

UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE CUENCA

CARRERA DE INGENIERÍA ELECTRÓNICA

**TRABAJO DE TITULACIÓN PREVIO A LA
OBTENCIÓN DEL TÍTULO DE INGENIERO
ELECTRÓNICO**

PROYECTO TÉCNICO:

**“PROCESAMIENTO DE SEÑALES DE VIBRACIÓN MEDIANTE
DESCOMPOSICIÓN MODAL EMPÍRICA (EMD) PARA LA EXTRACCIÓN DE
CARACTERÍSTICAS CON FINES DE DIAGNÓSTICO DE FALLOS EN
MAQUINARIA ROTATIVA”**

AUTOR:

CRISTIAN EFRÉN GRANDA ORDOÑEZ

TUTOR:

ING. DIEGO CABRERA MENDIETA; M.Sc.

Cuenca-Ecuador

2017

DERECHO DE AUTOR

Yo, Cristian Efrén Granda Ordoñez con documento de identificación CI. 0704548825, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del trabajo de titulación: “PROCESAMIENTO DE SEÑALES DE VIBRACIÓN MEDIANTE DESCOMPOSICIÓN MODAL EMPÍRICA (EMD) PARA LA EXTRACCIÓN DE CARACTERÍSTICAS CON FINES DE DIAGNÓSTICO DE FALLOS EN MAQUINARIA ROTATIVA”, mismo que ha sido desarrollado para optar por el título de: Ingeniero Electrónico, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

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

Cuenca, diciembre del 2017



CRISTIAN EFRÉN GRANDA ORDOÑEZ

CI: 0704548825

CERTIFICACIÓN

Yo declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: “PROCESAMIENTO DE SEÑALES DE VIBRACIÓN MEDIANTE DESCOMPOSICIÓN MODAL EMPÍRICA (EMD) PARA LA EXTRACCIÓN DE CARACTERÍSTICAS CON FINES DE DIAGNÓSTICO DE FALLOS EN MAQUINARIA ROTATIVA”, realizado por Cristian Efrén Granda Ordoñez, obteniendo el Proyecto Técnico con enfoque investigativo que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana.

Cuenca, diciembre del 2017

A handwritten signature in blue ink, reading "Diego Román Cabrera Mendieta", enclosed within a blue oval scribble.

ING. DIEGO ROMÁN CABRERA MENDIETA M. Sc.

CI: 010402830-3

DECLARATORIA DE RESPONSABILIDAD

Yo, Cristian Efrén Granda Ordoñez con número de cédula CI. 0704548825, autor del trabajo de titulación “PROCESAMIENTO DE SEÑALES DE VIBRACIÓN MEDIANTE DESCOMPOSICIÓN MODAL EMPÍRICA (EMD) PARA LA EXTRACCIÓN DE CARACTERÍSTICAS CON FINES DE DIAGNÓSTICO DE FALLOS EN MAQUINARIA ROTATIVA” certifico que el total contenido del Proyecto Técnico con enfoque investigativo, es de mi exclusiva responsabilidad y autoría.

Cuenca, diciembre del 2017



CRISTIAN EFRÉN GRANDA ORDOÑEZ

CI: 0704548825

AGRADECIMIENTO

En primer lugar agradezco a Dios que me brindó la paciencia suficiente para realizar este proyecto de titulación.

Agradezco también a mi tutor cuya tranquilidad, enseñanzas y consejos ayudaron a la culminación de este proyecto y además me gustaría extender mi agradecimiento al grupo de investigación GIDTEC por el apoyo y las facilidades que me brindaron de adquirir el conocimiento necesario para el desarrollo del proyecto.

CRISTIAN GRANDA

DEDICATORIA

A Dios por haberme guiado por el camino y darme la fuerza para lograr mis objetivos, además de su eterna bondad y amor. A mi madre Maeva Ordoñez Calle por haberme aconsejado en todo momento, y apoyado para la culminación de todos mis objetivos, también por sus enseñanzas, por la motivación constante que siempre me ha brindado permitiéndome ser una persona mucho más paciente y de bien, pero más que nada por darme su inmenso amor incondicional.

CRISTIAN GRANDA

RESUMEN

El estudio de la fiabilidad de los equipos industriales ha hecho posible la creación de los diferentes tipos de mantenimiento, como es el caso del mantenimiento predictivo que ha permitido la disminución de fallos. Para un buen mantenimiento predictivo en la maquinaria rotativa (engranajes y rodamientos) es necesario conocer acerca de sus señales de vibración ya que estas contienen información para diagnosticar tempranamente sus fallos, por esta razón existen algunos métodos que utilizan estas señales y son descompuestas para su procesamiento.

El método utilizado en este proyecto fue la Descomposición Modal Empírica (EMD) que descompone las señales de vibración en Funciones Modales Intrínsecas (IMF's) las cuales representan el modo de oscilación incrustado en los datos. Luego de esto las IMF's son convertidas en una estructura de GPUArray utilizando la computación paralela y distribuida para ser ejecutados en el GPU y mejorar los tiempos de procesamiento.

Se manejó una base de datos de rodamientos y otra para engranajes, las mismas que fueron obtenidas del grupo de investigación GIDTEC de la Universidad Politécnica Salesiana. Para la ejecución de la EMD, se utilizó el software matemático de Matlab en el cual se realizó la extracción de características en el dominio de tiempo y dominio de frecuencia, con estas características se creó un conjunto de datos que sirvió para entrenar el modelo de clasificador de Random Forest que permitió clasificar las clases encontradas en las bases de datos y seleccionar las mejores características. El rendimiento del clasificador se evaluó utilizando la matriz de confusión al sacar las métricas de exactitud, especificidad, sensibilidad o recall, precisión y F-score.

El mejor resultado se presentó en la base de datos de rodamientos con un error del 2.7% al entrenar el clasificador con las características en el dominio de tiempo en conjunto con la frecuencia y la clase de mejor rendimiento fue la P3 en rodamientos que corresponde a los fallos de pista externa y en engranajes la P2 que son los fallos de diente roto a 11.5%.

ÍNDICE GENERAL

AGRADECIMIENTO	V
DEDICATORIA	VI
RESUMEN	VII
ÍNDICE GENERAL	VIII
ÍNDICE DE FIGURAS	X
ÍNDICE DE TABLAS	XII
ANTECEDENTES	XIII
INTRODUCCIÓN	XV
OBJETIVOS	XVI
Objetivo General:	XVI
Objetivos específicos:	XVI
1. MARCO TEÓRICO	1
1.1. Sensores	1
1.1.1. Sensor de Aceleración o Acelerómetro	2
1.1.2. Sensores de Emisión Acústica	3
1.2. Dispositivos de Adquisición de Datos	4
1.3. Características de las Señales	6
1.3.1. Características en el Dominio de Tiempo	7
1.3.2. Características en el Dominio de Frecuencia	9
1.4. Matlab	10
1.4.1. Computación Paralela y Distribuida	10
1.5. EMD	13
1.6. Conjunto de Datos	16
1.7. Aprendizaje Automático	17
1.7.1. Tipos de Aprendizaje	17
1.8. Random Forest	18

1.9.	Métricas para Tareas de Clasificación	20
1.9.1.	Exactitud (Accuracy).....	21
1.9.2.	Especificidad	21
1.9.3.	Sensibilidad o Recall	21
1.9.4.	Precisión	21
1.9.5.	F-Score	22
2.	METODOLOGÍA	23
2.1.	Señales de Vibración	24
2.2.	Recorte de las Señales de Vibración.....	27
2.3.	Aplicación de la EMD y extracción de la IMF.....	28
2.3.1.	Ejecución de la EMD.....	28
2.3.2.	Conversión de las IMF en GPUArray	28
2.4.	Implementación de las Características.....	29
2.5.	Aplicación de Random Forest.....	30
3.	ANÁLISIS DE RESULTADOS	32
	CONCLUSIONES Y RECOMENDACIONES	44
	REFERENCIAS	45

ÍNDICE DE FIGURAS

Figura 1. Partes de un Acelerómetro Capacitivo.....	2
Figura 2. Diagrama de un Acelerómetro Piezoeléctrico.	3
Figura 3. Sensor de EA tipo diferencial de banda ancha.....	3
Figura 4. Sensor de EA tipo diferencial de baja frecuencia.	4
Figura 5. Esquema de Adquisición de Datos.....	4
Figura 6. Chasis NI cDAQ 9188.	5
Figura 7. Módulo NI 9234.....	6
Figura 8. Ejecución del Parfor.....	12
Figura 9. Tamizado EMD.....	14
Figura 10. Proceso de diagnóstico de fallos.	23
Figura 11. Posición de los acelerómetros en los rodamientos (a), posición de los acelerómetros en la caja de engranajes (b) [63] [64].....	26
Figura 12. Diagrama de la carga de archivos.	27
Figura 13. Recorte de las Señales.....	27
Figura 14. Diagrama de Flujo de process_file.....	29
Figura 15. Matriz de Características.....	30
Figura 16. Curvas de convergencia del proceso de entrenamiento de Random Forest en Rodamientos.	32
Figura 17. Curvas de convergencia del proceso de entrenamiento de Random Forest en Engranajes.	33
Figura 18. Mejores Características en el Dominio de Tiempo en Rodamientos.	38
Figura 19. Mejores Características en el Dominio de Frecuencia en Rodamientos.	39
Figura 20. Mejores Características en los Dominios de Tiempo en conjunto con la Frecuencia en Rodamientos.....	39
Figura 21. Mejores Características en el Dominio de Tiempo en Engranajes.	40
Figura 22. Mejores Características en el Dominio de Frecuencia en Engranajes.	40

Figura 23. Mejores Características en los Dominios de Tiempo en conjunto con la Frecuencia en Engranajes.	41
--	----

ÍNDICE DE TABLAS

Tabla 1. Características utilizadas para el diagnóstico de fallos.	6
Tabla 2. Matriz de Confusión de 2x2.	20
Tabla 3. Especificaciones de las Señales.	24
Tabla 4. Tipos de fallas en las bases de datos.	25
Tabla 5. Características de sensores utilizados en la obtención de las señales de vibración.	25
Tabla 6. Tamaño de las Matrices de Características.	30
Tabla 7. Resultados de oob-error obtenidos del clasificador Random Forest.	32
Tabla 8. Matriz de confusión de Rodamiento en dominio de tiempo.	33
Tabla 9. Matriz de confusión de Rodamiento en dominio de frecuencia.	34
Tabla 10. Matriz de confusión de Rodamiento en dominio de tiempo en conjunto con la frecuencia.	34
Tabla 11. Matriz de Confusión de Engranaje en dominio de tiempo.	34
Tabla 12. Matriz de confusión de Engranaje en dominio de frecuencia.	35
Tabla 13. Matriz de confusión de Engranaje en dominio de tiempo en conjunto con la frecuencia.	35
Tabla 14. Datos de Exactitud de cada clase.	36
Tabla 15. Datos de Especificidad de cada clase.	36
Tabla 16. Datos de Sensibilidad o Recall de cada clase.	37
Tabla 17. Datos de precisión de cada clase.	37
Tabla 18. Datos de F-Score de cada clase.	38
Tabla 19. Características más relevantes en el dominio de tiempo, dominio de frecuencia y dominio de tiempo en conjunto con la frecuencia en Rodamientos.	42
Tabla 20. Características más relevantes en el dominio de tiempo, dominio de frecuencia y dominio de tiempo en conjunto con la frecuencia en Engranajes.	43

ANTECEDENTES

Es necesario que en las industrias las máquinas se aseguren con un correcto funcionamiento, por esta razón J.P. Souris en 1992 comenta que el instrumento de producción debe responder a un objetivo fundamental: “*disponibilidad con una calidad de servicio óptimo*”[1].

Cuando hablamos de mantenimiento éste hace referencia a la conservación correcta del estado funcional de los equipos e instalaciones, ya que su buen funcionamiento hace que los niveles de vibración y de ruido disminuyan y tengan un periodo de vida útil más largo, lo que conlleva a que su seguridad sea propicia para laborar y por lo tanto mermar los riesgos laborales.

La adecuada normativa impuesta en las fábricas ha hecho que el mantenimiento sea una parte importante dentro de sus políticas lo cual resolverá ciertos costos como en averías imprevistas, indisponibilidad de equipos, costes de no calidad, baja mantenibilidad, entre otros [1].

El crecimiento exponencial de la industria a nivel mundial ha llevado a que los dueños de las fábricas tomen conciencia de un correcto plan de mantenimiento de sus maquinarias. Por este motivo uno de los tipos de mantenimiento implantado desde hace ya algún tiempo es el mantenimiento correctivo el mismo que por lo general es empleado en pequeñas y medianas industrias, este mantenimiento se da cuando el fallo ya fue producido en la máquina dándose así un mantenimiento correctivo con eliminación de averías en donde se realiza una reparación reemplazando los elementos deteriorados y, un mantenimiento correctivo con eliminación de causas es decir busca el motivo que originó el fallo [2]. Este tipo de mantenimiento dio paso al mantenimiento clásico preventivo que consta de una serie de revisiones para la detección de fallos, esto conlleva a que los costos sean elevados debido a que la empresa tiene que disponer de todos los repuestos en stock y, además el técnico de mantenimiento no realiza la verificación respectiva de sí las piezas están realmente desgastadas, sino que las cambia porque así se encuentra programado en el plan de mantenimiento.

En los años 50 comienza a evolucionar el mantenimiento industrial fijando dos ramas una de ellas son las técnicas de análisis de la fiabilidad de los equipos, y la otra las técnicas de verificación mecánicas cuya finalidad es la detección temprana de fallos, esta última dando paso al mantenimiento predictivo [1].

El mantenimiento predictivo surge para minimizar los costos generados por el mantenimiento correctivo y preventivo a largo plazo lo que permite realizar reacondicionamientos de acuerdo al estado real de la máquina, antes de que se presenten problemas de funcionamiento, evitando así inspecciones innecesarias y averías imprevistas [1]. Una de las opciones para este tipo de mantenimiento es el análisis de señales de vibración de maquinaria rotativa y si éste tipo de maquinaria sufre algún fallo también se ven afectadas dichas señales permitiendo de esta manera diagnosticar de una manera eficiente la avería para evitar que la máquina se detenga y con ella la producción [3][4]. Según estudios efectuados, cerca del 90% de las averías de la maquinaria rotativa se pueden diagnosticar mediante las señales de vibración [3] [5].

Como dato importante, en septiembre de 1968 en la conferencia de ASME Petroleum Mechanical Engineering Conference de Dallas se habla por primera vez sobre las manifestaciones vibratorias producidas por los defectos más frecuentes de la maquinaria rotativa mencionadas por J.S. Sohre dando paso a la detección de fallos mediante señales de vibración expuesto en este trabajo y en muchos otros [1].

Según la revisión bibliográfica realizada para este proyecto podemos decir que la extracción de características en señales de vibración es un método muy conocido, pero para la selección de las mejores características mediante la ejecución de Random Forest se ha encontrado muy pocas investigaciones sobre este tema, es por esto que el proyecto se ha desarrollado para el diagnóstico de fallos seleccionando las mejores características.

