

TESIS DOCTORAL

2017

DIAGNÓSTICO DE FALLOS EN CAJAS DE ENGRANAJES MEDIANTE LA APLICACIÓN DE DIFERENTES TÉCNICAS DE INTELIGENCIA ARTIFICIAL

Autor:

RENÉ VINICIO SÁNCHEZ LOJA

Ingeniero Mecánico por la Universidad Politécnica Salesiana

**PROGRAMA DE DOCTORADO EN
TECNOLOGÍAS INDUSTRIALES**

Director de Tesis: Dr. D. MARIANO ARTÉS GÓMEZ

TESIS DOCTORAL**2017****DIAGNÓSTICO DE FALLOS EN CAJAS DE
ENGRANAJES MEDIANTE LA APLICACIÓN DE
DIFERENTES TÉCNICAS DE INTELIGENCIA
ARTIFICIAL****Autor:****RENÉ VINICIO SÁNCHEZ LOJA**

Ingeniero Mecánico por la Universidad Politécnica Salesiana

**PROGRAMA DE DOCTORADO EN
TECNOLOGÍAS INDUSTRIALES****Director de Tesis: Dr. D. MARIANO ARTÉS GÓMEZ**

AGRADECIMIENTOS

A toda mi familia por su apoyo y comprensión, al Profesor Dr. Mariano Artés Gómez por todo el apoyo y sabiduría en saber guiar la presente tesis. A la Universidad Politécnica Salesiana sede Cuenca; de manera muy especial a todos los miembros del Grupo de Investigación y Desarrollo en Tecnologías Industriales (GIDTEC) por su apoyo en todo este proceso.

Madrid, mayo 2017.

Índice de contenidos

1.	INTRODUCCIÓN	1
1.1	Antecedentes.....	1
1.2	Objetivos del trabajo.....	3
1.3	Metodología.....	3
2.	DIAGNÓSTICO DE FALLOS EN ENGRANAJES BASADO EN EL ANÁLISIS DE SEÑALES DE VIBRACIÓN	6
2.1	Adquisición y acondicionamiento de la señal.....	6
2.2	Extracción de atributos.....	7
2.2.1	Análisis en el dominio del tiempo.....	7
2.2.2	Análisis en el dominio de la frecuencia	8
2.2.3	Análisis en el dominio tiempo – frecuencia.....	8
2.3	Selección y reducción de atributos.....	9
2.4	Clasificación	10
3.	CONCEPTOS TEÓRICOS	15
3.1	Cálculo del espectro con la FFT	15
3.2	Energía de la señal.....	16
3.3	Extracción de los atributos de condición en el dominio del tiempo	17
3.3.1	Media (<i>Mean Value, MV</i>).....	17
3.3.2	Varianza (<i>Variance, var</i>).....	18
3.3.3	Desviación estándar (<i>Standard Deviation, STD</i>).....	18
3.3.4	Raíz cuadrática media (<i>Root Mean Square, RMS</i>).....	18
3.3.5	Valor pico positivo máximo (<i>max</i>).....	19
3.3.6	Curtosis (<i>kurtosis, k</i>).....	19
3.3.7	Asimetría (<i>Skewness, s</i>)	19
3.3.8	Factor de cresta (<i>Crest Factor, CF</i>)	20
3.3.9	Operador de energía (<i>Energy Operator, EO</i>).....	20
3.3.10	Factor de forma (<i>Shape Factor, SF</i>)	21
3.3.11	Otros atributos de condición en el dominio del tiempo	21
3.4	Extracción de los atributos de condición en el dominio de la frecuencia.....	22
3.4.1	Frecuencia media (<i>Mean Frequency, MF</i>)	23
3.4.2	Varianza (<i>Variance, Varf</i>)	23
3.4.3	Asimetría (<i>Sf</i>).....	24

3.4.4	Curtosis (K_f).....	24
3.4.5	Frecuencia central (FC).....	24
3.4.6	Desviación estándar ($STDF$).....	24
3.4.7	Raíz media cuadrática ($RMSF$).....	24
3.4.8	Otros atributos de condición en el dominio de la frecuencia.....	25
3.4.9	División en sub-bandas de frecuencia.....	26
3.5	Extracción de los parámetros de condición en el dominio del tiempo-frecuencia... 27	
3.5.1	Descomposición Wavelet Packet.....	28
3.6	Métodos para selección y reducción de atributos.....	32
3.6.1	Selección de características por enfoque de aprendizaje supervisado.....	32
3.6.1.1	Bosques aleatorios.....	32
3.6.1.2	Análisis discriminante lineal.....	32
3.6.2	Selección de características por enfoque de aprendizaje no supervisado.....	33
3.6.2.1	Análisis del componente principal.....	33
3.6.2.2	Matriz de factorización no-negativa.....	34
3.6.2.3	Vecino más próximo.....	34
3.7	Redes neuronales artificiales.....	35
3.7.1	Redes neuronales artificiales clásicas.....	35
3.8	Algoritmos genéticos con redes neuronales artificiales.....	41
3.9	Redes neuronales convolucionales.....	42
3.10	Bosques aleatorios.....	43
3.11	Evaluación de los clasificadores.....	45
3.11.1	Medidas de desempeño en la clasificación.....	45
3.11.1.1	Matriz de confusión.....	45
3.11.1.2	Curva característica operativa del receptor.....	48
3.12	Pruebas estadísticas.....	49
3.12.1	Pruebas paramétricas.....	50
3.12.2	Pruebas no paramétricas.....	52
4.	EXPERIMENTACIÓN.....	56
4.1	Descripción del funcionamiento del banco.....	56
4.2	Adquisición de la señal de vibración.....	56
4.3	Configuración de la experimentación.....	58
4.3.1	Configuración del banco para pruebas con engranajes.....	58
4.3.2	Configuración de los parámetros del banco.....	58

4.3.3	Fallos simulados en la caja de engranajes.....	59
4.4	Planes experimentales	59
4.4.1	Plan experimental uno	59
4.4.2	Plan experimental dos.....	61
4.4.3	Plan experimental tres.....	63
4.4.4	Plan experimental cuatro	66
4.5	Bases de datos	67
4.5.1	Método para adquirir, procesar y almacenar las señales de vibración de las bases de datos	68
4.6	Bases de datos crudas.....	70
5.	IMPLEMENTACIÓN Y EVALUACIÓN DE PROPUESTAS DE SISTEMAS DE DIAGNÓSTICO DE FALLOS EN LAS CAJAS DE ENGRANAJES EMPLEANDO DIVERSAS TÉCNICAS DE INTELIGENCIA ARTIFICIAL.....	71
5.1	Propuesta general de los sistemas de diagnóstico.....	71
5.1.1	Adquisición y acondicionamiento de la señal de los sistemas de diagnóstico	71
5.1.2	Extracción de los atributos para los sistemas de diagnóstico	72
5.1.2.1	Extracción de atributos en el dominio de tiempo	72
5.1.2.2	Extracción de atributos en el dominio de la frecuencia.....	73
5.1.2.3	Extracción de atributos en el dominio del tiempo-frecuencia.....	74
5.1.2.4	Atributos extraídos de las bases de datos crudas	75
5.2	Sistema de diagnóstico uno (SD1)	75
5.2.1	Fase de selección y reducción de atributos.....	76
5.2.1.1	Eliminación de atributos altamente correlacionados	76
5.2.1.2	Selección de atributos por algoritmos genéticos (AG) y <i>random forest</i> (RF).....	77
5.2.1.3	Selección por análisis de componentes principales (PCA).....	79
5.2.1.4	Selección por Análisis de Discriminante Lineal (LDA).....	79
5.2.2	Fase de clasificación	79
5.2.3	Fase de evaluación.....	81
5.2.3.1	Gráficos Q-Q Normal	82
5.2.3.2	Pruebas paramétricas	83
5.2.3.3	Pruebas no paramétricas.....	83
5.2.3.3.1	Prueba de Friedman.....	84
5.2.3.3.2	Post-Hoc.....	84
5.2.3.3.3	Clasificación por Wilcoxon.....	85
5.2.3.3.4	Resultados de la evaluación de los clasificadores	85

5.2.4	Resultados del sistema de diagnóstico uno.....	85
5.3	Sistema de diagnóstico dos (SD2).....	86
5.3.1	Resultados del sistema de diagnóstico dos	89
5.4	Sistema de diagnóstico tres (SD3).....	90
5.4.1	Resultados del sistema de diagnóstico tres	93
6.	CONCLUSIONES Y TRABAJOS FUTUROS.....	95
6.1	Conclusiones	95
6.2	Trabajos futuros.....	96
	BIBLIOGRAFÍA.....	97

Índice de figuras

Figura 1. Esquema del proceso de diagnóstico de fallos en maquinaria basado en el análisis de señales de vibración.	2
Figura 2. Transformada Discreta de Fourier [146].....	15
Figura 3. Componentes del espectro de frecuencia.	22
Figura 4. Proceso de preparación para cómputo de indicadores en frecuencia.....	23
Figura 5. División del espectro en N sub-bandas iguales.....	26
Figura 6. Información detallada de la sub-banda.	27
Figura 7. Descomposición Wavelet Packet.	30
Figura 8. Análisis de discriminante lineal.	33
Figura 9. Componentes principales.	34
Figura 10. Red neuronal con propagación hacia adelante (FFNN).	36
Figura 11. Arquitectura de una FFNN [180].	36
Figura 12. Rede neuronal propagación en cascada (CFN).	37
Figura 13. Red neuronal de base radial (RBFN).	37
Figura 14. Diagrama de neurona de base radial [180].....	38
Figura 15. Arquitectura para RBFN.	38
Figura 16. Modelo de una red neuronal probabilística.....	39
Figura 17. Arquitectura simplificada de una PNN.	40
Figura 18. Mapas Auto-organizados (SOM).	41
Figura 19. Diagrama del uso de GA en redes neuronales.	42
Figura 20. Arquitectura de una red convolucional	43
Figura 21. Matriz de confusión clase binaria.	46
Figura 22. Curva ROC y AUC.	49
Figura 23. Grafico QQ-Normal.	50
Figura 24. Configuración para pruebas con engranajes.	56
Figura 25. Sistema de adquisición de señales de vibración a) acelerómetro b) tarjeta de adquisición c) chasis para el montaje de la tarjeta d) computador portátil.....	57
Figura 26. Configuración de la caja de engranajes. a) Configuración en dos etapas. b) Configuración para la primera etapa Z1 y Z2. c) Configuración para la segunda etapa Z3 y Z4.	58
Figura 27. a) Esquema de la caja de engranajes. b) Detalle de la parte interior de la caja de engranajes rectos y helicoidales.....	66

Figura 28. Proceso para adquirir, procesar y almacenar las señales de vibración.....	69
Figura 29. Fases generales de un sistema de diagnóstico.....	71
Figura 30. Proceso de extracción de atributos, las bases de datos crudas.....	72
Figura 31. Espectro completo de frecuencia de una señal de vibración.....	73
Figura 32. Proceso de extracción de atributos en el dominio de la frecuencia.....	74
Figura 33. Diagrama actividades de las fases del sistema de diagnóstico uno.....	77
Figura 34. Proceso de selección de atributos por AG y RF.....	78
Figura 35. Enfoque uno contra todos.	81
Figura 36. Diagrama de evaluación.....	82
Figura 37. Graficas Q-Q Normal para de los diferentes clasificadores utilizando DB_AG- RF. (a) FFNN, (b) FFPN, (c) RBFN-PNN, (d) OvA-FFPN, (e) SOM y (f) RF.....	83
Figura 38. Diagrama actividades de las fases del sistema de diagnóstico dos.	88
Figura 39. Diagrama actividades de las fases del sistema de diagnóstico tres.....	91
Figura 40. Exactitud del clasificador para las DB1 y DB2 en los diferentes dominios.	93
Figura 41. Exactitud del clasificador para las DB3 y DB4 en los diferentes dominios.	94

Índice de tablas

Tabla 1. Conjunto de Familias Wavelet.	31
Tabla 2. Características del acelerómetro.....	57
Tabla 3. Características de la tarjeta NI cDAQ-9234.....	57
Tabla 4. Características de los engranajes	58
Tabla 5. Configuración de los fallos para el plan experimental uno.....	60
Tabla 6. Configuración de los fallos para el plan experimental dos.....	62
Tabla 7. Configuración de los fallos para el plan experimental tres.	64
Tabla 8. Datos de la configuración de los fallos para la caja de engranajes	67
Tabla 9. Datos adicionales de la caja de engranajes.....	67
Tabla 10. Resumen de las bases de datos crudas.....	70
Tabla 11. Bases de datos crudas.....	72
Tabla 12. Bases de datos.....	75
Tabla 13. Parámetros de entrenamiento para los clasificadores.....	80
Tabla 14. Condiciones de normalidad usando la prueba de Shapiro Wilk ($\alpha = 0.05$), respecto a la exactitud.....	83
Tabla 15. Clasificación y p-valor de la prueba de Friedman.....	84
Tabla 16. Prueba de Holm y Shaffer, $\alpha=0.05$	84
Tabla 17. Clasificación de clasificadores por Wilcoxon.	85
Tabla 18. Clasificación de los clasificadores para diferentes métodos de selección....	85
Tabla 19. Exactitud de clasificación, para los conjuntos de datos con atributos normalizados.....	88
Tabla 20. Exactitud de clasificación datos no normalizados.....	89
Tabla 21. Exactitud de clasificación para los atributos extraídos normalizados y no normalizados.....	89
Tabla 22. Exactitud del clasificador para los subconjuntos en el dominio del tiempo..	92
Tabla 23. Exactitud del clasificador para los subconjuntos en el dominio de la frecuencia	92
Tabla 24. Exactitud del clasificador para los subconjuntos en el dominio tiempo- frecuencia.....	92
Tabla 25. Exactitud del clasificador para los subconjuntos en los tres dominios	92

Título

Diagnóstico de fallos en cajas de engranajes mediante la aplicación de diferentes técnicas de inteligencia artificial

Resumen

Las cajas de engranajes son vitales en la transmisión de movimiento en maquinaria industrial, un diagnóstico adecuado de las mismas es altamente demandado por la influencia en la economía de la empresa, al reducir costos operacionales, apoyar en las decisiones de mantenimiento y mejorar el nivel de seguridad.

En el presente trabajo se realizó el diagnóstico fallos en cajas de engranajes basados el análisis de señales de vibración mediante la aplicación de diferentes técnicas de inteligencia artificial. Para ello se establecieron cuatro bases de datos, tres bases de datos de señales de vibración se adquirieron en el laboratorio y una cuarta base de datos pública; dos bases fueron en engranajes rectos y una en engranajes helicoidales, la cuarta base de datos combina engranajes rectos y helicoidales. Posteriormente a cada señal de las bases de datos se le extrajo los atributos en los dominios del tiempo, frecuencia y tiempo-frecuencia; luego se desarrollaron tres sistemas de diagnóstico: el sistema uno, se evaluó mediante pruebas estadísticas el mejor clasificador entre cuatro redes neuronales y *random forest*, el sistema dos, se evaluó el mejor clasificador del sistema uno con el clasificador de máquinas de soporte vectorial y una red neuronal, el sistema tres, evaluó como los atributos y su dominio influye en el resultado del clasificador *random forest*.

Tras la ejecución de la pruebas en los sistemas de diagnóstico se determinó que *random forest* fue la técnica de inteligencia artificial que mejor desempeño tuvo para la clasificación de fallos en las cajas de engranajes

Palabras claves

Caja de engranajes, diagnóstico de fallos, señal de vibración, inteligencia artificial, monitoreo de la condición.

Title

Fault diagnosis of gearboxes using different artificial intelligence techniques

Abstract

Gearboxes are vital in the transmission of movement in industrial machinery, an adequate diagnosis is highly demanded by the influence in the economy of the company, to reduce operational costs, support maintenance decisions, and improve the safety level.

In the present work the diagnosis was made in gearbox failures based analysis of vibration signals through the application of different techniques of artificial intelligence. For this purpose, four databases were established, three databases of vibration signals were acquired in the laboratory and a fourth public database; two bases were in spur gears, and one in helical gears, the fourth database combines spur and helical gears. Subsequently to each signal of the databases the attributes were extracted in the domains of time, frequency and time-frequency; then three diagnostic systems were developed: system one, evaluated through statistical tests the best classifier between four neural networks and random forest, system two, evaluated the best classifier of system one with the classifier of vector support machines and a neural network, the system three evaluated as attributes and domain influences the result of the random forest classifier.

After the execution of the tests in the diagnostic systems it was determined that random forest was the technique of artificial intelligence that had better performance for the classification of faults in the gearboxes

Keywords

Gearbox, fault diagnosis, vibration signal, artificial intelligence, condition monitoring.

1. INTRODUCCIÓN

1.1 Antecedentes

La maquinaria rotativa, como ejes, rodamientos, engranajes, bombas, y otras, son parte fundamental de la maquinaria utilizada en la industria; a pesar de un buen diseño, estos elementos se deterioran con el paso del tiempo por ciertas condiciones de estrés y carga que se presentan de forma aleatoria [1], [2], en especial las cajas de engranajes que son las encargadas de transmitir potencia entre ejes en diferentes tipos de maquinaria industrial, por lo que su adecuado mantenimiento es importante para reducir los costes de producción y mejorar la calidad de los productos o servicios. Una de las medidas que se pueden adoptar para bajar el costo de producción consiste en reducir el costo de operación, incluyendo el costo de mantenimiento, que se reporta en algunas industrias como el segundo más alto o incluso el más alto de los costos de operación [3]–[6]. Además, el mantenimiento adecuado ayuda en la eficiencia energética[7], [8]. Como apoyo a las estrategias de mantenimiento surge el mantenimiento basado en la condición, entre cuyas principales características se pueden mencionar las siguientes: aviso previo del fallo inminente, mayor precisión en la predicción del fallo y ayuda en los procedimientos de diagnóstico, ya que se puede asociar el fallo de componentes específicos a través de los parámetros o atributos supervisados [3], [4], [9], [10]. Un programa de mantenimiento basado en la condición consta de tres pasos fundamentales: 1. Adquisición de datos; 2. Procesamiento de datos; 3. Toma de decisiones. La adquisición de datos es un paso fundamental para el monitoreo de la condición de la maquinaria, se debe recoger y almacenar la señal monitoreada; en el procesamiento de datos, la información obtenida se maneja y se analiza para una mejor comprensión e interpretación, incluyendo la validación de las señales de los sensores y los atributos de estimación o extracción; por último, en la fase de toma de decisiones, se deciden las acciones de mantenimiento basadas en los resultados de diagnóstico y/o de pronóstico de fallos que son dos aspectos muy importantes en un programa de mantenimiento basado en la condición [11].

En el mantenimiento preventivo de maquinaria rotativa, el diagnóstico de fallos es un proceso importante debido a que con él se pueden evitar daños graves si se realiza una detección temprana lo que puede impedir daños o paradas del sistema entero, en especial en las cajas de engranajes que son sistemas cerrados y en los que no se puede advertir fácilmente el fallo. El diagnóstico de una caja de engranajes mediante señales de vibración es el método

más comúnmente utilizado para la detección de fallos de maquinaria rotativa [12]. El proceso para el diagnóstico de fallos de maquinaria rotativa basado en el análisis de señales de vibración [13]–[16] se presenta en la Figura 1.



Figura 1. Esquema del proceso de diagnóstico de fallos en maquinaria basado en el análisis de señales de vibración.

En el proceso de diagnóstico, la etapa de clasificación, que también se denomina de reconocimiento de patrones, se realiza manualmente con base en la información obtenida en las etapas anteriores. Sin embargo, el reconocimiento manual de patrones requiere experiencia específica en el área de la aplicación del diagnóstico, lo que implica la necesidad de personal especializado. Por esta razón el reconocimiento automático de patrones es altamente deseable para una menor dependencia de personal especializado ya que la clasificación automática permitirá detectar e identificar fallos de forma que el personal poco cualificado pueda tomar las decisiones de mantenimiento sin la necesidad de un especialista. Es por esta razón por lo que existe una demanda creciente de técnicas de clasificación que sean automáticas y fiables [11] [17].

Para el diagnóstico se pueden aplicar diferentes clasificadores: redes neuronales artificiales [18]–[24], algoritmos genéticos [25]–[27], clasificadores basados en lógica difusa [28]–[30] y recientemente bosques aleatorios (*Random Forest, RF*) [31]–[34]. En el caso de diagnóstico de la condición de cajas de engranajes, el conocimiento de los fallos por lo general es ambiguo porque las relaciones entre síntomas y tipos de fallos no pueden ser fácilmente identificadas. Los principales motivos pueden explicarse de la siguiente manera: 1. Es difícil identificar los parámetros o atributos con los síntomas por lo cual todos los tipos de fallos no se pueden distinguir perfectamente. 2. En las primeras etapas de un fallo, en especial cuando es incipiente, los efectos del ruido de la señal son tan fuertes que los síntomas no son evidentes. 3. La maquinaria rotativa presenta respuestas dinámicas generadas por una amplia gama de posibles fuentes impulsivas, debido a que toda maquinaria rotativa trabaja acoplada a otras y pueden mezclarse las señales. 4. La información de fallo en algunos casos no es precisa o completa, debido al ruido de fondo o error de medición. 5. Los límites entre los diferentes modos de fallo se superponen o son ambiguos, por lo que son difíciles de distinguir [35],

[36]. Por estas razones, las técnicas de inteligencia artificial proporcionan herramientas para clasificar la información basadas en datos y han demostrado ser muy eficaces en el manejo de información compleja [25].

Por lo anteriormente indicado en el presente trabajo se abordará:

- La clasificación de diferentes tipos de fallos y en especial el fallo de rotura de diente en diferentes grados así como fallos incipientes.
- La selección de los parámetros de condición que permitan distinguir adecuadamente fallos incipientes y severos manteniendo la exactitud del clasificador.
- La evaluación de diferentes técnicas de inteligencia artificial para que la exactitud en la clasificación de fallos sea superior al 95%.

1.2 Objetivos del trabajo

Objetivo general

- Diagnosticar fallos en cajas de engranajes basados en el análisis de las señales de vibración mediante la aplicación de diferentes técnicas de inteligencia artificial.

Objetivos específicos

- Caracterizar los elementos necesarios para el diagnóstico de fallos en cajas de engranajes basado en el análisis de señales de vibración que empleen técnicas de inteligencia artificial.
- Construir tres bases de datos de señales de vibración de una caja de engranajes con diferentes tipos de fallos y en condiciones de laboratorio.
- Procesar las señales de vibración de la caja de engranajes con diferentes tipos de fallos en el dominio del tiempo, la frecuencia y tiempo-frecuencia.
- Implementar y evaluar una biblioteca de algoritmos para el diagnóstico de fallos en las cajas de engranajes basado en el análisis de señales de vibración procesadas empleando diversas técnicas de inteligencia artificial.

1.3 Metodología

Para desarrollar el trabajo se dividió en cuatro fases que se detallan a continuación:

Fase I.- Se inició con la revisión bibliográfica para caracterizar los elementos necesarios para el diagnóstico de fallos en cajas de engranajes que se basen en el análisis de señales de vibración y que empleen técnicas de inteligencia artificial para la clasificación.

Fase II.- Se establecieron cuatro bases de datos de señales de vibración crudas, tres bases de datos se levantaron en el laboratorio y la cuarta base de datos se obtuvo de una base de datos pública; de las tres bases de datos que se levantaron en el laboratorio, dos bases fueron en engranajes rectos y una en engranajes helicoidales, la cuarta base de datos combina engranajes rectos y helicoidales.

Fase III.- Las señales de vibración crudas obtenidas de cada una de las cuatro bases de datos se procesaron para que la señal pueda ser empleada en el clasificador; cada señal se procesó en los dominios del tiempo, frecuencia y tiempo-frecuencia; seguidamente se extrajeron los atributos de las señales y se procedió a seleccionar los atributos adecuados para ingresar al clasificador.

Fase IV.- Se diseñaron e implementaron tres algoritmos para el diagnóstico de fallos en cajas de engranajes; en la fase de clasificación se emplearon diferentes técnicas de inteligencia artificial como: redes neuronales, máquinas de soporte vectorial y *random forest*. Para evaluar el rendimiento de los clasificadores, se emplearon técnicas estadísticas y la base de datos uno, finalmente para evaluar si se podía generalizar los algoritmos se evaluaron los algoritmos con las cuatro bases de datos.

1.4 Organización del documento

La tesis está organizada en seis capítulos:

Capítulo uno, se presenta una introducción a la tesis, así como los objetivos generales y de manera sucinta se explica la metodología que se empleó para el desarrollo de la tesis.

En el capítulo dos, se desarrolla el estado del arte sobre los procesos de diagnóstico de fallos en engranajes; se orienta principalmente a presentar los elementos característicos de los sistemas de diagnóstico de fallos en cajas de engranajes en el contexto del mantenimiento predictivo.

En el capítulo tres, se presentan los conceptos teóricos de los elementos característicos de los sistemas de diagnóstico de fallos en cajas de engranajes.

En el capítulo cuatro, presenta como se desarrolló la experimentación para las diferentes pruebas realizadas en el laboratorio y los parámetros que se establecieron para cada experimentación.

En el capítulo cinco, se proponen y evalúan tres sistemas de diagnóstico de fallos en engranajes, para lo cual se valoran diferentes técnicas de clasificación en cada sistema y se presenta un análisis de resultados por cada sistema

Finalmente, el capítulo seis denominado conclusiones y trabajos futuros resume los principales resultados originales del trabajo y hacia donde se podrían direccionar los futuros trabajos en el diagnóstico de fallos.

2. DIAGNÓSTICO DE FALLOS EN ENGRANAJES BASADO EN EL ANÁLISIS DE SEÑALES DE VIBRACIÓN

El monitoreo de la condición y diagnóstico de máquinas es un tema de alto interés industrial para apoyar la productividad [38]–[40], tal es la importancia que se han establecido normas para su aplicación, algunas de ellas se indican a continuación: ISO 17359:2011 *Condition monitoring and diagnostics of machines - General guidelines*, ISO 18129:2015 *Condition monitoring and diagnostics of machines- Approaches for performance diagnosis*, ISO 2041 *Mechanical vibration, shock and condition monitoring – Vocabulary*, ISO 13372 *Condition monitoring and diagnostics of machines – Vocabulary*. Estas normas se centran en los procedimientos generales y requisitos que deben considerarse para crear un programa de monitoreo de la condición.

El diagnóstico de fallos es un proceso importante en el mantenimiento predictivo de las cajas de engranajes, debido a que se pueden evitar daños graves si la detección de fallos se la realiza en su fase inicial o, si se detectan fallos en funcionamiento, para realizar paradas programadas del sistema [2], [41]–[43]. El diagnóstico de un sistema de engranajes por el análisis de las señales de vibración es el método más comúnmente utilizado para la detección de fallos de los mismos [12], esto debido a que es posible determinar cuando la vibración sale de su patrón normal de comportamiento. Además de análisis de vibración para diagnóstico de fallos en engranajes se pueden utilizar otras técnicas como: emisiones acústicas [44]–[47], termografía [48], [49] análisis de aceites [50]–[53] corriente eléctrica [54]–[56] y audio [57]; todas estas técnicas se pueden utilizar en el monitoreo de la condición pero su instrumentación es costosa.

En la Figura 1 se presentó un esquema del proceso para el diagnóstico de fallos en engranajes basado en el análisis de señales de vibración, a continuación se presenta la revisión bibliográfica de cada uno de los pasos del proceso.

2.1 Adquisición y acondicionamiento de la señal

La señal de vibración de un sistema de engranajes se obtiene por medio de un acelerómetro, esta señal es acondicionada, amplificada y filtrada para posteriormente pasar a un convertidor analógico digital para ser procesada. Es importante que se consideren las normas para la ubicación del acelerómetro, como la norma ISO 10816-1:1995, para que el proceso tenga repetitividad y reproducibilidad cuando se aplique en otra máquina, también se deben seguir las indicaciones del fabricante del acelerómetro y del sistema de adquisición de datos. En esta fase es también importante el diseño del plan experimental para la toma de las señales

de vibración y garantizar una buena línea base de la condición de funcionamiento normal de la caja.

2.2 Extracción de atributos

Luego de la adquisición y acondicionamiento de la señal se hace necesario el tratamiento de la misma a fin de extraer sus rasgos y atributos principales para su análisis posterior. Existen numerosas técnicas de procesamiento disponibles que en función del dominio, se pueden clasificar en análisis en el dominio del tiempo, de la frecuencia y del tiempo – frecuencia [58]. Aunque los diferentes dominios de representación permiten analizar y observar comportamientos de las señales que se quieren analizar, comúnmente estas técnicas producen conjuntos de datos de alta dimensión que afectan el desarrollo computacional del proceso disminuyendo la velocidad de convergencia. Además, en algunos casos, los datos en el dominio de representación no son lo suficientemente discriminantes por lo que surge la necesidad de la extracción de atributos de condición a partir de las características de las señales de vibración. El proceso de calcular algunas medidas que se presentan en la señal se conoce como extracción de parámetros, características o atributos de condición.

Los atributos en los diferentes dominios deben ser sensibles al cambio de estado de la máquina, sin embargo, la extracción de atributos de condición es difícil debido a que las señales de vibración medidas en cualquier punto de la máquina a menudo contienen mucho ruido. Además, el conocimiento de diagnóstico es ambiguo porque las relaciones entre los síntomas y los tipos de fallos no pueden ser fácilmente identificadas, esto es debido a la complejidad de las condiciones de la maquinaria y a que el número de estados de fallo que se pueden identificar es muy grande; es difícil encontrar uno o varios atributos o indicadores de síntomas que puedan identificar perfectamente todos los fallos al mismo tiempo [59], por lo tanto una técnica efectiva de extracción de atributos de condición es muy importante y determina el éxito en el diagnóstico del fallo [60]. Las técnicas para la extracción de parámetros de condición en función de cada uno de los dominios de representación se presentan en [58]; los principales atributos de condición, también llamados indicadores de condición para cajas de engranajes, se presentan en [61]–[64]; en el capítulo 3 se presentará una descripción de cada uno de los indicadores que se emplearán en el presente trabajo.

2.2.1 Análisis en el dominio del tiempo

Los métodos de análisis en el dominio de tiempo tratan de analizar información de la amplitud y de la fase de la señal de vibración en el tiempo para detectar el fallo del sistema.

El dominio del tiempo es una perceptiva natural y facilita la interpretación física de la vibración mediante la estadística descriptiva de la señal. Es particularmente útil en el análisis de señales impulsivas de defectos de rodamientos y engranajes con características no estacionarias e impulsos transitorios cortos [61], [65]; los análisis temporales se han mostrado útiles, tanto en la detección como en la localización del daño y además son sensibles a la evolución de la degradación.

2.2.2 Análisis en el dominio de la frecuencia

Se refiere a realizar el análisis de las señales de vibración como una función de la frecuencia. La señal de vibración en el tiempo se transforma al dominio de la frecuencia mediante la aplicación de la transformada rápida de Fourier (*Fast Fourier Transform*, FFT), pero también se pueden aplicar otras técnicas como la transformada Hilbert, densidad de potencia espectral [66] y otras descritas en [58],[67]. La principal ventaja del análisis en el dominio de la frecuencia es que permite hacer un análisis espectral de los picos de frecuencia; esto a su vez permite detectar fallos que tienen una respuesta característica, además de que es posible percibir el comportamiento de la falla en el tiempo [68]. El análisis espectral es la técnica clásica en el diagnóstico de engranajes, ya que al comparar el espectro de una caja dañada con referencia al espectro de una caja de engranes en buenas condiciones, es posible detectar algunos fallos. Sin embargo, en sistemas complicados de engranajes es difícil identificar los fallos a partir del espectro debido al gran número de componentes implicados y que no hay variación en la amplitud del espectro cuando el fallo está en su etapa inicial [69].

Con respecto a la técnica más empleada para pasar al dominio de la frecuencia, que es la FFT, esta tiene desventajas, dado que no proporciona ninguna información acerca de la señal en el dominio del tiempo, lo cual genera un gran problema a la hora de evaluar señales no estacionarias, donde es particularmente beneficioso obtener una correlación entre el dominio de la frecuencia y del tiempo [70], [71].

2.2.3 Análisis en el dominio tiempo–frecuencia

Las técnicas de análisis de tiempo–frecuencia se han estudiado y aplicado al diagnóstico de fallos, debido a su capacidad de representar las señales en el dominio del tiempo y de la frecuencia, lo que permite el análisis de señales de vibración no estacionarias. Las condiciones no estacionarias se producen debido a que las cajas de engranajes a menudo operan bajo condiciones de carga y/o velocidad fluctuante durante el servicio o también a las características del material de los engranes. La técnica de diagnóstico de fallos se vuelve

más complicada cuando la caja de engranajes está sujeta a diversas condiciones de funcionamiento ya que el torque es la principal fuente de contribución a la energía a la señal de vibración [70], [71]. Técnicas como la transformada de Fourier en tiempo corto, *Winger-Ville distribution*, transformada Wavelet o NP4 [61] o *empirical mode decomposition* [72] han sido utilizadas para analizar en el dominio del tiempo-frecuencia. La técnica que más se ha empleado, y con gran éxito, en los últimos tiempos es la transformada de Wavelet (*Wavelet Transform*, WT) [70], [73] técnica con la que, combinada con diferentes técnicas de inteligencia artificial, se han obtenido buenos resultados para diagnosticar fallos en engranajes [74]–[77]. La utilización de la transformada Wavelet es compleja, debido a las múltiples maneras en que se puede aplicar y al elevado número de parámetros que es necesario definir para su uso [78], [79]. Con la WT se obtiene los coeficientes de correlación entre la señal y una función seleccionada denominada Wavelet madre, dependiendo del resultado de la escala y posición de esta función. Hay numerosos tipos de familias de Wavelets madre [78], [79], y en apartado 3.5, Extracción de los parámetros de condición en el dominio del tiempo-frecuencia, se explican las familias que se seleccionaron y las variaciones de la WT.

2.3 Selección y reducción de atributos

El conjunto de datos obtenidos en la extracción de atributos de condición son de alta dimensionalidad, en este conjunto de datos existen subconjuntos que contienen atributos irrelevantes o redundantes así como superiores; no todos son importantes para la clasificación. Si todos los parámetros se introducen directamente en un clasificador, harán que el proceso de clasificación sea lento y menos preciso [17]. Por lo tanto, para mejorar la exactitud de la clasificación y reducir la carga computacional del clasificador, se realiza el proceso de selección y/o reducción de los atributos de condición o también llamando reducción de la dimensión de datos, que consiste en eliminar atributos no deseados. Por tal motivo el atributo seleccionado debe ser lo suficientemente sensible al cambio de estado o condición del sistema de engranajes [80], [81]. Este procedimiento permite determinar qué atributos en los dominios del tiempo, la frecuencia y tiempo-frecuencia son las que aportan mayor información para realizar la clasificación. Adicionalmente los métodos de selección de atributos ofrecen los siguientes beneficios: facilitan la visualización y el entendimiento de los datos, reducen los requerimientos en medición y almacenamiento, y permiten mejorar la eficiencia del rendimiento de clasificación [82]. Por otra parte, la probabilidad de sobreentrenar un algoritmo de aprendizaje está relacionada de forma directa con el número de características, por lo cual utilizar algún método para la selección de atributos ayuda a

reducir el sobreajuste, además, se observa que el uso de selección de atributos ayuda a mejorar el rendimiento en la fase de entrenamiento de los algoritmos de aprendizaje [83].

