintenance. *International Journal of Production Research*, 44(14), 2869-2887.
- Randall, R. B. & Monitoring, V.-b. C. (2011). Industrial, aerospace and automotive applications. *VIBRATION-BASED CONDITON MONITORING*. West Sussex, 13-20.
- Romero Lozano, D. S. & Giraldo Ramos, J. A. (2018). Análisis de falla de un engranaje de la caja de cambios de una motocicleta.
- Ruiz, F. A., Isaza, C. V., Agudelo, A. F. & Agudelo, J. R. (2017). A new criterion to validate and improve the classification process of LAMDA algorithm applied to diesel engines. *Engineering Applications of Artificial Intelligence*, 60, 117-127.
- Sánchez, R.-V. (2018). *Diagnóstico de fallos en cajas de engranajes con base en la fusión de datos de señales de vibración, corriente y emisión acústica* (Tesis doctoral). Escuela de Ingeniería, Universidad Pontificia Bolivariana.
- Sánchez, R.-V., Lucero, P., Vásquez, R. E., Cerrada, M., Macancela, J.-C. & Cabrera, D. (2018). Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and KNN. *Journal of Intelligent & Fuzzy Systems*, 34(6), 3463-3473.
- Sharma, N. & Saroha, K. (2015). Study of dimension reduction methodologies in data mining. *International Conference on Computing, Communication & Automation*, 133-137.
- Sharma, V. (2021). A Review on Vibration-Based Fault Diagnosis Techniques for Wind Turbine Gearboxes Operating Under Nonstationary Conditions. *Journal of The Institution of Engineers (India): Series C*, 1-17.
- Sharma, V. & Parey, A. (2017). Frequency domain averaging based experimental evaluation of gear fault without tachometer for fluctuating speed conditions. *Mechanical Systems and Signal Processing*, 85, 278-295.
- Stadhouders, H., van Houtum (GJJAN), Tan, T. & Elwany, A. (2011). *A framework for implementing condition based maintenance based on operational data*. Technische Universiteit Eindhoven.
- Takens, F. (1981). Detecting strange attractors in turbulence. *Dynamical systems and turbulence, Warwick 1980* (pp. 366-381). Springer.

- Tang, J., Alelyani, S. & Liu, H. (2014). Feature selection for classification: A review. *Data classification: Algorithms and applications*, 37.
- Tang, S., Yuan, S. & Zhu, Y. (2019). Deep learning-based intelligent fault diagnosis methods toward rotating machinery. *IEEE Access*, 8, 9335-9346.
- Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168-192.
- Totten, G. E. (2017). *Friction, Lubrication, and Wear Technology*. ASM International. <https://doi.org/10.31399/asm.hb.v18.9781627081924>
- Vachtsevanos, G. J. & Vachtsevanos, G. J. (2006). *Intelligent fault diagnosis and prognosis for engineering systems* (Vol. 456). Wiley Hoboken.
- Van der Maaten, L. & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).
- Wu, S.-T., Campos, S. P. & de Aguiar, M. A. (1998). Scientific visualization of Poincaré maps. *Computers & Graphics*, 22(2-3), 209-216.
- Wu, S. & Chow, T. W. (2004). Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density. *Pattern Recognition*, 37(2), 175-188.
- Wulpi, D. J. (2013). *Understanding how components fail*. ASM international.
- Xu, Y., Sun, Y., Wan, J., Liu, X. & Song, Z. (2017). Industrial big data for fault diagnosis: Taxonomy, review, and applications. *IEEE Access*, 5, 17368-17380.
- Yan, R., Gao, R. X. & Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal processing*, 96, 1-15.
- Yin, S., Ding, S. X., Xie, X. & Luo, H. (2014). A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, 61(11), 6418-6428.
- Zhang, J., Hong, L. & Singh Dhupia, J. (2012). Gear fault detection in planetary gearbox using stator current measurement of ac motors. *Dynamic Systems and Control Conference*, 45295, 673-680.
- Zhang, Z. (2014). Data mining approaches for intelligent condition-based maintenance: a framework of intelligent fault diagnosis and prognosis System (IFDPS).