INTRODUCCIÓN

Los elementos fundamentales de las maquinarias rotativas son los rodamientos y engranajes. Para garantizar un correcto funcionamiento de las máquinas se requiere que estos dispositivos no presenten fallos [3] [6]. Estos fallos se pueden detectar a través de diferentes métodos como son: análisis de las señales de vibración, análisis de monitoreo de temperatura, análisis de emisiones acústicas y análisis del estado del lubricante. En este caso el proyecto se realizará mediante el análisis de señales de vibración las mismas que reconocen los siguientes tipos de fallas: fallas de rodamientos, fallas de engranajes, fallas de máquinas eléctricas y fallas de acoplamiento flexibles. Este método es el más utilizado por la razón de que toda máquina produce vibración y si ésta presenta fallos su funcionamiento cambia al igual que su espectro, ya que mediante el espectro de la señal se puede encontrar características en la detección de fallos [7].

Actualmente existen técnicas de extracción de características, ya sea en el dominio de tiempo utilizando características estadísticas, de frecuencia utilizando la transformada de Fourier, o a su vez se pueden utilizar tiempo-frecuencia al manejar alguna de las transformadas como es el método de wavelet y sus variantes [7]. Sin embargo, en el caso de la Transformada de Fourier se presentan ciertas debilidades debido a que esta transformada se utiliza en procesos estacionarios y periódicos, mientras que la EMD se usa en el análisis de señales no lineales y no estacionarias presentando los resultados en el dominio de tiempo en conjunto con la frecuencia permitiendo así la extracción de características en ambos dominios [8] [9]. De esta manera para este trabajo se propone utilizar el algoritmo de la descomposición modal empírica (EMD por sus siglas en inglés).

Se utilizará también uno de los algoritmos de Aprendizaje Automático como lo es Random Forest que permite clasificar un conjunto de datos con múltiples clases que los demás algoritmos existentes hasta el momento [10] [11]. Este algoritmo permite seleccionar las mejores características encontradas en el conjunto de datos [12].

Por otro lado, se explicará sobre la metodología utilizada y los resultados obtenidos mediante el proceso del diagnóstico de fallos para finalizar con las conclusiones y recomendaciones sobre este proyecto.

OBJETIVOS

Objetivo General:

Aplicar el método “Descomposición Modal Empírica” (EMD) para la extracción de características de señales de vibración con la finalidad de diagnosticar fallos en maquinaria rotativa.

Objetivos específicos:

- Conceptualizar mediante la revisión bibliográfica las técnicas del método de EMD.
- Implementar el algoritmo de EMD para procesar las señales de vibración.
- Obtener las Funciones Modales Intrínsecas (IMF) de las señales.
- Seleccionar los mejores atributos de las IMF's, para diagnosticar más eficientemente los fallos en las señales de vibración.

1. MARCO TEÓRICO

El diagnóstico de fallos mediante las señales de vibración comienza obteniéndose a través de los sensores de aceleración, convirtiendo las señales mecánicas a señales eléctricas y pasar a la etapa de adquisición de datos en el cual se utilizó los módulos de National Instruments para que las señales sean procesadas en el computador. Estas señales serán convertidas en IMF's mediante el algoritmo de EMD para extraer sus características en el dominio de tiempo y de frecuencia, las mismas servirán como un conjunto de entrenamiento para el aprendizaje automático y todo esto se ejecutará en el software matemático de Matlab que cuenta con herramientas para mejorar su rendimiento. A continuación se presentan los detalles teóricos de cada uno de los elementos en cuestión.

1.1. Sensores

Un sensor es un instrumento de medición que transforma cualquier tipo de variables físicas a una señal eléctrica [13].

Para una mejor elección del sensor se deben considerar los tipos de variables y las diferentes características de estas, entre las cuales se pueden considerar: las variables mecánicas, variables térmicas, variables eléctricas, variables magnéticas, variables ópticas y las variables químicas o moleculares; en este caso nos enfocaremos en las variables mecánicas las cuales se clasifican en: fuerza, longitud, área, flujo másico, presión, intensidad acústica, velocidad, aceleración, posición, volumen, longitud de onda acústica, entre otras [14].

Tomando en cuenta las características mencionadas anteriormente sobre las variables mecánicas podemos decir que existen diferentes tipos de sensores de medición. Nosotros nos enfocaremos en los sensores de aceleración y los sensores de emisión acústica, los cuales nos ayudarán para el análisis de las señales de vibración y las señales de emisión acústica.

1.1.1. Sensor de Aceleración o Acelerómetro

El acelerómetro es un dispositivo que mide la aceleración de los cuerpos y convierte el movimiento mecánico que produce la vibración en una señal eléctrica [15]. Los tipos más comunes de acelerómetro son:

- **Acelerómetro Capacitivo:** mide el cambio producido en la capacitancia entre dos o más conductores de un capacitor [14].

Un acelerómetro capacitivo está compuesto por un conjunto de capacidades fijas unidas al circuito integrado como se muestra en la Figura 1 a su lado izquierdo y, un conjunto de placas centrales incorporadas a unos anillos elásticos que permiten el movimiento de las mismas a través de la fuerza aplicada que se indica al lado derecho de la Figura 1 [16].

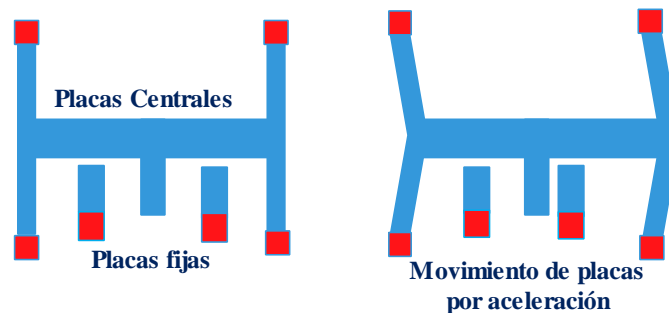


Figura 1. Partes de un Acelerómetro Capacitivo.

- **Acelerómetro Piezoeléctrico:** mide la diferencia de potencial en determinados cristales al ser sometidos a deformaciones mecánicas [14].

El cristal piezoeléctrico es producido por una masa sísmica que hace que la carga total almacenada sea proporcional a la fuerza variable y ésta será proporcional a la aceleración, provocando así una repartición de las cargas.

Este tipo de acelerómetro presenta como principales ventajas: circuitos menos complicados, capacidad de soportar mayores niveles de impacto y su funcionamiento tiene un bajo consumo de voltaje [17]. En la Figura 2 se presenta como se compone un Acelerómetro Piezoeléctrico.

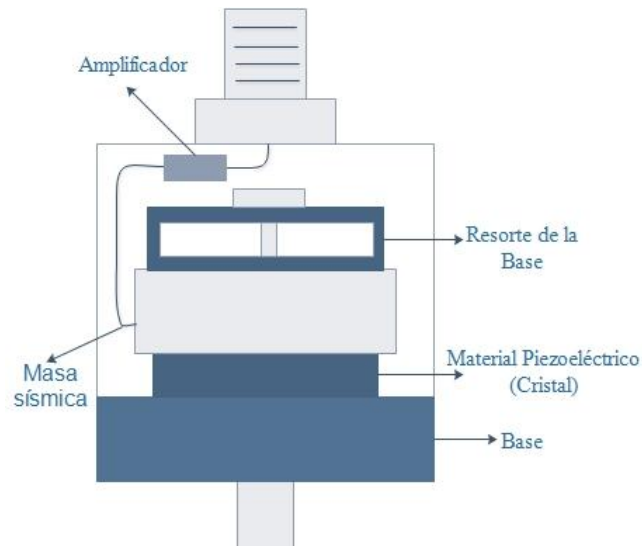


Figura 2. Diagrama de un Acelerómetro Piezoeléctrico.

1.1.2. Sensores de Emisión Acústica

Todo material produce emisión acústica (EA) al momento de fisurarse o de sufrir deformidades debido a la acción de fuerzas mecánicas las cuales generan ondas elásticas transitorias por el cambio de sus tensiones internas que se mueven a través del cuerpo hasta la superficie y son detectadas por los sensores que convierten las ondas mecánicas a eléctricas [18].

Entre algunos tipos de sensores de emisión acústica se puede considerar los siguientes:

- **Sensor diferencial de banda ancha.** Este sensor tiene una alta sensibilidad y un rango de frecuencia comprendido entre los 100 a 900 KHz, cuenta con una mayor inmunidad al ruido que otros tipos de sensores de EA, por esta razón es utilizado en aplicaciones de banda ancha para el análisis de frecuencia de las señales EA, y en aplicaciones de investigación que requieran una alta fidelidad [19]. A continuación, se muestra el sensor de EA diferencial de banda ancha en la Figura 3.



Figura 3. Sensor de EA tipo diferencial de banda ancha.

- **Sensor diferencial de baja frecuencia:** es utilizado para eliminar el ruido en aplicaciones que contengan mucho EMI (Interferencia Electromagnética) y transformadores con gran potencia. Estos sensores tienen un rango de frecuencia de 10 a 100 KHz [20]. En la Figura 4 se muestra este tipo de sensor de EA.

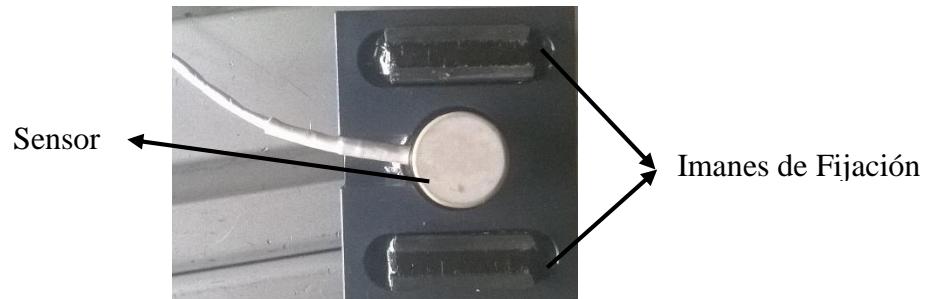


Figura 4. Sensor de EA tipo diferencial de baja frecuencia.

El estándar ISO 2372 – 1974 define la severidad de vibración entre una banda de frecuencia de 10 Hz a 1 KHz por lo tanto los sensores de aceleración pueden tener un rango medible de frecuencia de 10 Hz hasta un máximo de 50 KHz [21] en cambio, los sensores de EA según la Norma ASTM E1106 manejan un intervalo de frecuencia entre los 10 KHz a 1 MHz [22].

1.2. Dispositivos de Adquisición de Datos

Los sistemas de adquisición de datos convierten las señales analógicas del mundo real en su equivalente digital para ser utilizadas y manipuladas en un ordenador [23].

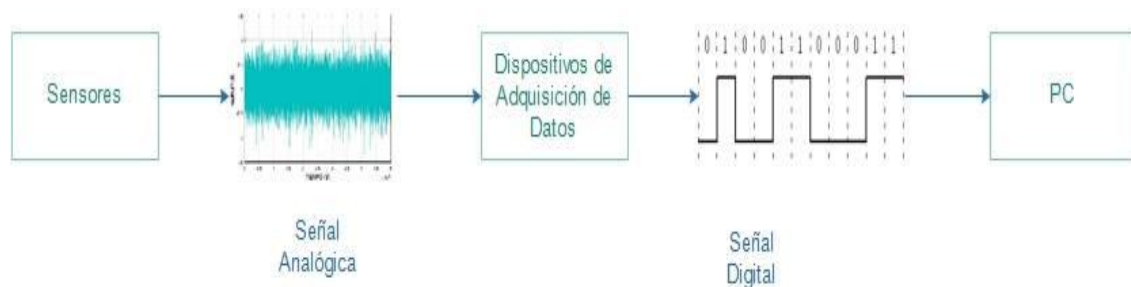


Figura 5. Esquema de Adquisición de Datos.

En la Figura 5 se muestra un esquema de conversión analógica a digital desde la obtención de las señales en los sensores que convierten la señal mecánica a una señal eléctrica medible, luego pasa por el dispositivo de adquisición de datos compuesto internamente por un circuito de muestreo y retención y un convertidor digital a analógico para ser enviados al computador y ser visualizados, procesados o almacenados [24].

Entre los dispositivos de adquisición de datos tenemos los de la marca National Instruments y basaremos nuestra explicación en el Chasis Modelo NI cDAQ 9188 y en el Módulo NI9234 que son los dispositivos utilizados en este proyecto.

Chasis NI cDAQ 9188

Consta de una entrada Gigabit Ethernet IEEE 802.3ab y tiene ocho ranuras para la inserción de módulos de la serie C. Los requerimientos de consumo en la entrada son de 9 v a 30 v, el consumo máximo de energía que tendrá este chasis será de 15 w y la temperatura normal de funcionamiento esta entre -20°C a 55°C pero, para mantener un rendimiento óptimo la temperatura no debe de ser más de los 45°C y si sobrepasa este rango se debe de colocar sobre una placa metálica como disipador utilizando el Kit de montaje [25]. A continuación, en la Figura 6 se presenta el diagrama del Chasis 9188 y sus partes:

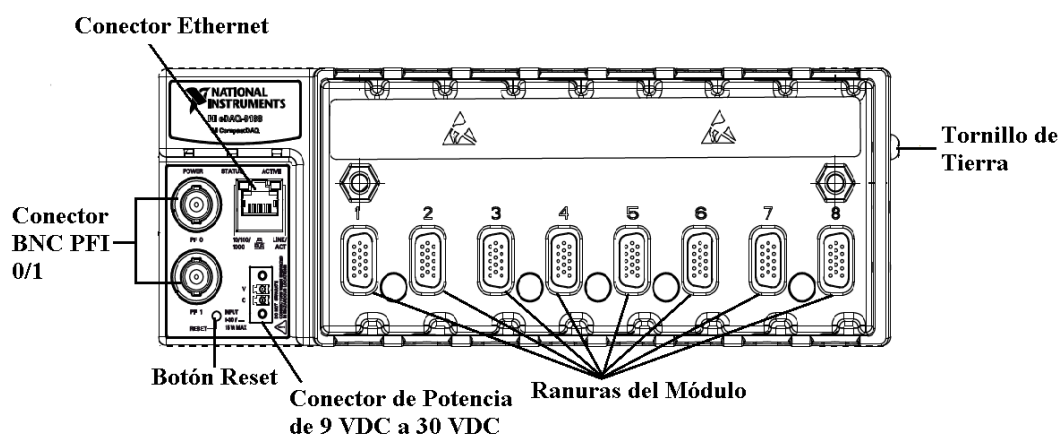


Figura 6. Chasis NI cDAQ 9188 [25].

Módulo NI 9234

Es un módulo de adquisición de señales dinámicas que maneja mediciones de alta precisión y consta de cuatro canales de entrada con velocidades de 2 a 50 KHz. Estos módulos son especializados para la adquisición de señales de vibración y cuenta con dos librerías: NI Asistente de Sonido y Vibración que es un software que minimiza el proceso de adquisición de las señales de vibración y el Análisis de Sonido y Vibración para LabVIEW que desarrolla aplicaciones para las señales de vibración [26].