Existen varios esquemas de selección de atributos y cada uno de ellos tiene sus ventajas y limitaciones, los esquemas de selección de características más comunes son los que están basados en estadística, en algoritmos clásicos y en inteligencia artificial [84].

Para realizar el proceso de selección de atributos se utilizan métodos de selección supervisados y no-supervisados; entre los métodos con enfoque de aprendizaje supervisado se pueden citar: el método de bosques aleatorios (*Random Forest, RF*) [85] y el método del análisis del discriminante lineal (*Linear Discriminant Analysis, LDA*) [86]; y como métodos de selección con enfoque de aprendizaje no supervisado están: el método del análisis de componente principal (*Principal Component Analysis, PCA*) [87], matriz de factorización no-negativa (*Non-negative Matrix Factorization, NMF*) [88] y vecino más próximo (*Most neighbors first MNF*) [89]. De los métodos indicados anteriormente el método de RF realiza la eliminación de los atributos a partir de los datos de entrada y se considera como método de selección [90], mientras que los otros métodos usan los datos de entrada para crear nuevos atributos, proceso que se puede conocer como proceso de reducción de atributos de alta dimensionalidad [91], [92].

2.4 Clasificación

El diagnóstico de fallos incluye tareas como la detección, el aislamiento y la identificación del fallo para tratar de relacionar los síntomas con el fallo [92], [93]. Los métodos de diagnóstico de fallos se pueden dividir en dos enfoques; el basado en modelos (*model-based*) y el basado en datos (*data-based methods* o *data-driven*) [92], [94], [95]. El enfoque basado en modelos utiliza modelos matemáticos y un conocimiento *a priori* del sistema físico para describir el comportamiento del sistema en estudio, lo que es difícil de lograr en un ámbito industrial debido a que las condiciones de la maquinaria por lo general no son estacionarias lo cual limita obtener un modelo muy cercano a la realidad [95]. Debido a esta limitación, el modelo basado en datos es un método eficiente, ya que puede extraer directamente de enormes cantidades de datos la información necesaria para el resultado del diagnóstico [96], algunas de estas técnicas han sido reportadas en los siguientes trabajos [60], [97]–[99]. Para el resultado del diagnóstico es necesario el reconocimiento de patrones que es un proceso de mapeo de la información obtenida en el proceso de selección de atributos en donde se obtienen los patrones en condiciones normales y condiciones con fallos; tradicionalmente, el reconocimiento de patrones se realiza manualmente por herramientas como el espectro de

la señal, diagrama de fase, *cepstrum*, frecuencia de engranaje o armónicos, o el análisis de los atributos en algunos de los dominios analizados [11], [68], [100], [101].

La Inteligencia Artificial (*Artificial Intelligence*, AI), por definición, es la habilidad de las computadoras u otras máquinas para realizar actividades que requieren inteligencia humana, el tema ha atraído a los investigadores sobre sus teorías y principios desde la conferencia de 1956 en Dartmouth y sus campos actuales de aplicación son muy amplios [102]. La AI se emplea generalmente para reducir la carga de trabajo humano y el aprendizaje automático (*Machine learning*) es una de las muchas áreas de la AI, que en especial se usa para la clasificación de patrones. Es ampliamente utilizado en el análisis de datos, ya que es capaz de asignar comandos a partir de los resultados del análisis de datos mediante conjuntos de algoritmos [103]. Para la tarea de clasificación de fallos en cajas de engranajes se han utilizado técnicas como redes neuronales artificiales (*Artificial Neural Network*, ANN), que representan sistemas de procesamiento de información interconectando unidades de procesamiento simples llamadas neuronas. Cada neurona es una unidad de procesamiento independiente que transforma su entrada a través de una llamada función de activación. Hay tres características importantes de las Redes neuronales que las convierten en una herramienta adecuada para utilizarla para clasificación: capacidad de generalización, tolerancia al ruido y respuesta rápida una vez entrenadas [104]. En [19] se evalúan tres tipos de redes neuronales para cuatro condiciones de la caja (una normal y tres de fallo), con ocho diferentes velocidades constantes y carga constante. En este trabajo se hace un interesante análisis, en donde las entradas de la capa de entrada de las redes son los atributos estadísticos del tiempo, los atributos de la frecuencia con su espectro dividido en bandas, todos ellos obtenidos de la señal de vibración, y, además, se utilizan como datos de entrada a la red la señal del acelerómetro y la velocidad del motor lo que implica tener dos instrumentos. En [105], [106] se clasifica cinco y cuatro tipos de fallos, respectivamente, pero para la capa de entrada se utiliza la desviación estándar de *wavelet packet coefficients* y la señal de tacómetro; obteniendo el 100% de exactitud en la clasificación pero necesita dos instrumentos de entrada. En los trabajos [107], [108] se clasifican tres condiciones de la caja a diferentes a velocidades constantes y se introduce en la experimentación dos cargas, en el desarrollo de los algoritmos de diagnóstico se demuestra la importancia del proceso de selección de atributos para lo cual se emplean algoritmos genéticos y se evalúa en ANN y máquinas de soporte vectorial con los cuales se obtienen porcentajes de exactitud en la clasificación del 100%, esto solo con procesar la señal de entrada en el dominio del tiempo del acelerómetro. En [109] se emplean redes neuronales para la detección de tres condiciones de fallos para condiciones no

estacionarias de la señal de vibración. En [110]–[112] se clasifican tres fallos empleando atributos en el dominio del tiempo, en [74], [113]–[115] cuatro condiciones son clasificadas analizando la señal en dominio tiempo-frecuencia con fallos incipientes y con carga. En los trabajos anteriores experimentan varias configuraciones de las redes neuronales, en donde se verifica que la configuración de la red es un proceso empírico y se pueden obtener buenos resultados, además, en estos trabajos se emplea tan solo la señal de vibración.

Por otra parte, los algoritmos genéticos se pueden utilizar para fines de diagnóstico de fallos, en la mayoría de los casos se utiliza como una técnica de soporte, especialmente para tareas de configuración de parámetros de la red neuronal [104] o configuración de otras técnicas como máquinas de soporte vectorial [116], [117]. En [15][118]–[120] se clasifican hasta nueve condiciones de la caja, los atributos que ingresan a la red neuronal son seleccionados por medio de algoritmos genéticos lo que optimiza el número de entradas a la red y se obtiene exactitud en la clasificación empleando los atributos procesados de la señal del acelerómetro. Los algoritmos genéticos también se han empleado para determinar la arquitectura de red neural [26], [27] y tratar que el proceso de configuración empírico sea el más optimizado en cuanto a la exactitud de la clasificación.

Para reconocimiento de patrones con redes neuronales es importante disponer de un buen número de muestras, debido que en la práctica es difícil obtener muestras de fallos suficientes se ha introducido en el diagnóstico de fallos de las máquinas la técnica de máquinas de soporte vectorial (*Support Vector Machines*, SVM) basada en la teoría del aprendizaje estadístico ya que puede trabajar con un menor número de muestras y tiene mejor generalización que ANN [121], en trabajos como [43], [122]–[127] se ha comprobado su efectividad para el diagnóstico de fallos.

Además de la ANN y la SVM, se utiliza también la técnica de los bosques aleatorios (*Random Forest*, RF), método propuesto por Breiman en 2001 [128]; básicamente, RF contiene múltiples clasificadores de árbol de decisión y no hay conexión entre dos árboles de decisión, cuando se envían los datos cada árbol de decisión toma la decisión de clasificación y al final el resultado depende de la mayoría. RF tiene un excelente rendimiento en el campo de reconocimiento de patrones en diferentes campos, esto se debe a su capacidad de generalización y su alta exactitud en la clasificación [129]. También se menciona que puede trabajar con extensas bases de datos [130] y mayor números de patrones a reconocer [131].

En esta revisión se han presentado métodos de: extracción de atributos, selección y reducción de atributos y clasificadores, todos estos importantes al momento de implementar el sistema

de diagnóstico de fallos mediante el análisis de señales de vibración. A partir de la revisión se determinó lo siguiente.

- Si bien los trabajos comparan dos o tres métodos de clasificación y/o selección de atributos la mayoría de los trabajos lo hace analizando una sola base de datos y comparan la exactitud del clasificador [88], [132]–[136] quedando pendiente un análisis que relacione los clasificadores y su método de selección con varias bases de datos.
- Pocos trabajos analizan múltiples fallos[87], [137]–[140] y fallos incipientes [43], [74] en las cajas de engranajes.
- Se determinan hasta nueve condiciones de la caja de engranajes y por lo general con carga y velocidad constante.
- Los trabajos obtienen excelentes resultados en la clasificación pero no se ha revisado como los atributos en cada dominio contribuyen a la clasificación.

Además para la clasificación de fallos de cajas de engranajes, técnicas como en lógica difusa [37], [141], y ANFIS (*Adaptive Neuro-Fuzzy Inference System*) [99], [142]–[144] se han reportado con gran éxito en la detección de fallos. La lógica difusa, que combina variables de entrada, definidas en términos de conjuntos difusos por medio de grupos de reglas que producen uno o varios valores de salida, y ANFIS que consiste en utilizar un método de aprendizaje adaptativo, es decir, partiendo de un conjunto de datos de entrada/salida, se elabora un sistema de inferencia difuso. En [28], [145] se clasifican fallos basados en un enfoque neuro-fuzzy, se emplea este enfoque por sus propiedades de aprendizaje y su capacidad de expresar el sistema de clasificación por reglas, en donde las señales de vibración en el dominio de la frecuencia se toman como las entradas al sistema de clasificación, el sistema consta de varios sistemas neuro-fuzzy para determinar el estado del elemento. En [28] se presta una atención especial al análisis de las reglas obtenidas por el sistema neuro-fuzzy, de hecho, se discute un proceso de extracción e interpretación de las reglas, en donde se determinó la importancia de las reglas en la clasificación. Se recalca que el conjunto de reglas obtenidas por el algoritmo neuro-fuzzy de aprendizaje podría ser completado con reglas proporcionadas por un experto en forma de reglas difusas para hacer el sistema de clasificación más eficiente. En [2] se presenta un esquema inteligente para la detección de defectos incipientes en engranajes contruidos de polímero. En donde se han medido varias señales de vibración y se han tomado sus características tiempo-frecuencia como entradas al sistema de diagnóstico. En este trabajo, se ha desarrollado una propuesta que combina diferentes paradigmas de inteligencia artificial denominado neuro-fuzzy genético en donde es posible en la fase final

del proceso de aprendizaje expresar el sistema de diagnóstico de fallos como un conjunto de reglas difusas, adicionalmente el sistema propuesto introduce mecanismos para reducir el número de reglas en sus algoritmos de aprendizaje y su estructura permite operar con un gran número de entradas. Esta última característica es muy apreciada en los problemas de clasificación, donde se genera un gran número de entradas al sistema neuro-fuzzy genético.

3. CONCEPTOS TEÓRICOS

3.1 Cálculo del espectro con la FFT

La transformada discreta de Fourier (*Discrete Fourier transform*, DFT) es la herramienta matemática que puede pasar la información de una señal del dominio del tiempo al dominio de la frecuencia. La Ecuación 1 que representa la transformada y la Ecuación 2 la transformada inversa, contienen la misma información en representaciones distintas [146].

$$X(k) = \sum_{j=1}^N x(j)\omega_N^{(j-1)(k-1)} \quad \text{Ecuación 1}$$

$$x(j) = \frac{1}{N} \sum_{k=1}^N X(k)\omega_N^{-(j-1)(k-1)} \quad \text{Ecuación 2}$$

Donde: $\omega_N = e^{(2\pi i)/N}$

El número total de muestras de una señal discretizada en el dominio del tiempo está representado por la letra N y en el dominio de la frecuencia es N/2. La escala de la frecuencia se suele representar como una fracción de la frecuencia de muestreo, esto se puede apreciar de una manera gráfica en la Figura 2.

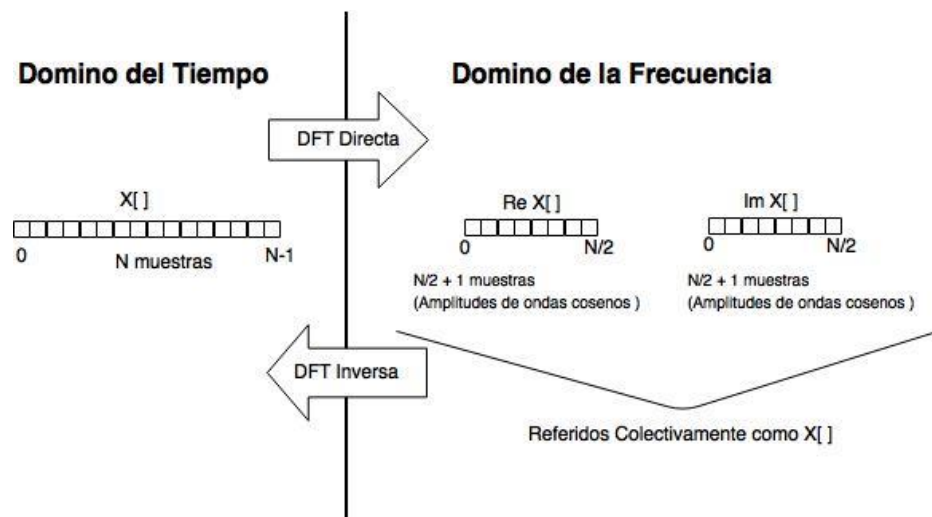


Figura 2. Transformada Discreta de Fourier [146]

Al aplicar la DFT se debe tener en cuenta lo siguiente, independientemente del número de muestras presentes en una señal discreta, el número de muestras en el espectro de frecuencia

se ve reducido a la mitad más uno. Una DFT de una señal con más muestras provee una mejor resolución en frecuencia, pero un mismo nivel de ruido.

La DFT representa la señal descompuesta en una serie de funciones sinusoidales, que al ser sumadas forman la señal original en el dominio del tiempo y puede calcularse de tres maneras diferentes:

- Con un conjunto de ecuaciones simultáneas, muy útil para entender la DFT pero ineficiente desde un punto de vista práctico
- Utilizando el principio de correlación que está basado en detectar una forma de onda conocida en otra señal
- Utilizando el algoritmo más eficiente que es la transformada rápida de Fourier, que descompone una DFT en n DFT's cada una con un simple punto. Esta llega a ser inclusive cientos de veces más rápida que la DFT y muy útil en aplicaciones de tiempo real.

Para calcular la DFT de un vector X usando un algoritmo de FFT se pueden utilizar diferentes paquetes de software, por ejemplo Matlab que dispone de una función (fft) adecuada.

3.2 Energía de la señal

Los términos de energía y potencia de una señal se usan para caracterizarla. La definición de energía y potencia hace referencia a cualquier señal $x(t)$. La energía de una señal en tiempo continuo viene definida por la siguiente expresión:

$$E = \int_{-\infty}^{+\infty} |x(t)|^2 dt, \quad \text{Ecuación 3}$$

Donde $|x(t)|^2$ representa la potencia instantánea de la señal. La energía se obtiene al evaluar la potencia instantánea a lo largo del tiempo. De una manera análoga, se puede obtener la energía en el dominio de la frecuencia transformando $x(t)$ en $x(f)$ mediante el uso de la FFT. La densidad de energía del espectro $E(f)$ se define en la ecuación 4.

$$E(f) = \int_{-\infty}^{+\infty} |x(f)|^2 dt \quad \text{Ecuación 4}$$

Para el caso de señales discretas, la energía se representa por la sumatoria de la potencia instantánea de cada muestra a lo largo del tiempo, como se indica en la ecuación 5.

$$E = \sum_{n=-\infty}^{\infty} |x[n]|^2 \quad \text{Ecuación 5}$$

El dominio del tiempo y el dominio de la frecuencia son representaciones equivalentes de la misma señal, por lo tanto ambas deben tener la misma energía. La Ecuación 6 se conoce como relación de Parseval.

$$\sum_{i=1}^N x_i^2 = \frac{2}{N} \sum_{k=0}^{N/2} |X_k|^2 \quad \text{Ecuación 6}$$

Donde x_i es la señal en el dominio del tiempo y X_k es el espectro de frecuencia modificado resultado de tomar la DFT de x_i , manteniendo la longitud N (en la notación se usa el tramo de muestras comprendidos entre 0 y N/2, por lo que se multiplica por 2 toda la ecuación). La energía es proporcional al cuadrado de la amplitud; la corrección 2/N se hace para considerar los valores RMS y no los valores pico.

3.3 Extracción de los atributos de condición en el dominio del tiempo

El análisis de señales en el dominio del tiempo encuentra su importancia en la información que se puede obtener a partir de la forma de onda que describe una señal. La inspección visual de la forma de onda permite discriminar si el origen de las distorsiones es mecánico o de otro tipo pero, además de una inspección visual de la forma de onda, existen indicadores basados en fundamentos matemáticos y estadísticos que se usan para reflejar las características de la señal que se denominan indicadores, parámetros o atributos de condición; dichos atributos aportan información puntual y mensurable de las diferentes características que existen dentro de las señales. Los indicadores que se describen a continuación comprenden un conjunto de medidas estadísticas y relaciones que se reportan en la revisión bibliográfica y que son consideradas relevantes, debido a su simplicidad de cálculo y se prestan para la interpretación.

3.3.1 Media (*Mean Value, MV*)

Con este valor se identifica el valor medio de la señal. El valor medio se define como la suma de todas las amplitudes de la señal x_i dividido para el número de muestras N.

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{Ecuación 7}$$

Donde N representa el número total de muestras y x_i la i-ésima muestra

La media es muy sensible a interferencias, o valores atípicos, que se obtienen en el proceso de adquisición de la señal.

3.3.2 Varianza (*Variance, var*)

Es una medida de dispersión, definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.

$$\sigma = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad \text{Ecuación 8}$$

3.3.3 Desviación estándar (*Standard Deviation, STD*)

Cuantifica la dispersión de la señal con respecto a la media, se basa en la potencia de la señal, y se calcula elevando al cuadrado cada una de las desviaciones con respecto a la media y obteniendo la raíz cuadrada:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad \text{Ecuación 9}$$

3.3.4 Raíz cuadrática media (*Root Mean Square, RMS*)

La aproximación más básica para la medición de defectos en el dominio del tiempo es usar la RMS. El valor RMS de una señal de vibración es una característica que mide el contenido de energía en la señal de vibración. Esta característica es muy efectiva en la detección de desequilibrio en maquinaria rotativa, pero esta no es bastante sensible para la detección de fallos incipientes; en particular, no proporciona información sobre el elemento que está defectuoso [58], [147]. De la definición de la RMS y de la Ecuación 10 se deduce que el valor no aumenta con los picos aislados en la señal. Por consiguiente, el parámetro no es sensible a un fallo incipiente del diente y su valor aumenta a medida que avanza el fallo del diente. Este parámetro es sensible a la carga y a los cambios de velocidad

en engranajes. El uso principal de este parámetro es para supervisar el nivel de vibración en general [64], [61]. Su expresión matemática es:

$$RMS = \sqrt{\frac{1}{N} \left[\sum_{i=1}^N (x_i)^2 \right]}$$

Ecuación 10

3.3.5 Valor pico positivo máximo (*max*)

Es el valor máximo de la señal en un intervalo de tiempo determinado [64]. No es totalmente confiable debido a que existen otros factores que pueden incrementar el nivel pico de la señal tales como: movimientos bruscos del sensor o interferencias cercanas [148], este parámetro por lo general no se utiliza solo. Se expresa por:

$$max = \max(x_i)$$

Ecuación 11

3.3.6 Curtosis (*kurtosis, k*)

La curtosis se define como el cuarto momento de la distribución y mide los picos o valles relativos de la distribución, comparada con una distribución normal. Proporciona una medida del tamaño de las colas de la distribución y se usa como un indicador de picos mayores en el conjunto de datos [147]. La curtosis describe la distribución como con picos o plana. Si una señal de vibración contiene picos agudos con un mayor valor, entonces su función de distribución será más nítida. Estos tipos de señales son los que se producen por una caja de engranajes con fallo, por lo tanto el valor de la curtosis será más alto para una caja de engranajes con fallo que para una caja en buenas condiciones [64]. La señal de un sistema de engranajes en buen estado se asocia con una distribución Gaussiana y tiene una curtosis alrededor de 3 [61]. La curtosis tiene la ventaja de que el valor calculado es independiente de las variaciones de carga y velocidad [148]. Su expresión matemática es:

$$k = \frac{N \sum_{i=1}^N (x_i - \mu)^4}{\left[\sum_{i=1}^N (x_i - \mu)^2 \right]^2}$$

Ecuación 12

3.3.7 Asimetría (*Skewness, s*)

La asimetría es una medida de la falta de simetría. Una distribución del conjunto de datos es simétrica si se ve igual a la izquierda y derecha del punto central [67]. Matemáticamente se describe como:

$$s = \frac{N \sum_{i=1}^N (x_i - \mu)^3}{\sigma^3}$$

3.3.8 Factor de cresta (*Crest Factor, CF*)

El factor de cresta se define como el valor pico de la señal dividida por la RMS. Cuando solamente un diente está dañado, no hay ningún cambio en el valor RMS, mientras que el valor de pico máximo si aumenta y, por lo tanto, el factor de cresta aumenta su valor. Este parámetro permite diagnosticar daños superficiales muy pequeños y es de uso frecuente para monitorizar la condición de la caja [149]. En condiciones de funcionamiento normal, el factor de cresta puede alcanzar entre 2 y 6; un valor superior a 6 generalmente se asocia con problemas; esta característica se utiliza para detectar cambios en el patrón de la señal debido a las fuentes de vibración impulsivas, como rotura de los dientes en un engranaje o un defecto en el exterior como la pista de un rodamiento. No obstante, el indicador factor de cresta, también llamado índice de impacto [148] no se considera una técnica muy sensible [147].

$$CF = \frac{\max}{RMS}$$

3.3.9 Operador de energía (*Energy Operator, EO*)

En un esfuerzo para reflejar la energía instantánea de las interacciones de flujo de vértice no lineales [150], James Teager observó que la energía para generar una señal sinusoidal sencilla varía con la amplitud y frecuencia [151]. Posteriormente, Kaiser postularía un método para calcular la energía de una señal; este método para calcular la energía instantánea de una señal toma el nombre de operador de energía Teager-Kaiser (*Teager-Kaiser Energy Operator TKEO*) como se describe en [152]. Este operador se usa como un rastreador de energía y se define con la siguiente ecuación:

$$\Psi_c[x(t)] = [\dot{x}(t)]^2 - x(t)\ddot{x}(t),$$

Donde $x(t)$ es la señal, $\dot{x}(t)$ y $\ddot{x}(t)$ son su primera y segunda derivada respectivamente. Para el caso discreto, el operador de energía Ψ_d en la Ecuación 16 puede ser aproximado mediante diferencias de tiempo como se ilustra en la Ecuación 17:

$$\Psi_d[x(n)] = x_n^2 - x_{n+1}x_{n-1} \quad \text{Ecuación 16}$$

Este operador presenta una propiedad local de la señal, puesto que abarca tres muestras adyacentes, lo que permite obtener una buena resolución. Para realizar el cómputo de este operador, se realiza un barrido a lo largo de toda la señal y posteriormente un promedio de todos los valores obtenidos. Además, el operador de energía se calcula como la curtosis normalizada de la señal resultante [153], [154]. Este parámetro puede cuantificar el impacto y se podría utilizar para aislar un fallo en particular [154]. Se expresa por:

Donde: $\Delta\bar{y} = \text{media de } \Delta y$

$$EO = \frac{N^2 \sum_{i=1}^N (\Delta y_i - \Delta\bar{y})^4}{[\sum_{i=1}^N (\Delta y_i - \Delta\bar{y})^2]^2} \quad \text{Ecuación 17}$$

$$\Delta y_i = x_{i+1}^2 - x_i^2$$

El operador de energía es un tipo de señal residual de la función de auto correlación [155], [156].

3.3.10 Factor de forma (*Shape Factor, SF*)

El factor de forma es el valor normalizado del nivel RMS mediante la media del valor de la señal. Este parámetro permite centralizar el valor RMS para reducir la imprecisión ocasionada por problemas en la adquisición.

$$SF = \frac{RMS}{\frac{1}{N} \sum_{i=1}^N |x_i|} \quad \text{Ecuación 18}$$

3.3.11 Otros atributos de condición en el dominio del tiempo

En los trabajos [157] y [17], se reportan también otros parámetros o atributos de condición (PC) con respecto al dominio del tiempo que se definen a continuación:

Media de valores absolutos

$$mean_{abs} = \overline{|x|} = \frac{1}{N} \sum_{i=1}^N |x_i| \quad \text{Ecuación 19}$$

Factor de impulso (*Impulse Factor, IF*)

$$IF = \frac{\max|x_i|}{\frac{1}{N} \sum_{i=1}^N |x_i|} \quad \text{Ecuación 20}$$

Factor de holgura (*Factor Clearance, FCl*)

$$FCl = \frac{\max(x_i)}{\frac{1}{N} \sum_{i=1}^N (x_i)^2} \quad \text{Ecuación 21}$$

3.4 Extracción de los atributos de condición en el dominio de la frecuencia

Para que una señal en el dominio del tiempo pueda migrar al dominio de la frecuencia es necesario el uso de la transformada de Fourier; la información contenida dentro de la señal es analizada a través del espectro, lo que permite observar diferentes componentes de frecuencia individuales que se encuentran inmersos. Los atributos o parámetros de condición se obtienen a partir del espectro de la señal de vibración; si los diferentes componentes individuales de frecuencia muestran un crecimiento en amplitud, entonces se puede decir que existe un problema que va en aumento, una visualización global de esto se puede apreciar en la Figura 3.

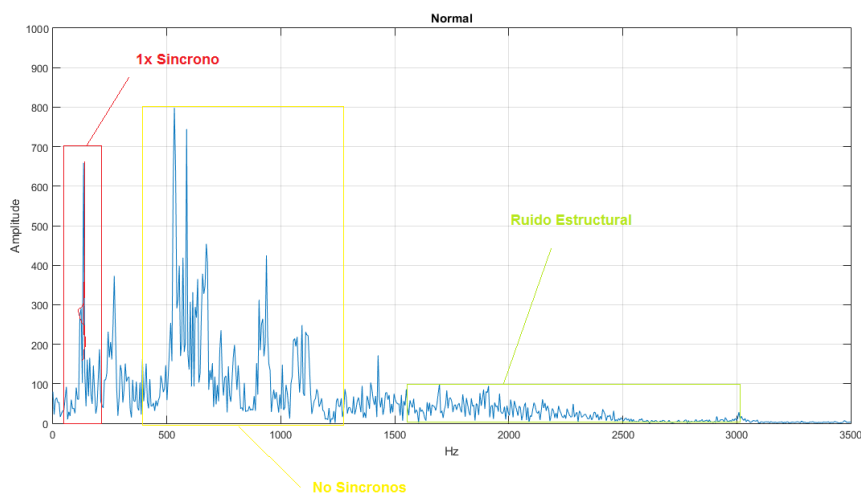


Figura 3. Componentes del espectro de frecuencia.

Para realizar el cómputo de los distintos atributos en el dominio de la frecuencia es necesario tener datos de una señal de vibración discretizada. A partir de esto se extrae un conjunto de características estadísticas donde $X(k)$ es el valor del espectro obtenido con la FFT, k es el número total de líneas o componentes de frecuencia que tiene el espectro, donde $k = 1, 2, \dots, K$ y f_k es la k -ésima componente de frecuencia. Para realizar esto, mediante el uso de Matlab se crea un vector con un número de términos igual al número de muestras que tiene la magnitud de la FFT y dicho vector contiene k partes correspondientes a componentes de espectro de frecuencia, la Figura 4 ilustra este proceso con la finalidad de enlazar el valor de la amplitud con su correspondencia en frecuencia real y para el posterior cómputo de todos los atributos de condición en el dominio de la frecuencia.

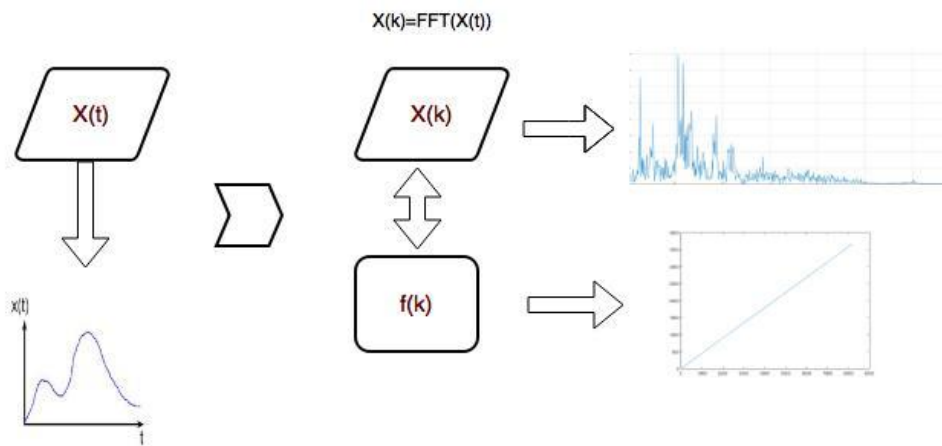


Figura 4. Proceso de preparación para cómputo de indicadores en frecuencia.

3.4.1 Frecuencia media (*Mean Frequency, MF*)

No es una característica muy representativa porque su valor es afectado por el ruido de la señal. Representa un promedio aritmético del espectro de vibración.

$$MF = \frac{\sum_{k=1}^K X(k)}{K}$$

Ecuación 22

3.4.2 Varianza (*Variance, Varf*)

Es un indicador de la variabilidad de la amplitud de la vibración en el espectro con respecto a la MF.

$$Varf = \frac{\sum_{k=1}^K (X(k) - MF)^2}{K - 1}$$

Ecuación 23

3.4.3 Asimetría (*Sf*)

Mide la estabilidad del espectro con respecto a la MF.

$$Sf = \frac{\sum_{k=1}^K (X(k) - MF)^3}{K(\sqrt{\text{Varf}})^3}$$

Ecuación 24

3.4.4 Curtosis (*Kf*)

Mide la estabilidad del espectro con respecto a la MF.

$$Kf = \frac{\sum_{k=1}^K (X(k) - MF)^4}{K(\text{Varf})^2}$$

Ecuación 25

3.4.5 Frecuencia central (*FC*)

Muestra los cambios de posición de las frecuencias principales y es útil para referenciar un rango de frecuencias.

$$FC = \frac{\sum_{k=1}^K f_k X(k)}{\sum_{k=1}^K X(k)}$$

Ecuación 26

3.4.6 Desviación estándar (*STDF*)

$$STDF = \sqrt{\frac{\sum_{k=1}^K (f_k - FC)^2 X(k)}{\sum_{k=1}^K X(k)}}$$

Ecuación 27

3.4.7 Raíz media cuadrática (*RMSF*)

La RMSF, al igual que la FC, muestra los cambios de posición de las frecuencias principales.

$$RMSF = \sqrt{\frac{\sum_{k=1}^K f_k^2 X(k)}{\sum_{k=1}^K X(k)}}$$

Ecuación 28

3.4.8 Otros atributos de condición en el dominio de la frecuencia

Además de los anteriores atributos en el dominio de la frecuencia, existen algunos reportados en la literatura como útiles y que revelan cierta información que no puede ser encontrada por otros atributos o en el dominio temporal. En [17], [157], [158], se pueden encontrar otros parámetros de condición (PC), los cuales pueden ser formados a partir de indicadores ya mencionados o independientes basados en momentos estadísticos.

$$PC\ 1 = \frac{\sum_{k=1}^K (f_k - FC)^3 X(k)}{STDF^2 K} \quad \text{Ecuación 29}$$

$$PC\ 2 = \sqrt{\frac{\sum_{k=1}^K f_k^4 X(k)}{\sum_{k=1}^K f_k^2 X(k)}} \quad \text{Ecuación 30}$$

$$PC\ 3 = \frac{\sum_{k=1}^K f_k^2 X(k)}{FC} \quad \text{Ecuación 31}$$

$$PC\ 4 = \frac{STDF}{\sqrt{\sum_{k=1}^K X(k) \sum_{k=1}^K f_k^4 X(k)}} \quad \text{Ecuación 32}$$

$$PC\ 5 = \frac{\sum_{k=1}^K (f_k - FC)^4 X(k)}{K (STDF)^4} \quad \text{Ecuación 33}$$

$$PC\ 6 = \frac{\sum_{k=1}^K (f_k - FC)^{\frac{1}{2}} X(k)}{K \sqrt{STDF}} \quad \text{Ecuación 34}$$

$$PC 7 = \frac{STDF}{FC}$$

Ecuación 35

$$PC 8 = \frac{\sum_{k=1}^K (f_k - FC)^3 X(k)}{K (STDF)^3}$$

Ecuación 36

3.4.9 División en sub-bandas de frecuencia

La información clave de las señales se encuentra comúnmente en la frecuencia, amplitud y fase. La FFT es la herramienta utilizada para extraer dicha información de las señales en la frecuencia. Se puede repetir la extracción de parámetros pertenecientes a la amplitud de las señales, como es el caso del valor RMS u otros, pero realizando dicho cálculo en un número establecido de sub-bandas del espectro total, para proporcionar más capacidad de diagnóstico [159]. En la literatura se encuentran diversos casos en donde se aplica una división en sub-bandas del espectro de frecuencia para la extracción de diferentes parámetros que sirvan de apoyo en el monitoreo de la condición, en [160] se pueden encontrar algunos ejemplos.

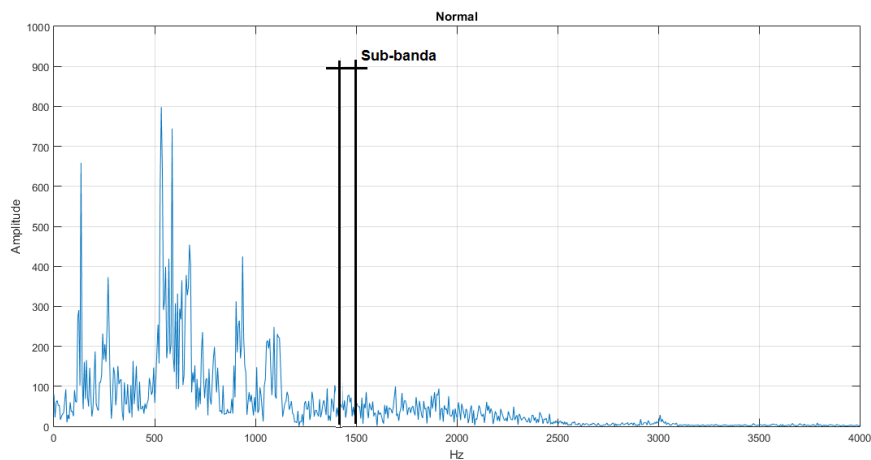


Figura 5. División del espectro en N sub-bandas iguales.