Una vez adquiridas las señales desde los sensores se transforman de una señal analógica a una digital con una tasa de muestreo de 24 bits para su resultado ser almacenado en un buffer de memoria. Este módulo tiene un rango de voltaje de $\pm 5\text{ V}$ y para la conversión

analógica digital utiliza la modulación delta-sigma es decir, si la velocidad de los datos es de 25 KS/s entonces cada conversión muestrea su señal de entrada a 3.2 MS/s [26]. En la Figura 7 se muestra el Módulo NI9234.



Figura 7. Módulo NI 9234.

1.3. Características de las Señales

Las características de una señal de vibración se pueden extraer en el dominio de tiempo, dominio de frecuencia y dominio de tiempo-frecuencia. A continuación se muestra las características utilizadas que pueden ser extraídas en los dos dominios: tiempo y frecuencia.

TIEMPO	FRECUENCIA
Valor Cuadrático Medio (RMS)	Total potencia (TTP)
Curtosis (K)	Momento espectral 1 (SM1)
Valor medio (MEAN)	Momento espectral 2 (SM2)
Punto máximo (MAX)	Momento espectral 3 (SM3)
Punto mínimo (MIN)	Curtosis espectral (KE)
Asimetría (SKEWNESS)	Frecuencia media (MNF)
Energía (ET)	Frecuencia mediana (MDF)
Desviación estándar (STD)	Variación de frecuencia central (VFC)
	Energía (EF)

Tabla 1. Características utilizadas para el diagnóstico de fallos.

1.3.1. Características en el Dominio de Tiempo

Las características estadísticas son las que más se utilizan para caracterizar las señales en el dominio de tiempo como se presenta en los siguientes puntos. Teniendo en cuenta que x representa la señal en el dominio de tiempo, x_i un punto en la señal y N el número de muestras.

1.3.1.1. Valor Cuadrático Medio (RMS)

El valor cuadrático medio mide la energía con que es emitida la señal y puede ser determinada promediando los valores cuadráticos de cada punto de la muestra en un intervalo de tiempo definido [27]. Se calcula de la siguiente manera:

$$x_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_N^2}{N}} \quad (1)$$

1.3.1.2. Curtosis (K)

La curtosis es una medida que permite caracterizar las distribuciones de probabilidad [28]. En una distribución normal el valor de la Curtosis es 3 [29] [30]. Su fórmula es:

$$k = \frac{[E(x - \mu)^4]}{\sigma^4} \quad (2)$$

Donde:

μ = media de x

σ = desviación estándar de x

E = representa el valor esperado de la cantidad t

1.3.1.3. Valor medio (MEAN)

El valor medio es el promedio de un conjunto de datos. Se define como [31] [32]:

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i = \frac{x_1 + x_2 + \dots + x_N}{N} \quad (3)$$

1.3.1.4. Punto máximo (MAX) y punto mínimo (MIN)

Son los valores máximos y los valores mínimos de cada señal.

$$\text{punto maximo} = \max (x(t))$$

$$\text{punto mínimo} = \min (x(t))$$

1.3.1.5. Asimetría (SKEWNESS)

La asimetría es una medida de los datos alrededor de la media de la muestra. Si la asimetría es negativa los datos se distribuyen más a la izquierda de la media que a la derecha. Si la asimetría es positiva, los datos se distribuyen más hacia la derecha. La asimetría de la distribución normal (o cualquier distribución perfectamente simétrica) es cero. Se determina a través de la siguiente fórmula [33]:

$$s = \frac{E(x - \mu)^3}{\sigma^3} \quad (4)$$

1.3.1.6. Energía (ET)

Las señales de energía están dadas por la siguiente ecuación:

$$E_x = \sum_{i=1}^N \frac{x^2(t)}{N} \quad (5)$$

Si la energía de una señal periódica es infinita su señal de potencia es finita y si su señal de energía es finita la señal de potencia es cero [34].

1.3.1.7. Desviación Estándar (STD)

La desviación estándar es la relación exacta entre la varianza, el valor medio cuadrático (el componente dinámico de la señal) y el valor medio (el componente estático de la señal), representada por [27]:

$$\sigma_x^2 = \overline{x^2} - \mu_x^2 \quad (6)$$

1.3.2. Características en el Dominio de Frecuencia

Para obtener las características en el dominio de frecuencia se utiliza el espectro de la señal calculando la transformada de Fourier y nuevas características como vemos a continuación.

1.3.2.1. Potencia Total

El total de potencia o momento espectral cero es una característica que se puede sacar a partir de la densidad espectral de potencia o del espectro de potencia (PSD).

$$TTP = \sum_{j=1}^M P_j = SM0 \quad (7)$$

Siendo P_j la densidad espectral de potencia de la señal original. La densidad espectral de potencia se calcula a partir de la transformada de Fourier de la función de autocorrelación [35].

1.3.2.2. Momento Espectral 1, 2 y 3

Los primeros 3 momentos espectrales son los más relevantes para determinar las características en el PSD [36].

$$SM1 = \sum_{j=1}^M P_j f_j \quad (8)$$

$$SM2 = \sum_{j=1}^M P_j f_j^2 \quad (9)$$

$$SM3 = \sum_{j=1}^M P_j f_j^3 \quad (10)$$

Siendo f_j el valor de la frecuencia del PSD de la señal [35].

1.3.2.3. Frecuencia Mediana

La frecuencia mediana (MDF) es una frecuencia en donde el espectro se divide en dos regiones con igual amplitud, es decir MDF es la mitad del total potencia [36]. Se la representa mediante la siguiente ecuación.

$$\sum_{j=1}^{MDF} P_j = \sum_{j=MDF}^M P_j = \frac{1}{2} \sum_{j=1}^M P_j \quad (11)$$

1.3.2.4. Frecuencia Media

Frecuencia media o frecuencia central (MNF) es la suma entre el producto del PSD por la frecuencia dividido para el PSD de la señal como se muestra en la ecuación 12 [35].

$$MNF = \frac{\sum_{j=1}^M P_j f_j}{\sum_{j=1}^M P_j} \quad (12)$$

1.3.2.5. Variación de Frecuencia Central

Variación de frecuencia central (VFC) se la calcula mediante los momentos espectrales SM0, SM1 y SM2 [35]. Se define como:

$$VFC = \frac{SM2}{SM0} - \left(\frac{SM1}{SM0} \right)^2 \quad (13)$$

1.3.2.6. Frecuencia Pico

Frecuencia pico (PKF) es la frecuencia en donde se produce la máxima cantidad de potencia y se representa con la siguiente ecuación [35]:

$$PKF = \max(P_j), \quad j = 1 \dots M \quad (14)$$

1.4. Matlab

Su nombre proviene de Matrix LABORatory debido a que gestiona arreglos de matrices [37]. Es un lenguaje diseñado específicamente para la computación técnica facilitando la solución de problemas a través del análisis, procesamiento de datos y gráfica de los mismos. Para esto el software consta de herramientas y librerías, estas últimas llamadas toolboxes (término común en inglés) que son funciones para la resolución de problemas específicos.

1.4.1. Computación Paralela y Distribuida

La computación en paralelo ayuda a resolver los problemas computacionales simultáneamente teniendo en cuenta que los problemas grandes se pueden dividir en problemas más pequeños [38].

La computación distribuida hace referencia a arquitecturas, algoritmos y modelos que permiten crear y gestionar sistemas distribuidos, entendiéndose por sistemas distribuidos la definición propuesta por Tanenbaum et. Al: *“Un sistema distribuido es una colección de computadoras independientes que se presenta a sus usuarios como un único sistema coherente”* [39].

Por lo tanto, podemos decir que la computación paralela ejecuta cualquier proceso en múltiples núcleos de la misma capacidad al mismo tiempo y se comunican entre sí a través de una memoria compartida y, la computación distribuida descompone los procesos para ser ejecutados en diferentes procesadores u ordenadores, esto conlleva a que la computación distribuida contenga una variedad de sistemas y aplicaciones que la computación paralela [39].

El toolbox de MATLAB para la computación paralela y distribuida denominada Parallel Computing Toolbox utiliza todo el procesamiento completo de los motores de cálculo al utilizar los núcleos del CPU y del GPU mediante la ejecución del comando Parfor y la estructura GPUArray [38] [40]. El Parallel Computing Toolbox ayudará en el ahorro de tiempo de ejecución y en la distribución de los datos mediante el uso de los procesadores multicore y clústeres de equipos.

Características principales del Parallel Computing Toolbox

- El comando Parfor ejecuta las tareas en paralelo en múltiples procesadores.
- Se utiliza la arquitectura CUDA para acceder a la GPU la cual se detallará más adelante.
- Ejecución interactiva y por lotes de aplicaciones paralelas.
- Matrices distribuidas para grandes conjuntos de datos.
- Uso completo de los procesadores multicore en los trabajadores que se ejecuten localmente.
- Matrices distribuidas y spmd (single-program-multiple-data) para manipulación de datos y algoritmos de datos paralelos [38] [40].

Matlab distribuye a cada uno de sus núcleos una tarea específica que puede ser realizada en paralelo utilizando uno de sus comandos como es el Parfor, siendo este una herramienta que ejecuta y mejora el rendimiento de muchos segmentos de código [38].

Las características del Parfor son las siguientes:

1. Los procesos del Parfor no son consecutivos debido a que las iteraciones se ejecutan en los diversos núcleos del procesador, por lo tanto, las iteraciones terminan de forma no sucesivas. Las instrucciones se procesan en paralelo como se muestra a continuación en el siguiente gráfico:

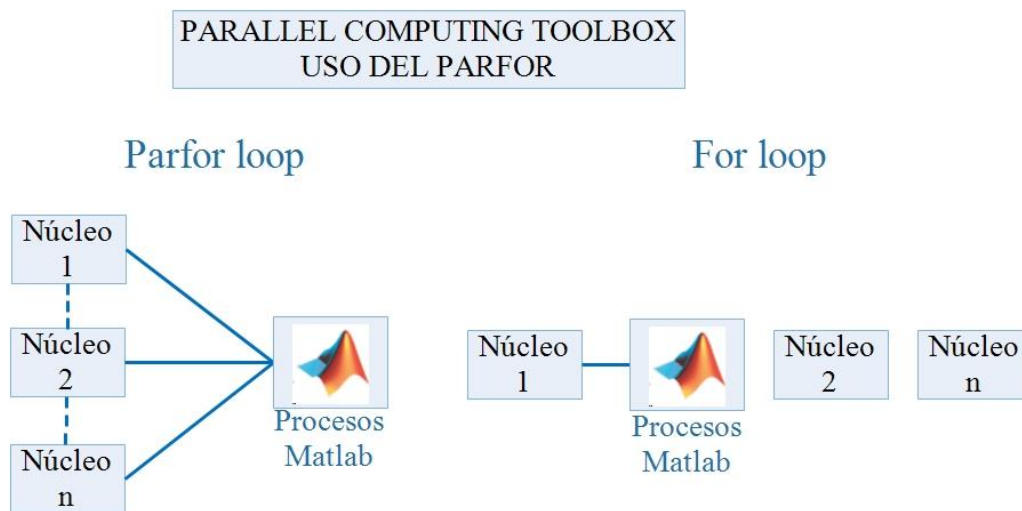


Figura 8. Ejecución del Parfor

2. Las iteraciones de bucle deben ser valores enteros de incremento consecutivos.
3. El cuerpo del Parfor debe ser independiente. Una iteración de bucle no puede depender de una iteración anterior, porque las iteraciones se ejecutan en un orden no determinístico.
4. No se puede usar un bucle parfor dentro de otro bucle parfor [41].

Como parte de la Computación Paralela tenemos la estructura GPUArray y para un mejor entendimiento de su ejecución se debe conocer acerca de que es una GPU y la arquitectura que este maneja como se explica a continuación.

- **La Unidad de Proceso Gráfico (GPU)** permite la ejecución de algunos procesos del computador, y consta de muchos más núcleos caracterizados por utilizar solo funciones matemáticas como suma, resta y multiplicación [42].
- **CUDA** (Compute Unified Device Architecture) permite el acceso a la GPU, el CUDA se ejecuta bajo diferentes lenguajes de programación como es C, C++, Fortran u OpenACC y utiliza cálculo en paralelo. Esta arquitectura fue desarrollada por NVIDIA, permitiendo el incremento de una manera

exponencial en el rendimiento del sistema y en algoritmos que ejecutan extensas cantidades de datos. Entre las GPU más utilizados que admite la arquitectura del CUDA están: GeForce, ION Quadro y Tesla GPUs [42].

Matlab tiene su propia estructura que ejecuta procesos matriciales en el GPU que es el GPUArray, esta estructura sirve para almacenar en el GPU matrices que deben ser de tipo numéricas o lógicas. Para recuperar estas matrices del GPU al espacio de trabajo del Matlab se utiliza la función *gather* [38].

1.5. EMD

La Descomposición Modal Empírica se creó a finales de los años 90 por Huang et. al. en 1998 para analizar las señales multicomponentes no lineales y no estacionarias [9][43][44]. La EMD descompone las señales originales en otras señales llamadas Funciones Modales Intrínsecas (IMF) [43][9].

Las características de las IMF's son descritas a continuación:

1. Se crean a partir de una función $x(t)$.
2. Son señales con componentes de alta y baja frecuencia [45]
3. Representan el modo de oscilación incrustada en los datos [9].

Las IMF son funciones que deben seguir dos criterios que son:

1. El número de extremos máximos y mínimos y el número de cruces por 0 no deben de diferir por más de 1.
2. En cualquier punto de una IMF el valor promedio entre las envolventes de los puntos máximos y mínimos debe ser cero o acercarse [43].

Para cumplir estos dos criterios se necesita un proceso iterativo de tamizado (sifting en su término en inglés), es decir se repite cada vez hasta que sea una función sin máximos ni mínimos [43].