La elección del tamaño de la sub-banda es un factor crucial, para llevarla a cabo es posible basarse en la detección de patrones o tendencias dentro del espectro o en la identificación de frecuencias instantáneas en puntos de cresta, entre otros. Sin embargo, el procedimiento seguido en nuestro caso es la división del espectro en un número N de partes iguales de manera que cada segmento tenga un ancho de banda entero, como se indica en la Figura 5. Por otra parte, el ancho de banda de cada sub-banda debe ser también lo suficientemente pequeño para poder discriminar componentes espectrales consecutivos [161]. La elección

del número de partes N en el cual se divide el espectro es netamente empírico. La información contenida en una sub-banda Figura 6 puede apreciarse con más detalle que analizando todo el espectro.

Para realizar la operación de la división del espectro en N sub-bandas se aplicó el siguiente procedimiento:

- Se obtiene la longitud de la magnitud de la señal transformada al dominio de la frecuencia mediante la FFT (espectro) y ésta se divide en el número de bandas establecido de manera empírica. Para garantizar que el resultado de esta división sea un número entero, se realiza un redondeo hacia menos infinito; mediante esta operación se obtiene un segmento proporcional del espectro.
- Con el segmento obtenido, se realiza un barrido a través de todo el espectro completo, desplazando dicho segmento N veces según la cantidad de sub-bandas establecidas; en cada desplazamiento y/o iteración se realiza el cálculo del RMS en dicha banda y se almacena el valor obtenido.

Mediante este proceso se pueden obtener los valores de los atributos de condición necesarios en cada sub-banda en que se ha dividido el espectro.

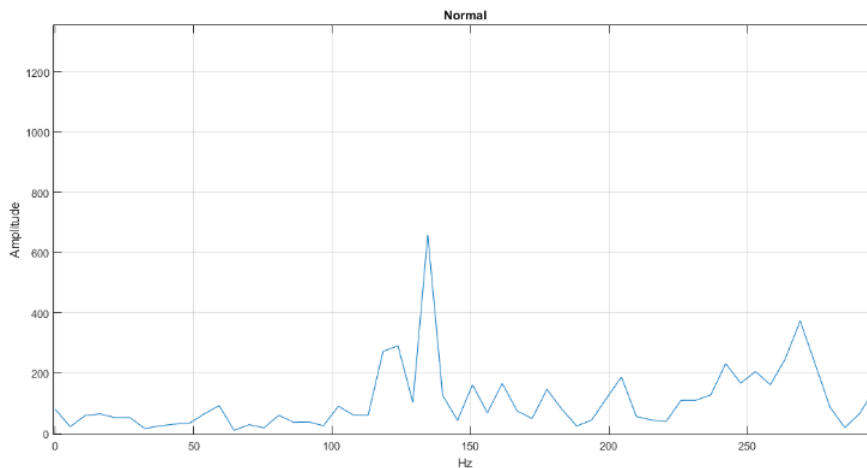


Figura 6. Información detallada de la sub-banda.

3.5 Extracción de los parámetros de condición en el dominio del tiempo-frecuencia

El análisis en tiempo-frecuencia (TF) surge a partir de la necesidad de representar cambios de condiciones de operación en las máquinas; en este espacio de representación, la meta principal es determinar la concentración de energía a lo largo del eje de frecuencia en un

instante de tiempo determinado [162], [163]. La transformación de TF en un caso ideal proveerá información directa acerca de los componentes de frecuencia que ocurren en cualquier tiempo dado mediante la combinación de información local de un espectro instantáneo de frecuencia con la información global del comportamiento temporal de la señal [163], [164]. En la literatura existen diferentes tipos de representaciones TF, por ejemplo: la distribución de Cohen, la transformada de Fourier de tiempo corto (STFT), el análisis cepstrum, entre otros [165]. Sin embargo, el uso de la transformada wavelet (TW) es aconsejable puesto que presentan un enfoque de representación TF a través de funciones derivadas de la traslación, modulación y escalamiento de una función base con un tiempo definido y frecuencia localizada [163]. Para una señal $x(t)$ la representación TF viene expresado por la siguiente ecuación:

$$TF_X(t, \omega) = \int_{-\infty}^{+\infty} x(\tau) \phi_{t,\omega}^* d\tau = (x, \phi_{t,\omega}), \quad \text{Ecuación 37}$$

Donde $\phi_{t,\omega}$ representa la función base. La transformada wavelet se utiliza debido a que entrega información en el dominio TF y de la cual se puede obtener la energía de la señal.

La TW es una transformación lineal cuyo patrón físico es el uso de una serie de funciones oscilantes con diferentes frecuencias como funciones de ventana para escanear y trasladar una señal $x(t)$ [75], [165]. Es por esta razón por lo que esta técnica se utiliza para el diagnóstico de fallos.

3.5.1 Descomposición Wavelet Packet

La TW descompone matemáticamente señales largas y complicadas en otras señales de menor dimensión, lo que facilita su interpretación [166]. La TW se clasifica en: transformada Wavelet continua (CWT), transformada Wavelet discreta (DWT) y transformada Wavelet Packet (WPT). Para una interpretación física, el módulo de la TW muestra como la energía de la señal varía con el tiempo y la frecuencia, en diversas aplicaciones [165], el cuadrado del módulo de la CWT es a menudo llamado escalograma.

La DWT puede ser implementada en términos de pares de filtros pasa-altos y filtros pasa-bajos denotados como $g(k)$ y $h(k)$ respectivamente, construidos a partir de una función wavelet seleccionada $\Psi(t)$ y su correspondiente función de escala $\phi(t)$ [73], [167], expresadas en las siguientes ecuaciones:

$$\phi(t) = \sqrt{2} \sum_k g(k)\phi(2t - k) \quad \text{Ecuación 38}$$

$$\Psi(t) = \sqrt{2} \sum_k h(k)\phi(2t - k) \quad \text{Ecuación 39}$$

Estos también son conocidos como filtros de espejo en cuadratura (*Quadrature Mirror Filters* QMF).

La WPT puede descomponer aún más la información detallada de la señal en la región de alta frecuencia [73]. Para realizar la WPT de una señal a un determinado nivel, las ecuaciones anteriores son unificadas como sigue:

$$u_{2n+1}(t) = \sqrt{2} \sum_k g(k)u_n(2t - k) \quad \text{Ecuación 40}$$

$$u_{2n}(t) = \sqrt{2} \sum_k h(k)u_n(2t - k) \quad \text{Ecuación 41}$$

Donde $u_0(t) = \phi(t)$ y $u_1(t) = \Psi(t)$. Consiguientemente la señal es descompuesta como se indica en las siguientes ecuaciones [73], [168]:

$$d_{j+1,2n+1} = \sum_m g(m - 2k)d_{j,n} \quad \text{Ecuación 42}$$

$$d_{j+1,2n} = \sum_m h(m - 2k)d_{j,n} \quad \text{Ecuación 43}$$

Donde $d_{j,n}$ denota los coeficientes wavelets en el j -ésimo nivel en la sub-banda n , $d_{j+1,2n}$ y $d_{j+1,2n+1}$ denotan los coeficientes en los niveles $j+1$, en las sub-bandas $2n$ y $2n+1$ respectivamente y m es el número de coeficientes wavelets.

La resolución tiempo-frecuencia de la TW depende de la frecuencia de la señal; a frecuencias elevadas, la TW alcanza una resolución temporal elevada pero una resolución de frecuencia baja; a bajas frecuencias, se puede obtener resolución alta en frecuencias y resolución

temporal baja; esta habilidad adaptativa es la que hace útil a la TW para el campo del diagnóstico de fallos [75], [165].

Dentro de la descomposición wavelet packet, WPT; una señal dada $X[n]$ se divide (se descompone) en la aproximación de baja frecuencia mediante el filtro $\Phi(t)$ y en un detalle de alta frecuencia usando el filtro $\Psi(t)$; esta división se realiza el número de veces que sea necesario según la aplicación. En la Figura 7 se presenta un diagrama de la descomposición y, como resultado de la descomposición, se obtienen los coeficientes wavelet en el último nivel.

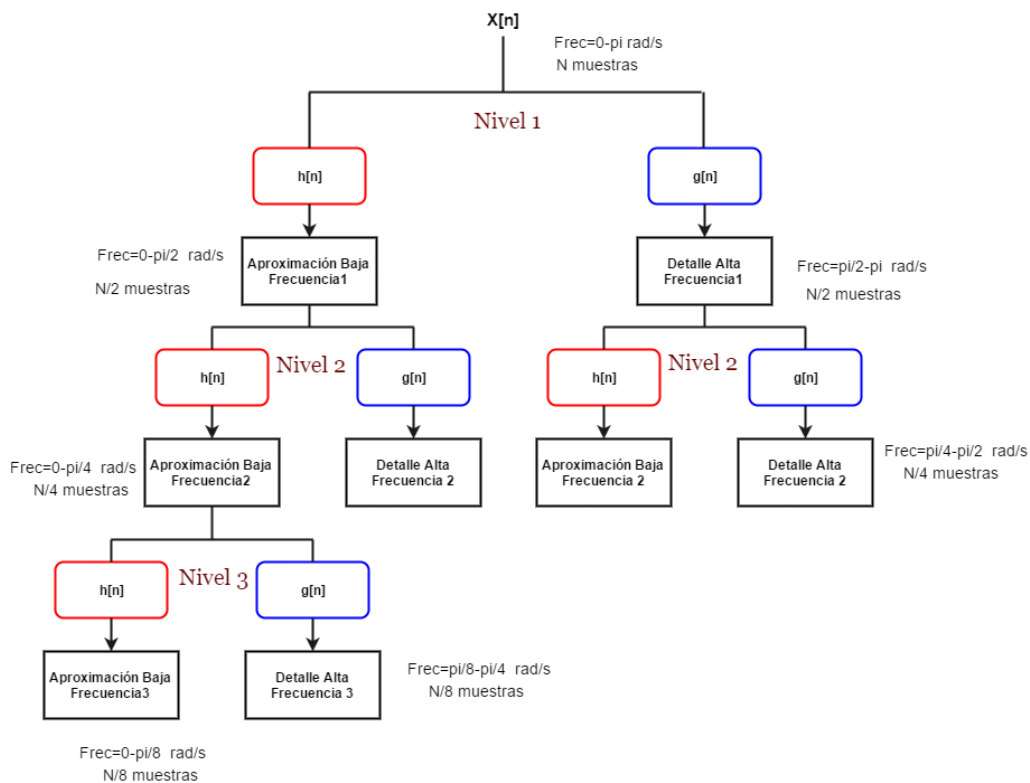


Figura 7. Descomposición Wavelet Packet.

Una TW puede enfocarse hacia estructuras de la señal localizadas mediante un procedimiento de “zoom” que reduce progresivamente el parámetro de escala, ya que usualmente las singularidades y estructuras irregulares contienen información esencial en una señal. Para detectar dichas singularidades no es suficiente seguir los módulos máximos wavelet a través de las escalas. Adaptar la resolución de la señal permite procesar únicamente los detalles relevantes para una tarea particular [161],[169]. La mayoría de las bases wavelets mejoran su capacidad para aproximar clases particulares de funciones con coeficientes wavelets diferentes de cero. Esto es válido tanto para la compresión de datos como para la eliminación de ruido [161].

Por lo expuesto anteriormente, la elección de las familias wavelet madre y los niveles que se usan dependen de la aplicación y de la resolución/descomposición que se desee obtener. En la Tabla 1 se presenta un resumen de las familias wavelets más representativas y usadas en la técnica por descomposición wavelet packet. El origen, la construcción y parámetros como los coeficientes de cada tipo de wavelet se explican en detalle en [170], [171].

Tabla 1. Conjunto de Familias Wavelet.

Familias Wavelet	Representación
Daubechies7	
Coiflet4	
Symlet	
Biort6.8	
Rbiort6.8	

El objetivo principal de la descomposición wavelet es obtener la energía presente en el último nivel para ser usado como parámetro de condición para el diagnóstico de fallos.

3.6 Métodos para selección y reducción de atributos

3.6.1 Selección de características por enfoque de aprendizaje supervisado

3.6.1.1 Bosques aleatorios

Los Bosques aleatorios (*Random Forest RF*) son una combinación de árboles predictores o árboles de decisión, de tal forma que cada árbol depende de un vector aleatorio muestreado independientemente y con la misma distribución para todos los árboles en el bosque [128]. El árbol de decisión divide el espacio de características dentro de regiones disjuntas que son asociadas a clases; cada nivel del árbol está construido usando las características que distinguen entre una clase y otra. Con el fin de construir el árbol, una métrica determina el grado de información contribuida por cada característica con respecto a las clases. Comúnmente la métrica utilizada para construir el árbol es la entropía. *RF* es un algoritmo basado en k árboles de decisión que son construidos y entrenados con versiones de muestras de arranque de los datos de entrenamiento originales. Entonces, dando una nueva entrada, la clase predictiva es obtenida a partir de un proceso de votación que es ejecutado sobre los procesos de cada árbol. *RF* es usado para obtener las características más representativas, debido a que cada árbol de decisión calcula el grado de información contribuido por cada característica de las clases. De esta forma en *RF*, la información de cada característica sobre todos los árboles puede ser promediada con el fin de obtener una categorización de las variables [85]. Los individuos fuera de bolsa (*out-of-bag*) pueden ser utilizados para estimar la importancia de atributos particulares. Un atributo recibe una puntuación de baja importancia si los valores de su permutación al azar no afectan a la capacidad de predicción del árbol, caso contrario recibe una puntuación de importancia alta [172].

3.6.1.2 Análisis discriminante lineal

El análisis discriminante lineal (*Linear Discriminant Analysis, LDA*) es un método supervisado empleado para la reducción de la dimensión de los datos [173]; consiste en determinar un sub-espacio de dimensión inferior en comparación con la dimensión de la muestra de datos original. Este método construye un menor número de características mediante la aplicación de una transformación lineal $G^T \in R^{m \times l}$ que mapea cada punto de datos $X = [x_1, \dots, x_n] \in$

$R^{m \times n}$ con m dimensiones a un menor espacio dimensional l usando la Ecuación 44. G^T se obtiene a través de una función objetivo que maximiza la distancia entre clases y minimiza la distancia dentro de las clases en el espacio dimensionalmente reducido, como se observa en la Figura 8. A menudo, el LDA para etiquetas de clase binarias se denomina Análisis Discriminante de Fisher (FDA, por sus siglas en inglés).

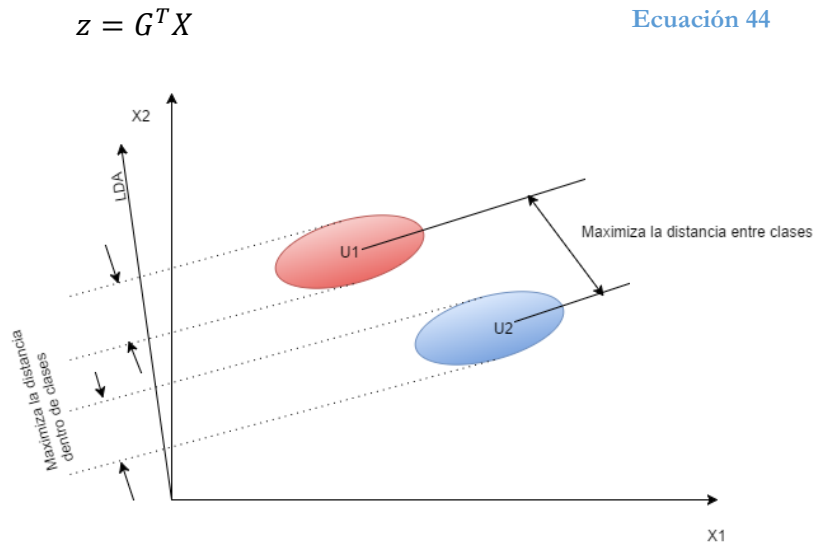


Figura 8. Análisis de discriminante lineal.

3.6.2 Selección de características por enfoque de aprendizaje no supervisado

3.6.2.1 Análisis del componente principal

El análisis del componente principal (*Principal Component Analysis*, PCA), trata de encontrar un espacio con una baja dimensión en el que se proyectan los datos, las así llamadas componentes principales son una combinación lineal de las variables originales y expresan la máxima variabilidad posible de los datos, además son independientes entre sí; estas nuevas componentes se ordenan de forma descendente con base a la variabilidad del conjunto de datos de entrada. Concretamente, PCA es una transformación lineal ortogonal que convierte un conjunto de datos $X \in R^{m \times n}$ en otro dentro de un nuevo espacio $z \in R^{k \times n}$. Generalmente para reducir datos de n -dimensiones dentro de un espacio de k -dimensiones, es necesario encontrar los vectores $U = [u_1, \dots, u_k]$ con el fin de proyectar los datos dentro de estos; estos son los vectores característicos de la matriz de covarianza de los datos normalizados [173]. En la Figura 9, se puede observar PC 1 (primera dimensión en el espacio PCA) que representa la mayor dispersión de los datos del espacio (x, y) , en cuanto a PC 2 esta representa el resto de variabilidad de los datos siendo menor a lo observado en PC 1.

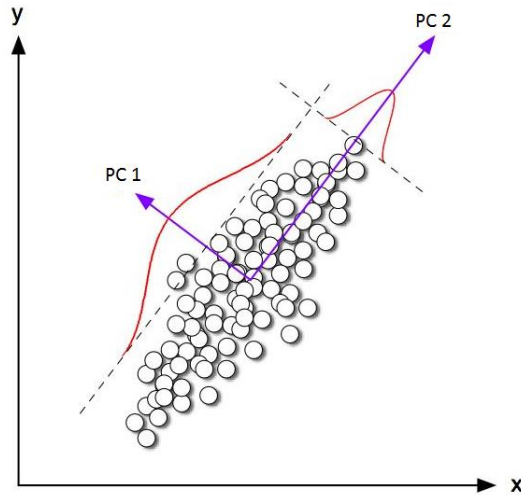


Figura 9. Componentes principales.

3.6.2.2 Matriz de factorización no-negativa

La matriz de factorización no-negativa (*Non-Negative Matrix Factorization* NMF) es un método de descomposición que representa una matriz no-negativa mediante dos matrices no-negativas de menor rango usando la Ecuación 45 reportada en [174].

$$\mathbf{V} = \mathbf{W}\mathbf{H} \quad \text{Ecuación 45}$$

Donde \mathbf{V} es la matriz no negativa con n características (columnas) y m muestras (filas), \mathbf{W} y \mathbf{H} son también matrices no negativas. El objetivo de la selección de características es encontrar matrices no-negativas $\mathbf{W} \in \mathbb{R}^{n \times r}$ y $\mathbf{H} \in \mathbb{R}^{r \times m}$, mediante la descomposición de la matriz $\mathbf{V} \in \mathbb{R}^{n \times m}$. Esto se realiza a través de un proceso de optimización que busca minimizar $F(\mathbf{V}, \mathbf{W}\mathbf{H})$, donde $F(\mathbf{V}, \mathbf{Z})$ es una función de costo que define la distancia entre \mathbf{V} y \mathbf{Z} se suele elegir la distancia euclidiana o la norma de Frobenius, donde $\mathbf{Z} = \mathbf{W}\mathbf{H}$. La dimensión de r es menor a n o m , por lo cual \mathbf{W} y \mathbf{H} son de menor dimensión que el original, siendo estos una versión comprimida de la matriz \mathbf{V} . Normalmente a \mathbf{W} se le conoce como matriz de pesos y a \mathbf{H} como matriz de características [174], [175].

3.6.2.3 Vecino más próximo

Vecino más próximo (*Most Neighbors First*, MNF) es un algoritmo no supervisado para selección de atributos basado en agrupación de los mismos, el objetivo es encontrar k agrupaciones que representan las características más significativas, donde una agrupación es un conjunto de atributos que son representados por un *medoide* (objeto para el cual la disimilitud promedio con todos los objetos en el conglomerado es mínima) como la

característica más importante [89]. Este proceso empieza seleccionando aleatoriamente k características como la agrupación inicial de *medoides*, luego, se calcula para cada característica una métrica de disimilitud, basada en el concepto de dependencia relativa, entre la característica de cada *medoide*; todas las características no-céntricas son asignadas a sus centroides más cercanos y se produce una reconfiguración de las agrupaciones, el proceso se repite hasta que se alcanzan los criterios de parada [176].

3.7 Redes neuronales artificiales

Las redes neuronales artificiales (*Artificial Neural Network*, ANN) son modelos computacionales inspirados en el sistema nervioso biológico. Al igual que un cerebro, una ANN se conforma por un gran número de neuronas interconectadas entre sí. Una red neuronal artificial está construida a base de capas, las cuales poseen diferentes números de neuronas, normalmente se diferencian tres capas; la capa de entrada, la capa oculta, que puede tener una o varias capas, y la capa de salida en donde se presentan los resultados del procesamiento. En una ANN el conocimiento no se almacena en instrucciones, el conocimiento de la red está en su topología y en la fuerza de conexión (sinapsis) entre las neuronas; el aprendizaje en una ANN se logra a través de algoritmos de entrenamiento particulares que mejor se desarrollan de acuerdo con las reglas de aprendizaje que tratan de emular los mecanismos de aprendizaje de un sistema biológico [177].

3.7.1 Redes neuronales artificiales clásicas

- a. Red neuronal con propagación hacia adelante (*Feed-forward Neural Network*, FFNN)

La FFNN es un tipo de red neuronal capaz de aproximar clases de funciones mediante funciones continuas e integrables. Las entradas $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ y los pesos \mathbf{w} , son utilizados por la FFNN para calcular la salida $\mathbf{y} = [y_1, y_2, \dots, y_m]^T$. En este tipo de redes, las entradas alimentan a las neuronas de la capa de entrada y las salidas de esta capa alimentan a las neuronas ocultas, y este proceso se repite hasta llegar a la capa de salida, Figura 10. Los valores de \mathbf{w} se calculan durante el proceso de entrenamiento supervisado de la red [178].

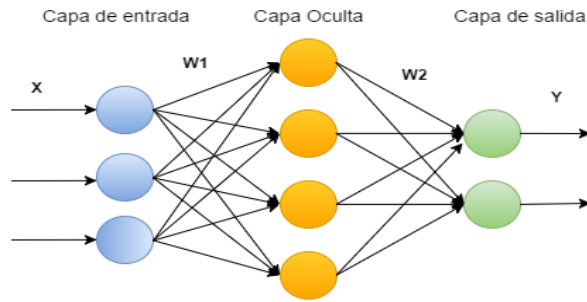


Figura 10. Red neuronal con propagación hacia adelante (FFNN).

Según Montana y Davis [179], esta red es de las más utilizadas por las siguientes razones: cuando se entrena con datos dispersos la red es capaz de entregar la salida correcta, y al utilizar el algoritmo de entrenamiento *backpropagation* es posible calcular los mejores valores de w .

En la Figura 11 se observa la arquitectura simplificada de las FFNNs, en la cual, la *tansig* (*hyperbolic tangent sigmoid transfer function*) es una de las funciones de activación más utilizadas en la capa oculta. Por otra parte, en la salida generalmente se utiliza una *purelin* (*linear transfer function*) si se trabaja como estimador o una *Log-Sigmoid* (*logistic-sigmoid transfer function*) si se trabaja como clasificador. Cabe mencionar que existen diversas funciones de activación que pueden ser utilizadas y que dependen de la ampliación de la red [180].

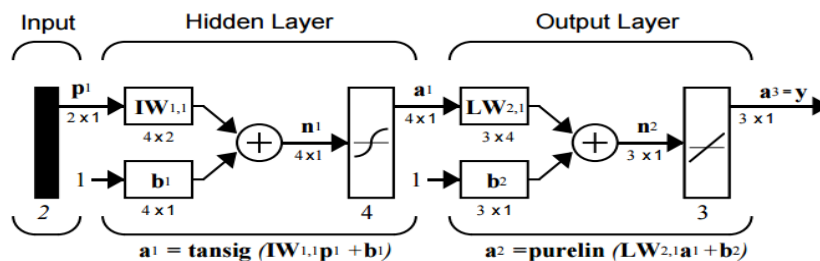


Figura 11. Arquitectura de una FFNN [180].

b. Red hacia adelante en cascada (*Cascade Forward Network, CFN*)

La CFN la cual es similar a las FFNN, pero incluye una de la entrada a cada capa y de cada capa a las capas sucesivas. En la Figura 12 se observa una red de tres capas que tiene conexiones desde la capa 1 a la capa 2, de la capa 2 a la capa 3, y de la capa 1 a la capa 3. La red de tres capas también tiene conexiones desde la entrada a las tres capas. Las conexiones adicionales podrían mejorar la velocidad a la que la red aprende la relación deseada utilizando

una función de activación como la *tansig* (tangente- sigmoide) [181]. Una red neuronal de este tipo puede tener varias capas de ocultas, aumentando la profundidad de la misma.

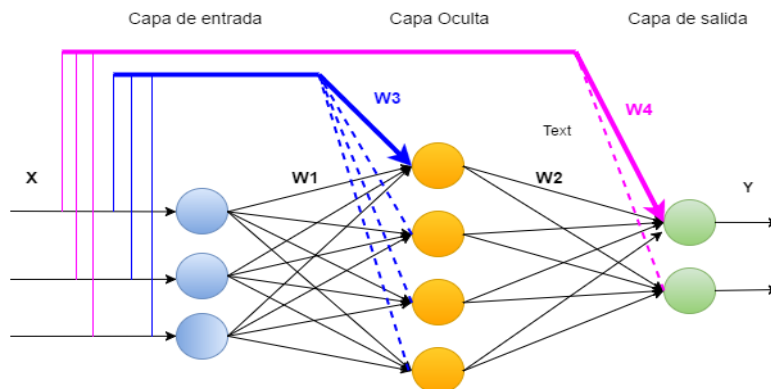


Figura 12. Red neuronal propagación en cascada (CFN).

c. Red de función de base radial (*Radial BasisFunction Network, RBFN*)

Es un tipo de red multicapa con alimentación hacia adelante, similar a las FFNN como se indica en la Figura 13. Tiene una característica distintiva, que es el hecho de poseer solo tres capas, la de entrada, la capa oculta (función de base radial $h(\mathbf{x})$) y la capa de salida (función lineal $f(\mathbf{x})$). Las neuronas en la capa de salida en una red de base radial simplemente realizan una combinación lineal de las activaciones de las neuronas ocultas [180].

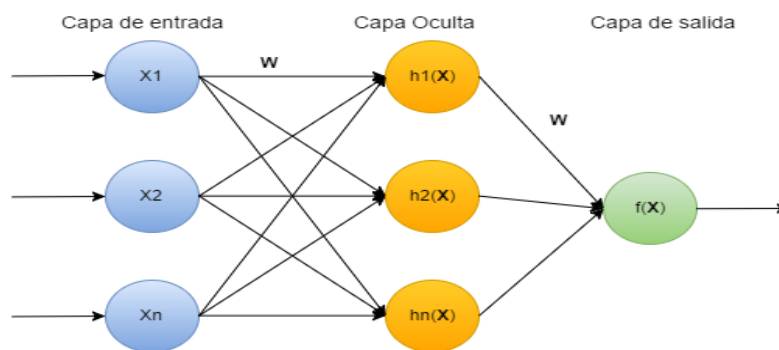


Figura 13. Red neuronal de base radial (RBFN).

Estas redes son diferentes a las redes vistas anteriormente; en la Figura 14 se observa una neurona de base radial, aquí la entrada de la red a la función de transferencia de base radial (*radbas*) es el vector de distancia entre el vector \mathbf{w} y el vector de entrada \mathbf{p} multiplicado por el *bias* (sesgo) b . La función de activación para esta neurona es de tipo Gaussiana.

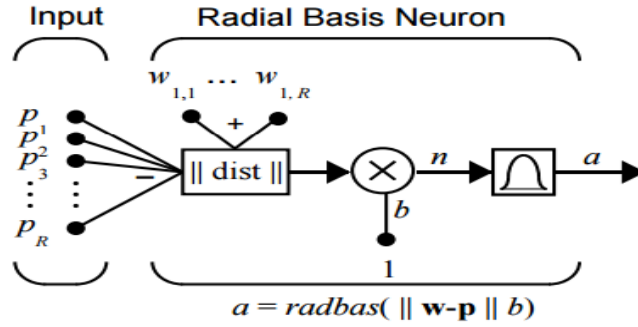


Figura 14. Diagrama de neurona de base radial [180].

En la Figura 15, tomada de [181], se observa se una RBFN donde la capa oculta utiliza una función de base radial y la capa de salida es una capa estándar lineal. La caja $\|dist\|$ calcula la distancia entre el vector de entrada \mathbf{p} y las filas de la matriz de pesos \mathbf{W} , luego se multiplica por el *bias* (sesgo) obteniendo el valor del vector \mathbf{n} que será evaluado en la función de activación S^1 del tipo radial. Esto da como resultado el vector \mathbf{a}^1 que ingresa a la capa lineal de salida donde se utiliza una función de activación S^2 del tipo *purelin* dando el vector de salida \mathbf{a}^2 [182].

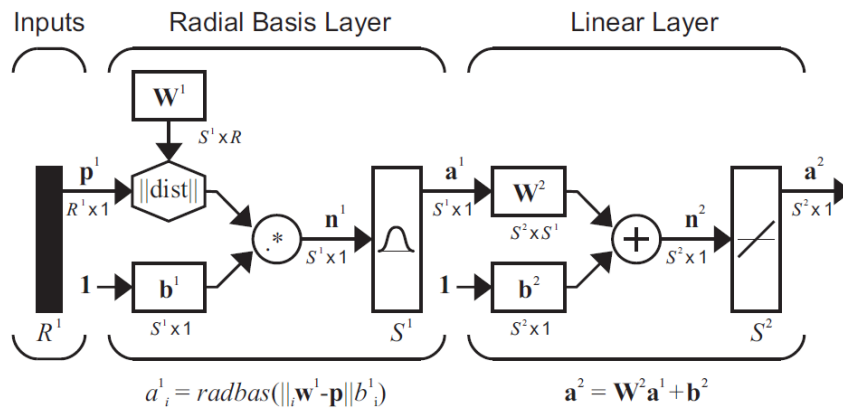


Figura 15. Arquitectura para RBFN.

El método de aprendizaje para este tipo de redes consta de dos etapas: en la primera, se definen los vectores de centros C_n de las capas ocultas lo que se puede realizar seleccionando los vectores aleatoriamente a partir de los datos de entrada x , o empleando un método no supervisado que permita dividir el espacio de patrones de entrada en clases empleando un algoritmo de k medias; en la segunda, se realiza una fase supervisada en donde se busca minimizar el error entre el valor de salida de la red y el de la salida deseada, que corresponde a su respectiva muestra de entrada p , para eso se puede emplear el método de mínimos cuadrados [183].

d. Red neuronal probabilística (*Probabilistic Neural Network, PNN*)

Es una variación de las RBFN, por lo que su estructura y entrenamiento son similares. Estos dos tipos de redes neuronales se diferencian en el hecho de que PNN está conformada por cuatro capas: capa de entrada, capa oculta, capa de sumatoria y capa de salida o decisión, véase la Figura 16, mientras que RBFN está conformada por tres capas.

En una PNN, la capa de entrada posee una neurona para cada caso, y se calcula la distancia entre la entrada y las muestras de los datos de entrenamiento obteniendo como resultado un vector donde sus elementos definen lo cerca que está una entrada de los datos de entrenamiento; la capa de sumatoria suma los resultados obtenidos para cada capa oculta y obtiene el valor de la densidad de probabilidad en cada clase y para la entrada. Finalmente, los resultados previos se dan como una entrada para una función de transferencia en la capa de salida, esta última capa selecciona el valor máximo de probabilidad [184].

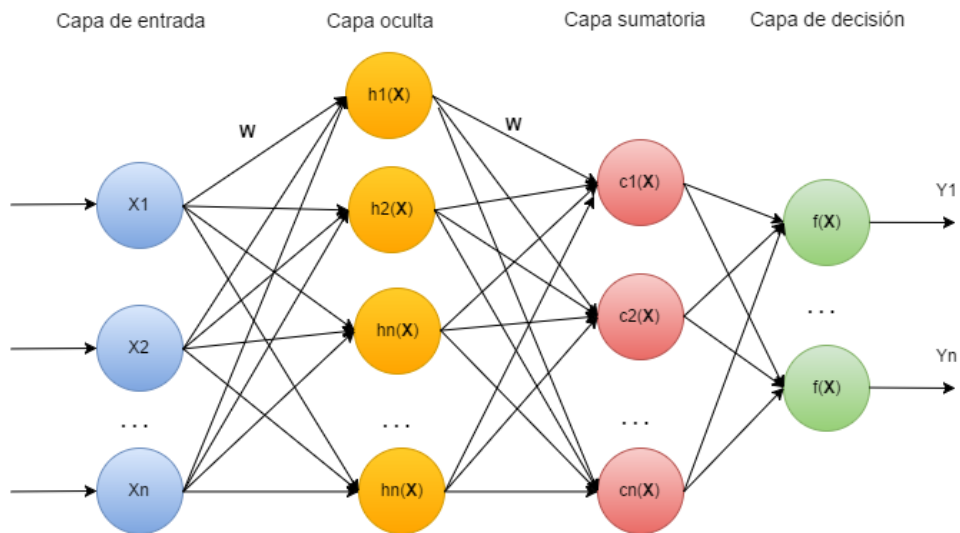


Figura 16. Modelo de una red neuronal probabilística.

En la Figura 17, se observa la arquitectura simplificada de la PNN tomada de [180], donde se observa la capa de entrada, la capa de *base radial* y la capa *competitive layer* que está formada por la capa *sumatoria* y la capa de *salida* del modelo presentado en la Figura 16. El valor de Q es el número de neuronas de la capa de *base radial* y K es el número de neuronas de la capa *sumatoria*.

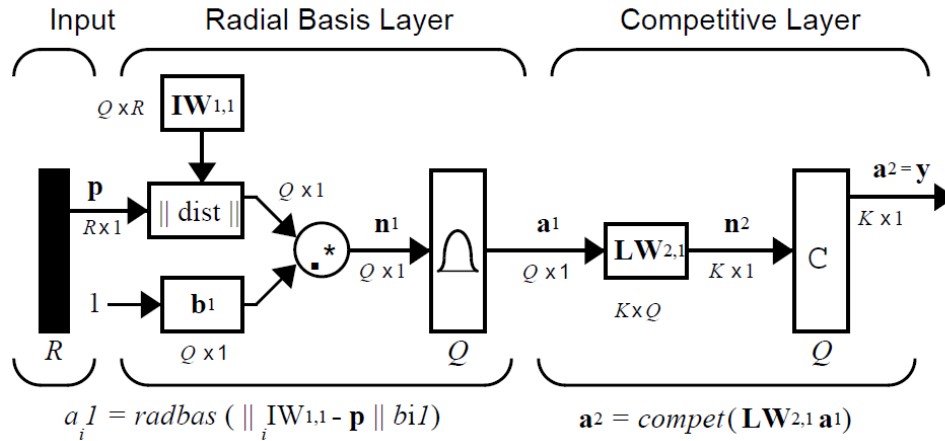


Figura 17. Arquitectura simplificada de una PNN.

e. Mapas auto-organizados (*Self-Organizing Maps*, SOM)

Las anteriores ANN analizadas utilizan un método de entrenamiento supervisado, el cual consiste en utilizar un conjunto de datos de entrenamiento en el cual se tiene una salida deseada para cada una de las entradas. Los SOM son un tipo de ANN que utiliza un método de entrenamiento no supervisado, clasifican los vectores de entrada en función de cómo están agrupados en el espacio; este tipo de ANN tiene una estructura en la cual las neuronas que tienen una relación estrecha de información se mantienen juntas con el fin de facilitar la interacción mediante conexiones sinápticas cortas. La estructura SOM más popular es la llamada red Kohonen, la cual tiene una estructura de una red de alimentación hacia adelante con una sola capa computacional dispuesta en filas y columnas. En términos generales, para crear una red SOM se debe establecer cuatro componentes fundamentales que son [180]:

- (1) Inicialización; se inicializan aleatoriamente los pesos de la red.
- (2) Competencia; para cada patrón de entrada las neuronas calculan sus respectivos valores de una función discriminante (distancia euclidiana entre el vector de entrada y el vector de pesos) que a su vez proporciona las bases para la competencia. En este caso, la neurona cuyo vector de pesos se acerca más al vector de entrada es el ganador.
- (3) Cooperación; la neurona ganadora se coloca en una localización espacial del vecindario de las neuronas excitadas, proporcionando con ello la base para la cooperación entre las neuronas vecinas.
- (4) Adaptación; el resto de las neuronas se ven afectadas por la neurona ganadora, experimentando un ajuste en los pesos de conexión asociados.