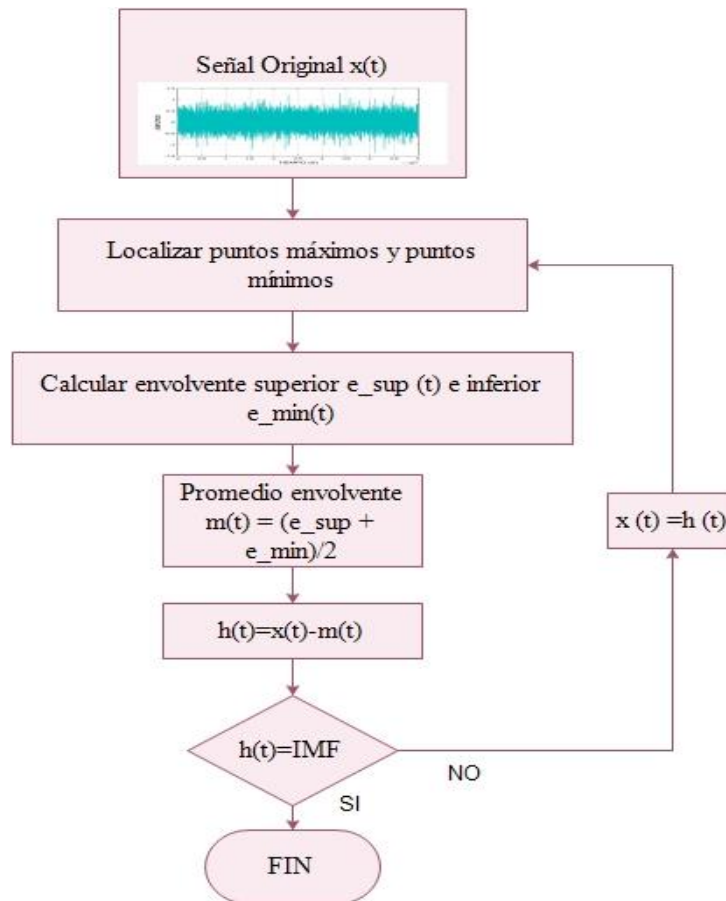


Figura 9. Tamizado EMD

En la Figura 9 del proceso de tamizado en primer lugar se localizan los puntos máximos y puntos mínimos, los mismos que son interpolados para obtener la envolvente de la señal superior $e_{sup}(t)$ y la envolvente inferior $e_{min}(t)$, posteriormente se calcula el promedio de las envolventes mediante la fórmula:

$$m(t) = \frac{(e_{sup} + e_{min})}{2} \quad (15)$$

Luego se obtiene la diferencia entre la señal original y el promedio de la envolvente mediante:

$$h(t) = x(t) - m(t) \quad (16)$$

El resultado de la ecuación (16) debe cumplir los dos criterios dichos anteriormente para que sea una IMF es decir $h(t)$ es IMF, en caso de que esto no se cumpla se vuelve a realizar el mismo procedimiento hasta que se convierta en una IMF [43].

Para el cálculo de las envolventes tanto superior como inferior se utiliza el *spline cúbico* que es el tipo de interpolación más utilizada para la creación de las envolventes [44]. Sin embargo, en el documento “Analysis of Transformer Winding Vibration Based on Modified Empirical Mode Decomposition” [46] se utiliza el *B-spline* como un mejor ajuste para las envolventes y como otro método que se puede manejar es el *spline taut* que ajusta la tensión de la curva spline [9].

El segundo criterio del cálculo del promedio de las envolventes es debido a que el spline cúbico no se ajusta lo suficiente y si el valor promedio es muy bajo ocasiona que el número de iteraciones sea mucho mayor y a su vez provoque una degeneración de la IMF, por lo que Huang et. al. plantea que deben ser contados el número de extremos y número de cruces por cero en cada iteración y estos números sean constantes durante un número definido de iteraciones [43].

A continuación, se presentarán las aplicaciones de la EMD para diferentes áreas:

El método CEEMD-EMD [45] es utilizado para el análisis de las señales de vibración defectuosas de los rodamientos, este método híbrido tiene un mejor rendimiento que los métodos convencionales de EMD y EEMD debido a que resuelve los problemas de confusión de modos de oscilación y sobre todo elimina el ruido residual de las IMF ahorrando muchos recursos a la hora del cálculo computacional.

Otra aplicación de la EMD es su utilización para el diagnóstico de fallos en un motor a diésel como se muestra en el documento “Fault diagnosis for engine based on EMD and wavelet packet BP neural network” [47]. Este utiliza los paquetes de descomposición de wavelet para eliminar el ruido contenido en las señales provenientes del motor a diésel y luego ser procesadas con el algoritmo de EMD que descomponen sus señales de vibración en IMF y así sacar características en el dominio de frecuencia para luego ser utilizadas como datos de entrenamiento de una red neuronal back propagation.

La EEMD es superior a la EMD por que corrige la mezcla de modos de oscilación añadiendo ruido blanco para proporcionar una escala uniforme y simétrica en tiempo-frecuencia que luego se anula mediante el promedio de los conjuntos de datos. Al obtener las IMF se utiliza en ellas la transformada de Hilbert marginal para convertir las señales en el dominio de tiempo-frecuencia y reflejar las características de las señales de vibración que pueda diagnosticar los fallos de los motores a diésel [48].

Las señales de vibración provenientes del devanado de los transformadores es otro método en donde se aplica la EMD descomponiendo sus señales para ver el estado del funcionamiento de estos. Para evitar los errores de ajustes en la parte de la interpolación de los puntos máximos y mínimos se utiliza un nuevo algoritmo que es el B-spline no uniforme, con este algoritmo la interpolación tiene un mejor ajuste a la hora de utilizar el tamizado de la EMD [46].

Otra aplicación de la EMD es el análisis de los datos de precipitaciones de las ciudades, estos a su vez son datos no lineales y no estacionarios ideales para el análisis con la EMD, la conclusión a la que llegan al descomponer los datos es: la no estacionariedad se va debilitando y los componentes de baja frecuencia permanecen cerca de las señales estacionarias [49].

1.6. Conjunto de Datos

Un conjunto de datos se crea a partir de una matriz de características y se puede utilizar en la ejecución de algoritmos de aprendizaje automático [50], las columnas del conjunto de datos comúnmente representan una variable en particular y las filas las clases dentro de estos. El conjunto de datos incluye información que a menudo no está disponible con los datos de campo [51] [52].

Para poder evaluar algunos de los algoritmos de aprendizaje se puede realizar sólo con muestras fuera del conjunto de entrenamiento [53]. Por esta razón se dividirá los datos de entrenamiento en dos conjuntos, el conjunto de entrenamiento y el conjunto de prueba.

El conjunto de entrenamiento siempre se representará con la mayor parte de los datos obtenidos es decir con un 70% y la parte del conjunto de validación se tomará el porcentaje restante. El conjunto de prueba nos da el error del modelo, por lo que el modelo que tenga menor error al momento de ser entrenado con el conjunto de entrenamiento y ser validado con el conjunto de prueba es el modelo que nos convendrá utilizar [53].

Las clases son un conjunto de datos con características específicas y el número de clases esperadas afecta la elección del algoritmo. La clase que representa la categoría de interés se denomina clase “positiva” y las clases de fondo se denominan “negativas” y

cuando existen más de dos clases se denomina multiclasas, algunos algoritmos de aprendizaje pueden tratar estos problemas de multiclasas, entre estos tenemos: Árboles de decisiones, Random Forest, etc.

1.7. Aprendizaje Automático

El aprendizaje automático (Machine Learning) es parte de la inteligencia artificial (IA) y su función principal es hacer que las computadoras puedan modelar procesos y predecirlos por sí solas a través de un conjunto de datos [53][54], gracias a esto y al avance de la computación los métodos de aprendizaje automático se utilizan en un sin número de aplicaciones.

Los métodos de aprendizaje automático proporcionan técnicas para la realización de modelos computacionales de fenómenos tanto naturales como artificiales, el proceso para crear modelos se llama “entrenamiento” y los datos utilizados para los entrenamientos se denominan “datos de entrenamiento” [53]. Para que un entrenamiento sea preciso se necesita una gran cantidad de datos de entrenamiento siendo el primer paso la recolección de estos.

Para el aprendizaje automático se debe probar diferentes tipos de algoritmos o de modelos [53], con el fin de encontrar el que de mejor rendimiento debido a que según el teorema No Free Lunch indica: " Si un algoritmo es bueno en clasificar un tipo de datos no puede ser bueno para clasificar otro tipo de datos" [55]. El rendimiento dependerá del costo computacional (tiempo de entrenamiento y memoria disponible), la cantidad, calidad de los datos de entrenamiento y de las propias características del algoritmo.

1.7.1. Tipos de Aprendizaje

El aprendizaje es un campo muy complejo y amplio por lo cual se ha dividido en varios tipos de aprendizajes entre los cuales se consideran los siguientes:

1.7.1.1. Aprendizaje Supervisado

Basa su algoritmo en la experiencia, contiene información característica para lograr hacer predicciones o tomar decisiones [56]. Uno de los ejemplos que podemos indicar de estos tipos de algoritmos es la detección de spam en los correos electrónicos para esto se proporciona una etiqueta de SPAM y NO SPAM, este algoritmo debe encontrar

una regla para poder etiquetar bien si un nuevo correo puede ser spam o no [53] [57] [58].

Los algoritmos de aprendizaje supervisado utilizan el valor de las variables de entrada para elaborar un modelo y así predecir el valor de las variables de salida (que son conocidas en el conjunto de entrenamiento). Si las variables de salida son valores continuos entonces el modelo predictivo se denomina “función de regresión” y si las variables de salida toman un conjunto discreto de valores entonces el modelo predictivo se denomina “clasificador [53].

1.7.1.2. Aprendizaje no Supervisado

Los algoritmos del aprendizaje no supervisado se basan en el agrupamiento de los datos con el fin de encontrar algún patrón o poder organizarlos. El algoritmo descubre las características o patrones entre los grupos de datos para dividirlos coherentemente [53]. Un ejemplo es aplicar estos tipos de algoritmos para poder clasificar el comportamiento social de una población específica (humanos, roedores, aves) entre otros [58].

1.8. Random Forest

Random Forest llamado también bosque aleatorio es un conjunto de árboles de decisiones en donde cada uno de los árboles depende del valor de un vector aleatorio con el mismo tamaño y cada árbol vota unitariamente por la clase más popular en la entrada [59].

Para saber cómo funciona Random Forest primero conoceremos como es su algoritmo [59]:

- Sea N el número de casos en el conjunto de entrenamiento, M es el número de variables en el clasificador.
- Sea m el número de variables de entrada a ser usado para determinar la decisión en un nodo dado; m debe ser mucho menor que M . El valor de m se mantiene constante durante el crecimiento del bosque.
- Elegir un conjunto de entrenamiento para un árbol y usar el resto de los casos de prueba para estimar el error.

- Para cada nodo del árbol, elegir aleatoriamente m variables en las cuales basar la decisión. Calcular la mejor partición del conjunto de entrenamiento a partir de las m variables.

Una vez que el algoritmo es entrenado, la evaluación de cada nueva entrada se realizará con un conjunto de árboles. El resultado se obtendrá por medio de las coincidencias que tengan cada árbol, y en caso de regresión por el valor promedio de los resultados [59]. Por otra parte, la tasa de error que se obtiene de este clasificador Random Forest depende de dos condiciones:

- La fuerza de los clasificadores individuales en el bosque y,
- La correlación entre ellos en términos de funciones del margen bruto [59].

Se puede considerar que Random Forest es superior a otros clasificadores como el perceptrón multicapa y las redes neuronales artificiales en configuración del perceptrón con propagación hacia atrás [60] debido a las siguientes características:

- Funciona para grandes bases de datos por esta razón es muy utilizable en Big data.
- Puede manejar muchas variables en los datos de entrenamiento.
- Da estimaciones sobre que variables son importantes en la clasificación.
- Calcula las proximidades entre pares de casos que se pueden utilizar en la agrupación.
- Tiene métodos para equilibrar el error en conjuntos de datos desequilibrados de población de clase.

Matlab también tiene su propia herramienta para la ejecución de Random Forest que es el comando `Treebagger` el cual combina el resultado de los árboles de decisión para mejorar la generalización de estos. El comando `Treebagger` consta de parámetros que deben ser definidos previamente por el usuario. Entre estos describiremos los parámetros utilizados para este proyecto [61].

1. La cantidad de árboles de decisiones seleccionados para utilizar el comando `Treebagger`.

2. El parámetro ‘OOBPrediction’ que por su defecto se encuentra desactivada, pero en este caso fue utilizada para calcular la probabilidad de clases previstas para cada árbol.
3. El parámetro ‘options’ que esta a su vez cuenta con el subparámetro 'UseParallel' que permite utilizar el parallel computing toolbox para la ejecución de los árboles de decisiones en el comando Treebagger.
4. El parámetro ‘OOBVarImp’ sirve para encontrar la estimación de la mejor característica y se debe inicializar en el comando Treebagger para poder calcular la matriz de 'OOBPredictorDeltaError' que luego de un proceso de desorden de las mismas se produce un menor desempeño del clasificador [62].

1.9. Métricas para Tareas de Clasificación

Para un mejor rendimiento del aprendizaje supervisado se utiliza la matriz de confusión la misma que muestra el rendimiento de los sistemas de clasificación. Su construcción se realiza con la información obtenida desde un conjunto de datos de referencia y la salida del clasificador. La diagonal de esta matriz representa el porcentaje de aciertos de cada clase y los demás valores representan el porcentaje de confusión entre cada clase [63], sus columnas representan la salida del clasificador y sus filas la clase verdadera como se muestra en la Tabla 2.

VP: número de verdaderos positivos

FN: número de falsos negativos

FP: números de falsos positivos

VN: números de verdaderos negativos

Número total de muestras		Salida del Clasificador	
		Positiva	Negativa
Clase Verdadera	Positiva	VP	FN
	Negativa	FP	VN

Tabla 2. Matriz de Confusión de 2x2 [64].

El número de Falsos Positivos se determina de la siguiente manera [65]:

$$FP = \frac{\# \text{ Decisiones Falsas Positivas}}{\# \text{ Clases Verdaderas Negativas}} \quad (17)$$

El número de Falsos Negativos es igual a [65]:

$$FN = \frac{\# \text{ Decisiones Falsas Negativas}}{\# \text{ Clases Verdaderas Positivas}} \quad (18)$$

Es necesario conocer sobre las diferentes métricas utilizadas las mismas que permiten medir el rendimiento del algoritmo de aprendizaje [66]. Existen algunas métricas de las cuales se utilizaron las siguientes.

1.9.1. Exactitud (Accuracy)

Indica la cercanía de la lectura con respecto a una entrada o valor conocido, es igual al límite de concordancia entre el resultado de una medición y el valor verdadero [67]. Está relacionado con la Especificidad y Sensibilidad. Su ecuación para el cálculo se determina como [65]:

$$exactitud = \frac{VP + VN}{VP + VN + FP + FN} \quad (19)$$

1.9.2. Especificidad

Es el total de verdaderos negativos para el número total de clases verdaderas negativas, es decir, establece los casos realmente negativos [65].

$$Especificidad = \frac{\#VN}{\# \text{ Clases Verdaderas Negativas}} \quad (20)$$

1.9.3. Sensibilidad o Recall

Se utiliza para determinar los casos realmente positivos, es decir, consiste en la relación entre el número de Verdaderos Positivos con el número de Clases Verdaderas Positivas [53] [65]:

$$Sensibilidad = \frac{\#VP}{\# \text{ Clases Verdaderas Positivas}} \quad (21)$$

1.9.4. Precisión

Se define como el número de verdaderos positivos para la suma del número de falsos positivos más los verdaderos positivos como se muestra en la siguiente fórmula [53]:

$$Precisión = \frac{\#VP}{(\#FP + \#VP)} \quad (22)$$

1.9.5. F-Score

Es una medida de comparación entre la Precisión y el Recall, es decir la media armónica entre ambas métricas que se resuelve como [53].

$$F_1 = \frac{2 \times Precisión \times Recall}{Precisión + Recall} \quad (23)$$

Lo correcto es mantener tanto el recall como la precisión altas [53].

2. METODOLOGÍA

En este proyecto para el diagnóstico de fallos en señales de vibración de rodamiento y engranajes, se utilizó el algoritmo de EMD y se dividió en varias etapas para su mejor rendimiento y disminución del costo computacional como podemos ver en la siguiente figura:

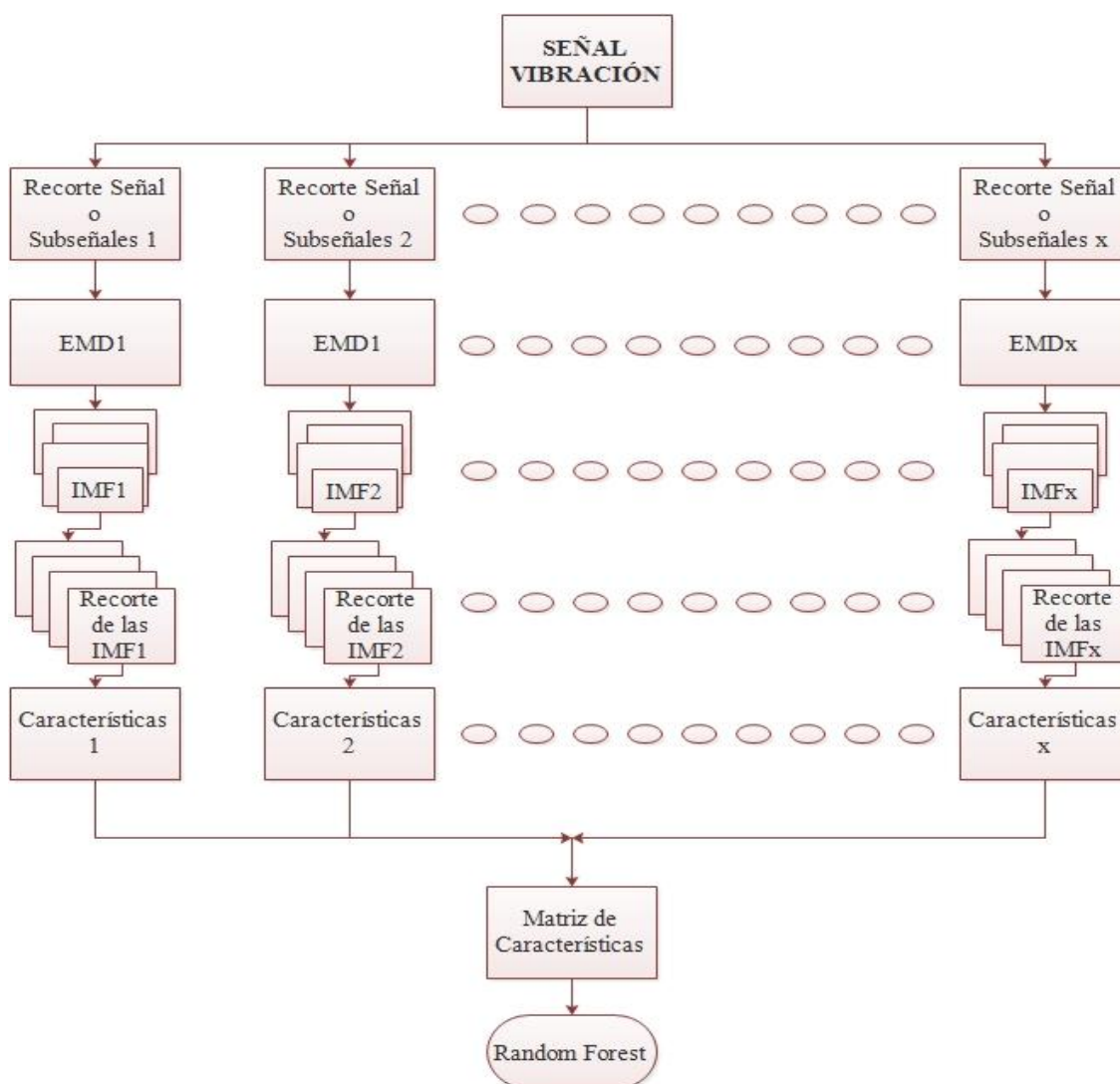


Figura 10. Proceso del diagnóstico de fallos.

Para el diagnóstico de fallos lo primero es la obtención de las señales de vibración para ser procesadas en el software matemático de Matlab recortando las señales originales en subseñales con un número de muestras menores, con estas subseñales se utiliza el algoritmo de EMD que descompone las señales en IMF para sacar las características en el dominio de tiempo y de frecuencia las cuales se describieron en el subsección 1.8 y se

crea una matriz de todas las características mencionadas para entrenar el clasificador Random Forest. A continuación se explicará detalladamente el procedimiento utilizado.

2.1. Señales de Vibración

La base de datos de las señales de vibración fue facilitada por el grupo de investigación GIDTEC y consta de características diferentes tanto en rodamientos como en engranajes mostrados en la tabla 3.

SEÑALES	RODAMIENTOS	ENGRANAJES
Tiempo de duración	20 sg.	10 sg.
Tamaño Total de la Muestra	1000000	500000
Repeticiones	5 (R1, R2, R3, R4, R5)	5 (R1, R2, R3, R4, R5)
Frecuencia	8, 10, 15 Hz (F1, F2, F3)	Constante: 8, 10, 15 Hz (F1, F2, F3) y 0.5 Hz (F4)
		Variable: 8Hz-15Hz sin onda seno 0.5 Hz (F4)
Tipos de cargas	L1, L2, L3	L1, L2, L3

Tabla 3. Especificaciones de las Señales [68].

El total de las señales obtenidas en los rodamientos son de 315 y en los engranajes son 750, cada una caracterizada por el mismo número de repeticiones en frecuencias, cargas y fallas. Tanto en los rodamientos como en los engranajes la frecuencia es la velocidad del eje y las fallas son los estados en los que se encuentran los elementos mecánicos, mientras que en los rodamientos la carga es la carga total en el eje y en los engranajes es la carga total en el freno magnético [68].

Los diferentes tipos de fallas en engranajes y en rodamientos en las señales de vibración propuestos por el grupo de investigación para su análisis se encuentran en la tabla 4:

FALLOS	RODAMIENTOS	ENGRANAJES
P1	Normal	Normal
P2	Pista interna	Diente roto (11.5%)
P3	Pista externa	Diente roto (19.5%)
P4	Bola	Diente roto (28 %)
P5	Pista interna	Diente roto (37.2%)
P6	Pista interna	Diente roto (51.7%)
P7	Pista externa	Diente roto (60.5%)
P8		Diente roto (70.2%)
P9		Diente roto (85.6%)
P10		Diente roto (100%)

Tabla 4. Tipos de fallas en las bases de datos [68].

Se utilizaron las muestras de un solo acelerómetro (Acc1). Los acelerómetros que se montaron se muestran en la siguiente tabla:


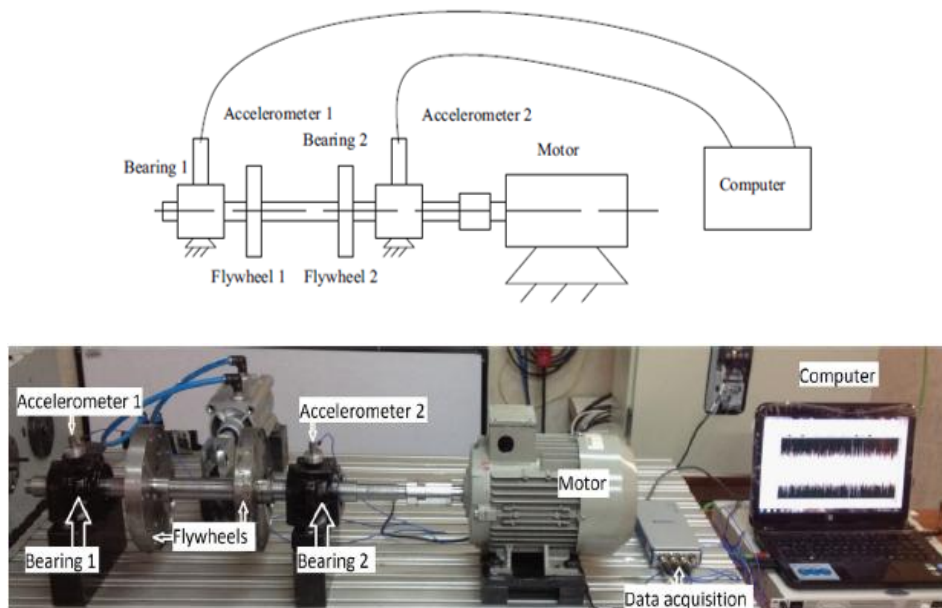
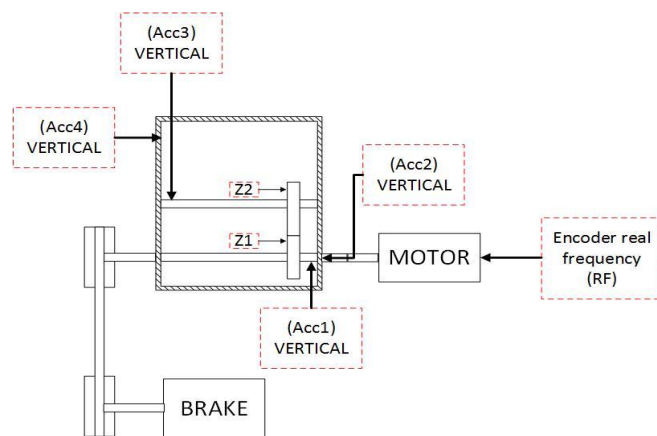
Especificaciones	ACC1	Fotos
Rodamiento	Uniaxial PCB 352C03, sensibilidad 10.06 mV/g	
Engranaje	IMI sensor modelo: 603C01 sensibilidad 100 mV / g	

Tabla 5. Características de sensores utilizados en la obtención de las señales de vibración.

En la figura 11 se muestra la posición en donde se encuentran colocados los sensores en la caja de engranajes y en la chumacera de los rodamientos.



(a)



(b)

Figura 11. Posición de los acelerómetros en los rodamientos (a), posición de los acelerómetros en la caja de engranajes (b) [68].

Para seleccionar las señales de vibración se utilizó una función (`list_file`) en Matlab, la cual se encarga de tomar las rutas de cada una de las señales individuales de vibración, siendo necesario cargar la ruta en donde se ha guardado la base de datos ya sea de rodamientos o de engranajes, el diagrama de flujo del función `list_file` es el siguiente.



Figura 12. Diagrama de la carga de archivos.

2.2. Recorte de las Señales de Vibración

Para el mejor procesamiento de las señales y la optimización del tiempo del algoritmo de EMD se recortó la señal original en 100 subseñales, es decir cada subseñal constará de 5.000 muestras en el caso de rodamientos y 10.000 muestras en engranajes. En la Figura 13 se muestra el recorte de la señal original en varias subseñales.

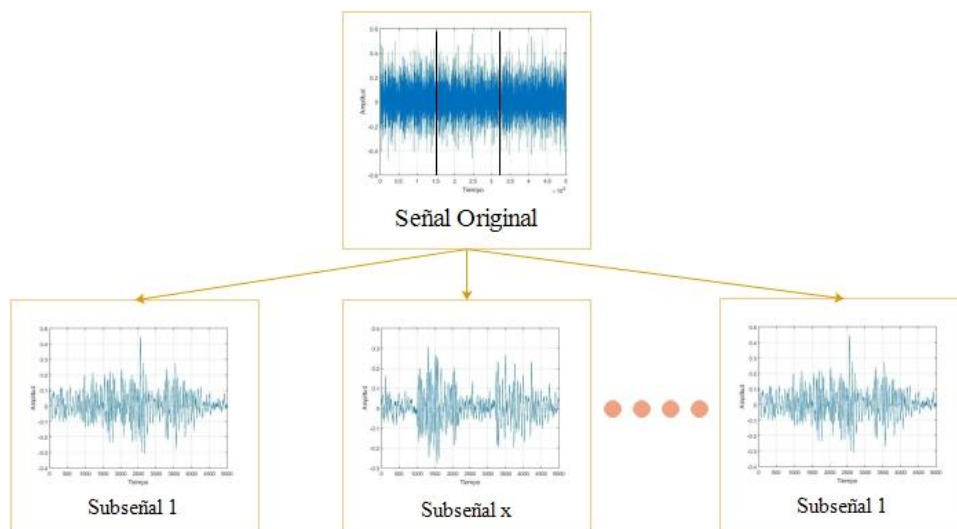


Figura 13. Recorte de las Señales.

2.3. Aplicación de la EMD y extracción de la IMF

Después de obtener las subseñales resultantes el siguiente paso es sacar las IMF para esto se utilizó el toolbox de EMD de Matlab desarrollado por Gabriel Rilling, Patrick Flandrin y Paulo Goncalves [44].

2.3.1. Ejecución de la EMD

Para la optimización de recursos computacionales se utilizó el Parallel Computing Toolbox, y se usó el comando Parfor.

La cantidad de señales en toda la base de datos se utilizó como iteraciones para el Parfor. Cada subseñal da como resultado las IMF que se configuran con el máximo número de iteraciones de 2000 para evitar el sobre tamizado [44].

Las IMF se calcularon mediante el proceso de tamizado explicado en el capítulo 1 en la sección 1.4, por esta razón el número de IMF de cada subseñal serán distintas. Como la cantidad de IMF por señal de vibración es diferente, el paso previo es encontrar el mínimo de las IMF's y recortarlas para que coincidan entre sí con la misma cantidad. El mínimo de IMF's como resultado de la base de datos para rodamientos es de 6 y en la de engranajes es 8.

2.3.2. Conversión de las IMF en GPUArray

Para la aplicación de la EMD se desarrolló una función (`process_file`) que se encarga de generar las IMF, convertirlas en GPUArray y conseguir las características de todas las IMF como se muestra en la Figura 14. La estructura "GPUArray" cambia el tipo de variable de las IMF de tipo numérico doble a tipo GPUArray que permite su ejecución en la GPU. A continuación, se muestra el Diagrama de flujo de esta función.