Las redes SOM utilizan la arquitectura de la Figura 18, tomada de [180], en la cual la caja *ndist* calcula la distancia entre el vector de entrada \mathbf{p} y matriz de pesos \mathbf{IW} (formada por la entrada de la matriz de pesos) obteniendo el vector \mathbf{S} , luego se suma el *bias* \mathbf{b} y se obtiene el vector \mathbf{n} que entra a la función de transferencia de competitividad y devuelve un valor de 0 para todas las neuronas excepto la ganadora.

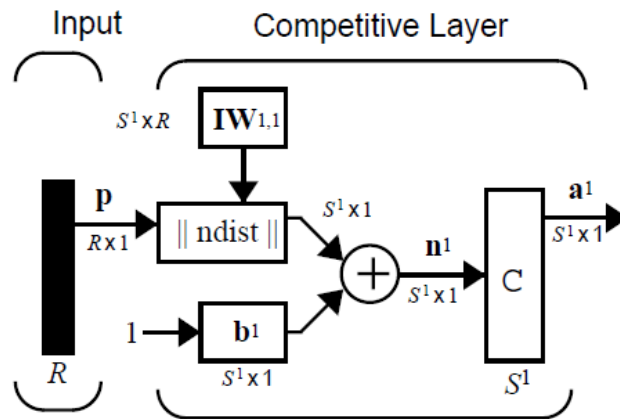


Figura 18. Mapas Auto-organizados (SOM).

3.8 Algoritmos genéticos con redes neuronales artificiales

Un algoritmo genético (*Genetic Algorithm, GA*), opera sobre una población de individuos donde cada uno de ellos es un posible candidato para la solución del problema a resolver. Cada uno de estos individuos está compuesto por uno o más cromosomas los cuales a su vez son vistos como un conjunto de genes, codificados como una cadena binaria. Asociado con cada individuo hay una función de aptitud (*fitness function*), la cual indica qué tan buena es la solución del candidato. El algoritmo en sí mismo es un procedimiento interactivo estocástico donde en cada iteración los individuos son estocásticamente seleccionados basados en su aptitud; los individuos seleccionados (progenitores) tendrán una oportunidad de cruzar sus genes entre sí, por lo tanto, propagarán parte de sus genes a nuevos individuos (descendencia). Durante este proceso ocurren algunas mutaciones. La descendencia puede reemplazar total o parcialmente a la población vieja y todo este proceso se repite hasta que se alcance el criterio de detención establecido [185].

En una ANN el número de capas ocultas, entre otros parámetros, necesita ser definido a priori. Diferentes capas ocultas resultan en diferentes ANN con diferentes capacidades de aproximación. Los GA pueden ser utilizados para encontrar una arquitectura adecuada para la ANN.

Los algoritmos genéticos se utilizan generalmente con una ANN para reducir el número de atributos y encontrar el número de capas más eficientes para la red. El proceso de los GA presentado en la Figura 19 empieza creando una población aleatoria de individuos, cada uno de estos codifica una estructura de red neuronal candidata para determinados subconjuntos de atributos, acto seguido estas se entrenan y evalúan. Cuando se alcanzan los criterios de parada, el proceso finaliza y se obtienen los mejores cromosomas, caso contrario el proceso continúa seleccionando una nueva población para crear una nueva ANN y prosigue. Una vez que el proceso de GA finaliza, se ejecuta una fase de clasificación con los parámetros obtenidos en los pasos previos.

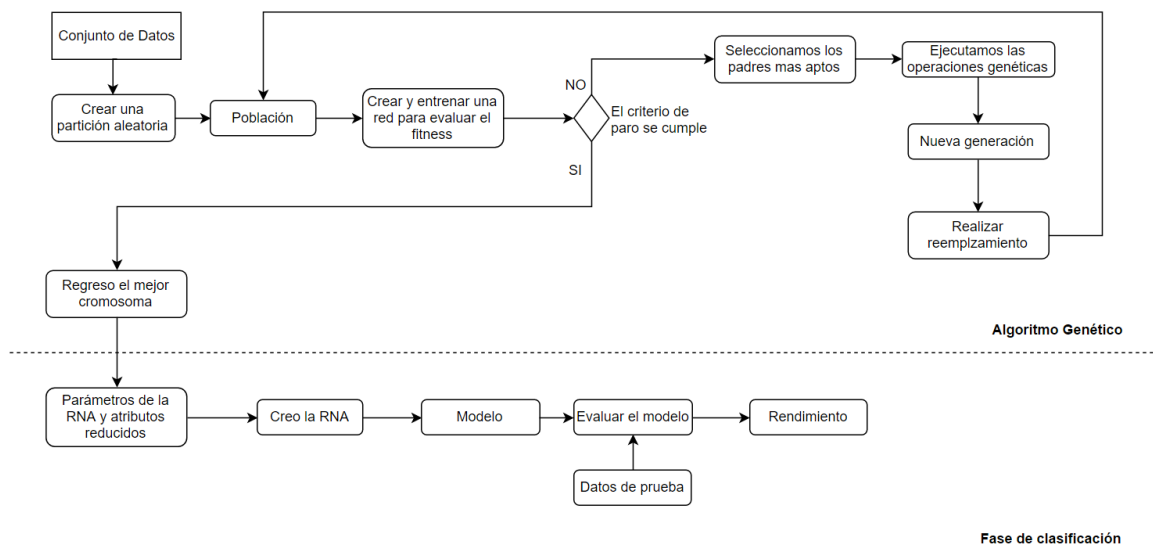


Figura 19. Diagrama del uso de GA en redes neuronales.

3.9 Redes neuronales convolucionales

Las redes convolucionales (*Convolutional Neural Networks* CNN's) son redes bio-inspiradas en la corteza visual de animales y son una variante de los perceptrones multi capa (MLP's). Las CNN's son redes neuronales que usan la operación de convolución en lugar de multiplicaciones de matrices sus capas [186]. De una manera práctica, el uso de la operación de convolución con entradas bidimensionales (matrices y/o imágenes), se puede definir como sigue:

$$s(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n) \quad \text{Ecuación 46}$$

Donde I es la matriz de entrada y K es el *kernel* o núcleo que se usa como filtro para la convolución.

Una CNN está compuesta por una o varias capas convolucionales, cada una de estas capas convolucionales está compuesta por varios *kernels*, los cuales al realizar la convolución con la entrada generan diferentes mapas de características a la salida. Estos mapas pasan a una etapa de sub muestreo que generalmente usa la técnica de *max-pooling* la cual consiste en tomar pequeños grupos de valores provenientes de los mapas de características y realiza un promediado de dichos valores para lograr reducir los parámetros dando como resultado una agrupación de atributos robustos frente a cambios como la traslación local. Al juntar toda la secuencia de procesos descritas anteriormente en capas y etapas, se forma lo que se conoce como la arquitectura de una red convolucional, la misma que se puede observar en la Figura 20[187].

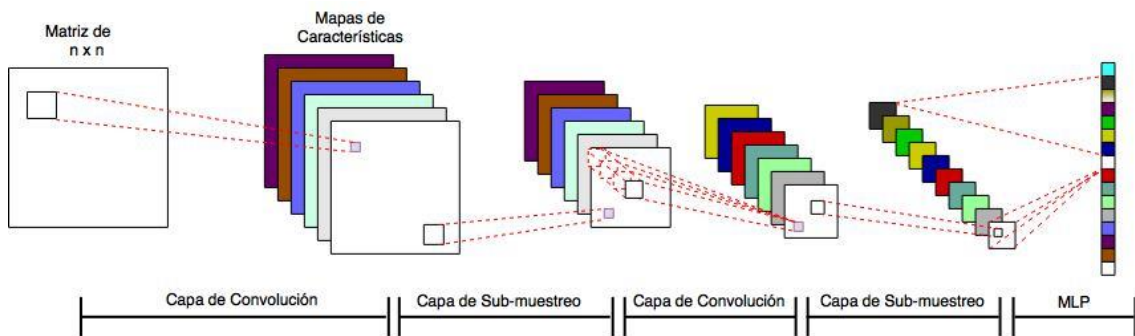


Figura 20. Arquitectura de una red convolucional

3.10 Bosques aleatorios

Los bosques aleatorios o RF es una técnica utilizada para tareas de clasificación o predicción, la cual se basa en la división de un espacio de atributos mediante el uso de árboles de decisión. Estos árboles se basan en una serie de preguntas que se formulan sobre las variables que definen a cada elemento n de un vector x ; a fin de asignarle una clase C [188].

- **Formalismo matemático**

Dado que m es la cantidad de pares de muestras definidas como (x_i, y_i) , siendo $i = 1, \dots, m$.

Donde:

x_i : representan un vector de n atributos, $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$.

y_i : indica que el vector de atributos x_i está asociado a una clase C , $y_i = C$.

Se seleccionan un atributo j y un punto de división s y a continuación se establecen los semiplanos R_1 y R_2 :

$$R_1(j, s) = X | X_j \leq s, \quad R_2(j, s) = X | X_j > s \quad \text{Ecuación 47}$$

Sea \hat{p}_{kc} la proporción de observaciones de clases C en la región R_k respecto del total de observaciones N_k .

El problema de optimización está dado por:

$$\arg \max_c \hat{p}_{kc} \quad \text{Ecuación 48}$$

$$\text{Sea } \hat{p}_{kc} = \left(\frac{1}{N_k} \right) \sum x_i \in R_k I(y_i = C),$$

En donde I es el indicador de pertenencia de x_i a una región R .

\hat{p}_{kc} : es la función de impureza de los nodos hijos, también se utiliza el índice Gini o la ganancia de información.

- **Criterio de paro**

Se puede definir un nivel de profundidad con base al número de nodos o alguna función de costo.

- **Técnica de Bagging**

Se aplica para disminuir la varianza del modelo basado en árboles. Esta técnica divide el conjunto de datos de entrenamiento en diferentes subconjuntos generando así diferentes árboles de decisión $T_b, b = 1, \dots, B$ [188]. Este clasificador selecciona la clase con el mayor número de votos entregados desde cada clasificador, mediante:

$$\hat{C}_{bag}(x_i) = \arg \max_c \hat{f}_{bag}(x_i) \quad \text{Ecuación 49}$$

Donde:

c : clase

\hat{f}_{bag} : clasificador que propone la clase c para x_i

\hat{C}_{bag} : clase estimada

$\hat{f}_{bag}(x_i)$: vector $p_c(x_i)$, proporción de estimadores que propone la clase c

- **Selección aleatoria de atributos y muestras (con reemplazo)**

Permite disponer de árboles de decisión con variación controlada, consiguiendo así una discriminación estocástica; esta selección representa alrededor del 2/3 de los datos originales.

- **OOB-error**

Es el error promedio de clasificación asociado a cada árbol T_b , $OOB_b = m - m_b$, se usa durante el proceso de entrenamiento, se conoce también como el error fuera de bolsa.

3.11 Evaluación de los clasificadores

3.11.1 Medidas de desempeño en la clasificación

Una medida de desempeño de un clasificador o clasificadores representa los resultados que el clasificador obtiene en la fase de pruebas, estos resultados a su vez pueden ser analizados para que se realicen las correcciones en la arquitectura del clasificador o atributos elegidos.

En el aprendizaje automático existen varias medidas de desempeño, las cuales entregan información para ser analizada, estas medidas son totalmente independientes del número de muestras utilizadas para prueba. Para determinar estas medidas de desempeño hace falta definir algunas variables como: verdaderos positivos (TP) que es el número de muestras que se clasifican correctamente como pertenecientes a una clase; verdaderos negativos (TN) que es el número de muestras que correctamente fueron clasificadas como no pertenecientes a una clase; falsos positivos (FP) que es el número de muestras que incorrectamente se asignaron como pertenecientes a una clase y falsos negativos (FN) que es el número de muestras, que perteneciendo a una clase no se incluyeron como parte de esta [189].

3.11.1.1 Matriz de confusión

Es una herramienta visual, la cual permite observar el desempeño de un algoritmo empleado en un sistema de clasificación. La matriz de confusión expone las funciones discriminantes de cada clase, comparando para el conjunto de individuos en la tabla de aprendizaje, la predicción dada versus la clase a la que realmente pertenece. En la matriz de confusión las columnas muestran el número de predicciones de cada clase y las filas representan a las instancias en la clase real [190].

En la Figura 21 se observa un ejemplo de una matriz de confusión de dos filas y dos columnas.

		Valores Predecidos		Total:
		p	n	
Valores Reales	p'	Verdaderos Positivos	Falsos Negativos	P'
	n'	Falsos Positivos	Verdaderos Negativos	N'
Total:		P	N	

Figura 21. Matriz de confusión clase binaria.

A partir de la matriz de confusión se puede calcular algunas métricas que son utilizadas para la clasificación binaria, y que han sido analizadas en otros estudios [191], [192] :

- Exactitud (*Accuracy*, ACC), que mide la efectividad del clasificador e indica la proporción del número total de predicciones que fueron correctas. Para determinar la ACC se utiliza la Ecuación 50:

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \quad \text{Ecuación 50}$$

- Precisión (p): estimación de que la predicción correcta sea positiva, Ecuación 51.

$$P = \frac{|TP|}{|TP| + |FP|} \quad \text{Ecuación 51}$$

- Susceptibilidad (*Recall*) eficacia de un clasificador para identificar etiquetas positivas, Ecuación 52.

$$R = \frac{|TP|}{|TP| + |FN|} \quad \text{Ecuación 52}$$

- F1 Score: relación entre etiquetas positivas de los datos y aquellos conseguidos por el clasificador, Ecuación 53. La constante β controla la compensación entre la precisión y el *recall*.

$$F = \frac{(\beta^2 + 1) * P * R}{\beta^2 * P + R} \quad \text{Ecuación 53}$$

- Especificidad: cómo efectivamente un clasificador identifica etiquetas negativas, Ecuación 54.

$$S = \frac{TN}{FP + TN}$$

Ecuación 54

- Área bajo la curva (AUC): habilidad del clasificador para evitar clasificaciones falsas, Ecuación 55.

$$AUC = \frac{1}{2} * \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$$

Ecuación 55

Las métricas utilizadas para clasificación multiclase se exponen a continuación. Donde TP_i son los valores positivos para de la i -ésima clase C_i , FP_i son los falsos positivos, FN_i los falsos negativos y TN_i los verdaderos negativos. Por otra parte, el índice M representa el promediado macro y l es el total de clases [192]:

- Exactitud promedio (*average accuracy*, AA): eficacia media por clase de un clasificador, Ecuación 56.

$$AA = \frac{\sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i}}{l}$$

Ecuación 56

- Tasa de error (*error rate* ER): promedio del error de clasificación por clase, Ecuación 57.

$$ER = \frac{\sum_{i=1}^l \frac{FP_i + FN_i}{TP_i + FN_i + FP_i + TN_i}}{l}$$

Ecuación 57

- $Precision_M$: promedio por clase de la coincidencia de las etiquetas de las clases de datos con aquellas del clasificador, Ecuación 58.

$$Precision_M = \frac{\sum_{i=1}^l \frac{FP_i}{TP_i + FP_i}}{l}$$

Ecuación 58

- $Recall_M$: promedio por clase de la eficacia de un clasificador para identificar las etiquetas positivas de las clases, Ecuación 59.

$$\text{Recall}_M = \frac{\sum_{i=1}^l \frac{TP_i}{TP_i + FN_i}}{l}$$

Ecuación 59

- Fscore_M : relación entre las etiquetas positivas de los datos y aquellas conseguidas por un clasificador basado en un promediado por clase, Ecuación 60.

$$\text{Fscore}_M = \frac{(\beta^2 + 1) \text{Precision}_M \text{Recall}_M}{\beta^2 \text{Precision}_M + \text{Recall}_M}$$

Ecuación 60

3.11.1.2 Curva característica operativa del receptor

La curva característica operativa del receptor (*Receiver Operating Characteristic*, ROC) es una representación gráfica que permite observar el desempeño de un clasificador binario según se varía el umbral de discriminación.

Si la clasificación es correcta se asigna 1, mientras que si la clasificación es errónea se asigna 0. Con estos valores se puede determinar el número de los verdaderos positivos (TP), falsos negativos (FN), verdaderos negativos (TN) y falsos positivos (FP). La exactitud de un clasificador también puede ser medida en términos del área bajo la curva (AUC) a partir de la curva ROC [193], véase Figura 22.

Para representar gráficamente la curva ROC es necesario determinar la razón de verdaderos positivos (sensibilidad) y la razón de falsos positivos (especificidad), las cuales se determinan por la Ecuación 61 y la Ecuación 62, respectivamente.

$$TPR = \frac{TP}{TP + FN}$$

Ecuación 61

$$FPR = \frac{FP}{FP + TN}$$

Ecuación 62

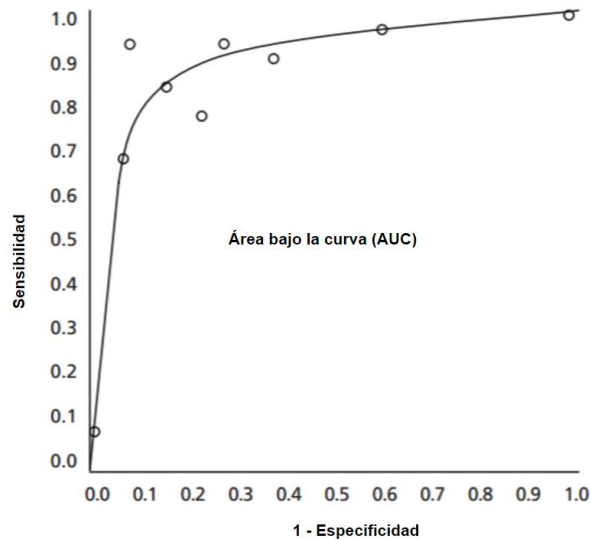


Figura 22. Curva ROC y AUC.

La razón de verdaderos positivos (TPR) determina hasta qué punto un clasificador es capaz de clasificar correctamente los casos positivos, de entre todos los casos positivos disponibles. Y la razón de falsos positivos (FPR) determina cuantos resultados positivos son incorrectos de entre todos los casos negativos posibles.

3.12 Pruebas estadísticas

Las pruebas (*test*) estadísticas, son pruebas que se aplican a las estadísticas obtenidas para determinar su grado de certeza y su significancia, estas pruebas se utilizan para contrastar un conjunto de datos para un problema en estudio.

Existen varios tipos de pruebas estadísticas, el tipo de prueba a utilizar dependerá de la distribución de la variable que se va a comparar; si la distribución de datos cumple ciertas condiciones como normalidad, homocedasticidad¹ y/u otras condiciones se utilizan pruebas paramétricas, y si no cumple dichos criterios se aplican pruebas no paramétricas o de distribución libre que se basan en los rangos de distribución de la variable. Las pruebas no paramétricas son más exigentes al rechazar la hipótesis nula de igualdad y por lo tanto tienen menos posibilidades de acertar cuando no la rechazan. Cabe recalcar que la eficacia de una prueba estadística aumenta cuando el tamaño de la muestra también lo hace. Por lo tanto, para conseguir mayor eficacia con una prueba no paramétrica es necesario aumentar el tamaño de la muestra [194], [195].

¹ La hipótesis de homocedasticidad supone igual varianza para distribuciones y análisis o que el error cometido por el modelo tiene siempre la misma varianza.

3.12.1 Pruebas paramétricas

Las pruebas paramétricas son herramientas comprensivas para contrastar un conjunto de observaciones en un problema en estudio y asumen la existencia de distribuciones estadísticas subyacentes a los datos. Para poder emplear pruebas paramétricas el conjunto de datos a evaluar debe cumplir tres condiciones: (1) independencia, dos eventos son independientes si la ocurrencia de uno de ellos no modifica la probabilidad de la ocurrencia del otro; (2) normalidad, se consigue cuando los datos colectados siguen una distribución normal con la media y la varianza, (3) homocedasticidad, igual varianza para distribuciones y análisis. Para comparar múltiples distribuciones uno de los procedimientos paramétricos más populares es el análisis de varianza (ANOVA), la diferencia entre dos o más medias de las muestras relacionadas se obtiene con mediciones ANOVA repetitivas. Para el caso de evaluar el clasificador, las medias de las muestras son las tasas de clasificación; que fueron obtenidas con el mismo conjunto de datos y la misma partición para el entrenamiento y el conjunto de pruebas [195].

a. Normalidad

- **Gráfico de cuantiles (QQ-Normal)**

Estos gráficos son utilizados para analizar la normalidad, donde en el eje horizontal se ubican cada valor (cuantil) de la función de distribución de probabilidad observada y en el eje vertical la prevista por el modelo de distribución normal. Si ambas distribuciones coinciden los puntos se ubican sobre una recta, y en caso de que la distribución de la muestra difiera de la distribución normal estos puntos se ubican fuera de dicha recta como se indica en la Figura 23 [196].

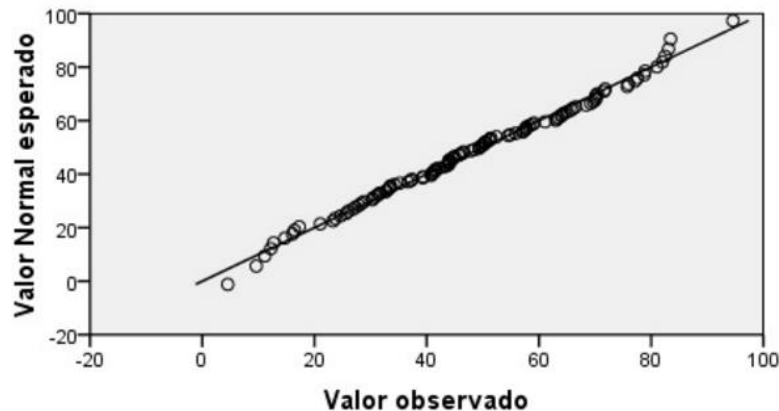


Figura 23. Grafico QQ-Normal.

- **Shapiro–Wilk Test**

Esta prueba permite conocer si los datos $(x_1 + x_2, \dots, x_n)$ provienen de una población normalmente distribuida. En [197] se presenta el estudio completo realizado por Samuel Shapiro y Martin Wilk. En la Ecuación 63, se presenta su cálculo:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{Ecuación 63}$$

Donde: $x_{(i)}$ es el número que ocupa la i -ésima posición en la muestra,

$\bar{x} = (x_1 + x_2, \dots, x_n)/n$ es la media muestral.

Las variables a_i se calculan con:

$$(a_1, a_2, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

Donde:

$$m = (m_1, m_2, \dots, m_n)^T$$

siendo m_1, m_2, \dots, m_n los valores medios del estadístico ordenado, de variables idénticamente e independientemente distribuidas. V es la matriz de covarianzas de dicho estadístico ordenado.

Los rangos de W pueden estar entre 0 y 1. Si el valor de W es muy bajo se rechaza la hipótesis nula, teniendo en cuenta que la hipótesis nula es que la población está distribuida normalmente. En la actualidad ya existen diferentes algoritmos, desarrollados para Matlab, que implementan esta prueba.

b. Homocedasticidad

- **Levene's Test**

Esta prueba es utilizada para evaluar la homogeneidad de varianzas. Esta prueba no asume que todas las poblaciones se distribuyen normalmente y se recomienda cuando la hipótesis de normalidad no es viable.

Su cálculo se observa en la Ecuación 64, cada grupo k tiene una distribución normal con posibilidad de medias diferentes y desviación estándar $\sigma_1, \sigma_2, \dots, \sigma_k$. Sea n_1, n_2, \dots, n_k el números de sujetos de cada grupo, Z_{ij} los valores de respuesta y N el tamaño total de las muestras. Esta prueba asume que los datos se obtienen tomando una muestra aleatoria simple de cada una de las poblaciones k [198] [199].

$$W = \frac{(N - k) \sum_{i=1}^k N_i (Z_i - Z)^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - Z_i)^2}$$

Ecuación 64

Donde:

- W es el resultado de la prueba.
- k es el número de diferentes grupos a los que pertenecen los casos muestreados.
- N es el número total de casos en todos los grupos.
- N_i es el número de casos en el grupo i .
- Y_{ij} es el valor de la variable mediada para el j -ésimo caso de i -ésimo grupo.
- $Z_{ij} = \begin{cases} |Y_{ij} - \bar{Y}_i|, \\ |Y_{ij} - \tilde{Y}_i|, \end{cases}$ (\bar{Y}_i es la media del i -ésimo grupo y \tilde{Y}_i es la mediana del i -ésimo grupo).
- $Z = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{N_i} Z_{ij}$, es la media de Z_{ij} .
- $Z = \frac{1}{N} \sum_{i=1}^{N_i} Z_{ij}$, es la media de Z_{ij} para el grupo i .

Se debe tener en cuenta que si p -valor resultante de la prueba de Levene es menor a un umbral de significación (típicamente 0.05) se rechaza la hipótesis nula. Esta hipótesis nula hace referencia a la igualdad de varianzas, concluyendo así que existe una diferencia entre las variaciones en la población; al igual que para la prueba de Shapiro para esta prueba también se pueden encontrar diferentes algoritmos para Matlab.

3.12.2 Pruebas no paramétricas

Las pruebas no paramétricas permiten probar una hipótesis, con la diferencia que para las pruebas no paramétricas de que no deben ajustarse a ninguna distribución, es decir, no es necesario que la distribución de la población cumpla determinados parámetros como la independencia, normalidad y homocedasticidad como pasa con las pruebas paramétricas.

Hay dos tipos de comparaciones que pueden realizarse mediante estos enfoques: (1) comparaciones por parejas, esta es un proceso estadístico de comparación intrínseca entre dos algoritmos y (2) comparaciones múltiples, ya sea un contraste entre un algoritmo contra

un grupo de N algoritmos, o una comparación de todos contra todos los algoritmos [195]. Una ventaja de la comparación por parejas frente a las comparaciones múltiples es que los resultados de la comparación por parejas son fáciles de analizar; mientras que los resultados de comparaciones múltiples son más complejos de analizar, por lo que se requiere realizar un procedimiento *post-hoc* con el fin de discriminar las diferencias entre los algoritmos evaluados [200]. Los procedimientos *post-hoc* son pruebas posteriores usadas para detectar pares diferentes que no fueron detectados por el método de comparación múltiple. Por ejemplo, la comparación múltiple con pruebas no paramétricas, como las desempeñadas por las pruebas de Friedman, solo pueden detectar una diferencia significativa entre las observaciones mediante rechazo de la hipótesis nula, pero esto no contribuye con más información.

- **Prueba de Friedman**

Desarrollada por Milton Friedman, esta prueba sirve para probar la diferencia entre varias muestras relacionadas y se usa cuando el mismo parámetro se ha medido en condiciones diferentes sobre los mismos temas [201].

Sea $\{x_{ij}\}_{m \times n}$ una tabla datos donde m son las filas (bloques) y n las columnas (tratamientos). Se calcula el orden dentro de cada bloque y luego se reemplaza por una nueva tabla $\{r_{ij}\}_{m \times n}$ donde el valor r_{ij} es el orden de x_{ij} en cada bloque i , [202].

El cálculo de varianzas intra e inter grupo se puede expresar como:

$$\bar{r}_j = \frac{1}{m} \sum_{i=1}^m r_{ij}$$

$$\bar{r} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n r_{ij}$$

$$SS_t = n \sum_{j=1}^n (\bar{r}_j - \bar{r})^2$$

$$SS_e = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (r_{ij} - \bar{r})^2$$

El estadístico viene dado por la Ecuación 65:

$$Q = \frac{SS_t}{SS_e}$$

Ecuación 65

Cuando n o k son muy largas ($n > 15$ o $k > 4$) la distribución de probabilidad Q se puede aproximar por una chi-cuadrado y los p-valores se consiguen a través de $P(\chi_{k-1}^2 \geq Q)$.

- **Análisis Post-Hoc**

Consiste en observar los patrones de datos que no se especificaron *a priori* después de que el experimento ha concluido. El análisis Post-Hoc normalmente se preocupa por la búsqueda de patrones y/o relaciones entre los subgrupos de poblaciones muestreadas que de otro modo permanecerían sin ser detectados. Existen diferentes métodos que se utilizan para este análisis. A continuación, se analizará la prueba de Holm.

Prueba de Holm: Es un método utilizado para contrarrestar el problema de comparaciones múltiples [203]. Los pasos para realizarla se resumen con el siguiente procedimiento:

- Sea $H_1 \dots H_m$ las familias de hipótesis asociadas y $P_1 \dots P_m$ los correspondientes p-valores.
- Ordenar los p-valores de menor a mayor $P_{(1)} \dots P_{(m)}$ y dejar que las hipótesis asociadas sean $H_{(1)} \dots H_{(m)}$.
- Para que un nivel de significancia α sea k el índice mínimo es tal que $P_k > \frac{\alpha}{m+1-k}$
- Rechazar las hipótesis nulas $H_{(1)} \dots H_{(k-1)}$, si $k=1$ entonces no se rechaza ninguna hipótesis.

- **Prueba de signos de Wilcoxon**

Es una prueba no paramétrica que permite comparar la media de dos muestras relacionadas, muestras emparejadas, o mediciones repetidas y determinar si existe diferencia entre ellas [204] [205].

Siendo N el tamaño de la muestra de número de pares, existe un total de $2N$ puntos de datos, para $i = 1, \dots, N$, siendo $x_{1,i}$ y $x_{2,i}$ las muestras tomadas.

- H_0 (Hipótesis nula): establece que las distribuciones poblacionales x_1 y x_2 son idénticas.
- H_1 (Hipótesis alternativa): establece que existe una diferencia de la tendencia central de las poblaciones (dos colas) o bien la frecuencia relativa poblacional para las $x_{1,i}$ se desplaza a la derecha de $x_{2,i}$ (una cola).

Su procedimiento es el siguiente:

1. Prueba por pares

- a. Establecer la hipótesis nula H_0 ; en este caso las distribuciones poblacionales x_1 y x_2 son idénticas, y su alternativa H_1 establece que existe una diferencia de la tendencia central de las poblaciones.
- b. Para $i = 1, \dots, N$, calcular la diferencia $D_i = x_{1,i} - x_{2,i}$, las diferencias iguales a cero se eliminan.
- c. Clasificar los valores absolutos de D_i , asignando 1 al más pequeño, un 2 al segundo más pequeño y así sucesivamente.
- d. Etiquetar con el signo a cada valor clasificado con base al signo que tenía en D_i .
- e. Calcular W^+ (suma de rangos positivos) y W^- (suma de rangos negativos). Tener en cuenta que $W^+ + W^- = \frac{n(n+1)}{2}$.
- f. Elegir $W = \min(W^-, W^+)$. Revisar las tablas de Wilcoxon para comparar el valor estadístico W , y si el valor es menor o igual a algún W_0 se rechaza la hipótesis. Para hipótesis alternativas; si $W^- \leq W_0$ la distribución de x_1 se desplaza a la derecha de x_2 por otra parte si $W^+ \leq W_0$ la distribución x_2 se desplaza a la derecha de x_1 .

Existen tablas que entregan valores de probabilidad que permiten hallar las regiones de rechazo para esta prueba [205].

4. EXPERIMENTACIÓN

4.1 Descripción del funcionamiento del banco

La experimentación se desarrolló en el laboratorio de vibraciones de la Universidad Politécnica Salesiana, sede Cuenca (Ecuador). En la Figura 24 se presenta el banco y sus posibles configuraciones para pruebas en engranajes. El movimiento se genera por un motor marca SIEMENS, modelo 1LA7 090-4YA60, con alimentación trifásica de 220 VCA, de 2 HP, 1700 RPM, que mediante un acoplamiento flexible transmite el movimiento al eje de entrada de la caja de engranajes. En el eje de salida de la caja de engranajes se ubica una polea que transmite por medio de una banda flexible el movimiento al freno electromagnético ROSATI de 8.83 kW para simular la carga. La caja de engranes se puede configurar en una o dos etapas, ya sea para trabajar con engranajes rectos o helicoidales. Es posible también combinar fallos en engranajes y rodamientos de los ejes de la caja. La velocidad de giro del motor se controla mediante el variador de frecuencia DANFOSS VLT 1.5 kW y la carga del freno se controla manualmente mediante la fuente de voltaje de corriente continua TDK-Lambda, GEN 100-15-IS510. Para la adquisición de datos se emplean acelerómetros PCB ICP 352C03, la tarjeta de NI cDAQ-9234 y la computadora portátil HP Pavilion g4-2055la.



Figura 24. Configuración para pruebas con engranajes.

4.2 Adquisición de la señal de vibración

El sistema de adquisición de datos de las señales de vibración se presenta en la Figura 25. Está constituido por un acelerómetro PCB ICP 352C03 que va montado en la caja de engranajes cuyas características principales se indican en la

Tabla 2. La señal del acelerómetro se adquiere a 50 kS/s por la tarjeta NI cDAQ-9234 cuyas características se indican en la

Tabla 3. La tarjeta se acopla en el chasis NI cDAQ 9191 que envía los datos a la computadora portátil HP Pavilion g4-2055la para ser almacenados según la configuración del software de adquisición de datos.

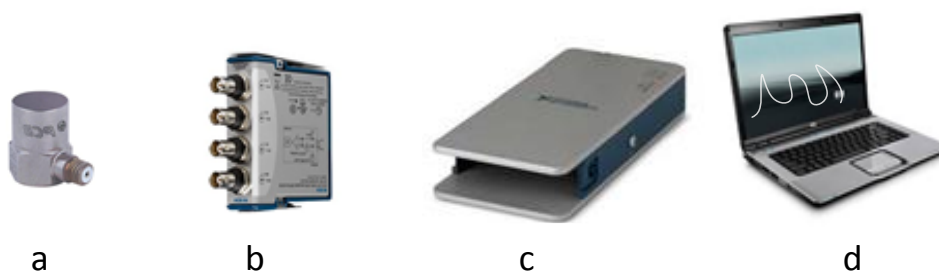


Figura 25. Sistema de adquisición de señales de vibración a) acelerómetro b) tarjeta de adquisición c) chasis para el montaje de la tarjeta d) computador portátil

Tabla 2. Características del acelerómetro.