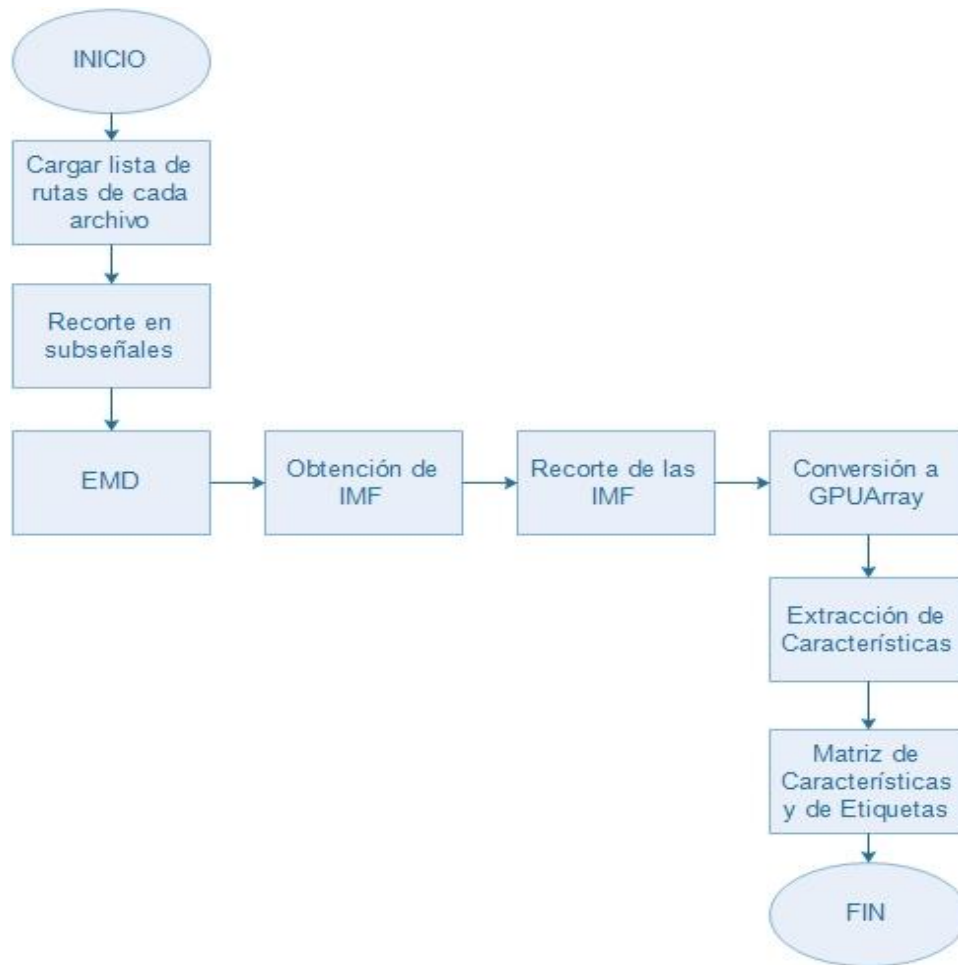


Figura 14. Diagrama de Flujo de process_file.

2.4.Implementación de las Características

Toda señal cuenta con características que pueden ser extraídas. Para este proyecto se implementaron las características en el dominio de frecuencia como una condición que el grupo de investigación requería y de tiempo porque son las características más comunes. De esta manera las características fueron programadas para que cumplan las especificaciones para ser ejecutadas en la GPU mediante la utilización de la estructura GPUArray de Matlab.

A continuación, se creó una matriz de todas las características en donde las columnas representan la cantidad de características que se obtienen con la multiplicación del número de IMF's por el número de características programadas y las filas representan la cantidad de subseñales obtenidas dando como resultado el tamaño de las matrices explicados en la tabla 6.

MATRIZ DE CARACTERÍSTICAS

Característica 1.1	Característica 1.2	Característica 1.3	Característica 1.x
Característica 2.1	Característica 2.2	Característica 2.3	Característica 2.x
Característica 3.1	Característica 3.2	Característica 3.3	Característica 3.x
Característica q.1	Característica q.2	Característica q.3	Característica q.x

Figura 15. Matriz de Características

Tamaño de las Matrices de Características		
	Rodamientos	Engranajes
Tiempo	31500 x 48	75000 x 64
Frecuencia	31500 x 54	75000 x 72
Tiempo y Frecuencia	31500 x 102	75000 x 136

Tabla 6. Tamaño de las Matrices de Características.

La matriz de características se dividió en una matriz de entrenamiento que consta del 70% de la matriz original y una de prueba con el 30% restante, teniendo en cuenta que lo mismo sucede con la matriz de etiquetas la cual representa las señales con fallos y sin fallos dentro de la matriz de características. Aquí se utilizó el comando `crossvalind` que permite sesgar los datos y tomar aleatoriamente el 70% de estos.

2.5. Aplicación de Random Forest

Después se utilizó el comando `Treebagger` que se configuró con el número de árboles de decisión y el conjunto de entrenamiento de la matriz de características y de las etiquetas. Para ver el porcentaje de error de entrenamiento de Random Forest se utilizó 100 árboles de decisión y la opción del `parallel computing toolbox` que permitió mejorar los tiempos de procesamiento de Random Forest, ya que entre más árboles de decisión que se utilicen mayor es el costo computacional.

Una vez ejecutado el comando Treebagger con el conjunto de datos de entrenamiento y mostrar el error se volvió a ejecutar el comando pero con el conjunto de prueba para evaluar a través de la matriz de confusión si Random Forest clasificaba correctamente los tipos de fallos y de esta manera sacar el valor de todas las métricas utilizadas.

La parte del conjunto de datos de entrenamiento en cada árbol de decisión de Random Forest devuelve el tipo de clase con el cual el árbol fue entrenado, se saca el promedio de la salida de todos los árboles dando un conjunto de errores y con este conjunto se vuelve a entrenar cada árbol para sacar una diferencia de errores con el subparámetro 'OOBPredictorDeltaError' y mediante esto se calculan las mejores características identificadas a través de las tablas que se muestra en el análisis de resultados.

3. ANÁLISIS DE RESULTADOS

Como hablamos en el punto anterior utilizamos dos bases de datos para la extracción de características y los diagnósticos de fallos los cuales ejecutamos en el dominio de tiempo y en el dominio de frecuencia y con la suma de los dos dominios anteriores dando seis diferentes resultados y estos son procesados para ser evaluados con el clasificador de Random Forest.

A continuación, se muestra el *OOBError* a partir de la ejecución del clasificador con el conjunto de entrenamiento:

Dominios	Rodamientos	Engranajes
Tiempo	4.794%	41.75%
Frecuencia	7.946%	51.74%
Tiempo & frecuencia	2.698%	39.19%

Tabla 7. Resultados de OOBError obtenidos del clasificador Random Forest.

En la Figura 16 se muestra las curvas de convergencia del proceso de entrenamiento de Random Forest en Rodamientos y en la Figura 17 en Engranajes.

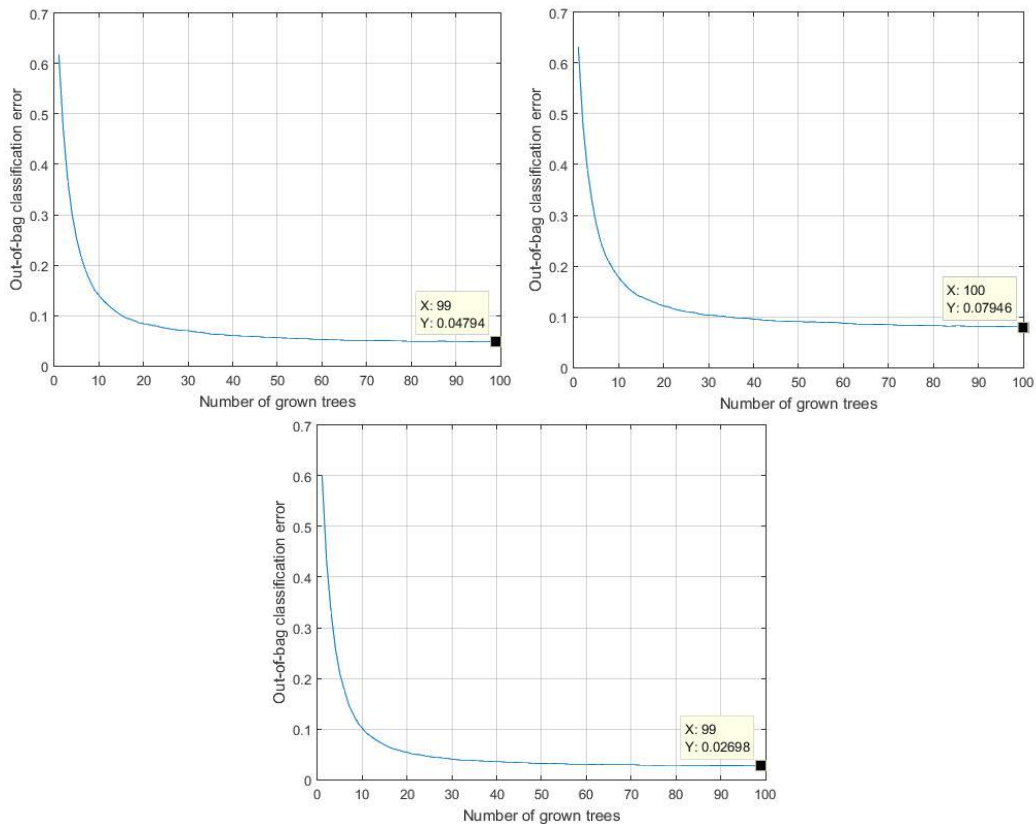


Figura 16. Curvas de convergencia del proceso de entrenamiento de Random Forest en Rodamientos.

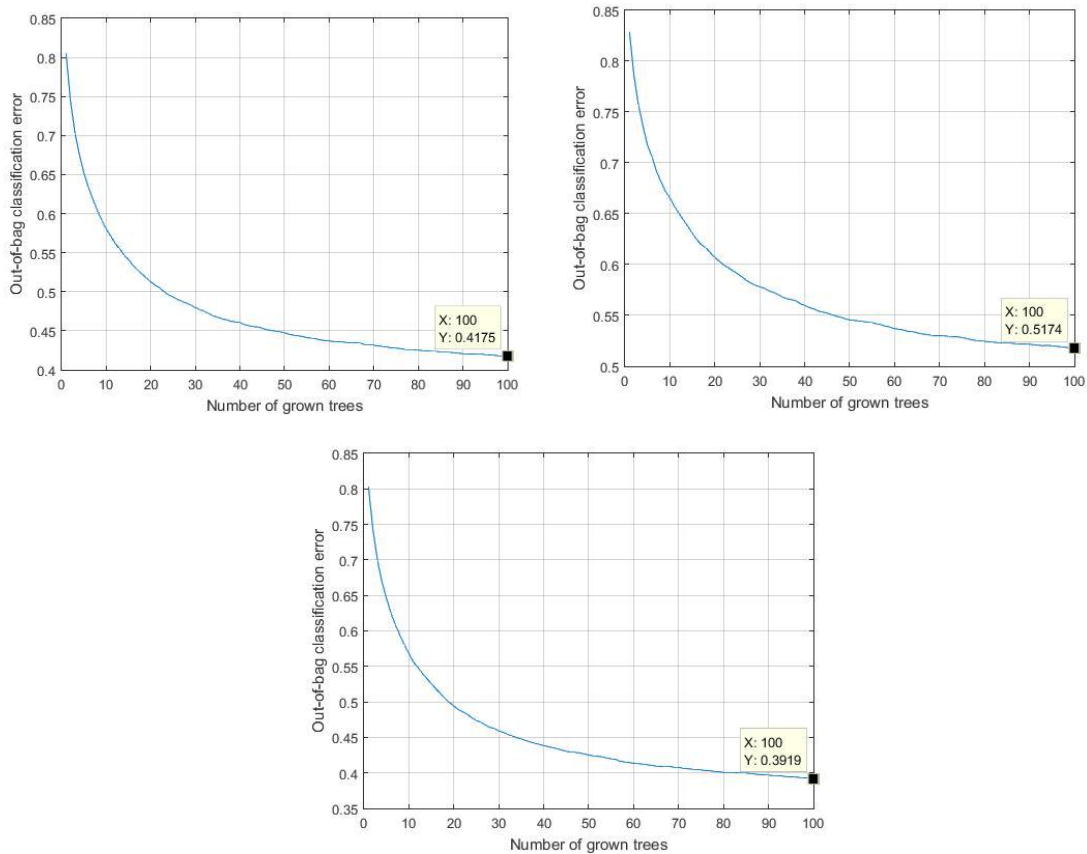


Figura 17. Curvas de convergencia del proceso de entrenamiento de Random Forest en Engranajes.

En las siguientes tablas se muestran las matrices de confusión que fueron creadas a partir del conjunto de prueba en sus diferentes dominios tanto de Rodamientos como de Engranajes que fue obtenida de las bases de datos:

		SALIDA DEL CLASIFICADOR						
CLASES		1	2	3	4	5	6	7
CLASE VERDADERA	P1	1293	0	18	2	0	37	0
	P2	2	1297	2	4	21	18	6
	P3	8	1	1328	0	0	0	13
	P4	13	33	25	1216	8	4	51
	P5	0	27	0	6	1273	40	4
	P6	0	21	0	3	33	1292	1
	P7	2	3	35	6	2	3	1299

Tabla 8. Matriz de confusión de Rodamiento en dominio de tiempo.

		SALIDA DEL CLASIFICADOR						
		1	2	3	4	5	6	7
CLASES								
CLASE VERDADERA	P1	1307	5	3	2	6	26	1
	P2	14	1241	9	6	39	28	13
	P3	5	3	1331	1	1	2	7
	P4	22	29	10	1195	21	31	42
	P5	2	24	1	2	1227	53	41
	P6	3	49	4	2	59	1217	16
	P7	0	33	30	4	50	42	1191

Tabla 9. Matriz de confusión de Rodamiento en dominio de frecuencia.

		SALIDA DEL CLASIFICADOR						
		1	2	3	4	5	6	7
CLASES								
CLASE VERDADERA	P1	1312	1	5	1	2	29	0
	P2	2	1318	2	2	11	13	2
	P3	1	3	1345	0	0	0	1
	P4	5	7	11	1301	7	5	14
	P5	0	4	1	1	1310	25	9
	P6	0	15	0	1	28	1305	1
	P7	1	7	5	5	3	3	1326

Tabla 10. Matriz de confusión de Rodamiento en dominio de tiempo en conjunto con la frecuencia.

		SALIDA DEL CLASIFICADOR									
		1	2	3	4	5	6	7	8	9	10
CLASES											
CLASE VERDADERA	P1	1319	183	92	81	41	102	111	94	61	166
	P2	142	1602	68	117	30	62	64	63	63	39
	P3	67	68	1284	178	145	147	104	33	140	84
	P4	64	74	111	1278	147	149	183	67	144	33
	P5	64	91	195	196	1165	167	110	38	123	101
	P6	121	89	168	167	180	1087	139	70	116	113
	P7	73	78	103	175	91	147	1296	87	111	89
	P8	130	69	42	85	58	79	107	1363	107	210
	P9	61	72	109	143	85	103	91	61	1427	98
	P10	158	69	81	32	63	104	66	143	69	1465

Tabla 11. Matriz de Confusión de Engranaje en dominio de tiempo.

		SALIDA DEL CLASIFICADOR									
		1	2	3	4	5	6	7	8	9	10
	CLASES										
CLASE VERDADERA	P1	1121	309	77	85	54	84	192	138	80	110
	P2	324	1162	114	92	52	93	122	109	129	53
	P3	56	98	1057	230	173	170	90	29	201	146
	P4	64	101	228	1037	176	163	169	53	165	94
	P5	95	87	229	261	824	231	139	71	159	154
	P6	82	90	183	168	220	883	217	101	162	144
	P7	139	108	88	194	105	164	1075	194	111	72
	P8	189	102	60	109	73	102	165	1199	98	153
	P9	35	68	130	136	67	69	80	25	1502	138
	P10	136	102	120	93	89	94	79	119	170	1248

Tabla 12. Matriz de confusión de Engranaje en dominio de frecuencia.