Modelo	352C03
Sensibilidad	10.06 mV/g
Rango de medición	± 500 g pk
Rango de frecuencia	($\pm 5\%$) 0.5 a 10000 Hz
Rango de frecuencia	($\pm 5\%$) 0.3 a 15000 Hz
Resolución de ancho de banda	0.0005 g rms
Frecuencia de resonancia	≥ 50 KHz

Tabla 3. Características de la tarjeta NI cDAQ-9234.

Máxima velocidad de muestreo	51.2 kS/s por canal; entrada de ± 5 V
Número de canales	4
Resolución	24 bits; rango dinámico de 102 dB; filtros anti-aliasing
Condicionamiento de señales	IEPE seleccionable por software (0 ó 2 mA)
Acoplamiento de AC/DC	Seleccionable por software; acoplado en AC (0.5 Hz)

4.3 Configuración de la experimentación

4.3.1 Configuración del banco para pruebas con engranajes

La caja de engranajes es configurable para trabajar en una etapa o dos etapas tanto en engranajes rectos o helicoidales, como se indica en la Figura 26, también se pueden simular fallos de excentricidad en el eje de salida o fallo en el rodamiento 6005LLUC3/2AS sobre

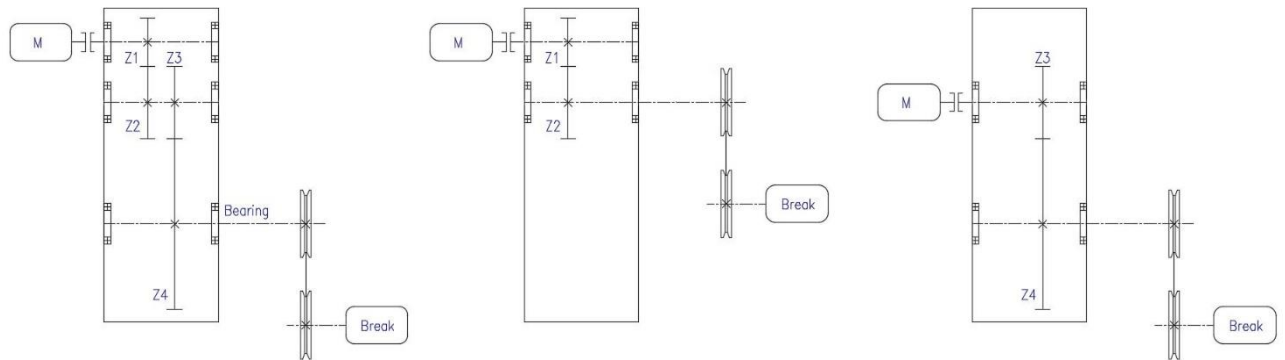


Figura 26. Configuración de la caja de engranajes. a) Configuración en dos etapas. b) Configuración para la primera etapa Z1 y Z2. c) Configuración para la segunda etapa Z3 y Z4.

los cuales giran los ejes de la caja. Las características de los engranajes se detallan en la Tabla 4, el espesor de los dientes es de 20 mm, construidos de material SAE/AISI 4140, módulo 2.25, con ángulo de presión $\Phi = 20^\circ$ y un ángulo de hélice $\psi = 20^\circ$ de para el caso de los engranajes helicoidales; para las pruebas se emplearon engranajes nuevos.

Tabla 4. Características de los engranajes

Posición del engranaje	Número de dientes engranajes rectos	Número de dientes engranajes helicoidales
Engranaje (Z1)	27	30
Engranaje (Z2)	53	45
Engranaje (Z3)	53	45
Engranaje (Z4)	80	80

4.3.2 Configuración de los parámetros del banco

Para adquirir las señales de vibración, los parámetros variables en el banco son:

- 1) Velocidad de rotación del motor que se puede regular con el DANFOSS VLT 1.5 kW y es configurable para velocidad constante o variable.
- 2) Carga a la caja se aplica con el freno electromagnético ROSATI de 8.83 kW la carga es controlada manualmente mediante la fuente de voltaje de corriente continua TDK-Lambda, GEN 100-15-IS510 y se puede variar la carga de 0 hasta 2 HP.

- 3) Para la lubricación se utilizó el método de lubricación por salpicadura para todas las pruebas se mantuvo el 35% nivel de la altura del depósito que asegura la lubricación de los rodamientos y los engranajes, el aceite utilizado fue el Gulf EP Lubricant HD 220.
- 4) La configuración se puede realizar para pruebas en engranajes rectos o helicoidales en una o dos etapas.
- 5) Seleccionar los fallos en engranajes que se detalla en los planes experimentales.

4.3.3 Fallos simulados en la caja de engranajes

Para simular los defectos en la caja de engranajes se maquinó engranaje y piñón según el caso, los fallos inducidos se detallan en cada uno de los planes experimentales; se trabajó con fallos en los engranajes, múltiples fallos en engranajes y fallos en los rodamientos de la caja.

4.4 Planes experimentales

Se diseñaron tres planes experimentales que se levantaron en el laboratorio y un cuarto plan experimental se adecuó de las señales de vibración de la base de datos pública disponible en <https://www.phmsociety.org/references/datasets> [206].

4.4.1 Plan experimental uno

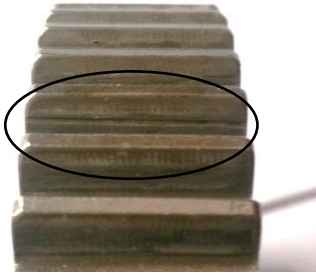
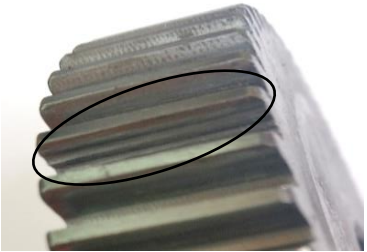
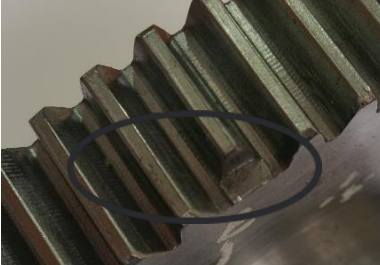

La configuración de la caja de engranajes y condiciones específicas de la prueba se detalla a continuación:

- a. Numero de etapas: 1
- b. Engranajes montados: Z3 y Z4 con dientes rectos
- c. Número de condiciones: 10
- d. Número de velocidades de la prueba (6):
 - i. Velocidades constantes: 8 Hz, 12Hz, 15 Hz
 - ii. Velocidades variables: 5-12 Hz, 12-18 Hz, 8-15Hz
- e. Cargas de las pruebas (3): 0V, 10V, 30V
- f. Número de repeticiones de cada prueba: 5
- g. Total de señales adquiridas: 900
- h. Frecuencia de muestreo: 50 kS/s
- i. Tiempo de muestreo: 10 segundos
- j. Nombre de la base de datos: DB1
- k. Sensibilidad del acelerómetro: 100 mV/g

- Equipo empleado para la adquisición: NI 9234 y NI CompactDAQ-9191

En la siguiente tabla se detalla el montaje de la caja de engranajes y la descripción de los fallos.

Tabla 5. Configuración de los fallos para el plan experimental uno.

Condiciones o Clases	Z3	Z4	Fotografía del fallo
P1	Normal	Normal	
P2	Gripado Una línea Todo el ancho del diente Profundidad: 0.5mm Ancho: 0.8mm	Normal	
	Gripado Dos líneas Todo el ancho del diente Profundidad: 0.4mm Ancho: 0.7mm	Normal	
P4	Rotura del diente 25%	Normal	
P5	Rotura del diente 50%	Normal	

P6	Rotura del diente 100%	Normal	
P7	Normal	Grieta en el diente Profundidad: 4mm Longitud: 3mm Ancho: 1mm	
P8		Grieta en el diente Todo el ancho del diente Profundidad: 3mm Ancho: 1mm Ángulo: 45°	
P9	Normal	Gripado Una línea Todo el ancho del diente Profundidad: 0.5mm Ancho: 0.7mm	
P10	Rotura del diente 25%	Grieta en el diente profundidad: 4mm Longitud: 5mm	Ver condición P4 y P7

4.4.2 Plan experimental dos





La configuración de la caja de engranajes y condiciones específicas de la prueba se detalla a continuación:

- a. Numero de etapas: 2
- b. Engranajes montados: Z1, Z2, Z3 y Z4 con dientes helicoidales
- c. Número de condiciones: 11
- d. Número de velocidades de la prueba (3): 8Hz, 12Hz, 15Hz
- e. Cargas de la prueba (3): 0V, 10V, 30V
- f. Número de repeticiones de cada prueba: 5
- g. Total de señales adquiridas: 495

- h. Frecuencia de muestreo: 50 kS/s
- i. Tiempo de muestreo: 10 segundos
- j. Nombre de la base de datos: DB2
- k. Sensibilidad del acelerómetro: 100 mV/g
- l. Equipo empleado para la adquisición: NI 9234 y NI CompactDAQ-9191

En la siguiente tabla se detalla el montaje de la caja de engranajes y la descripción de los fallos.

Tabla 6. Configuración de los fallos para el plan experimental dos.

Condiciones o Clases	Localización del fallo	Descripción del fallo	Fotografía del fallo
P1	Normal	Normal	
P2	Z1	Gripado Una raya Todo el ancho del diente Profundidad 0.2 mm Ancho 1mm	
P3	Z2	Gripado: Una raya Todo el ancho del diente Largo: 10mm Ancho: 1.2mm Profundidad: 0.25mm	
P4	Z3	Picaduras 40% de los dientes Diámetro: 0.8mm	
P5	Z3	Gripado Dos rayas Todo el ancho del diente Ancho: 0.7mm Profundidad: 0.4mm	

P6	Z4	<p>Gripado</p> <p>Una raya</p> <p>Todo el ancho del diente</p> <p>Ancho: 1.2 mm</p> <p>Profundidad: 0.25mm</p>	
P7	Z4	<p>Grieta</p> <p>Largo: 2mm</p> <p>Angulo 45°</p> <p>Profundidad 15% del diente</p> <p>Ancho 5mm</p>	
P8	B1	<p>Picadura en la pista interna</p> <p>Diámetro: 2.0mm</p> <p>Profundidad 0.2mm</p>	
P9	B1	<p>Picadura en la pista externa</p> <p>Diámetro 2.0mm</p> <p>Profundidad: 0.2mm</p>	
P10	B1	<p>Picadura en la bola</p> <p>Diámetro: 2.0mm</p> <p>Cono: 120°</p>	
P11	Eje de Z4	<p>Excentricidad del eje de Z4</p>	

4.4.3 Plan experimental tres



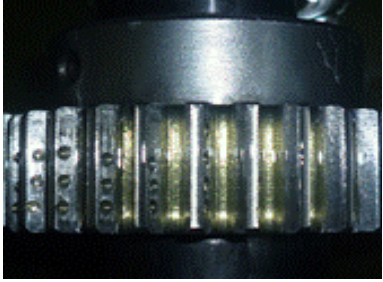
La configuración de la caja de engranajes y condiciones específicas de la prueba se detalla a continuación:







- a. Numero de etapas: 1
- b. Engranajes montados: Z1 y Z2 con dientes rectos

- c. Número de condiciones: 10
- d. Número de velocidades de la prueba (5): 5Hz, 10Hz, 15Hz, 20Hz y 25 Hz
- e. Cargas de la prueba (3): 10%, 50% y 90%
- f. Número de repeticiones de cada prueba: 10
- g. Total de señales adquiridas: 1500
- h. Frecuencia de muestreo: 11045 S/s
- i. Tiempo de muestreo: 10 segundos
- j. Nombre de la base de dato: DB3
- k. Sensibilidad del acelerómetro: 330 mV/g
- l. Equipo empleado para la adquisición: Digivive MX-300

En la siguiente tabla se detalla el montaje de la caja de engranajes y la descripción de los fallos.

Tabla 7. Configuración de los fallos para el plan experimental tres.

Condiciones o Clases	Z1	Z2	Fotografía del fallo
P1	Normal	Normal	
P2	Normal	Grieta en el diente Profundidad: 4mm Longitud: 3mm Ancho: 1mm	
P3	Normal	Rotura del diente 10%	
P4	Picadura 100% de los dientes picados, agujeros de Diámetro 2.5m y cono a 120°	Normal	

P5	Desgaste de 0.5mm en la cara de un diente	Normal	
P6	Normal	Desalineamiento angular de 1° entre ejes	
P7	Normal	Desalineamiento angular de 2° entre ejes	
P8	Normal	Desalineamiento angular de 3° entre ejes	
P9	Normal	Rotura del diente 50%	
P10	Normal	Rotura del diente 100%	

4.4.4 Plan experimental cuatro

Para este plan experimental los datos se han tomado de <https://www.phmsociety.org/references/datasets> [206].

En la Figura 27 se indica un esquema de la caja de engranajes de la base de datos pública, misma que se puede configurar para engranajes rectos y helicoidales.

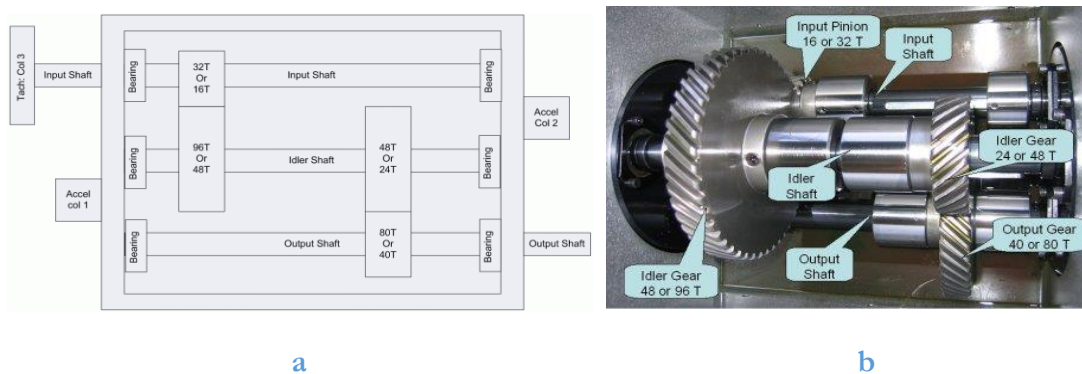


Figura 27. a) Esquema de la caja de engranajes. b) Detalle de la parte interior de la caja de engranajes rectos y helicoidales.

Los datos de la configuración se indican a continuación:

- Numero de etapas: dos etapas, trabaja con dientes rectos y helicoidales
- Número de condiciones: 14
- Número de velocidades de la prueba (5): 30Hz, 35Hz, 40Hz, 45Hz, 50Hz
- Cargas de la prueba (2): baja, alta
- Número de repeticiones de cada prueba: 2
- Total de señales adquiridas: 280
- Frecuencia de muestreo: 66.66 kS/s
- Tiempo de muestreo: 4 segundos
- Nombre de la base de datos: DB4
- Sensibilidad del acelerómetro: 10 mV/g

En las tablas 8 y 9 se presentan los fallos para la caja configurada para engranajes rectos y helicoidales con la combinación de fallos para rodamientos, ejes y engranajes.

Tabla 8. Datos de la configuración de los fallos para la caja de engranajes

Fault	Spur Gear				Bearing						Shaft	
	32T	96T	48T	80T	IS:IS	ID:IS	OS:IS	IS:OS	ID:OS	OS:OS	Input	Output
P1	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good
P2	Chipped	Good	Eccentric	Good	Good	Good	Good	Good	Good	Good	Good	Good
P3	Good	Good	Eccentric	Good	Good	Good	Good	Good	Good	Good	Good	Good
P4	Good	Good	Eccentric	Broken	Ball	Good	Good	Good	Good	Good	Good	Good
P5	Chipped	Good	Eccentric	Broken	Inner	Ball	Outer	Good	Good	Good	Good	Good
P6	Good	Good	Good	Broken	Inner	Ball	Outer	Good	Good	Good	Imbalance	Good
P7	Good	Good	Good	Good	Inner	Good	Good	Good	Good	Good	Good	Keyway Sheared
P8	Good	Good	Good	Good	Good	Ball	Outer	Good	Good	Good	Imbalance	Good
	Helical Gear				Bearing						Shaft	
	16T	48T	24T	40T	IS:IS	ID:IS	OS:IS	IS:OS	ID:OS	OS:OS	Input	Output
P9	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good
P10	Good	Good	Chipped	Good	Good	Good	Good	Good	Good	Good	Good	Good
P11	Good	Good	Broken	Good	Good	Good	Good	Combinati on	Inner	Good	Bent Shaft	Good
P12	Good	Good	Good	Good	Good	Good	Good	Combinati on	ball	Good	Imbalance	Good
P13	Good	Good	Broken	Good	Good	Good	Good	Good	Inner	Good	Good	Good
P14	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Bent Shaft	Good

Tabla 9. Datos adicionales de la caja de engranajes

Part of Bearing	
IS	Input Shaft
ID	Idler Shaft
OS	Output Shaft
:IS	Input Side
:OS	Output Side

4.5 Bases de datos

Por cada plan experimental se siguió un proceso metódico, para que las señales de vibración muestreadas sean válidas para el desarrollo del trabajo. A continuación se presenta el método que se diseñó para levantar las bases de datos.

4.5.1 Método para adquirir, procesar y almacenar las señales de vibración de las bases de datos

En la Figura 28 se presentan las fases para obtener los atributos o indicadores de condición necesarios que ingresan a un clasificador.

1. Se inicia con el montaje del sistema mecánico e instrumentos como se diseñó el plan experimental. Es importante verificar la línea base de la señal de vibración para que la base de datos sea confiable, ya que con estos datos se entrenaran las diferentes técnicas de inteligencia artificial como condición normal.
2. La siguiente fase es la adquisición y acondicionamiento de las señales; la información captada por el acelerómetro es digitalizada a través del dispositivo NI COMPAC DAQ y enviada a una PC mediante un bus Ethernet como se detalló en el apartado 4.2.
3. Para una siguiente fase, la señal digitalizada ingresa a una PC a través del programa LABVIEW, el cual es el encargado de guardar la señal discreta en disco duro con un formato binario propio de LABVIEW conocido como TDMS, posteriormente se realiza una conversión de formato y la reestructuración de la información con el fin de manipular los datos; estas últimas acciones se realizan con el programa MATLAB.
4. Cada base de datos está organizada según el plan experimental, en la etapa 4 se presenta la tabla del plan experimental en donde se tienen cuatro columnas que representan lo siguiente, para el ejemplo presentado:

- Repeticiones de la prueba denotado como R1 hasta R5 en total cinco repeticiones de la prueba.
- Frecuencia de giro representada como F o V en total seis velocidades.
- Carga de la prueba representada con L, en total tres cargas L1, L2 y L3.
- El fallo estudiado como P en total diez clases de P1 hasta P10.

Cada carpeta P que representa un fallo o clase en estudio va a contener la experimentación de la prueba a seis velocidades distintas, tres cargas y cinco repeticiones, esto implica un total de 90 muestras por cada P, para esta base de datos se tienen 900 muestras.

5. Con la base de datos cruda completa, se procede a efectuar un paso de extracción de características en los dominios de tiempo, frecuencia y tiempo-frecuencia, mediante el cálculo de parámetros estadísticos de cada una de las señales almacenadas en la base de datos como se puede apreciar en la fase 5 del esquema. Todas las características extraídas de cada una de las señales se almacenan en una estructura

organizada por diferentes dominios para facilitar el acceso al valor correspondiente de cada atributo individual. El resultado de este proceso es una estructura de datos denominada *corpus builder* la cual cuenta con todos los atributos extraídos de cada señal

- Para poder utilizar los atributos extraídos de cada señal en un clasificador es necesario modificar la presentación en la que se encuentran los datos y organizarlos en una matriz de atributos, en donde cada fila de la matriz representa una muestra (señal) y cada columna representa un atributo. Es importante adjuntar a la matriz de datos las etiquetas con el nombre/tipo de atributo a cual pertenece cada columna de la matriz y la clase a la cual corresponde cada fila. Este proceso de organización de los datos en una sola matriz se lo denomina *matriz builder* en donde se encuentran los datos con sus respectivas etiquetas; tanto de los atributos como la clase a la cual pertenecen.

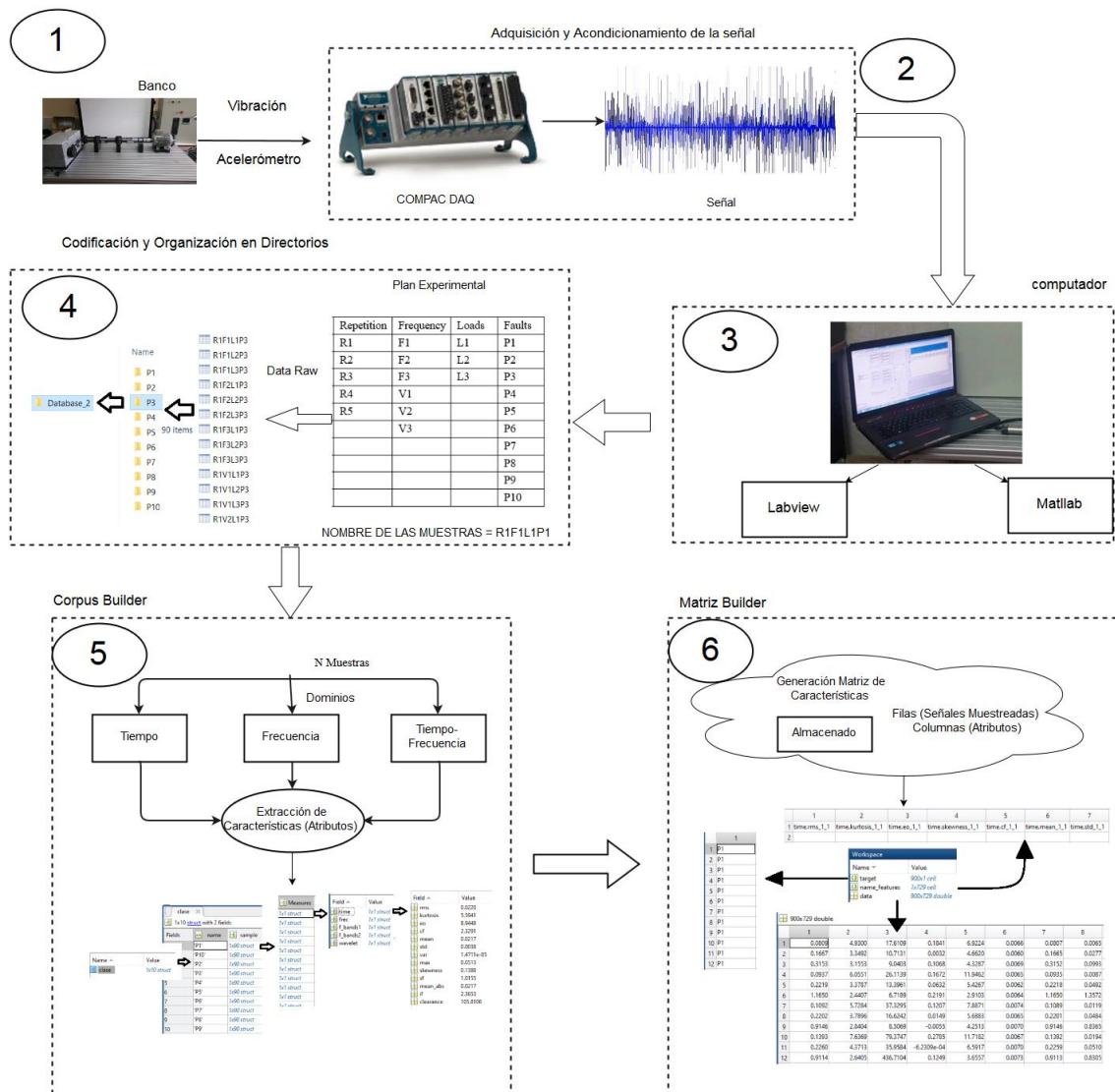


Figura 28. Proceso para adquirir, procesar y almacenar las señales de vibración.

4.6 Bases de datos crudas

Se desarrollaron cuatro planes experimentales. En la Tabla 10 se presenta un resumen de las bases de datos crudas obtenidas.

Tabla 10. Resumen de las bases de datos crudas.

Número de plan experimental	Nombre de la base de datos cruda	Número de clases o condiciones	Número de muestras
1	DB1	10	900
2	DB2	11	495
3	DB3	10	1500
4	DB4	14	280

5. IMPLEMENTACIÓN Y EVALUACIÓN DE PROPUESTAS DE SISTEMAS DE DIAGNÓSTICO DE FALLOS EN LAS CAJAS DE ENGRANAJES EMPLEANDO DIVERSAS TÉCNICAS DE INTELIGENCIA ARTIFICIAL

5.1 Propuesta general de los sistemas de diagnóstico

Los sistemas de diagnóstico de fallos de las cajas de engranajes tienen por lo general cinco fases como se presenta en la Figura 29. Para el presente trabajo se han diseñado tres sistemas de diagnóstico: el sistema uno (SD1), determina el mejor clasificador mediante técnicas estadísticas, para las pruebas del sistema se emplea una base de datos; el sistema dos (SD2), evalúa el mejor clasificador del sistema uno con otras técnicas de clasificación, se emplean cuatro bases de datos para las pruebas; el sistema tres (SD3), evalúa como el dominio de los atributos influye en el resultado del clasificador, para esto se emplea cuatro bases de datos.

Los tres sistemas de diagnóstico propuestos mantienen las fases 1 (adquisición y acondicionamiento de la señal) y 2 (extracción de atributos), el resto de fases tiene variaciones que se explicaran en cada sistema. Por lo cual a continuación se presentan las fases 1 y 2.

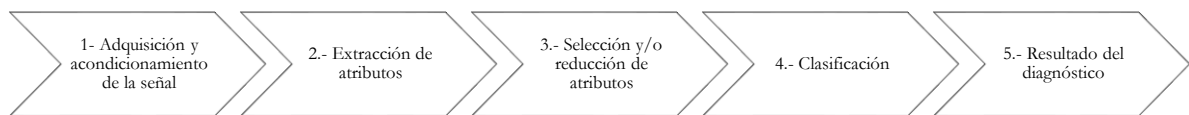


Figura 29. Fases generales de un sistema de diagnóstico.

5.1.1 Adquisición y acondicionamiento de la señal de los sistemas de diagnóstico

En la Tabla 11 se presentan las cuatro bases de datos de las señales de vibración crudas que serán utilizadas para probar los sistemas de diagnóstico, las condiciones de adquisición y acondicionamiento se presentaron en el capítulo 4; en la tabla también se indica el número de clases para cada base de datos y si la base tiene multifallo.

Tabla 11. Bases de datos crudas.

Número de plan experimental	Nombre de la base de datos cruda	Número de clases o condiciones	Número de muestras	Multifallo
1	DB1	10	900	si
2	DB2	11	495	no
3	DB3	10	1500	no
4	DB4	14	280	si

5.1.2 Extracción de los atributos para los sistemas de diagnóstico

Para cada base de datos de las señales crudas de vibración obtenidas, se procede a la extracción de atributos para encontrar parámetros representativos de las señales para caracterizar cada condición de la máquina; en la Figura 30, se presentan los tres dominios en los cuales se extrajeron los atributos, 13 atributos en el dominio del tiempo, 396 en el dominio de la frecuencia y 320 en el dominio tiempo-frecuencia, en total para cada base de datos crudos se obtuvieron 729 atributos.

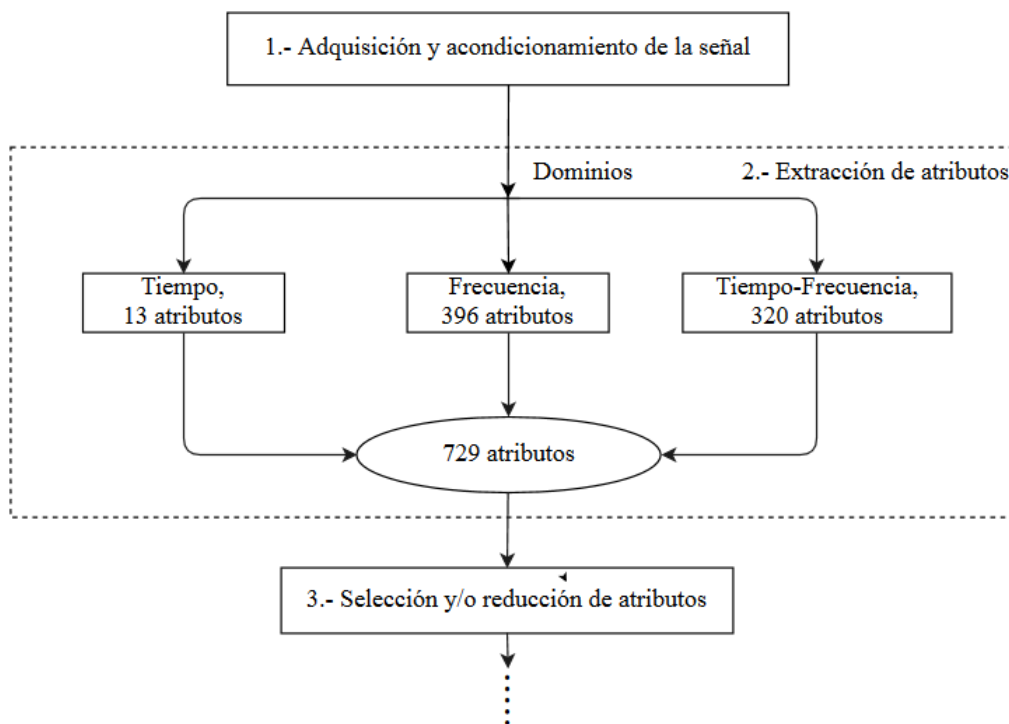


Figura 30. Proceso de extracción de atributos, las bases de datos crudas.

5.1.2.1 Extracción de atributos en el dominio de tiempo

Los atributos extraídos en el dominio del tiempo fueron: raíz cuadrática media, curtosis, operador de energía, factor de cresta, media, desviación estándar, varianza, valor pico

positivo máximo, asimetría, factor de forma, media de valores absolutos, factor de impulso y factor de holgura, todos estos explicados en el capítulo 3.

5.1.2.2 Extracción de atributos en el dominio de la frecuencia

Para obtener las características en el dominio de la frecuencia se aplica la FFT a cada uno de los datos crudos en el dominio del tiempo. En la Figura 31 se puede observar el espectro completo de frecuencia de una señal de vibración luego de aplicar la FFT.

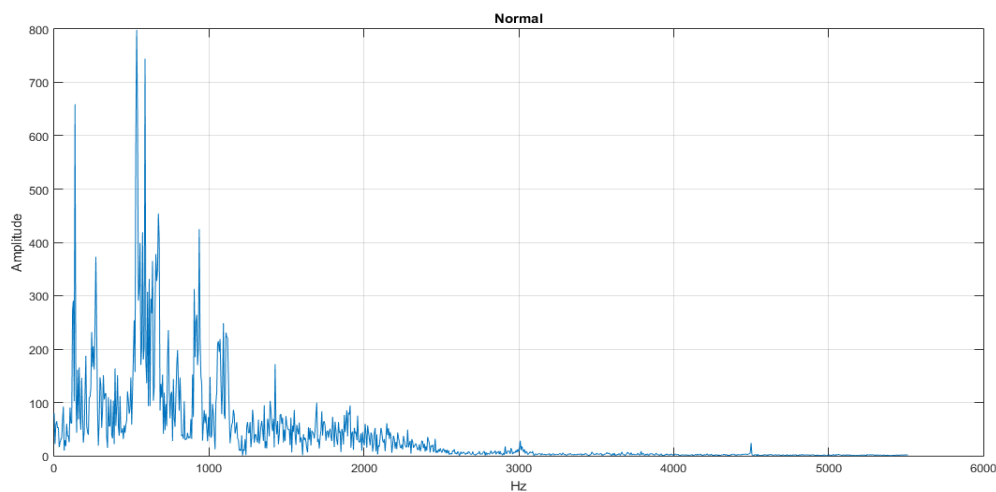


Figura 31. Espectro completo de frecuencia de una señal de vibración.

Como se puede apreciar, el espectro de la señal de vibración alcanza frecuencias máximas alrededor de 5600 Hz, sin embargo, alrededor de 3600Hz se puede observar que la amplitud de las componentes de frecuencia es cero o muy cercano, esto se mantiene para todas las bases de datos. Si se realizaran los cálculos de los atributos de condición en frecuencia utilizando el espectro completo muy posiblemente reflejarían valores erróneos debido a que, después de alcanzar ciertas frecuencias, el valor de la amplitud del espectro tiende a cero; por esta razón, se ha realizado una segmentación del espectro útil de la señal de vibración, hasta los 3600Hz.

Luego de obtener el espectro útil, en la Figura 32 se presenta el proceso que se empleó para obtener los atributos estadísticos en el dominio de la frecuencia; el proceso fue el siguiente: 1.- El espectro útil se divide en 80 y 16 bandas iguales, esta división en bandas se hace con el propósito de detectar cambios en la amplitud en el espectro vibración en una banda, generalmente estos cambios de amplitud pueden representar un fallo, estos cambios posiblemente no serían perceptibles si se analizará todo el espectro; 2.- De las 80 y 16 bandas se calcula: raíz media cuadrática, curtosis, media y desviación estándar, con lo que se obtienen

320 y 64 atributos para cada división de bandas; 3.- Del espectro útil se calculan doce atributos que son: media, varianza, coeficiente de asimetría, curtosis, frecuencia central, raíz media cuadrática, desviación estándar, PC1, PC2, PC3, PC4 y PC5. Por lo que se obtienen 396 atributos en el dominio de la frecuencia. La determinación del número de bandas fue un proceso empírico basado en obtener la mejor información para la clasificación.

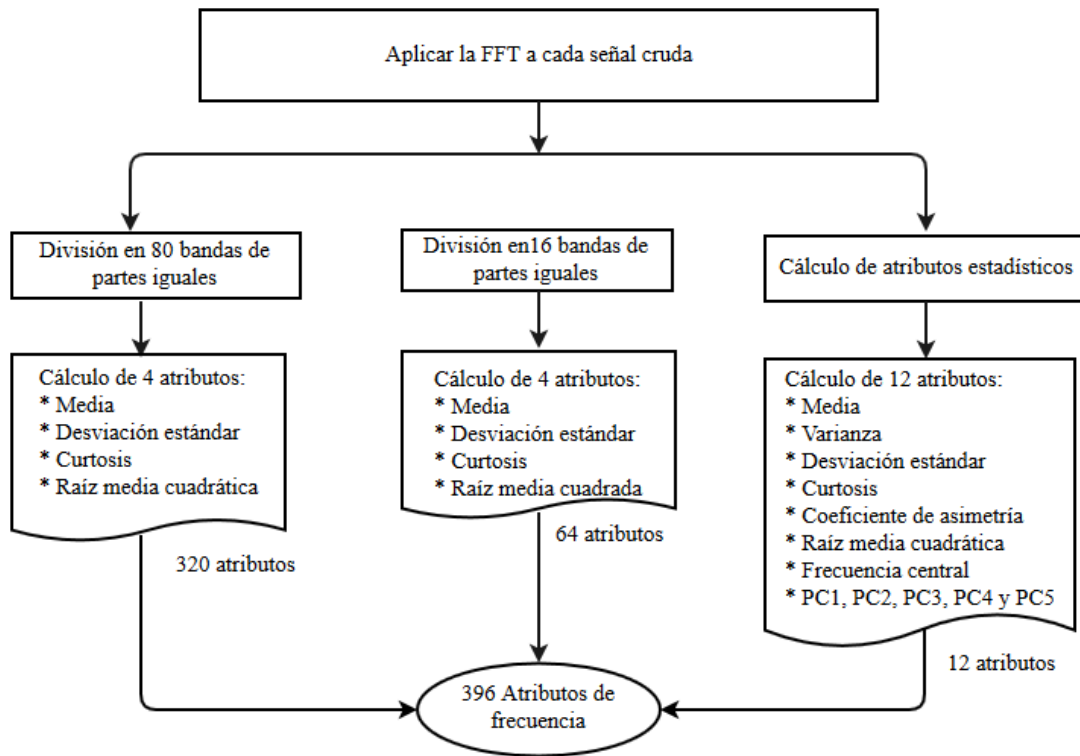


Figura 32. Proceso de extracción de atributos en el dominio de la frecuencia.