		SALIDA DEL CLASIFICADOR									
		1	2	3	4	5	6	7	8	9	10
	CLASES										
CLASE VERDADERA	P1	1442	188	78	67	38	64	106	110	59	98
	P2	176	1620	58	77	27	65	59	62	62	34
	P3	52	89	1273	220	139	139	89	27	141	81
	P4	56	84	103	1377	168	123	136	43	111	49
	P5	69	84	225	196	1148	173	106	43	116	90
	P6	95	67	145	140	214	1140	137	78	116	118
	P7	98	67	76	156	77	137	1357	117	74	91
	P8	126	65	42	68	59	69	114	1457	78	172
	P9	36	60	81	110	53	77	83	29	1584	137
	P10	127	43	74	27	57	79	48	96	110	1589

Tabla 13. Matriz de confusión de Engranaje en dominio de tiempo en conjunto con la frecuencia.

A partir de las matrices de confusión se obtienen las métricas de exactitud, especificidad, sensibilidad o recall, precisión y F-Score para evaluar el clasificador mediante el conjunto de prueba por cada clase. De acuerdo a los resultados obtenidos, como podemos ver en la Tabla 14 y 15 como mejor prueba tenemos la de Rodamientos en el dominio de tiempo en conjunto con la frecuencia con una Exactitud de 99.3 %y en

el caso de la especificidad con 99.59% respectivamente, siendo estos valores los promedios de todas las clases y considerando que mientras más se acerque a 100% más óptimo será su resultado.

EXACTITUD						
	RODAMIENTOS			ENGRANAJES		
	T&F	F	T	T&F	F	T
Clasificador	99.3%	97.76%	98.63%	92.43%	89.87%	91.81%

Tabla 14. Datos de Exactitud del clasificador.

ESPECIFICIDAD						
	RODAMIENTOS			ENGRANAJES		
	T&F	F	T	T&F	F	T
Clasificador	99.59%	98.69%	99.20%	95.8%	94.37%	95.44%

Tabla 15. Datos de Especificidad del clasificador.

En los resultados de cada prueba como mejor rendimiento con respecto a la métrica de sensibilidad o recall y F-Score se da en rodamientos en la clase P3 con 99.63% y 98.93% y en engranajes la P2 con 72% y 70.18% reflejados en la Tabla 16 y 18 respectivamente, en el caso de la precisión en la Tabla 17 tenemos en rodamientos la clase P1 con 99.32% y en engranajes P8 con 70.66%, siendo que los mejores resultados pertenecen a los Rodamientos en el dominio de tiempo en conjunto con la frecuencia que en comparación a los Engranajes su rendimiento es mucho mayor, dando como mejores pruebas mediante Random Forest la clase de pista externa que corresponde a P3 en rodamientos y en engranajes P2 que es Diente roto a un 11.5%.

SENSIBILIDAD O RECALL						
CLASES	RODAMIENTOS			ENGRANAJES		
	T&F	F	T	T&F	F	T
P1	97.19%	96.81%	95.78%	64.09%	49.82%	58.62%
P2	97.63%	91.93%	96.07%	72.00%	51.64%	71.20%
P3	99.63%	98.59%	98.37%	56.58%	46.98%	57.07%
P4	96.37%	88.52%	90.07%	61.20%	46.09%	56.80%
P5	97.04%	90.89%	94.30%	51.02%	36.62%	51.78%
P6	96.67%	90.15%	95.70%	50.67%	39.24%	48.31%
P7	98.22%	88.22%	96.22%	60.31%	47.78%	57.60%
P8				64.76%	53.29%	60.58%
P9				70.40%	66.76%	63.42%
P10				70.62%	55.47%	65.11%

Tabla 16. Datos de Sensibilidad o Recall de cada clase.

PRECISIÓN						
CLASES	RODAMIENTOS			ENGRANAJES		
	T&F	F	T	T&F	F	T
P1	99.32%	96.60%	98.10%	63.33%	50.02%	59.98%
P2	97.27%	89.67%	93.85%	68.44%	52.18%	66.89%
P3	98.25%	95.89%	94.32%	59.07%	46.24%	56.99%
P4	99.24%	98.60%	98.30%	56.48%	43.12%	52.12%
P5	96.25%	87.46%	95.21%	57.98%	44.95%	58.10%
P6	94.57%	86.99%	92.68%	55.18%	43.01%	50.63%
P7	98%	90.85%	94.54%	60.45%	46.18%	57.07%
P8				70.66%	58.83%	67.51%
P9				64.63%	54.09%	60.44%
P10				64.62%	53.98%	61.09%

Tabla 17. Datos de precisión de cada clase.

F-SCORE						
CLASES	RODAMIENTOS			ENGRANAJES		
	T&F	F	T	T&F	F	T
P1	98.24%	96.71%	96.93%	63.71%	49.92%	59.29%
P2	97.45%	90.78%	94.95%	70.18%	51.91%	68.98%
P3	98.93%	97.22%	96.30%	57.80%	46.60%	57.03%
P4	97.78%	93.29%	94.01%	58.75%	44.55%	54.36%
P5	96.64%	89.14%	94.75%	54.28%	40.36%	54.76%
P6	95.60%	88.54%	94.17%	52.83%	41.04%	49.44%
P7	98.11%	89.52%	95.37%	60.38%	46.96%	57.33%
P8				67.58%	55.92%	63.86%
P9				67.39%	59.76%	61.90%
P10				67.49%	54.71%	63.04%

Tabla 18. Datos de F-Score de cada clase.

A continuación, en las siguientes Figuras se muestra las mejores características de acuerdo al error que presentan, las mismas en el dominio de tiempo, dominio de frecuencia y dominio de tiempo en conjunto con la frecuencia tanto en Rodamientos como en Engranajes al entrenar el clasificador Random Forest con el conjunto de entrenamiento.

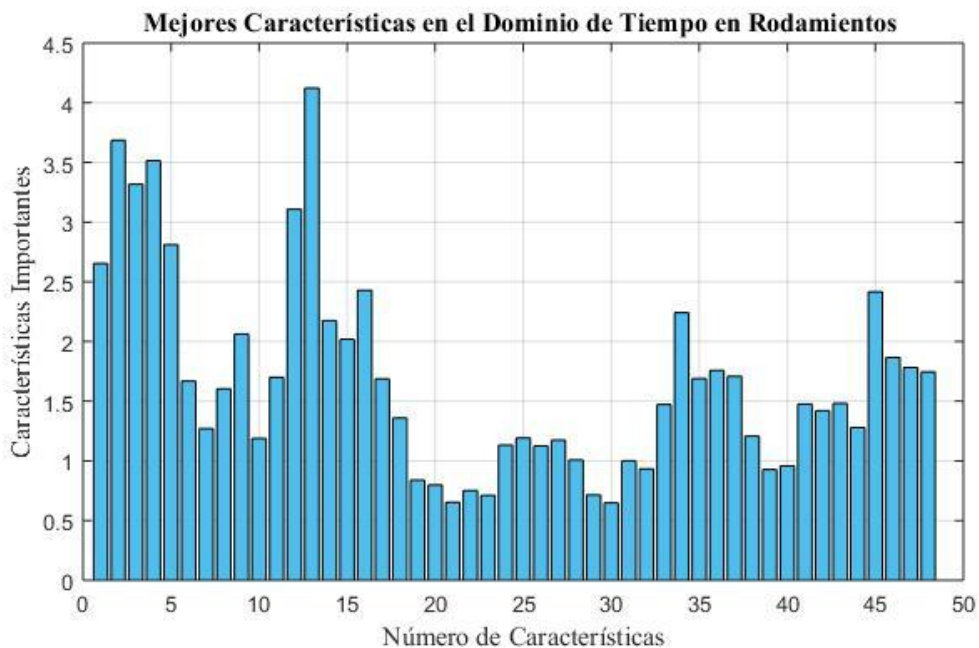


Figura 18. Mejores Características en el Dominio de Tiempo en Rodamientos.

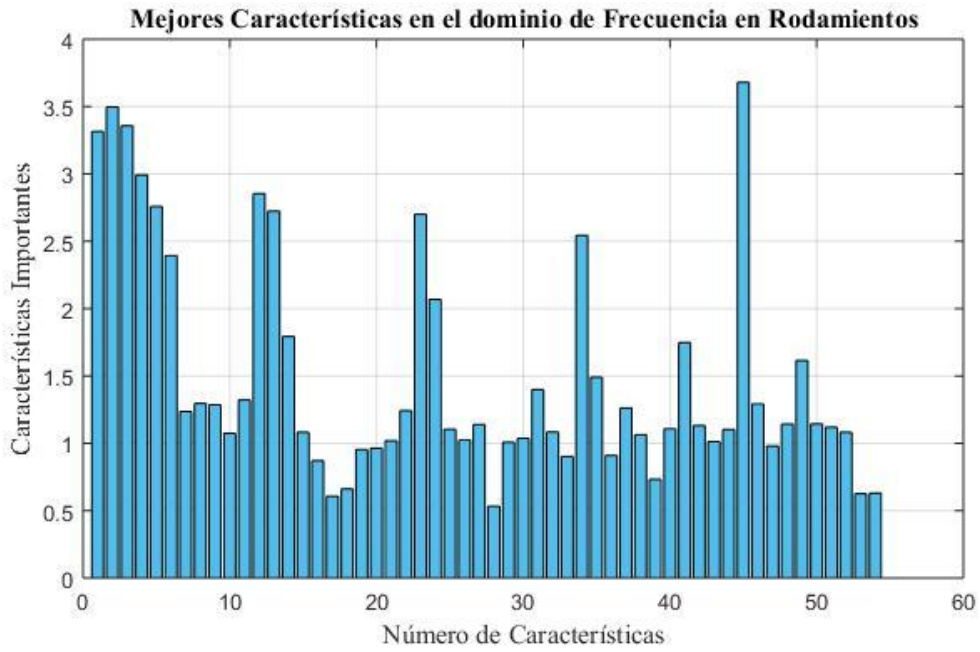


Figura 19. Mejores Características en el Dominio de Frecuencia en Rodamientos.

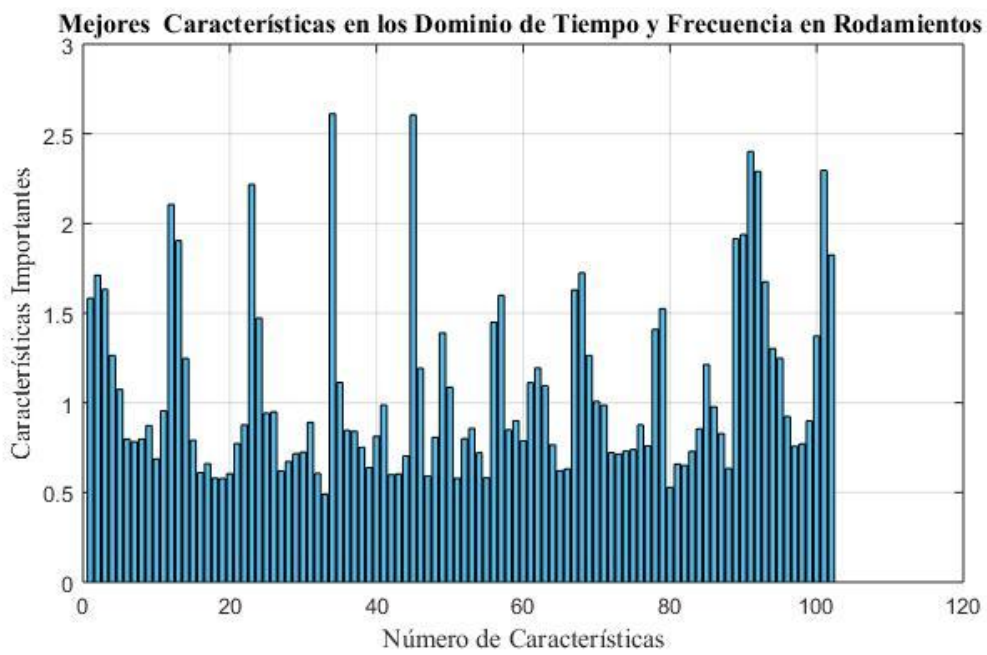


Figura 20. Mejores Características en los Dominios de Tiempo en conjunto con la Frecuencia en Rodamientos.

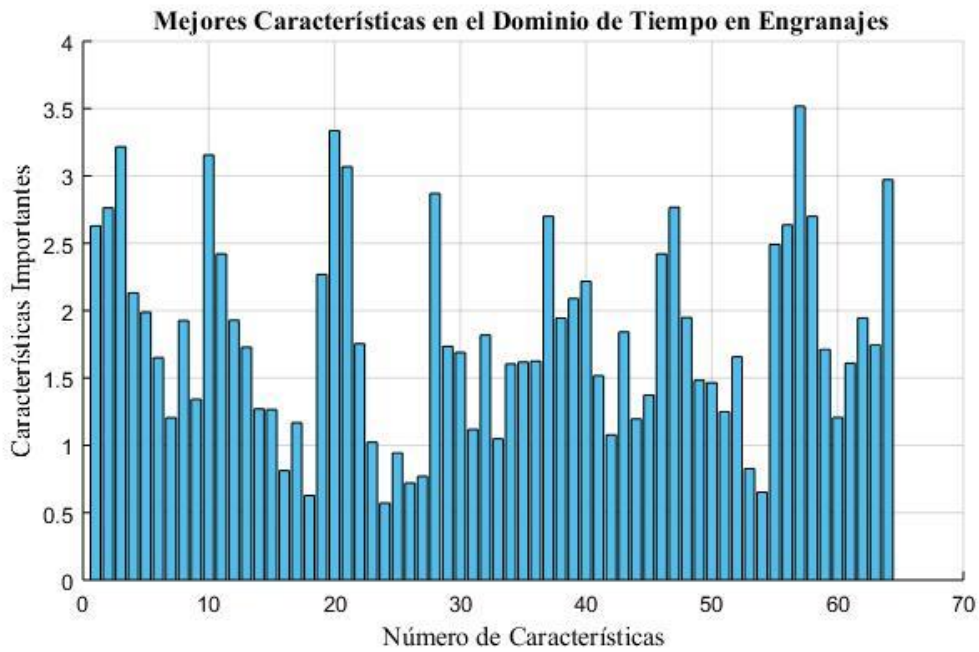


Figura 21. Mejores Características en el Dominio de Tiempo en Engranajes.