5.1.2.3 Extracción de atributos en el dominio del tiempo-frecuencia

Para extraer los atributos en el dominio del tiempo-frecuencia se utilizó la técnica de descomposición *wavelet packet (WPD)* con la finalidad de analizar la energía obtenida en los últimos niveles, para ser usados como atributos de condición para el diagnóstico de fallos.

El *Toolbox Wavelet* que proporciona MatLabR2013a fue utilizado para realizar el análisis de paquetes de wavelet. Las señales de vibración crudas en el tiempo se descomponen usando *Wavelet Packet Transform*; cinco wavelets madre fueron consideradas por este análisis: *Daubechies (db7)*, *Symlet(sym)*, *Coifler(coif4)*, *Biortogonal(bior6.8)* y *Reverse Birtogonal (rbior6.8)*. La descomposición de *wavelet* fue realizada sobre 4 niveles para cada wavelet madre obteniéndose 64 coeficientes por cada una; lo que da un total de 320 atributos en el tiempo-frecuencia.

5.1.2.4 Atributos extraídos de las bases de datos crudas

En la Tabla 12 se presenta el nombre de la base de datos de los atributos por cada plan experimental.

Tabla 12. Bases de datos.

Número de plan experimental	Nombre de la base de datos cruda	Número de condiciones o clases	Nombre de la base de datos de los atributos	Número de muestras obtenidas
1	DB1	10	DB1_729	900
2	DB2	11	DB2_729	495
3	DB3	10	DB3_729	1500
4	DB4	14	DB4_729	280

5.2 Sistema de diagnóstico uno (SD1)

En este sistema se planteó evaluar diferentes clasificadores para la detección de fallos provenientes de una caja de engranes, mediante el uso de diferentes técnicas de inteligencia artificial entre ellas diferentes redes neuronales artificiales y *random forest*.

Se emplearon las siguientes redes neuronales:

- i. Redes de reconocimiento de patrones con propagación hacia adelante (*Feed Forward Pattern Network, FFPN*).
- ii. Redes neuronales con propagación hacia adelante (*Feed Forward Neural Network, FFNN*).
- iii. Redes de reconocimiento de patrones con propagación hacia adelante, enfoque uno contra todos (*Feed Forward Pattern Network, OvA-FFPN*).
- iv. Redes neuronales probabilísticas (*Probabilistic Neural Network, RBFN-PNN*).
- v. Mapas autorganizados (*Self-organizing Maps, SOM*).

En la Figura 33 se presenta el sistema de diagnóstico uno (SD1) propuesto, el sistema está dividido en cinco fases: 1.- Adquisición y acondicionamiento de la señal de vibración, 2.- Extracción de los atributos, en esta fase se obtiene la base de datos de los atributos denominada DB1_729 que contiene 729 atributos; 3.- Selección y reducción de atributos, se remueven los atributos altamente correlacionados, obteniéndose la base de datos BD1_CORR a esta base de datos se le aplican tres métodos de selección de atributos y se obtienen las bases de datos DB1_AG-RF, DB1_LD, y DB1_PCA; 4.- fase de clasificación, se entrena y evalúan seis clasificadores: RF, FFNN, FFPN, OvA-FFPN, RBFN-PNN, SOM,

30 veces cada uno, se obtiene la métrica de exactitud en cada prueba y se guarda en un vector de resultados, luego estas métricas se usan en la fase 5 de evaluación, en donde, mediante pruebas paramétricas, no paramétricas y un procedimiento de clasificación se determinará el mejor clasificador.

A continuación se explica con detalle de la fase tres en adelante, ya que las fases uno y dos son genéricas y se explicaron en los puntos 5.1.1 y 5.1.2 de este capítulo.

5.2.1 Fase de selección y reducción de atributos

La fase 3, selección y reducción de atributos, se presenta en el diagrama de la Figura 33 y en ella se realizan cuatro actividades; (i) eliminación de atributos altamente correlacionados, (ii) selección de atributos por algoritmos genéticos y *random forest*, (iii) selección de atributos por análisis de componentes principales y (iv) selección por análisis discriminante lineal.

5.2.1.1 Eliminación de atributos altamente correlacionados

La base de datos de atributos DB1_729, que contiene 729 atributos de los tres dominios, ingresa al proceso de eliminación por correlación, de los cuales se retiraron los que superaron un umbral del 95% de correlación (546 atributos), obteniéndose así, 183 atributos no correlacionados; a este conjunto se le ha denominado DB1_CORR, los atributos obtenidos tienen la siguiente distribución en los diferentes dominios: 10 atributos pertenecientes al dominio del tiempo, 158 al dominio de la frecuencia y 15 al dominio tiempo-frecuencia.

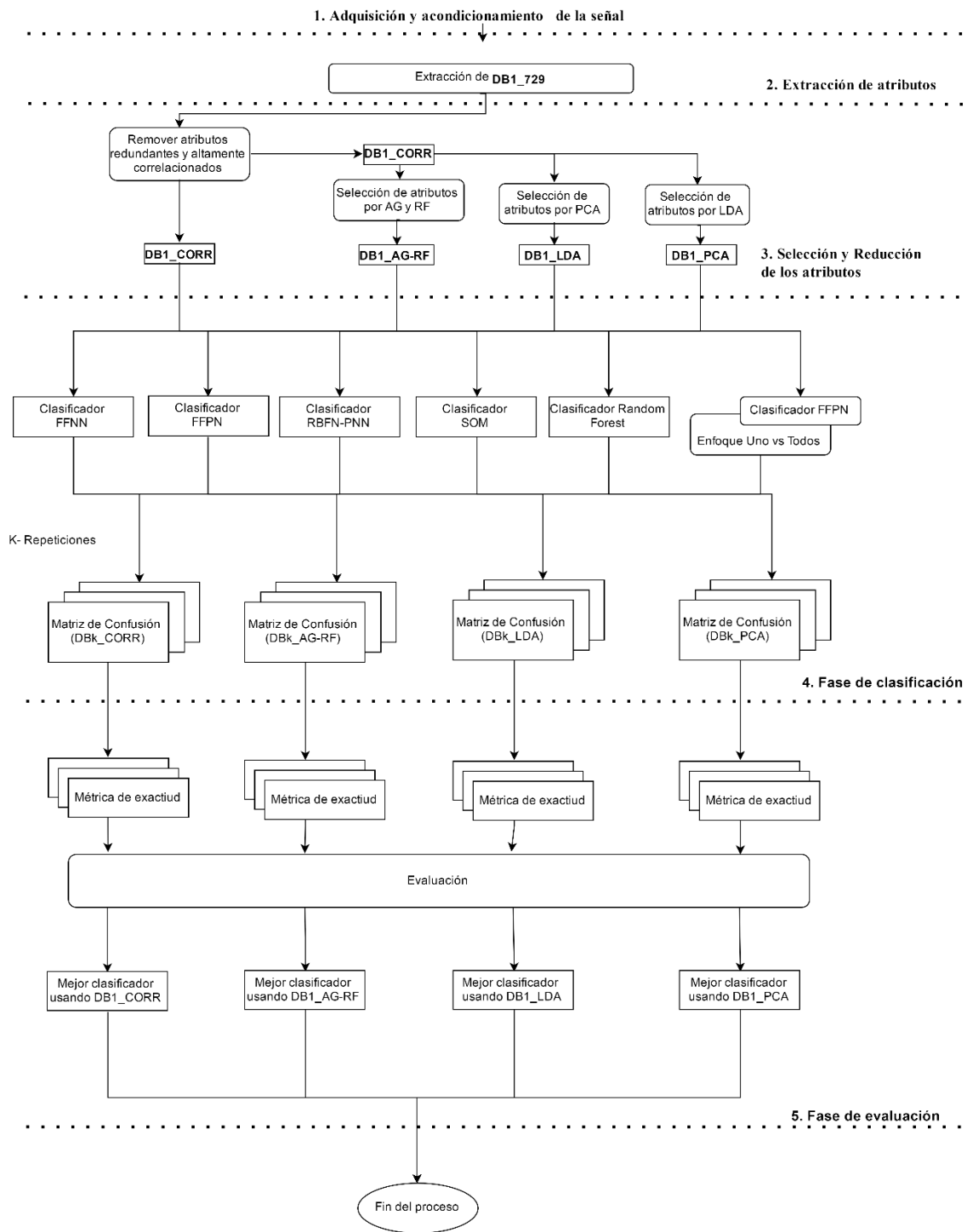


Figura 33. Diagrama actividades de las fases del sistema de diagnóstico uno.

5.2.1.2 Selección de atributos por algoritmos genéticos (AG) y *random forest* (RF)

Al conjunto de datos DB1_CORR se aplica la selección de atributos mediante el uso de algoritmos genéticos y *random forest*, para este proceso se utilizó la metodología presentada en [207]; con la ayuda de la Figura 34, se explica el procedimiento empleado.

1. Población inicial: es codificada como una población inicial de 20 individuos, donde cada uno representa un vector binario, cada elemento del vector representa un atributo.
2. Reproducción: se utilizó una probabilidad del 80%, eligiendo dos padres para formar un hijo y cuyos genes fueron seleccionados a partir de un vector binario mediante el proceso de cruce disperso.
3. Mutación: se eligió una tasa de 0.05.
4. Función *Fitness*:
 - a) *OobError*: se evalúa el error utilizando las muestras que no fueron utilizadas en el entrenamiento de cada árbol, al final se promedia el error de cada uno para obtener el error del bosque.
 - b) *FScore*: es una métrica obtenida a partir de la matriz de confusión del clasificador generada mediante el promedio ponderado del *Recall* y la precisión. Esta métrica permite evaluar el desempeño del clasificador.
 - c) Número de árboles: este valor se determina mediante la curva de convergencia del *OobError* y se verifica el número de árboles para el cual se obtuvo el menor error.

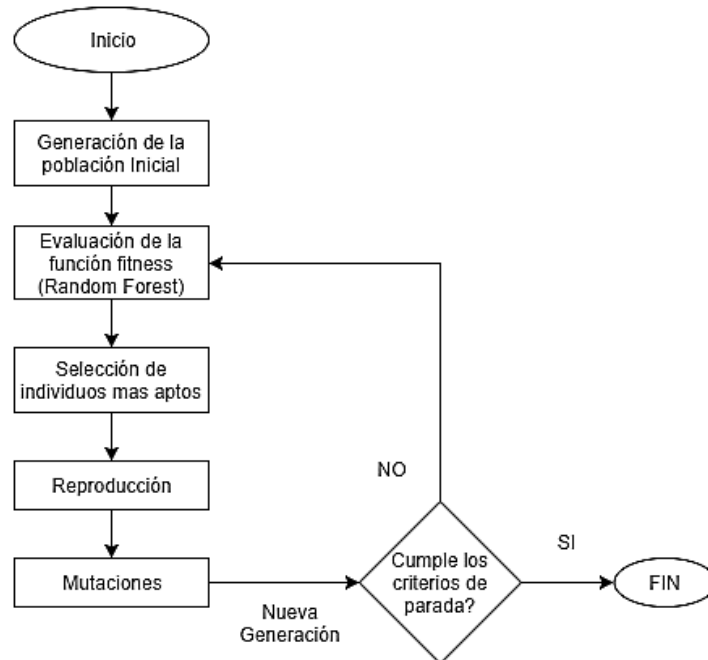


Figura 34. Proceso de selección de atributos por AG y RF.

Aplicando la selección de atributos por algoritmos genéticos y *random forest* al conjunto DB1_CORR se obtiene el conjunto DB1_AG-RF1, el cual contiene 102 atributos; 7

pertenecientes al dominio del tiempo, 87 al dominio de la frecuencia y 8 al dominio de tiempo-frecuencia.

5.2.1.3 Selección por análisis de componentes principales (PCA)

El análisis PCA se aplicó a la DB1_CORR y se obtuvo un nuevo conjunto de datos del cual se utilizaron sus 100 primeros componentes principales ya que estos representan el 95% de variabilidad de los datos originales, generando así DB1_PCA.

5.2.1.4 Selección por Análisis de Discriminante Lineal (LDA)

Se aplicó el LDA al conjunto de datos DB1_CORR, se obtuvieron 9 atributos, generando así el nuevo conjunto de datos DB1_LDA.

5.2.2 Fase de clasificación

En la fase de clasificación se entrenaron seis clasificadores: RF, FFNN, FFPN, OvA-FFPN, RBFN-PNN, SOM, se utilizaron cuatro conjuntos de datos: DB1_CORR, BD1_AG-RF, DB1_LDA y DB1_PCA; cada uno de los conjuntos de datos fueron divididos en tres subconjuntos de la siguiente manera: 70%, entrenamiento, 15% validación y 15% prueba. Los clasificadores RF, RBFN-PNN y SOM utilizaron solo los subconjuntos de entrenamiento, 70 %, y prueba, 15%, debido que estos no tienen un proceso de validación. Cada uno de los clasificadores fue entrenado un total de $k=30$ veces y en cada entrenamiento se obtuvo su matriz de confusión de la exactitud, obteniendo así un conjunto de datos de 1×30 por cada clasificador, denominado métricas de exactitud en la Figura 33, estas métricas se emplearán en la fase de evaluación.

Las redes neuronales utilizadas para este entrenamiento tienen la característica de ser redes con propagación hacia adelante, por otra parte, el número de capas y neuronas se seleccionaron de forma empírica hasta conseguir un buen rendimiento; esto se consiguió variando el número de capas de cada red y el número de neuronas de las mismas; los parámetros de arquitectura se presentan en la

Tabla 13 en la que también se detallan los parámetros de RF.

Tabla 13. Parámetros de entrenamiento para los clasificadores.

Clasificador	Parámetros
FFPN	<ul style="list-style-type: none"> • Número de capas ocultas: 1 • Neuronas: 45 • Método de entrenamiento: Gradiente desecinte estocástico • Función de transferencia: Tangente Sigmoidal Hiperbólica • Función de rendimiento: Error mínimo cuadrático • Conjunto de entrenamiento: 70% de los datos • Conjunto de validación: 15% de los datos • Conjunto de prueba: 15% de los datos
FFNN	<ul style="list-style-type: none"> • Número de capas ocultas: 3. • Neuronas: 40 neuronas por capa. • Método de entrenamiento: Gradiente desecinte estocástico • Función de transferencia: Tangente Sigmoidal Hiperbólica • Función de rendimiento: Error mínimo cuadrático • Conjunto de entrenamiento: 70% de los datos • Conjunto de validación: 15% de los datos • Conjunto de prueba: 15% de los datos
OvA-FFPN	<ul style="list-style-type: none"> • Número de capas ocultas: 1 • Número de neuronas: 45 • Entrenamiento: Gradiente desecinte estocástico. • Función de transferencia: Tangente Sigmoidal Hiperbólica • Función de rendimiento: Error mínimo cuadrático • Conjunto de entrenamiento: 70% de los datos • Conjunto de validación: 15% de los datos • Conjunto de prueba: 15% de los datos
RBFN-PNN	<ul style="list-style-type: none"> • Diversificación: valor aleatorio entre 0 y 1. • Conjunto de entrenamiento: 70% de los datos • Conjunto de prueba: 15% de los datos
SOM	<ul style="list-style-type: none"> • Tamaño de los vecinos iniciales: 10 • Distancia Función: Distancia Euclidiana • Función Topología: Hexagonal. • Número de operaciones de entrenamiento para la convergencia inicial del espacio de entrada: 500 • Conjunto de entrenamiento: 70% de los datos • Conjunto de prueba: 15% de los datos
RF	<ul style="list-style-type: none"> • Número de árboles: 300 • Error fuera de bolsa: Activado • Importancia de atributos: Activado • Selección con reemplazo: Activado • Conjunto de entrenamiento: 70% de los datos • Conjunto de prueba: 15% de los datos

Enfoque unos contra todos con ANN (OvA-FFNN)

Este enfoque permite convertir un problema de K clases a uno de K clases binarias, mediante la creación de un modelo $y_k(x)$ para cada clase, donde las muestras pertenecientes a la clase C_k son etiquetadas como positivas y las muestras del resto de clases como negativas, véase

la Figura 35. Para la evaluación de una muestra x se utiliza la Ecuación 66 donde se elige como ganador al modelo que haya generado el máximo valor de clasificación respecto al resto [208].

$$Y(x) = \max_k \{y_k(x)\}$$

Ecuación 66

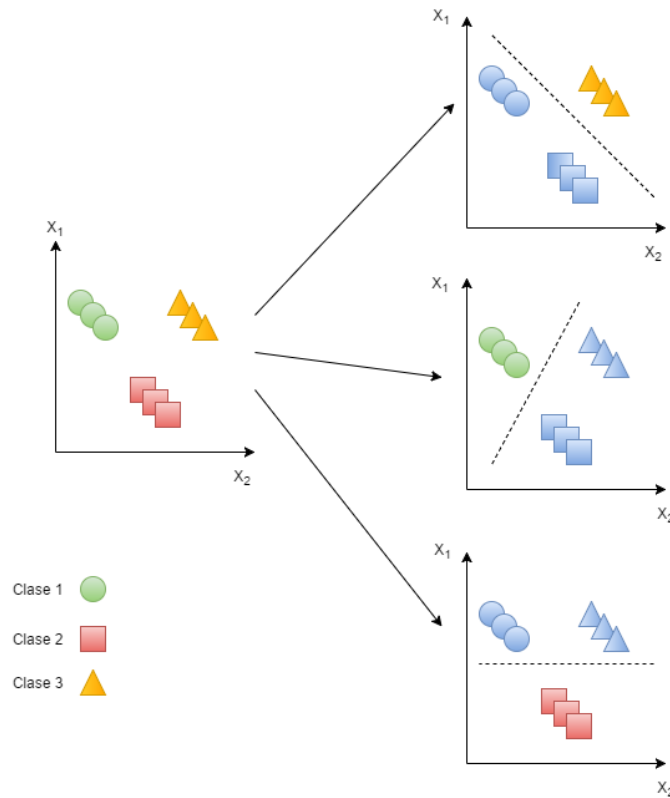


Figura 35. Enfoque uno contra todos.

5.2.3 Fase de evaluación

En esta etapa se utilizaron las cuatro bases datos (DB_AG-RF, DB_CORR, DB_LDA, DB_PCA), pero se toma como referencia el conjunto de datos DB1_AG-RF para la explicación de la metodología que se presenta en la Figura 36. Los diferentes conjuntos de datos (cada uno con sus respectivas matrices de confusión) obtenidos en la etapa de clasificación permiten calcular la métrica de exactitud de cada prueba, obteniendo así un vector de exactitudes ACC [1x30], que se representa mediante gráficos Q-Q Normal; cada una de estas poblaciones son utilizadas como parámetros de comparación entre los clasificadores, el proceso de evaluación es el siguiente:

- Aplicando pruebas paramétricas para comparar si cada población pertenece a una distribución normal.
- Se procede al uso de pruebas no paramétricas; se inicia identificando si todas las poblaciones son diferentes, en caso de rechazarse esta hipótesis.
- Un enfoque Post-Hoc es aplicado para identificar los pares de poblaciones que son diferentes.
- Finalmente, un proceso de clasificación por pares se selecciona la población ganadora, que sería el mejor clasificador.

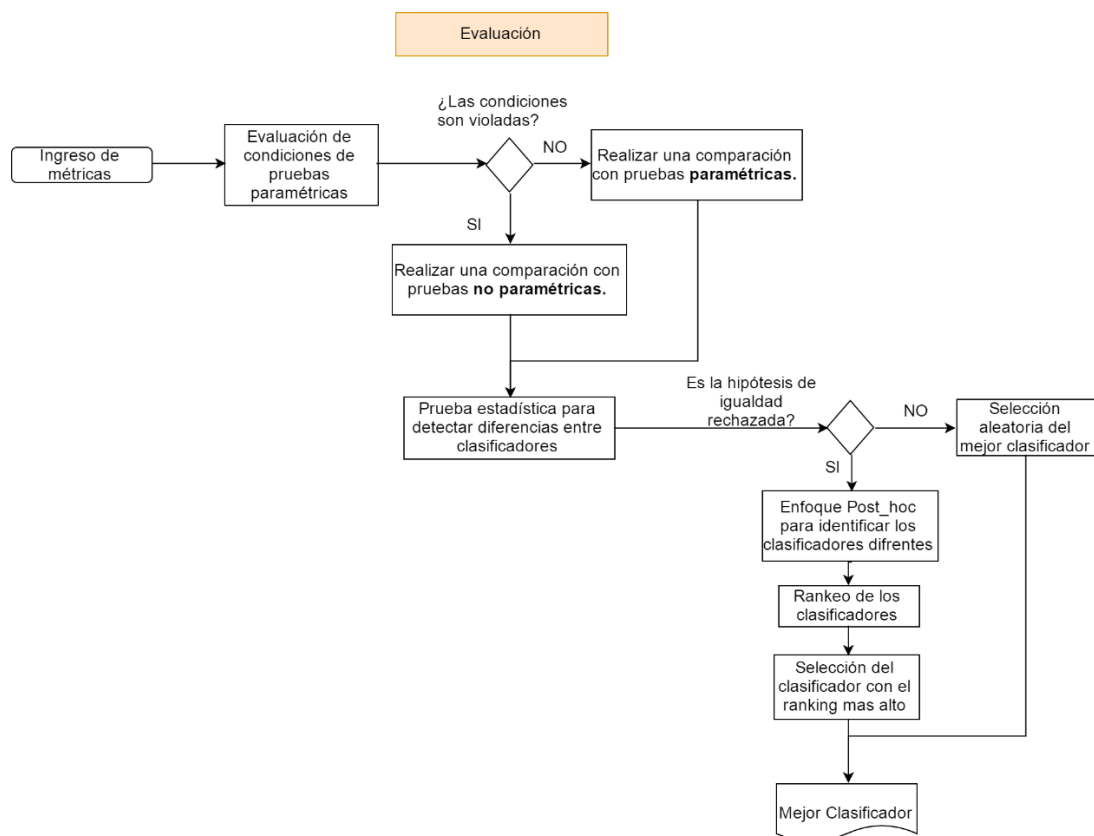


Figura 36. Diagrama de evaluación.

5.2.3.1 Gráficos Q-Q Normal

El vector ACC de cada clasificador se representa por una gráfica Q-Q Normal, la cual es un método gráfico que permite diferenciar entre una población con muestras aleatorias y una población normal. En la Figura 37, se presentan las gráficas Q-Q Normal para los 6 clasificadores utilizados en el sistema de diagnóstico, siendo el clasificador SOM (Figura 37-e) el único cuya distribución tiene un comportamiento de distribución normal; el comportamiento del resto de clasificadores no pertenece a una distribución normal.

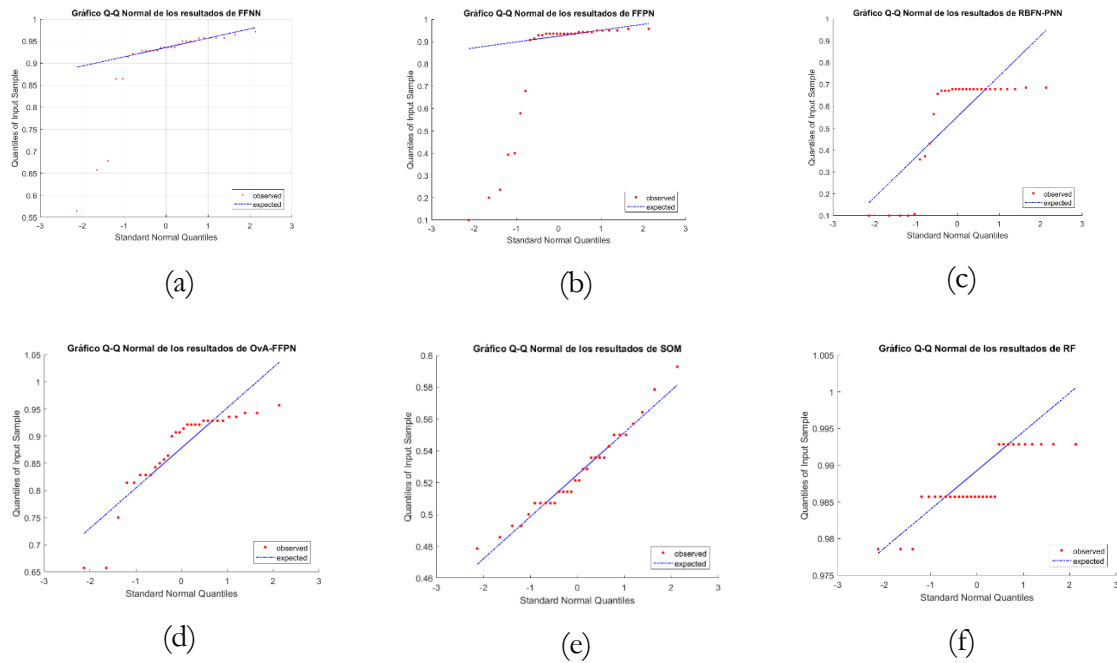


Figura 37. Graficas Q-Q Normal para de los diferentes clasificadores utilizando DB_AG-RF. (a) FFNN, (b) FFPN, (c) RBFN-PNN, (d) OvA-FFPN, (e) SOM y (f) RF.

5.2.3.2 Pruebas paramétricas

En este apartado se utilizó la normalidad a fin de evaluar las condiciones de las pruebas paramétricas. En la Tabla 14, se observan los resultados de la prueba de Shapiro Wilk, el clasificador SOM acepta la hipótesis nula validando la condición de normalidad, el resto de clasificadores rechazan la misma; por lo cual se procede a aplicar pruebas no paramétricas.

Tabla 14. Condiciones de normalidad usando la prueba de Shapiro Wilk ($\alpha = 0.05$), respecto a la exactitud.

Clasificadores	p-valor
FFNN	2.2077e-07
FFPN	6.3374e-07
RBFN-PNN	1.1985e-06
OvA-FFPN	1.6163e-04
SOM	0.5312
RF	2.0924e-05

5.2.3.3 Pruebas no paramétricas

Este apartado se inicia aplicando la prueba de Friedman que permite identificar si un grupo de poblaciones son iguales, es decir tienen las mismas distribuciones, y por otra parte su hipótesis alternativa dice que por lo menos una población difiere del resto; luego se aplica un enfoque Post-Hoc que permitirá identificar los pares de poblaciones que son diferentes,

dichas poblaciones entran a una prueba de clasificación por Wilcoxon que permitirá identificar la población ganadora, o sea el mejor clasificador.

5.2.3.3.1 Prueba de Friedman

Los valores de la clasificación y su p-valor se presentan en la Tabla 15. En consecuencia la hipótesis de igualdad es rechazada dado que su *p-valor* es inferior a 0.05. Esto quiere decir que existe como mínimo una o más poblaciones diferentes.

Tabla 15. Clasificación y p-valor de la prueba de Friedman.

Clasificador	Clasificación
RF	17
FFNN	11.95
FFPN	10.15
OvA-FFPN	9.6
RBFN-PNN	5.05
SOM	3.3
p-valor	6.876e-26

5.2.3.3.2 Post-Hoc

Este enfoque permite detectar la diferencia entre dos clasificadores mediante las pruebas de Holm y Shaffer, cuyos resultados se presentan en la Tabla 16. Los p-valores menores a 0.05, que van desde el 1 al 11 serán utilizados para la etapa de clasificación por Wilcoxon dado que estas son los pares de poblaciones diferentes. Por otra parte, los pares 12, 13, 14 y 15 se eliminan dado que estas poblaciones son iguales.

Tabla 16. Prueba de Holm y Shaffer, alpha=0.05

Pares	Clasificador	Holm	Shaffer	p-valor
1	SOM vs RF	0.00333	0.00333	3.264417E-21
1	RBFN-PNN vs RF	0.00357	0.005	1.633968E-16
3	FFNN vs SOM	0.00384	0.005	5.507836E-9
4	FFPN vs RF	0.00416	0.005	1.144788E-6
5	OvA-FFPN vs RF	0.00454	0.005	1.144788E-6
6	FFPN vs RBFN-PNN	0.005	0.005	3.774293E-6
7	OvA-FFPN vs SOM	0.00555	0.00714	4.455110E-6
8	FFPN vs SOM	0.00625	0.00714	2.913813E-4
9	FFPN vs RF	0.007142	0.00714	7.213815E-4
10	RBFN-PNN vs OvA-FFPN	0.00833	0.00833	7.213815E-4
11	FFPN vs RBFN-PNN	0.01	0.0125	7.213815E-4
12	FFNN vs FFPN	0.0125	0.0125	0.214193
13	FFNN vs OvA-FFPN	0.01666	0.01666	0.214193
14	RBFN-PNN vs SOM	0.025	0.025	0.227195
15	FFPN vs OvA-FFPN	0.05	0.05	0.999999

5.2.3.3.3 Clasificación por Wilcoxon

Para elegir el mejor clasificador se utilizó la clasificación mediante la prueba de Wilcoxon, esta prueba implementa una comparación entre pares de poblaciones identificando un ganador y posteriormente se realiza un conteo de las veces ganadas de cada clasificador; los resultados de esta prueba se presentan en la Tabla 17, teniendo como ganador al clasificador *Random Forest*.

Tabla 17. Clasificación de clasificadores por Wilcoxon.

Clasificador	Ranking
FFNN	2
FFPN	3
RBFN-PNN	0
OvA-FFPN	2
SOM	0
RF	5

5.2.3.3.4 Resultados de la evaluación de los clasificadores

En este apartado se presentan los resultados obtenidos utilizando los conjuntos de datos restantes DB1_CORR, DB1_LDA y DB1_PCA, pertenecientes a la fase de reducción (Figura 33). El proceso de evaluación es el mismo mencionado en los puntos anteriores y los resultados finales de la clasificación se presentan en la Tabla 18 donde el mejor clasificador fue *Random Forest* para todos los conjuntos de datos obtenidos mediante diferentes métodos de selección y reducción de atributos.

Tabla 18. Clasificación de los clasificadores para diferentes métodos de selección.

Clasificador	Conjunto de datos			
	DB1_CORR	DB1_PCA	DB1_LDA	DB_AG-RF
FFNN	1	1	2	2
FFPN	2	1	2	3
RBFN-PNN	1	2	0	0
OvA-FFPN	1	1	2	2
SOM	0	0	0	0
RF	5	5	4	5

5.2.4 Resultados del sistema de diagnóstico uno

Se seleccionaron y evaluaron seis clasificadores por medio de la matriz de confusión, usando como métrica la exactitud del clasificador para las pruebas estadísticas; de la caja de engranajes se adquirió una base de datos cruda, la misma de la que se extrajo el conjunto de

características para luego seleccionar las características por cuatro distintos métodos de selección, obteniéndose cuatro subconjuntos, uno por cada método de selección; cada subconjunto es ingresado en los seis clasificadores, cada clasificador es entrenado 30 veces para cada subconjunto y se obtiene la matriz de confusión de la exactitud que fue evaluada mediante pruebas estadísticas que determinaron como mejor clasificador a RF.

En la sección 5.2.3 se realizó la evaluación de los seis clasificadores con los cuatro subconjuntos de atributos obtenidos por los cuatro métodos de selección de atributos. Los atributos, cómo se indicó en el capítulo 3, son indicadores basados en fundamentos matemáticos y estadísticos que se usan para reflejar las características de la señal.

Con las gráficas Q-Q normal se determinó que el clasificador SOM (Mapas auto-organizados) fue el único que pertenecía a una distribución normal; con la prueba de Shapiro Wilk se comprobó que el clasificador SOM acepta la hipótesis nula validando la condición de normalidad, el resto de clasificadores rechazan la misma; por lo cual se procede a aplicar pruebas no paramétricas para la evaluación de los clasificadores. El hecho de que se deban emplear pruebas no paramétricas concuerda con los resultados presentados por Luengo et al. [209] que evaluaron diferentes ANN.

Para el análisis no paramétrico se aplicó la prueba de Friedman que determinó que como mínimo una población es diferente; el enfoque Post-Hoc permitió detectar la diferencia entre clasificadores mediante las pruebas de Holm y Shaffer, y los clasificadores que resultaron diferentes se clasificaron mediante la prueba de Wilcoxon y se determinó como mejor clasificador a *Random Forest (RF)* para los cuatro conjuntos de datos, lo que demuestra que RF es un clasificador adecuado para clasificar patrones de atributos de señales de vibración, este es un resultado importante debido que hasta el momento RF había sido encontrado como el mejor clasificador en otros campos [210] y ahora se determinó que fue el mejor clasificador en el diagnóstico de fallos en cajas de engranajes mediante la evaluación de pruebas estadísticas.

5.3 Sistema de diagnóstico dos (SD2)

En el sistema dos se busca evaluar cómo afecta la normalización y la selección de atributos a la exactitud de tres técnicas de clasificación; para esta evaluación se emplearon como técnicas de clasificación: redes neuronales de propagación hacia adelante (FFNN), Máquinas de Soporte Vectorial (SVM) y *Random Forest (RF)*. Se utilizaron cuatro bases de datos diferentes para una mejor evaluación.

El diagrama de actividades de las fases del sistema de diagnóstico dos se presenta en la Figura 38 y se detalla a continuación:

Fase 1, adquisición y acondicionamiento de la señal. Esta fase se explicó para cada uno de los planes experimentales en el capítulo 4; para evaluar este sistema se emplearon cuatro bases de datos (DBn) (n = 1, 2, 3 4).

Fase 2, extracción de atributos. Se generaron un total de 729 atributos, pertenecientes a los dominios de tiempo, frecuencia y tiempo-frecuencia. Posteriormente se realizó la normalización de los atributos (x) a la media (u) y la desviación estándar (std), Ecuación 67.

$$Z = \frac{x - u}{std} \quad \text{Ecuación 67}$$

En esta fase se obtienen dos conjuntos de datos DBn_729N, que es el conjunto de datos normalizados(N), y DBn_729NN, que es el conjunto de atributos no normalizados (NN).

Fase 3, selección de atributos., Los dos conjuntos de datos anteriores DBn_729N y DBn_729NN se ingresan al proceso de eliminación por correlación, se retiraran los que superaran un umbral del 95% y se obtienen los dos nuevos subconjuntos de datos DBn_1N y DBn_1NN; a estos dos nuevos subconjuntos de datos se le aplicó la selección de atributos por *Random Forest* identificando los 14 mejores atributos y se obtuvieron dos nuevos subconjuntos de datos DBn_2N y DBn_2NN.

Fase 4, clasificación. Se utilizan tres tipos de clasificadores: redes neuronales de propagación hacia adelante (FFPN), *random forest* y máquinas de soporte vectorial, la configuración de los dos primeros clasificadores se presentó en la Tabla 13 y para SVM se emplea *kernel* gaussiano, la distancia euclidiana y el parámetro C de 1; los subconjuntos de entrenamiento utilizados fueron DBn_1N, DBn_1NN, DBn_2N y DBn_2NN.

1. Adquisición y acondicionamiento de la señal (cuatro base de datos DB1, DB2, DB3 Y DB4)

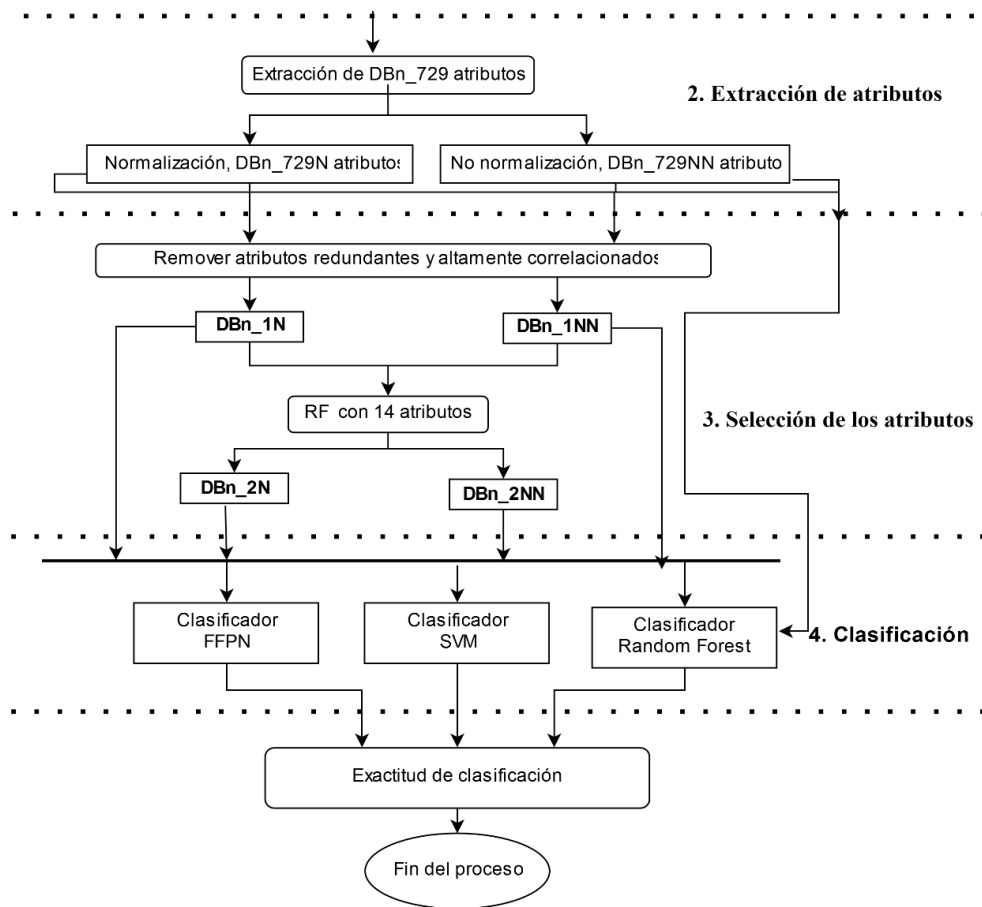


Figura 38. Diagrama actividades de las fases del sistema de diagnóstico dos.

En la Tabla 19 y Tabla 20 se presentan los resultados de la exactitud de la clasificación para los diferentes subconjuntos de datos normalizados y no normalizados, respectivamente. Los porcentajes más altos de exactitud en la clasificación se presentaron para el clasificador RF, también tuvo buen rendimiento el clasificador FFPN y el clasificador con menor rendimiento fue SVM, que no fue el mejor clasificador para ninguno de los subconjuntos de datos.

Tabla 19. Exactitud de clasificación, para los conjuntos de datos con atributos normalizados.

Subconjunto de datos	Clasificador		
	FFPN	RF	SVM
DBn_1N, selección por correlación			
DB1_1N	0.95	1	0.1
DB2_1N	1	1	0.909
DB3_1N	0.8956	0.9533	0.1000
DB4_1N	0.8675	0.9277	0.0964
DBn_2N, selección por RF			
DB1_2N	0.9929	1	0.9857
DB2_2N	1	1	0.974
DB3_2N	0.633	0.8067	0.73 56
DB4_2N	0.8824	0.8824	0.8039

Tabla 20. Exactitud de clasificación datos no normalizados.

Subconjunto	Clasificador		
	FFPN	RF	SVM
DBn_1N, selección por correlación			
DB1_1NN	0.5929	1	0.1
DB2_1NN	1	1	0.1039
DB3_1NN	0.90	0.9600	0.100
DB4_1NN	0.8554	0.8916	0.0723
DBn_2N, selección RF			
DB1_2NN	1	1	0.7714
DB2_2NN	1	0.9870	0.9221
DB3_2NN	0.7111	0.8311	0.1422
DB4_2NN	0.7229	0.8795	0.0723

Para las bases de datos DB3 y DB4 ninguno de los clasificadores alcanzó el 100% de exactitud; por esa razón se intentó con los conjuntos DB3_729_N y DB3_729_NN, conjuntos que no tienen ningún proceso de selección de atributos; los resultados se presentan en la Tabla 21.

Tabla 21. Exactitud de clasificación para los atributos extraídos normalizados y no normalizados

Conjunto	RF
DB1_729_N	0.9786
DB2_729_N	1
DB3_729_N	0.9422
DB4_729_N	0.8431
DB1_729_NN	0.9786
DB2_729_NN	1
DB3_729_NN	0.9489
DB4_729_NN	0.8039

5.3.1 Resultados del sistema de diagnóstico dos

En el sistema dos, se evaluó de qué manera afecta la normalización y la selección de atributos (por dos métodos) a la exactitud de tres clasificadores: FFNN, SVM y RF; para esto se tomaron cuatro conjuntos de datos de diferentes cajas de engranajes. Con base a los resultados presentados en las tablas 19 - 21 se realiza el siguiente análisis.

El clasificador RF tuvo mejor desempeño en su exactitud para todos los subconjuntos de datos, su rendimiento se ve afectado cuando el subconjunto de datos es menor, como sucede al aplicar el método de selección por RF, en donde los subconjuntos de datos tienen 14

atributos; cuando se aplica selección por correlación que entrega un mayor número de atributos y se normalizan los atributos, se tiene el mejor rendimiento del clasificador.

El clasificador FFPN fue el segundo mejor desempeño, su rendimiento es mejor cuando se normalizan los atributos, y su mejor rendimiento se obtuvo cuando se aplicó la selección por correlación que entrega un mayor número de atributos y se normalizan los atributos; el rendimiento disminuye cuando los datos no se normalizan.

En el caso del clasificador SVM, el mayor rendimiento se obtuvo cuando tiene menos atributos y los datos están normalizados y el menor rendimiento cuando el número de atributos es mayor y no están normalizados.

Que RF sea el mejor clasificador, seguido de FFPN y, por último, SVM, son resultados similares a los obtenidos en el trabajo de Liu et al. [211], que compara RF, BPNN y SVM para la clasificación de bebidas, lo que indica que RF es un clasificador apropiado para varios campos.

5.4 Sistema de diagnóstico tres (SD3)

En los sistemas de diagnóstico uno y dos, el clasificador por RF ha sido el que mejores resultados ha obtenido para las diferentes bases de datos y métodos de selección de atributos; en el sistema tres se busca identificar cómo seis atributos por cada dominio (tiempo, frecuencia, tiempo-frecuencia) y de forma conjunta (los tres dominios) pueden afectar al clasificador.

El diagrama de la Figura 39 se presenta el sistema tres, que se describe a continuación de manera detallada.

Fase 1, adquisición y acondicionamiento de la señal. Esta fase se explicó en cada uno de los planes experimentales del capítulo 4; para evaluar este sistema se utilizaron las cuatro (bases de datos (DBn) ($n = 1, 2, 3, 4$)).

Fase 2, extracción de atributos. En la Figura 39 se puede observar que se obtuvieron cuatro conjuntos para cada base de datos; DBn_T corresponde a 13 atributos en el dominio del tiempo, DBn_F corresponde a 396 atributos en el dominio de la frecuencia, DBn_TF corresponde a 320 atributos en el dominio de tiempo-frecuencia, y, DBn_729 corresponde al conjunto de los atributos en los tres dominios con 729 atributos; los cuatro conjuntos anteriores pasan por un proceso de normalización (N) con lo cual se obtienen los subconjuntos DBn_T_N, DBn_F_N, DBn_TF_N y DBn_729_N; y los subconjuntos de

atributos no normalizados (NN) se denominan DBn_T_NN, DBn_F_NN, DBn_TF_NN y DBn_729_NN, por lo tanto en esta fase se obtuvieron ocho subconjuntos de atributos.

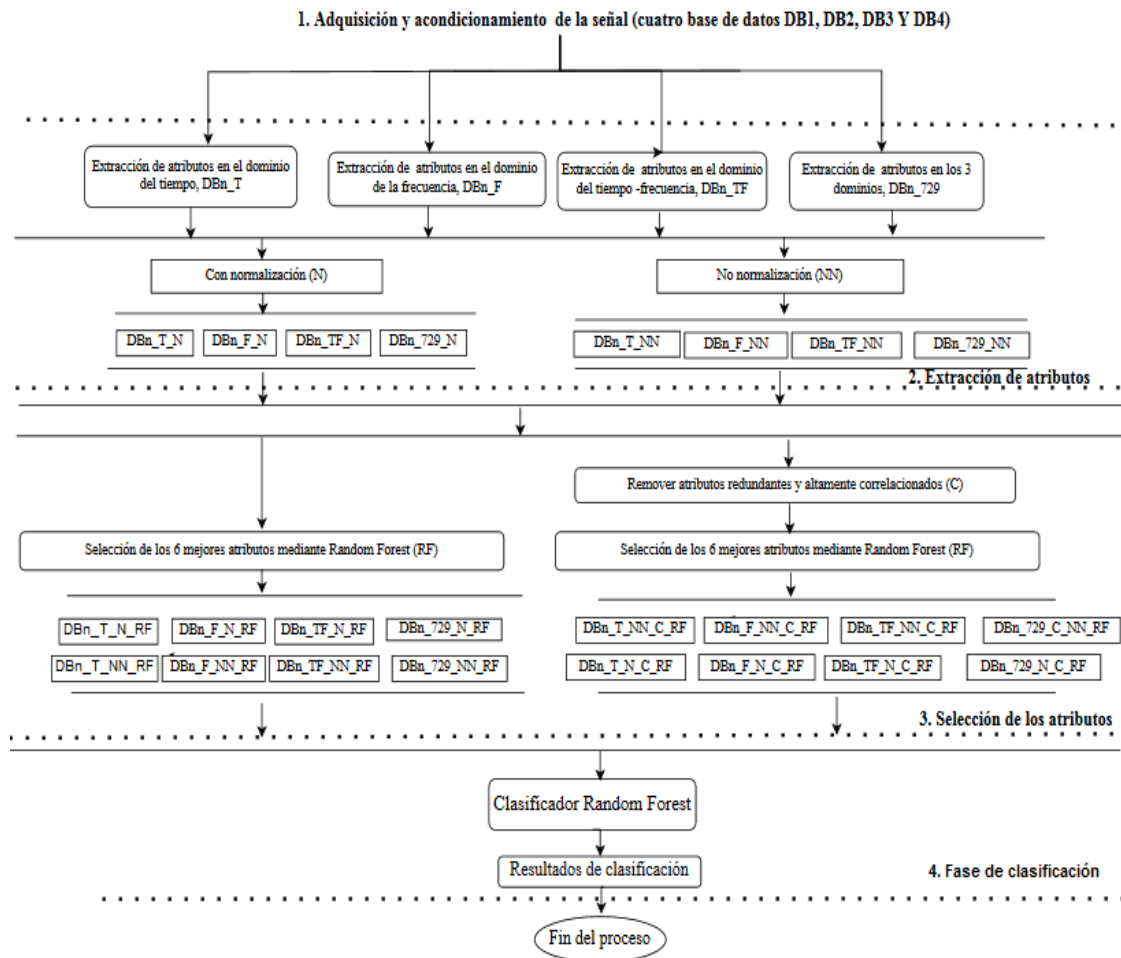


Figura 39. Diagrama actividades de las fases del sistema de diagnóstico tres.

Fase 3, selección de atributos. Se obtuvieron en esta fase 16 subconjuntos de atributos presentados en la Figura 39; los ocho primeros subconjuntos (DBn_T_NN_RF, DBn_F_NN_RF, DBn_TF_NN_RF y DBn_729_NN_RF, DBn_T_N_RF, DBn_F_N_RF, DBn_TF_N_RF y DBn_729_N_RF) se obtuvieron al aplicar la selección por RF a los ocho conjuntos obtenidos en la fase 2; los siguientes ocho subconjuntos (DBn_T_NN_C_RF, DBn_F_NN_C_RF, DBn_TF_NN_C_RF y DBn_729_NN_C_RF, DBn_T_N_C_RF, DBn_F_N_C_RF, DBn_TF_N_C_RF y DBn_729_N_C_RF) se obtuvieron al aplicar la correlación (C) y la selección por RF a los ocho conjuntos obtenidos en la fase 2.

En la fase 4, de clasificación. En esta fase se utilizó *random forest* y posteriormente se obtuvo la métrica de exactitud del entrenamiento para cada subconjunto de datos.

En las tablas 22 a 25 se presentan los resultados de la exactitud del clasificador para los diferentes dominios; para interpretar las tablas se tiene la siguiente nomenclatura:

T: subconjunto de atributos en el dominio del tiempo

F: subconjunto de atributos en el dominio de la frecuencia

TF: subconjunto de atributos en el dominio tiempo-frecuencia

729: para el subconjunto de atributos de los tres dominios anteriores

N: los atributos han sido normalizados

C: los atributos han sido reducidos por correlación antes del proceso de selección

NN: los atributos no han sido normalizados

RF: para indicar que se aplicó selección de atributos por *random forest*

Tabla 22. Exactitud del clasificador para los subconjuntos en el dominio del tiempo

	T_N_RF	T_N_C_RF	T_NN_C_RF	T_NN_RF
DB1	0,95	0,9714	0,9714	0,9429
DB2	0,9091	0,9351	0,9091	0,8961
DB3	0,6435	0,7	0,7	0,6696
DB4	0,7647	0,7451	0,7451	0,7647

Tabla 23. Exactitud del clasificador para los subconjuntos en el dominio de la frecuencia

	F_N_RF	F_N_C_RF	F_NN_C_RF	F_NN_RF
DB1	0,7857	0,8429	0,8429	0,8
DB2	0,8701	0,8831	0,8701	0,8831
DB3	0,6273	0,7578	0,7081	0,7453
DB4	0,7647	0,7843	0,8235	0,8235

Tabla 24. Exactitud del clasificador para los subconjuntos en el dominio tiempo-frecuencia

	TF_N_RF	TF_N_C_RF	TF_NN_C_RF	TF_NN_RF
DB1	0,9429	0,9643	0,9714	0,9286
DB2	0,8701	0,987	0,987	0,8571
DB3	0,4286	0,6211	0,6211	0,6832
DB4	0,8039	0,8627	0,7059	0,7451

Tabla 25. Exactitud del clasificador para los subconjuntos en los tres dominios

	729_N_RF	729_N_C_RF	729_NN_C_RF	729_NN_RF
DB1	0,9643	0,9571	0,9643	0,95
DB2	0,8831	0,961	0,9481	0,8442
DB3	0,5963	0,5714	0,5901	0,5839
DB4	0,7647	0,8235	0,8235	0,8627

5.4.1 Resultados del sistema de diagnóstico tres

En el sistema tres, como ya se ha indicado, se buscó identificar cómo seis atributos por cada dominio (tiempo, frecuencia, tiempo-frecuencia) y de forma conjunta (los tres dominios) pueden afectar al clasificador RF.

En la Figura 40 se presenta la exactitud del clasificador para las bases de datos 1 (DB1) y 2 (DB2) por cada uno de los dominios; la exactitud es menor para los atributos en el dominio de la frecuencia, para el resto de dominios la exactitud es similar.

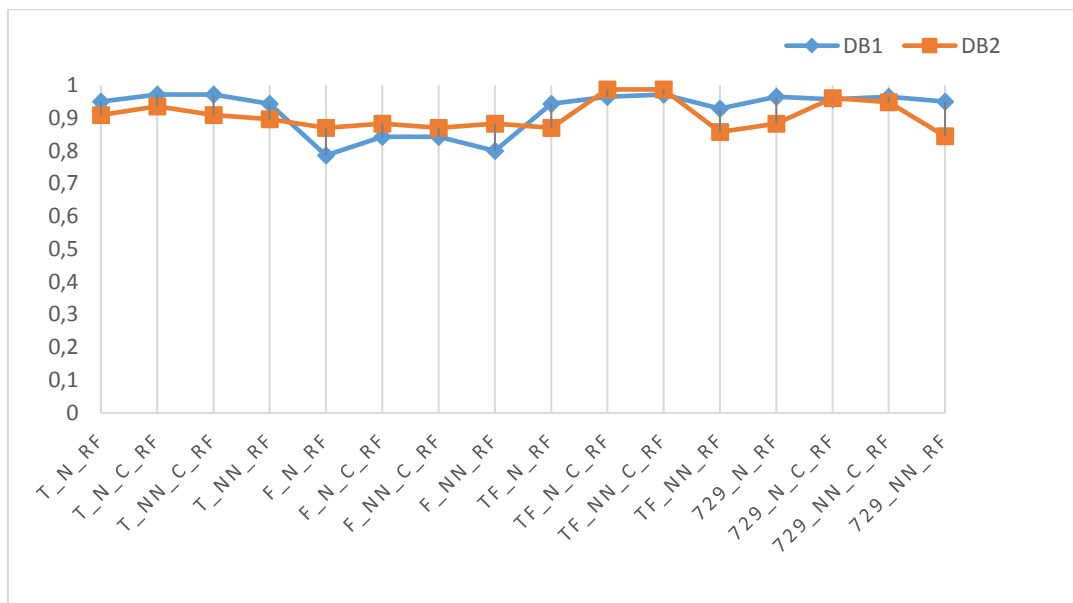


Figura 40. Exactitud del clasificador para las DB1y DB2 en los diferentes dominios.

En la

Figura 41 se presenta para las bases de datos 3 (DB3) y 4 (DB4) su exactitud por cada uno de los dominios; en el dominio de la frecuencia la DB3 tiene mejor exactitud; para los atributos de los dominios T-F y los tres dominios tiene la menor exactitud. Para la DB4 los atributos en los dominios de la frecuencia y en los tres dominios tienen mejor exactitud.

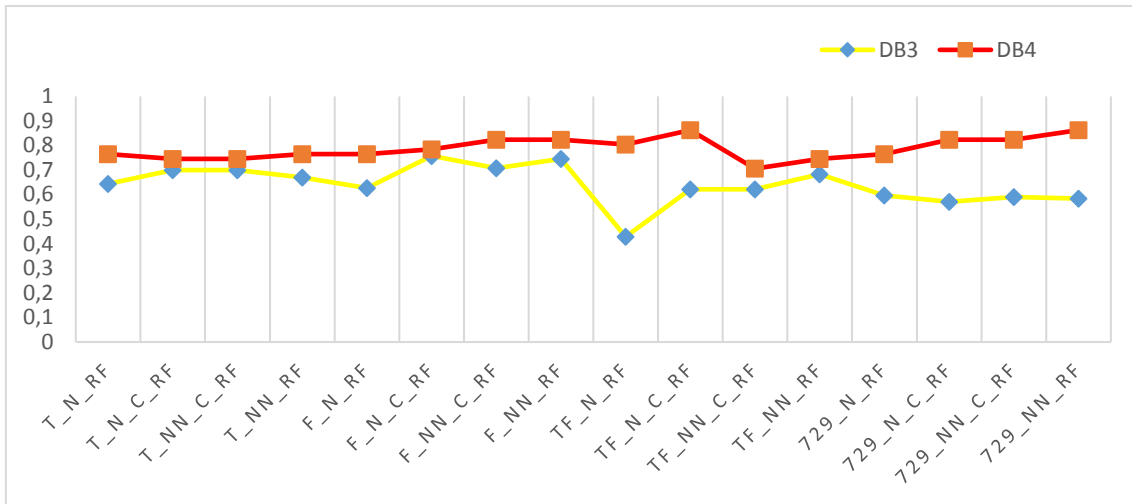


Figura 41. Exactitud del clasificador para las DB3 y DB4 en los diferentes dominios.

Los mejores resultados se presentan para los atributos en el dominio del tiempo, pero los atributos deben haber pasado por un proceso de selección por correlación; esto se nota tanto para atributos normalizados como no normalizados. En segundo lugar están los atributos en el dominio de la frecuencia; igual que en el dominio del tiempo los atributos deben haber pasado por un proceso de selección por correlación y hay una buena exactitud para atributos normalizados y no normalizados. Para los atributos del dominio tiempo-frecuencia y los atributos de los tres dominios las exactitudes son menores, pero igualmente su exactitud es mayor cuando los atributos han pasado por un proceso de correlación tanto para atributos normalizados como no normalizados.

6. CONCLUSIONES Y TRABAJOS FUTUROS

6.1 Conclusiones

En esta tesis se diagnosticaron diversos tipos de fallos en cajas de engranajes, basándose en el análisis de las señales de vibración, mediante la aplicación de diferentes técnicas de inteligencia artificial.

- Se construyó tres bases de datos de señales de vibración de una caja de engranajes con diferentes tipos de fallos y con diferente grado de severidad; una cuarta base de datos fue externa.
- Se procesaron las señales de vibración de las cuatro bases de datos en los dominios del tiempo, la frecuencia y tiempo-frecuencia y se obtuvieron los atributos (parámetros de condición) para cada dominio.
- Se diseñaron y analizaron tres sistemas de diagnóstico. En el primero, SD1, se seleccionaron y evaluaron seis clasificadores (FFNN, FFPN, RBFN-PNN, OvA-FFPN, SOM, RF) por medio de la matriz de confusión, usando como métrica la exactitud del clasificador para las pruebas estadísticas. En el segundo, SD2, evaluó cómo afecta la normalización y la selección de atributos a la exactitud de tres técnicas de clasificación: redes neuronales de propagación hacia adelante (FFNN), Máquinas de Soporte Vectorial (SVM) y *Random Forest* (RF). En el sistema tres, SD3, se buscó identificar cómo seis atributos por cada dominio (tiempo, frecuencia, tiempo-frecuencia) y de forma conjunta (los tres dominios) pueden afectar al clasificador RF.
- La técnica RF, (bosques aleatorios), cuando se trabajó con seis atributos, no proporcionó una buena exactitud para la DB3 y la DB4, esto se debe a que estas bases de datos contienen información de distintos tipos de fallos y algunos son fallos poco diferenciables; una solución para esto se encontró extrayendo los atributos y, a continuación, sometiéndolos a un proceso de correlación y normalización tras lo cual la técnica RF presentó la mayor exactitud.
- La técnica de bosques aleatorios (*Random Forest*) es la técnica que ha proporcionado los mejores resultados para los diferentes sistemas de diagnóstico evaluados y para las cuatro bases de datos ensayadas cuando los atributos se normalizan y se aplica la correlación como método de selección a los atributos extraídos. Este es un resultado importante, debido que esta técnica había demostrado su eficacia como clasificador

en otros campos pero es la primera vez que se determina, mediante la evaluación con pruebas estadísticas que de entre todos los clasificadores estudiados RF ha sido el mejor clasificador en el diagnóstico de fallos en cajas de engranajes.

6.2 Trabajos futuros

Como consecuencia del presente trabajo pueden señalarse varias líneas de investigación futuras entre las que pueden señalarse las siguientes:

- El análisis y optimización de la división en bandas del espectro FFT de la señal de vibración para la detección de fallos en engranajes.
- La utilización de otros clasificadores como los basados en lógica difusa u otros, para poder tener una mejor interoperabilidad de las relaciones entre el atributo, el síntoma y el fallo.
- Experimentar y analizar diferentes niveles de severidad empezando desde un fallo incipiente hasta un fallo severo, para varios tipos de fallos en engranajes, y determinar la relevancia de los atributos en los diferentes dominios para los distintos tipos de fallos y nivel de severidad.
- Integrar el sistema de diagnóstico desarrollado a una aplicación para realizar un diagnóstico en línea

BIBLIOGRAFÍA

- [1] S. Edwards, A. W. Lees, y M. I. Friswell, «Fault diagnosis of rotating machinery», *Shock Vib. Dig.*, vol. 30, n.º 1, pp. 4-13, 1998.
- [2] G. N. Marichal, M. L. Del Castillo, J. López, I. Padrón, y M. Artés, «An Artificial Intelligence Approach for Gears Diagnostics in AUVs», *Sensors*, vol. 16, n.º 4, p. 529, abr. 2016.
- [3] J. Kim, Y. Ahn, y H. Yeo, «A comparative study of time-based maintenance and condition-based maintenance for optimal choice of maintenance policy», *Struct. Infrastruct. Eng.*, pp. 1–12, 2016.
- [4] H. Li, E. Deloux, y L. Dieulle, «A condition-based maintenance policy for multi-component systems with Lévy copulas dependence», *Reliab. Eng. Syst. Saf.*, vol. 149, pp. 44–55, 2016.
- [5] A. Mérigaud y J. V. Ringwood, «Condition-based maintenance methods for marine renewable energy», *Renew. Sustain. Energy Rev.*, vol. 66, pp. 53–78, 2016.
- [6] B. Al-Najjar, A. Ingwald, y M. Kans, «Maintenance in Real Estate and Manufacturing Industries: Differences, Problems, Needs and Potentials-Four Case Studies», en *Proceedings of the 10th World Congress on Engineering Asset Management (WCEAM 2015)*, 2016, pp. 13–27.
- [7] W. Xu y L. Cao, «Energy efficiency analysis of machine tools with periodic maintenance», *Int. J. Prod. Res.*, vol. 52, n.º 18, pp. 5273-5285, sep. 2014.
- [8] A. Certa, M. Enea, G. Galante, y T. Lupo, «A multi-objective approach to optimize a periodic maintenance policy», *Int. J. Reliab. Qual. Saf. Eng.*, vol. 19, n.º 06, p. 1240002, dic. 2012.
- [9] D. Goyal y B. S. Pabla, «Condition based maintenance of machine tools—A review», *CIRP J. Manuf. Sci. Technol.*, vol. 10, pp. 24-35, Agosto 2015.
- [10] R. Ahmad y S. Kamaruddin, «An overview of time-based and condition-based maintenance in industrial application», *Comput. Ind. Eng.*, vol. 63, n.º 1, pp. 135–149, 2012.
- [11] A. K. S. Jardine, D. Lin, y D. Banjevic, «A review on machinery diagnostics and prognostics implementing condition-based maintenance», *Mech. Syst. Signal Process.*, vol. 20, n.º 7, pp. 1483-1510, oct. 2006.
- [12] N. Saravanan y K. I. Ramachandran, «A case study on classification of features by fast single-shot multiclass PSVM using Morlet wavelet for fault diagnosis of spur bevel gear box», *Expert Syst. Appl.*, vol. 36, n.º 8, pp. 10854-10862, oct. 2009.
- [13] P. Gupta y S. Wadhvani, «Feature Selection by Genetic Programming, And Artificial Neural Network-based Machine Condition Monitoring», *Int. J. Eng. Innov. Technol. IJEIT*, vol. 1, n.º 4, pp. 177-181, 2012.
- [14] N. R. Sakthivel, V. Sugumaran, y S. Babudevasenapati, «Vibration based fault diagnosis of monoblock centrifugal pump using decision tree», *Expert Syst. Appl.*, vol. 37, n.º 6, pp. 4040-4049, jun. 2010.
- [15] Z. Yang, W. I. Hoi, y J. Zhong, «Gearbox fault diagnosis based on artificial neural network and genetic algorithms», en *System Science and Engineering (ICSSE), 2011 International Conference on*, 2011, pp. 37–42.
- [16] Y. Lei, Z. He, Y. Zi, y Q. Hu, «Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs», *Mech. Syst. Signal Process.*, vol. 21, n.º 5, pp. 2280-2294, jul. 2007.
- [17] Y. Lei, Z. He, y Y. Zi, «A new approach to intelligent fault diagnosis of rotating machinery», *Expert Syst. Appl.*, vol. 35, n.º 4, pp. 1593-1600, nov. 2008.
- [18] J. Yin, W. Wang, Z. Man, y S. Khoo, «Statistical modeling of gear vibration signals and its application to detecting and diagnosing gear faults», *Inf. Sci.*, vol. 259, pp. 295-303, feb. 2014.

- [19] I. A. Abu-Mahfouz, «A comparative study of three artificial neural networks for the detection and classification of gear faults», *Int. J. Gen. Syst.*, vol. 34, n.º 3, pp. 261-277, jun. 2005.
- [20] H. Li, J. Zhao, X. Zhang, y H. Teng, «Gear Crack Level Classification Based on EMD and EDT», *Math. Probl. Eng.*, 2014.
- [21] R. F. M. Brandão, J. A. B. Carvalho, y F. P. M. Barbosa, «Neural Networks for Condition Monitoring of Wind Turbines Gearbox», *J. Energy Power Eng.*, vol. 6, n.º 4, pp. 638–644, 2012.
- [22] A. Prieto, M. Atencia, y F. Sandoval, «Advances in artificial neural networks and machine learning», *Neurocomputing*, vol. 121, pp. 1-4, dic. 2013.
- [23] S. Ai y H. Li, «Application of order cepstrum and neural network to gear fault detection», en *Computational Engineering in Systems Applications, IMACS Multiconference on*, 2006, vol. 2, pp. 1822–1827.
- [24] Z. Chen y X. Lian, «Fault diagnosis of gear box based on multi-weight neural network», en *Computational Intelligence and Industrial Applications, 2009. PACILA 2009. Asia-Pacific Conference on*, 2009, vol. 1, pp. 178–181.
- [25] M. A. Abed, A. N. Ismail, y Z. M. Hazi, «Pattern recognition using genetic algorithm», *Int. J. Comput. Electr. Eng.*, vol. 2, n.º 3, pp. 583–588, 2010.
- [26] E. Ebrahimi *et al.*, «Developing an Intelligent Fault Diagnosis of MF285 Tractor Gearbox Using Genetic Algorithm and Vibration Signals», *Mod. Mech. Eng.*, vol. 03, n.º 04, pp. 152-160, 2013.
- [27] J. Fang, «Gear fault diagnosis method of intelligence based on genetic algorithm to optimize the BP neural network», *Adv. Mater. Res.*, vol. 756, pp. 3674–3679, 2013.
- [28] G. Marichal, M. Artes, J. García Prada, y O. Casanova, «Extraction of rules for faulty bearing classification by a Neuro-Fuzzy approach», *Mech. Syst. Signal Process.*, 2011.
- [29] A. K. Dash, «Analysis of Adaptive Fuzzy Technique for Multiple Crack Diagnosis of Faulty Beam Using Vibration Signatures», *Adv. Fuzzy Syst.*, vol. 2013, pp. 1-16, 2013.
- [30] A. M. Acilar y A. Arslan, «A novel approach for designing adaptive fuzzy classifiers based on the combination of an artificial immune network and a memetic algorithm», *Inf. Sci.*, vol. 264, pp. 158-181, abr. 2014.
- [31] W. Yan, «Application of random forest to aircraft engine fault diagnosis», en *Computational Engineering in Systems Applications, IMACS Multiconference on*, 2006, vol. 1, pp. 468–475.
- [32] L. I. Xin-Hai, «Using “random forest” for classification and regression», *Chin. J. Appl. Entomol.*, vol. 4, p. 041, 2013.
- [33] Q. Zhou, H. Zhou, Q. Zhou, F. Yang, y L. Luo, «Structure damage detection based on random forest recursive feature elimination», *Mech. Syst. Signal Process.*, vol. 46, n.º 1, pp. 82-90, may 2014.
- [34] B.-S. Yang, X. Di, y T. Han, «Random forests classifier for machine fault diagnosis», *J. Mech. Sci. Technol.*, vol. 22, n.º 9, pp. 1716–1725, 2008.
- [35] K. Mollazade, H. Ahmadi, M. Omid, y R. Alimardani, «An intelligent model based on data mining and fuzzy logic for fault diagnosis of external gear hydraulic pumps», *Insight - Non-Destr. Test. Cond. Monit.*, vol. 51, n.º 11, pp. 594-600, nov. 2009.
- [36] X. Liu, L. Ma, y J. Mathew, «Machinery fault diagnosis based on fuzzy measure and fuzzy integral data fusion techniques», 2009.
- [37] N. Saravanan, S. Cholairajan, y K. I. Ramachandran, «Vibration-based fault diagnosis of spur bevel gear box using fuzzy technique», *Expert Syst. Appl.*, vol. 36, n.º 2, Part 2, pp. 3119-3135, mar. 2009.
- [38] H. Czichos, Ed., *Handbook of technical diagnostics: fundamentals and application to structures and systems*. Berlin: Springer, 2013.

- [39] A. J. Guillén, A. Crespo, J. F. Gómez, y M. D. Sanz, «A framework for effective management of condition based maintenance programs in the context of industrial development of E-Maintenance strategies», *Comput. Ind.*, vol. 82, pp. 170-185, oct. 2016.
- [40] A. J. G. López, A. C. Márquez, J. F. G. Fernández, y A. G. Bolaños, «Towards the Industrial Application of PHM: Challenges and Methodological Approach».
- [41] C. Lin y V. Makis, «Optimal Bayesian maintenance policy and early fault detection for a gearbox operating under varying load», *J. Vib. Control*, p. 1077546314554475, 2014.
- [42] C. Rahmoune y D. Benazzouz, «Early detection of pitting failure in gears using a spectral kurtosis analysis», *Mech. Ind.*, vol. 13, n.º 4, pp. 245-254, 2012.
- [43] M. Namdari y H. Jazayeri-Rad, «Incipient fault diagnosis using support vector machines based on monitoring continuous decision functions», *Eng. Appl. Artif. Intell.*, vol. 28, pp. 22-35, feb. 2014.
- [44] Y. H. Ali, R. A. Rahman, y R. I. R. Hamzah, «Acoustic emission signal analysis and artificial intelligence techniques in machine condition monitoring and fault diagnosis: a review», *J. Teknol.*, vol. 69, n.º 2, 2014.
- [45] V. Saxena, N. Chowdhury, y S. Devendiran, «Assessment of gearbox fault detection using vibration signal analysis and acoustic emission technique», *IOSR J. Mech. Civ. Eng. IOSRJMCE*, vol. 7, pp. 52–60, 2013.
- [46] G. Siracusano *et al.*, «A framework for the damage evaluation of acoustic emission signals through Hilbert–Huang transform», *Mech. Syst. Signal Process.*, vol. 75, pp. 109–122, 2016.
- [47] Y. Qu, D. He, J. Yoon, B. Van Hecke, E. Bechhoefer, y J. Zhu, «Gearbox Tooth Cut Fault Diagnostics Using Acoustic Emission and Vibration Sensors — A Comparative Study», *Sensors*, vol. 14, n.º 1, pp. 1372-1393, ene. 2014.
- [48] L.-S. Law, J. Kim, W. Y. H. Liew, y S.-K. Lee, «An approach to monitoring the thermomechanical behavior of a spindle bearing system using acoustic emission (AE) energy», *Int. J. Precis. Eng. Manuf.*, vol. 14, n.º 7, pp. 1169–1175, 2013.
- [49] A. M. Younus, A. Widodo, y B.-S. Yang, «Evaluation of thermography image data for machine fault diagnosis», *Nondestruct. Test. Eval.*, vol. 25, n.º 3, pp. 231-247, sep. 2010.
- [50] J. Salgueiro, G. Peršin, J. Vižintin, M. Ivanovič, y B. Dolenc, «On-line Oil Monitoring and Diagnosis», *Stroj. Vestn.-J. Mech. Eng.*, vol. 59, n.º 10, pp. 604–612, 2013.
- [51] R. K. Biswas, M. C. Majumdar, y S. K. Basu, «Vibration and Oil Analysis by Ferrography for Condition Monitoring», *J. Inst. Eng. India Ser. C*, vol. 94, n.º 3, pp. 267–274, 2013.
- [52] T. H. Loutas, D. Roulias, E. Pauly, y V. Kostopoulos, «The combined use of vibration, acoustic emission and oil debris on-line monitoring towards a more effective condition monitoring of rotating machinery», *Mech. Syst. Signal Process.*, vol. 25, n.º 4, pp. 1339-1352, may 2011.
- [53] J. Phillips, E. Cripps, J. W. Lau, y M. R. Hodkiewicz, «Classifying machinery condition using oil samples and binary logistic regression», *Mech. Syst. Signal Process.*, vol. 60, pp. 316–325, 2015.
- [54] B. R. Rajendra y S. V. Bhaskar, «Condition Monitoring of Gear Box by Using Motor Current Signature Analysis», *Int. J. Sci. Res. Publ.*, vol. 3, ago. 2013.
- [55] P. Boshkoski y A. Urevc, «Monitoring pitting gear faults using electrical current and sound emissions», en *9th International PhD Workshop on Systems and Control*, Izola, Slovenia, 2008.
- [56] C. Kar y A. R. Mohanty, «Monitoring gear vibrations through motor current signature analysis and wavelet transform», *Mech. Syst. Signal Process.*, vol. 20, n.º 1, pp. 158–187, 2006.

- [57] P. Henriquez, J. B. Alonso, M. A. Ferrer, y C. M. Travieso, «Review of Automatic Fault Diagnosis Systems Using Audio and Vibration Signals», *Syst. Man Cybern. Syst. IEEE Trans. On*, vol. 44, n.º 5, pp. 642–652, 2014.
- [58] H. Yang, J. Mathew, y L. Ma, «Vibration feature extraction techniques for fault diagnosis of rotating machinery: a literature survey», en *Asia-Pacific Vibration Conference*, Gold Coast, Australia., 2003.
- [59] H. Wang, P. Chen, y S. Wang, «Intelligent diagnosis methods for plant machinery», *Front. Mech. Eng. China*, vol. 5, n.º 1, pp. 118-124, nov. 2009.
- [60] H. Yang, J. Mathew, y L. Ma, «Intelligent diagnosis of rotating machinery faults-A review», 2002.
- [61] A. S. Sait y Y. I. Sharaf-Eldeen, «A Review of Gearbox Condition Monitoring Based on vibration Analysis Techniques Diagnostics and Prognostics», *Rotating Mach. Struct. Health Monit. Shock Vib. Vol. 5*, pp. 307–324, 2011.
- [62] V. Sharma y A. Parey, «Fault Diagnosis of Gearbox Using Various Condition Monitoring Indicators for Non-Stationary Speed Conditions: A Comparative Analysis».
- [63] M. Lebold, K. McClintic, R. Campbell, C. Byington, y K. Maynard, «Review of vibration analysis methods for gearbox diagnostics and prognostics», en *Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology*, 2000, vol. 634, p. 16.
- [64] P. Večř, M. Kreidl, y R. Šmíd, «Condition indicators for gearbox condition monitoring systems», *Acta Polytech.*, vol. 45, pp. 35-43, 2005.
- [65] Javier Sanz Corretge, y Juan José Benito, «Diagnosis de modos de fallo en máquinas rotativas mediante técnicas basadas en redes neuronales artificiales y en la transformación con ondículas.», Tesis Doctoral, UNED, 2007.
- [66] A. A. Mohammed, R. D. Neilson, W. F. Deans, y P. MacConnell, «Crack detection in a rotating shaft using artificial neural networks and PSD characterisation», *Meccanica*, vol. 49, n.º 2, pp. 255-266, feb. 2014.
- [67] A. Aherwar, «An Investigation on Gearbox Fault Detection Using Vibration Analysis Techniques a Review», *Aust. J. Mech. Eng.*, vol. 10, n.º 2, 2012.
- [68] P. Jayaswal, A. Wadhvani, y K. Mulchandani, «Machine fault signature analysis», *Int. J. Rotating Mach.*, vol. 2008, 2008.
- [69] W. Q. Wang, F. Ismail, y M. Farid Golnaraghi, «Assessment of gear damage monitoring techniques using vibration measurements», *Mech. Syst. Signal Process.*, vol. 15, n.º 5, pp. 905-922, sep. 2001.
- [70] H. Bendjama, S. Bouhouche, y M. S. Boucherit, «Application of Wavelet Transform for Fault Diagnosis in Rotating Machinery», *Int. J. Mach. Learn. Comput.*, vol. 2, n.º 1, pp. 82–87, 2012.
- [71] X. Wang, V. Makis, y M. Yang, «A wavelet approach to fault diagnosis of a gearbox under varying load conditions», *J. Sound Vib.*, vol. 329, n.º 9, pp. 1570–1585, 2010.
- [72] Y. Lei, J. Lin, Z. He, y M. J. Zuo, «A review on empirical mode decomposition in fault diagnosis of rotating machinery», *Mech. Syst. Signal Process.*, vol. 35, n.º 1-2, pp. 108-126, feb. 2013.
- [73] R. Yan, R. X. Gao, y X. Chen, «Wavelets for fault diagnosis of rotary machines: A review with applications», *Signal Process.*, vol. 96, pp. 1-15, mar. 2014.
- [74] N. Saravanan y K. I. Ramachandran, «Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN)», *Expert Syst. Appl.*, vol. 37, n.º 6, pp. 4168-4181, jun. 2010.
- [75] J. Wang, D. Kong, S. Dong, y C. Wang, «The Gearbox Fault Diagnosis Based on Wavelet Transform», 2015.
- [76] H. Chen, Y. Lu, y L. Tu, «Fault identification of gearbox degradation with optimized wavelet neural network», *Shock Vib.*, vol. 20, n.º 2, pp. 247–262, 2013.

- [77] X. Fan y M. J. Zuo, «Gearbox fault detection using Hilbert and wavelet packet transform», *Mech. Syst. Signal Process.*, vol. 20, n.º 4, pp. 966-982, may 2006.
- [78] M. J. Gómez García, «Metodología de selección de indicadores óptimos para el análisis y diagnóstico del estado de la máquina: aplicación a elementos mecánicos rotativos», Thesis Doctoral, Universidad Carlos III de Madrid, Madrid, España, 2014.
- [79] C. Castejón, M. J. Gómez, J. C. Garcia-Prada, y A. Junior, «Automatic Selection of the WPT Decomposition Level for Condition Monitoring of Rotor Elements Based on the Sensitivity Analysis of the Wavelet Energy», *Int. J. Acoust. Vib.*, vol. 20, n.º 2, pp. 95–100, 2015.
- [80] K. Medjaher, F. Camci, y N. Zerhouni, «Feature extraction and evaluation for Health Assessment and Failure prognostics.», en *Proceedings of First European Conference of the Prognostics and Health Management Society, PHM-E'12.*, 2012, pp. 111–116.
- [81] A. Saxena y G. Vachtsevanos, «Optimum feature selection and extraction for fault diagnosis and prognosis», en *Proceedings of the 2007 AAAI Fall Symposium on Artificial Intelligence for Prognostics, Arlington, VA*, 2007.
- [82] X. Geng, T.-Y. Liu, T. Qin, y H. Li, «Feature selection for ranking», en *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007, pp. 407–414.
- [83] I. Guyon y A. Elisseeff, «An introduction to variable and feature selection», *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.
- [84] G. G. Yen y W. F. Leong, «Fault classification on vibration data with wavelet based feature selection scheme», en *31st Annual Conference of IEEE Industrial Electronics Society, 2005. IECON 2005.*, 2005, p. 6 pp.-.
- [85] R. Genuer *et al.*, «Variable selection using Random Forests. Pattern Recognition Letters», vol. 31, n.º 14, 2010.
- [86] B. Li, P. Zhang, D. Liu, S. Mi, y P. Liu, «Classification of time–frequency representations based on two-direction 2DLDA for gear fault diagnosis», *Appl. Soft Comput.*, vol. 11, n.º 8, pp. 5299-5305, dic. 2011.
- [87] S. Devendiran y K. Manivannan, «Vibration Signal Based Multi-fault Diagnosis of Gears using Roughset Integrated PCA and Neural Networks», *Int. J. Mech. Mechatron. Eng.*, vol. 15, n.º 01, feb. 2015.
- [88] B. Li, P. Zhang, H. Tian, S. Mi, D. Liu, y G. Ren, «A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox», *Expert Syst. Appl.*, vol. 38, n.º 8, pp. 10000-10009, ago. 2011.
- [89] T.-P. Hong, Y.-L. Liou, S.-L. Wang, y B. Vo, «Feature selection and replacement by clustering attributes», *Vietnam J. Comput. Sci.*, vol. 1, n.º 1, pp. 47–55, 2014.
- [90] G. Chandrashekar y F. Sahin, «A survey on feature selection methods», *Comput. Electr. Eng.*, vol. 40, n.º 1, pp. 16-28, ene. 2014.
- [91] M. H. Gharavian, F. Almas Ganj, A. R. Ohadi, y H. Heidari Bafroui, «Comparison of FDA-based and PCA-based features in fault diagnosis of automobile gearboxes», *Neurocomputing*, vol. 121, pp. 150-159, dic. 2013.
- [92] N. Bayar, S. Darmoul, S. Hajri-Gabouj, y H. Pierreval, «Fault detection, diagnosis and recovery using Artificial Immune Systems: A review», *Eng. Appl. Artif. Intell.*, vol. 46, pp. 43-57, nov. 2015.
- [93] G. J. Vachtsevanos, Ed., *Intelligent fault diagnosis and prognosis for engineering systems*. Hoboken, NJ: Wiley, 2006.
- [94] M. S. Kan, A. C. C. Tan, y J. Mathew, «A review on prognostic techniques for non-stationary and non-linear rotating systems», *Mech. Syst. Signal Process.*, vol. 62-63, pp. 1-20, oct. 2015.

- [95] S. Yin, S. X. Ding, X. Xie, y H. Luo, «A Review on Basic Data-Driven Approaches for Industrial Process Monitoring», *IEEE Trans. Ind. Electron.*, vol. 61, n.º 11, pp. 6418-6428, nov. 2014.
- [96] V. T. Tran y B.-S. Yang, «Machine Fault Diagnosis and Prognosis: The State of The Art», *Int. J. Fluid Mach. Syst. IJFMS*, vol. 2, n.º 1, pp. 61–71, 2009.
- [97] Z. Y. Han, Y. L. Liu, H. Y. Jin, y H. Y. Fu, «A Review of Methodologies Used for Fault Diagnosis of Gearbox», *Appl. Mech. Mater.*, vol. 415, pp. 510-514, sep. 2013.
- [98] Xiaofeng Liu, «Machinery Fault Diagnostics Based on Fuzzy Measure and Fuzzy Integral Data Fusion Techniques».
- [99] G. M. Gharehbolagh, S. Farokhzad, M. R. A. A. Abad, y M. Ranjbarkohan, «Condition Monitoring and Fault Diagnosis of Gearbox Using Vibration Monitoring and Adaptive Neuro-Fuzzy Inference System», 2013.
- [100] J. Liu, C. S. Yang, y Q. F. Lou, «Vibration Analysis Based Feature Extraction for Bearing Fault Detection», *Appl. Mech. Mater.*, vol. 197, pp. 124-128, sep. 2012.
- [101] P. Jayaswal, A. K. Wadhvani, y K. B. Mulchandani, «Machine Fault Signature Analysis», *Int. J. Rotating Mach.*, vol. 2008, pp. 1-10, 2008.
- [102] J. Mohd Ali, M. A. Hussain, M. O. Tade, y J. Zhang, «Artificial Intelligence techniques applied as estimator in chemical process systems – A literature survey», *Expert Syst. Appl.*, vol. 42, n.º 14, pp. 5915-5931, Agosto 2015.
- [103] K. W. Kow, Y. W. Wong, R. K. Rajkumar, y R. K. Rajkumar, «A review on performance of artificial intelligence and conventional method in mitigating PV grid-tied related power quality events», *Renew. Sustain. Energy Rev.*, vol. 56, pp. 334-346, Abril 2016.
- [104] V. Palade, C. D. Bocaniala, y L. C. Jain, Eds., *Computational intelligence in fault diagnosis*. London: Springer, 2006.
- [105] J. Rafiee, F. Arvani, A. Harifi, y M. H. Sadeghi, «Intelligent condition monitoring of a gearbox using artificial neural network», *Mech. Syst. Signal Process.*, vol. 21, n.º 4, pp. 1746-1754, may 2007.
- [106] X. Zhang, L. Xiao, y J. Kang, «Application of an improved Levenberg-Marquardt back propagation neural network to gear fault level identification», *J. Vibroengineering*, vol. 16, n.º 2, pp. 855–868, 2014.
- [107] B. Samanta, «Gear fault detection using artificial neural networks and support vector machines with genetic algorithms», *Mech. Syst. Signal Process.*, vol. 18, n.º 3, pp. 625–644, 2004.
- [108] B. Samanta, «Artificial neural networks and genetic algorithms for gear fault detection», *Mech. Syst. Signal Process.*, vol. 18, n.º 5, pp. 1273-1282, sep. 2004.
- [109] H. Li, Y. Zhang, y H. Zheng, «Gear fault detection and diagnosis under speed-up condition based on order cepstrum and radial basis function neural network», *J. Mech. Sci. Technol.*, vol. 23, n.º 10, pp. 2780-2789, oct. 2009.
- [110] P. V. Srihari, «A Method To Improve Reliability Of Gearbox Fault Detection With Artificial Neural Networks», *Int. J. Automot. Mech. Eng. IJAME*, vol. 2, pp. 221–230, 2010.
- [111] K. Heidarbeigi, H. Ahmadi, M. Omid, y A. Tabatabaeefar, «Evolving an artificial neural network classifier for condition monitoring of massy ferguson tractor gearbox», *Int. J. Appl. Eng. Res.*, vol. 5, n.º 12, pp. 2097–2107, 2010.
- [112] S. Farokhzad, M. R. A. A. Abad, y M. RanjbarKohan, «Fault Diagnosis of Massy Ferguson Tractor Gearbox Using Vibration Signal With Neural Network», *Int. J. Agric. Crop Sci.*, pp. 1675–1681, 2013.
- [113] A. Soleimani, «Gear Fault Identification Using Artificial Neural Network and Adaptive Neuro-Fuzzy Inference System», *Appl. Mech. Mater.*, vol. 110-116, pp. 2562-2569, oct. 2011.

- [114] G. B. Zhang, Y. J. Ge, y Y. J. Liu, «Fault Diagnosis of Gearbox Based on Wavelet Transforms and Neural Networks», *Appl. Mech. Mater.*, vol. 148-149, pp. 919-922, dic. 2011.
- [115] M. Akbari, H. Homaei, y M. Heidari, «An Intelligent Fault Diagnosis Approach for Gears and Bearings Based on Wavelet Transform as a Preprocessor and Artificial Neural Networks», *Int. J. Math. Model. Comput.*, vol. 4, n.º 4, 2014.
- [116] J. Xuan, H. Jiang, T. Shi, y G. Liao, «Gear Fault Classification Using Genetic Programming and Support Vector Machines», *Int. J. Inf. Technol.*, vol. 11, n.º 9, 2005.
- [117] D. J. Bordoloi y R. Tiwari, «Optimum multi-fault classification of gears with integration of evolutionary and SVM algorithms», *Mech. Mach. Theory*, vol. 73, pp. 49-60, mar. 2014.
- [118] M. Cerrada, R. V. Sánchez, D. Cabrera, G. Zurita, y C. Li, «Multi-Stage Feature Selection by Using Genetic Algorithms for Fault Diagnosis in Gearboxes Based on Vibration Signal», *Sensors*, vol. 15, n.º 9, pp. 23903-23926, sep. 2015.
- [119] A. Hajnayeb, A. Ghasemloonia, S. E. Khadem, y M. H. Moradi, «Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis», *Expert Syst. Appl.*, vol. 38, n.º 8, pp. 10205-10209, ago. 2011.
- [120] C. Rajeswari, B. Sathiyabhama, S. Devendiran, y K. Manivannan, «A Gear Fault Identification using Wavelet Transform, Rough set Based GA, ANN and C4.5 Algorithm», *Procedia Eng.*, vol. 97, pp. 1831-1841, 2014.
- [121] A. Widodo y B.-S. Yang, «Support vector machine in machine condition monitoring and fault diagnosis», *Mech. Syst. Signal Process.*, vol. 21, n.º 6, pp. 2560-2574, ago. 2007.
- [122] J. Yang, Y. Zhang, y Y. Zhu, «Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension», *Mech. Syst. Signal Process.*, vol. 21, n.º 5, pp. 2012-2024, jul. 2007.
- [123] C. Jing y J. Hou, «SVM and PCA based fault classification approaches for complicated industrial process», *Neurocomputing*, vol. 167, pp. 636-642, nov. 2015.
- [124] L. Jiang, H. Yin, X. Li, y S. Tang, «Fault Diagnosis of Rotating Machinery Based on Multisensor Information Fusion Using SVM and Time-Domain Features», *Shock Vib.*, vol. 2014, pp. 1-8, 2014.
- [125] B. Chouri, M. Fabrice, y A. Dandache, «Bearing Fault Diagnosis Based on Alpha-Stable distribution feature extraction and SVM Classifier», *2014 Int. Conf. Multimed. Comput. Syst. Icmcs*, pp. 1545-1550, 2014.
- [126] Z. Zhang, L. Gu, y Y. Zhu, «Intelligent Fault Diagnosis of Rotating Machine Based on SVMs and EMD Method», *Open Autom. Control Syst. J.*, vol. 5, pp. 219-230, 2013.
- [127] P. Konar y P. Chattopadhyay, «Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs)», *Appl. Soft Comput.*, vol. 11, n.º 6, pp. 4203-4211, sep. 2011.
- [128] L. Breiman, «Random forests», *Mach. Learn.*, vol. 45, n.º 1, pp. 5-32, 2001.
- [129] T. Han y D. Jiang, «Rolling Bearing Fault Diagnostic Method Based on VMD-AR Model and Random Forest Classifier», *Shock Vib.*, vol. 2016, 2016.
- [130] R. Satishkumar y V. Sugumaran, «Vibration based Health Assessment of Bearings using Random Forest Classifier», *Indian J. Sci. Technol.*, vol. 9, n.º 10, mar. 2016.
- [131] R. K. Patel y V. K. Giri, «Feature selection and classification of mechanical fault of an induction motor using random forest classifier», *Perspect. Sci.*, abr. 2016.
- [132] I. A. Abu-Mahfouz, «A comparative study of three artificial neural networks for the detection and classification of gear faults», *Int. J. Gen. Syst.*, vol. 34, n.º 3, pp. 261-277, jun. 2005.
- [133] N. Saravanan, V. Kumar Siddabattuni, y K. Ramachandran, «A comparative study on classification of features by SVM and PSVM extracted using Morlet wavelet for fault diagnosis of spur bevel gear box», *Expert Syst. Appl.*, vol. 35, n.º 3, pp. 1351-1366, 2008.

- [134] S. Natarajan, «Fault Gear Categorization: A Comparative Study on Feature Classification using Rough Set Theory and ID3», *International J. Artif. Intell. Appl. Smart Devices*, vol. 1, pp. 41-64, 2013.
- [135] L. Nacib, S. Saad, y S. Sakhara, «A Comparative Study of Various Methods of Gear Faults Diagnosis», *J. Fail. Anal. Prev.*, vol. 14, n.º 5, pp. 645-656, sep. 2014.
- [136] Pooja Gupta y Sulochana Wadhvani, «Feature Selection by Genetic Programming, And Artificial Neural Network-based Machine Condition Monitoring», *Int. J. Eng. Innov. Technol. IJEIT*, vol. 1, n.º 4, pp. 177-181, 2012.
- [137] W. Teng, X. Ding, X. Zhang, Y. Liu, y Z. Ma, «Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform», *Renew. Energy*, vol. 93, pp. 591-598, Agosto 2016.
- [138] Z. Liu, W. Guo, Z. Tang, y Y. Chen, «Multi-Sensor Data Fusion Using a Relevance Vector Machine Based on an Ant Colony for Gearbox Fault Detection», *Sensors*, vol. 15, n.º 9, pp. 21857-21875, ago. 2015.
- [139] R. Shao, W. Hu, y J. Li, «Multi-Fault Feature Extraction and Diagnosis of Gear Transmission System Using Time-Frequency Analysis and Wavelet Threshold De-Noising Based on EMD», *Shock Vib.*, vol. 20, n.º 4, pp. 763-780, 2013.
- [140] Z. Li, X. Yan, C. Yuan, Z. Peng, y L. Li, «Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method», *Mech. Syst. Signal Process.*, 2011.
- [141] S. G. Jolandan, H. Mobli, H. Ahmadi, M. Omid, y S. S. Mohtasebi, «Fuzzy-Rule-Based Faults Classification of Gearbox Tractor.», *WSEAS Trans. Appl. Theor. Mech.*, vol. 7, n.º 2, 2012.
- [142] K. Kampouropoulos, F. Andrade, A. Garcia, y L. Romeral, «A Combined Methodology of Adaptive Neuro-Fuzzy Inference System and Genetic Algorithm for Short-term Energy Forecasting», *Adv. Electr. Comput. Eng.*, vol. 14, n.º 1, pp. 9-14, 2014.
- [143] I. Asilturk, «On-line surface roughness recognition system by vibration monitoring in CNC turning using adaptive neuro-fuzzy inference system (ANFIS)», *Int. J. Phys. Sci.*, vol. 6, n.º 22, pp. 5353–5360, 2011.
- [144] E. Ebrahimi y K. Mollazade, «Intelligent fault classification of a tractor starter motor using vibration monitoring and adaptive neuro-fuzzy inference system», *Insight - Non-Destr. Test. Cond. Monit.*, vol. 52, n.º 10, pp. 561-566, oct. 2010.
- [145] G. N. Marichal, M. L. del Castillo Zas, J. López, I. Padrón, y M. Artés, «Diagnóstico de engranajes en sistemas de propulsión marinos utilizando técnicas ANFIS», *Rev. Iberoam. Ing. Mecánica*, vol. 20, n.º 1, pp. 37–47, 2016.
- [146] S. W. Smith, *The Scientist & Engineer's Guide to Digital Signal Processing*, 1st ed. California Technical Pub, 1997.
- [147] M. Lebold, K. McClintic, R. Campbell, C. Byington, y K. Maynard, «Review of vibration analysis methods for gearbox diagnostics and prognostics», en *Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA*, 2000, pp. 623–634.
- [148] K. N. Gupta, «Vibration—A tool for machine diagnostics and condition monitoring», *Sadhana*, vol. 22, n.º 3, pp. 393–410, 1997.
- [149] P. Večeř, M. K. y Kreidl, R. Šmíd, «Condition Indicators for Gearbox Condition Monitoring Systems», 2005.
- [150] A. Neustein y H. A. Patil, Eds., *Forensic Speaker Recognition*. New York, NY: Springer New York, 2012.
- [151] I. Antoniadou, G. Manson, N. Dervilis, T. Barszcz, W. J. Staszewski, y K. Worden, «Use of the Teager-Kaiser energy operator for condition monitoring of a wind turbine gearbox».

- [152] J. F. Kaiser, «Some useful properties of Teager's energy operators», en *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, 1993, vol. 3, pp. 149–152.
- [153] W. B. H. J. Decker, *Crack Detection for Aerospace Quality Spur Gears*. National Aeronautics and Space Administration, Glenn Research Center, 2002.
- [154] H. Al-Atat, D. Siegel, y J. Lee, «A Systematic Methodology for Gearbox Health Assessment and Fault Classification», *Int. J. Progn. Health Manag.*, 2011.
- [155] Y. Qu, B. Van Hecke, D. He, J. Yoon, E. Bechhofer, y J. Zhu, «Gearbox fault diagnostics using AE sensors with low sampling rate», *J Acoust. Emiss.*, vol. 31, p. 67, 2013.
- [156] R. B. Randall, *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. Chichester: Wiley, 2011.
- [157] X. Zhao, M. J. Zuo, y Z. Liu, «Diagnosis of pitting damage levels of planet gears based on ordinal ranking», en *Prognostics and Health Management (PHM), 2011 IEEE Conference on*, 2011, pp. 1–8.
- [158] H. Q. Wang y P. Chen, «Fault diagnosis of centrifugal pump using symptom parameters in frequency domain», *Agric. Eng. Int. CIGR EJournal*, 2007.
- [159] A. K. Nandi, C. Liu, y M. D. Wong, «Intelligent Vibration Signal Processing for Condition Monitoring», en *Proceedings of the International Conference Surveillance 7*, 2013, pp. 1–15.
- [160] W. Dargie, «Analysis of time and frequency domain features of accelerometer measurements», en *Computer Communications and Networks, 2009. ICCCN 2009. Proceedings of 18th International Conference on*, 2009, pp. 1–6.
- [161] S. G. Mallat y G. Peyré, *A wavelet tour of signal processing: the sparse way*, 3. ed. Amsterdam: Elsevier [u.a.], 2009.
- [162] L. Debnath, Ed., *Wavelets and Signal Processing*. Boston, MA: Birkhäuser Boston, 2003.
- [163] E. Sejdić, I. Djurović, y J. Jiang, «Time–frequency feature representation using energy concentration: An overview of recent advances», *Digit. Signal Process.*, vol. 19, n.º 1, pp. 153-183, ene. 2009.
- [164] K. Gröchenig, *Foundations of Time-Frequency Analysis*. Boston, MA: Birkhäuser Boston, 2001.
- [165] Z. K. Peng y F. L. Chu, «Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography», *Mech. Syst. Signal Process.*, vol. 18, n.º 2, pp. 199-221, mar. 2004.
- [166] G. F. Bin, J. J. Gao, X. J. Li, y B. S. Dhillon, «Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network», *Mech. Syst. Signal Process.*, vol. 27, pp. 696-711, feb. 2012.
- [167] S. G. Mallat, «A theory for multiresolution signal decomposition: the wavelet representation», *Pattern Anal. Mach. Intell. IEEE Trans. On*, vol. 11, n.º 7, pp. 674–693, 1989.
- [168] R. R. Coifman y M. V. Wickerhauser, «Entropy-based algorithms for best basis selection», *Inf. Theory IEEE Trans. On*, vol. 38, n.º 2, pp. 713–718, 1992.
- [169] L. Cohen, *Time-frequency analysis*. Englewood Cliffs, N.J: Prentice Hall PTR, 1995.
- [170] D. L. Fugal, «Conceptual wavelets in digital signal processing: an in-depth, practical approach for the non-mathematician: Space & Signals Technical Pub», 2009.
- [171] L. Chun-Lin, «A tutorial of the wavelet transform», *NTUEE Taiwan*, 2010.
- [172] D. M. Reif, A. A. Motsinger, B. A. McKinney, J. E. Crowe, y J. H. Moore, «Feature Selection using a Random Forests Classifier for the Integrated Analysis of Multiple Data Types», 2006, pp. 1-8.
- [173] P. Xanthopoulos, P. M. Pardalos, y T. B. Trafalis, *Robust Data Mining*, 1.ª ed. Springer-Verlag New York, 2013.

- [174] D. D. Lee y H. S. Seung, «Algorithms for non-negative matrix factorization», en *Advances in neural information processing systems*, 2001, pp. 556–562.
- [175] N. D. Ho, P. Van Dooren, y V. Blondel, «Weighted nonnegative matrix factorization and face feature extraction», *Submitt. Image Vis. Comput.*, 2007.
- [176] T.-P. Hong y Y.-L. Liou, «Attribute clustering in high dimensional feature spaces», en *Machine Learning and Cybernetics, 2007 International Conference on*, 2007, vol. 4, pp. 2286–2289.
- [177] E. Assareh, M. A. Behrang, M. Ghalambaz, A. R. Noghrehabadi, y A. Ghanbarzadeh, «An Analysis of Wind Speed Prediction Using Artificial Neural Networks: A Case Study in Manjil, Iran», *Energy Sources Part Recovery Util. Environ. Eff.*, vol. 34, n.º 7, pp. 636-644, abr. 2012.
- [178] Q. Zhang y K. C. Gupta, *Neural networks for RF and microwave design (Book+ Neuromodeler Disk)*. Artech House, Inc., 2000.
- [179] D. J. Montana y L. Davis, «Training Feedforward Neural Networks Using Genetic Algorithms.», en *IJCAI*, 1989, vol. 89, pp. 762–767.
- [180] L. J. Farrugia, «ORTEP-3 for Windows—a version of ORTEP-III with a Graphical User Interface (GUI)», *J. Appl. Crystallogr.*, vol. 30, n.º 5, pp. 565–565, 1997.
- [181] D. S. Badde, A. Gupta, y V. K. Patki, «Cascade and Feed Forward Back propagation Artificial Neural Network Models for Prediction of Compressive Strength of Ready Mix Concrete», pp. 1-6.
- [182] Martin T. Hagan, Howard B. Demuth, Mark Hudson Beale, y Orlando De Jesús, *Neural Network Desing*, 2.ª ed. .
- [183] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.
- [184] J.-S. Wang, J.-D. Song, y J. Gao, «Rough Set-Probabilistic Neural Networks Fault Diagnosis Method of Polymerization Kettle Equipment Based on Shuffled Frog Leaping Algorithm», *Information*, vol. 6, n.º 1, pp. 49-68, feb. 2015.
- [185] D. E. Goldberg y others, *Genetic algorithms in search optimization and machine learning*, vol. 412. Addison-wesley Reading Menlo Park, 1989.
- [186] I. G. Ian Goodfellow, Yoshua Bengio, y Aaron Courville, *Deep Learning. Chapter 9. Convolutional Networks*. 2016.
- [187] David Leverington, «A Basic Introduction to Feedforward Backpropagation Neural Networks». Texas Tech University.
- [188] W. E. Deming y S. L. Morgan, «Preface to the Second Edition», *Data Handl. Sci. Technol.*, vol. 11, p. XV, 1993.
- [189] Banerjee A, Chitnis UB, Jadhav SL, Bhawalkar JS, Chaudhury S., «Hypothesis testing, type I and type II errors», *Ind Psychiatry J*, pp. 127-131, jun. 2016.
- [190] V. A. Barrera Núñez, X. Berjaga Moliner, G. Ordóñez Plata, y J. Meléndez Frigola, «Strategies for Voltage Sag Source Relative Location in Power Distribution Networks», *Rev. Fac. Ing. Univ. Antioquia*, n.º 50, pp. 159–169, 2009.
- [191] E. Costa, A. Lorena, A. Carvalho, y A. Freitas, «A review of performance evaluation measures for hierarchical classifiers», en *Evaluation Methods for Machine Learning II: papers from the AAAI-2007 Workshop*, 2007, pp. 1–6.
- [192] M. Sokolova y G. Lapalme, «A systematic analysis of performance measures for classification tasks», *Inf. Process. Manag.*, vol. 45, n.º 4, pp. 427-437, jul. 2009.
- [193] J. Fan, S. Upadhye, y A. Worster, «Understanding receiver operating characteristic (ROC) curves», *CJEM*, vol. 8, n.º 01, pp. 19-20, 2006.
- [194] J. Demšar, «Statistical Comparisons of Classifiers over Multiple Data Sets», *J. Mach. Learn. Res.*, vol. 7, pp. 1-30, 2006.
- [195] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures*. crc Press, 2003.
- [196] V. A. Riera, E. J. Garrido, C. M. Vaquer, E. T. Puig, y N. V. Canela, *Guía para el análisis estadístico con R Commander*, vol. 391. Edicions Universitat Barcelona, 2014.

- [197] S. S. Shapiro y M. B. Wilk, «An analysis of variance test for normality (complete samples)», *Biometrika*, vol. 52, n.º 3/4, pp. 591–611, 1965.
- [198] NCSS, «Levene Test of Variances (Simulation)». [En línea]. Disponible en: http://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/PASS/Levene_Test_of_Variiances-Simulation.pdf. [Accedido: 17-nov-2016].
- [199] Levene Howard, Harold Hotelling, y Ingram Olkin, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press: Alia.
- [200] S. García y F. Herrera, «An Extension on “Statistical Comparisons of Classifiers over Multiple Data Sets” for all Pairwise Comparisons», *J. Mach. Learn. Res.*, vol. 9, pp. 2677-2694, 2008.
- [201] P. Sprent y N. C. Smeeton, *Applied nonparametric statistical methods*, 3rd ed. Boca Raton: Chapman & Hall/CRC, 2001.
- [202] M. Friedman, «The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance», *J. Am. Stat. Assoc.*, vol. 32, n.º 200, p. 675, dic. 1937.
- [203] Sture Holm, «A Simple Sequentially Rejective Multiple Test Procedure», *Scandinavian Journal of Statistics*, vol. 6, pp. 65-70, 10-sep-2013.
- [204] Frank Wilcoxon, «Individual Comparisons by Ranking Methods», vol. 1, pp. 80-83, dic. 1945.
- [205] D. D. Wackerly, W. Mendenhall, y R. L. Scheaffer, *Mathematical statistics with applications*, 7th ed. Belmont, CA: Thomson Brooks/Cole, 2008.
- [206] «Public Data Sets | PHM Society». [En línea]. Disponible en: <https://www.phmsociety.org/references/datasets>. [Accedido: 24-ago-2016].
- [207] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, y C. Li, «Fault diagnosis in spur gears based on genetic algorithm and random forests», *Mech. Syst. Signal Process.*, vol. 70-71, pp. 87-103, mar. 2016.
- [208] C. M. Bishop, *Pattern recognition and machine learning*, Corrected at 8. printing 2009. New York, NY: Springer, 2009.
- [209] J. Luengo, S. García, y F. Herrera, «A study on the use of statistical tests for experimentation with neural networks: Analysis of parametric test conditions and non-parametric tests», *Expert Syst. Appl.*, vol. 36, n.º 4, pp. 7798-7808, may 2009.
- [210] M. Fernández-Delgado, E. Cernadas, S. Barro, y D. Amorim, «Do we need hundreds of classifiers to solve real world classification problems», *J Mach Learn Res*, vol. 15, n.º 1, pp. 3133–3181, 2014.
- [211] M. Liu, M. Wang, J. Wang, y D. Li, «Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar», *Sens. Actuators B Chem.*, vol. 177, pp. 970-980, feb. 2013.