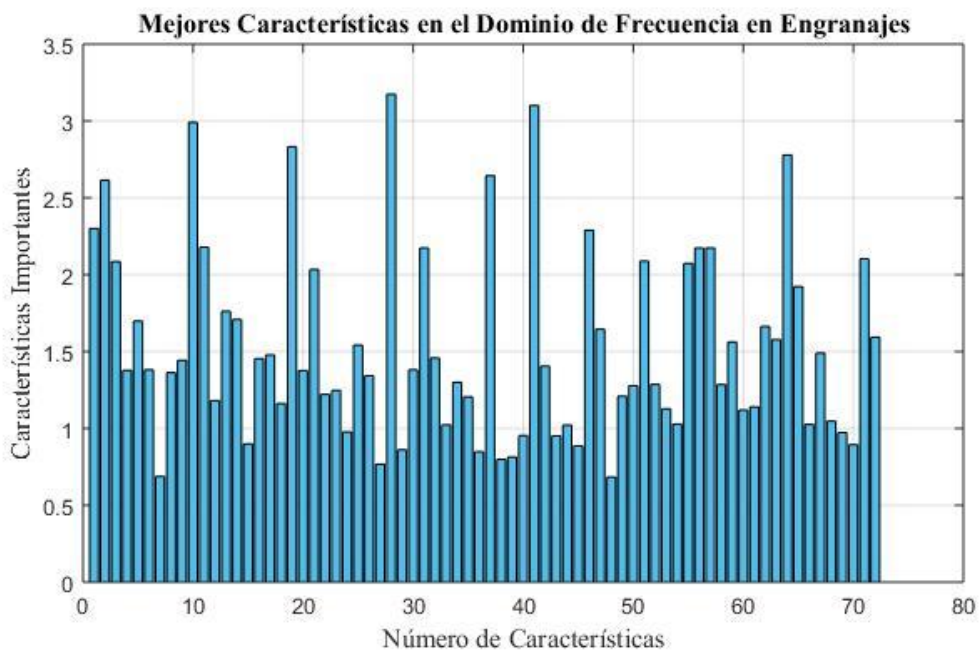


Figura 22. Mejores Características en el Dominio de Frecuencia en Engranajes.

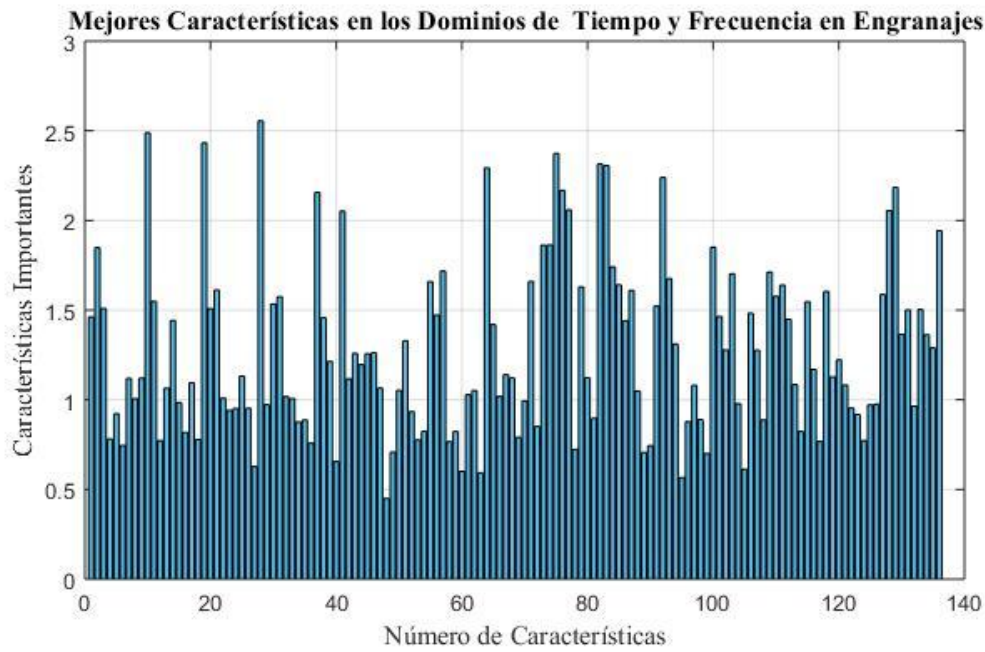


Figura 23. Mejores Características en los Dominios de Tiempo en conjunto con la Frecuencia en Engranajes.

De acuerdo a las Figuras 18, 19 y 20; se presentarán las características más importantes en los Rodamientos con relación a cada dominio, todas estas fueron seleccionadas por encima del 80% de la característica más alta que se muestra en la siguiente tabla:

RODAMIENTOS								
TIEMPO			FRECUENCIA			TIEMPO Y FRECUENCIA		
CARACTERÍSTICAS		Nivel de Importancia	CARACTERÍSTICAS		Nivel de Importancia	CARACTERÍSTICAS		Nivel de Importancia
N°	Nombre		N°	Nombre		N°	Nombre	
13	Valor Medio	4.1228	45	Variación de Frecuencia Total	3.6798	34	Frecuencia Media	2.6111
2	RMS	3.6865	2	Total Potencia	3.4967	45	Variación de Frecuencia Central	2.6059
4	RMS	3.514	3	Total Potencia	3.3573	91	Energía en Frecuencia	2.4019
3	RMS	3.3175	1	Total Potencia	3.3153	92	Energía en Frecuencia	2.2898
			4	Total Potencia	2.993	101	Desviación Estándar	2.2966
						23	SM3	2.2176
						12	SM1	2.1053

Tabla 19. Características más relevantes en el dominio de tiempo, dominio de frecuencia y dominio de tiempo en conjunto con la frecuencia en Rodamientos.

Con respecto a los Engranajes en las Figuras 21, 22 y 23 se determinan así mismo como características más relevantes las seleccionadas por encima del 80% de la característica más alta en sus diferentes dominios mostrando como resultados los siguientes:

ENGRANAJES								
TIEMPO			FRECUENCIA			TIEMPO Y FRECUENCIA		
CARACTERÍSTICAS		Nivel de Importancia	CARACTERÍSTICAS		Nivel de Importancia	CARACTERÍSTICAS		Nivel de Importancia
N°	Nombre		N°	Nombre		N°	Nombre	
57	Desviación Estándar	3.5174	28	SM3	3.1741	28	SM3	2.5551
20	Valor Medio	3.3352	41	Frecuencia Media	3.1004	10	SM1	2.4882
3	RMS	3.2163	10	SM1	2.99	19	SM2	2.4328
10	Curtosis	3.1559	19	SM2	2.833	75	Curtosis	2.3739
21	Valor Medio	3.069	64	Variación de Frecuencia Central	2.779	82	Valor Medio	2.3152
64	Desviación Estándar	2.972	37	Curtosis Espectral	2.644	83	Valor Medio	2.3055
28	Punto Máximo	2.87	2	Total Potencia	2.616	64	Variación de Frecuencia Central	2.2936
						92	Punto Máximo	2.2391
						129	Desviación Estándar	2.1842
						76	Curtosis	2.1673
						37	Curtosis Espectral	2.1567
						77	Curtosis	2.0587
						128	Energía en Frecuencia	2.0539
						41	Frecuencia Media	2.0526

Tabla 20. Características más relevantes en el dominio de tiempo, dominio de frecuencia y dominio de tiempo en conjunto con la frecuencia en Engranajes.

CONCLUSIONES Y RECOMENDACIONES

En base al desarrollo del proyecto y de acuerdo a los resultados obtenidos, el diagnóstico de fallos en maquinaria rotativa con la aplicación de EMD en señales de vibración y el entrenamiento de Random Forest da como resultado que EMD para las señales de rodamientos es eficiente pero para las señales de engranajes tiene una carencia muy significativa, debido a que las señales de vibración de engranajes son difíciles de reconocer en el espectro porque están distribuidos en bandas de frecuencia amplias y pueden superponerse con algunos componentes de fondo.

Para mejorar los resultados obtenidos en los engranajes es necesario la utilización de la EMD en conjunto con la Transformada de Hilbert, ya que con esto se podrá calcular la frecuencia instantánea de la Transformada de Hilbert y así utilizarla como una característica para ser entrenada en el clasificador de Random Forest.

El algoritmo de EMD es un método iterativo, en consiguiente calcular la interpolación entre sus puntos y el proceso del tamizado genera mucho costo computacional, por lo cual aplicando los conceptos de la computación paralela y distribuida se logra generar tiempos de ejecución aceptables.

La extracción de características para este proyecto es de suma importancia debido a que con estas se formó un conjunto de datos que sirvió para entrenar el clasificador de Random Forest mostrando los resultados de error obtenidos en la Tabla 7, y por otro lado medir su rendimiento mediante las métricas utilizadas Random Forest es un tipo de inteligencia artificial que mejor se adapta para determinar los fallos en rodamientos y engranajes.

Una de las mejores clases obtenidas en la clasificación al momento de utilizar EMD es la de pista externa, dando a entender que las características mencionadas son las más óptimas para clasificar este tipo de clase.

Las mejores características seleccionadas mediante Random Forest para el diagnóstico de fallos para Rodamientos tanto en tiempo como en frecuencia y en conjunto las dos se muestran en la Tabla 19 y en Engranajes se encuentra en la Tabla 20, siendo recomendable que éstas sean utilizadas para verificar si el error de clasificación que tiene Random Forest disminuye.

REFERENCIAS

- [1] F. Gómez, Tecnología del mantenimiento industrial, 1st ed. Murcia: Universidad de Murcia, 1998.
- [2] M. Cabanas, M. Melero, G. Orcajo, J. Cano, and J. Sariego, Técnicas para el mantenimiento y diagnóstico de máquinas eléctricas rotativas. Barcelona: Marcombo, 1998.
- [3] V. Sánchez and D. Arpi, “Diagnóstico De Fallas En Cajas De Engranajes Rectos Con Redes Neuronales Artificialesperceptrón Multicapa Y Análisis Discriminante Lineal,” vol. 1. pp. 1–9, 2012.
- [4] S. Rafael, “Análisis De Filtros Para La Detección De Fallos En Rodamientos Industriales A Partir De Su Señal De Vibración,” Universidad Carlos III De Madrid, 2010.
- [5] V. Sanchez, “Diagnóstico de fallos en engranajes basado en el análisis de señales de vibración empleando una red neuronal perceptrón multicapa con retropropagación del error,” Universidad Nacional De Educación A Distancia, 2012.
- [6] G. Zurita, V. Sánchez, and D. Cabrera, “A Review of Vibration Machine Diagnostics by Using Artificial Intelligence Methods,” UPB - Investigación Desarro., vol. 1, no. 1, pp. 102–114, Sep. 2016.
- [7] P. Jayaswal, A. K. Wadhwani, and K. B. Mulchandani, “Machine fault signature analysis,” Int. J. Rotating Mach., vol. 2008, pp. 1–10, Mar. 2008.
- [8] D. Kbaier Ben Ismail, P. Lazure, and I. Puillat, “Application of Hilbert-Huang decomposition to temperature and currents data in the Réunion island,” Ocean. 2016 MTS/IEEE Monterey, OCE 2016, vol. 1, pp. 1–9, 2016.
- [9] Norden E. Huang et al., “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” vol. 454, pp. 903–995, 1998.
- [10] Q. Yao, J. Wang, L. Yang, H. Su, and G. Zhang, “A fault diagnosis method of engine rotor based on Random Forests,” 2016 IEEE International Conference on Prognostics and Health Management (ICPHM). IEEE, Beijing, pp. 1–4, Jun-2016.
- [11] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, “Fault diagnosis in spur gears based on genetic algorithm and random forest,” Mech. Syst. Signal Process., vol. 70–71, pp. 87–103, Mar. 2016.

- [12] F. Pacheco et al., “A statistical comparison of neuroclassifiers and feature selection methods for gearbox fault diagnosis under realistic conditions,” *Neurocomputing*, vol. 194, pp. 192–206, Jun. 2016.
- [13] J. Guillen and M. Hernandez, *Informática industrial - Joaquín Guillén Rodríguez, Mauricio Hernández Hernández - Google Libros*, 1st ed. Monterrey: Editorial Digital Tecnológico de Monterrey, 2013.
- [14] Diego Aranda, *ELECTRÓNICA: Técnicas digitales y microcontroladores*, 1st ed. Buenos Aires: Fox Andina y Dálaga, 2014.
- [15] F. M. Díez, *Higiene industrial: manual para la formación del especialista*. España: Lex Nova, 2009.
- [16] N. Quizhpe and L. Brito, “Aplicaciones de un acelerómetro para la medición de inclinaciones horizontales y verticales utilizando microcontroladores avanzados y comunicación serial datalogger e interfaz gráfica. Fuente de energía 4 pilas recargables,” Escuela Superior Politecnica Del Litoral, 2010.
- [17] P. Ochoa, M. Villegas, J. Pons, M. Bemgochea, and J. Fernández, “Piezocomposites metal-cerámica como elementos activos en acelerómetros,” *Boletín la Soc. Española Cerámica y Vidr.*, vol. 36, no. 5, pp. 525–528, 2002.
- [18] J. Otegui and E. Rubertis, *Cañerías y recipientes de presión*, 1st ed. Balcarce: EUDEM.
- [19] Physical Acoustics Corporation, *WD Sensor Product Data Sheet*. MISTRAS, 2011, p. 1.
- [20] Physical Acoustics Corporation, “D9241A Sensor - 30 kHz Frequency Acoustic Emission Sensor.” MISTRAS, p. 1, 2011.
- [21] J. Pintor, “Normativa sobre Vibraciones.” Navarra.
- [22] ASTM International, *Método estándar para la calibración primaria de sensores de emisión acústica*. IEEE, 2002, p. 11.
- [23] R. Oliva, “Eficacia de los agrotexiles como método de protección de cultivos,” Universidad de Almería, 2017.
- [24] H. López García, *Control por computador: diseño y realización práctica*, 1st ed. Oviedo: Universidad Oviedo, Servicio de Publicaciones, 1993.
- [25] N. Instruments, “NI cDAQ™ -9188,” National Instruments, Texas, 2017.
- [26] N. Instruments, “Ni 9234.” National Instruments, Texas, pp. 1–14, 2015.
- [27] G. Mosquera, M. Piedra, and R. Armas, *Las vibraciones mecánicas y su aplicación al mantenimiento predictivo*, vol. 1. Caracas: Graficas Leon, 2000.

- [28] R. B. Gutiérrez and G. C. Pere, 55 respuestas a dudas típicas de estadística, 2 da. Madrid: Diaz de Santos, 2010.
- [29] MathWorks, “Kurtosis - MATLAB,” 2006, 2017. [Online]. Available: https://www.mathworks.com/help/stats/kurtosis.html?searchHighlight=kurtosis&s_tid=doc_srchtile. [Accessed: 27-Nov-2017].
- [30] R. Cai, F. Xie, W. Chen, and Z. Hao, “An Efficient Kurtosis-based Causal Discovery Method for Linear Non-Gaussian Acyclic Data,” IEEE, Jun. 2017.
- [31] R. Cáceres, Estadística aplicada a las ciencias de la salud, 1st ed. España: Díaz de Santos, 2007.
- [32] MathWorks, “Average or mean value of array - MATLAB,” 2006, 2017. [Online]. Available: https://www.mathworks.com/help/matlab/ref/mean.html?searchHighlight=mean&s_tid=doc_srchtile. [Accessed: 27-Nov-2017].
- [33] MathWorks, “Skewness - MATLAB,” 2006, 2017. [Online]. Available: https://www.mathworks.com/help/stats/skewness.html?searchHighlight=SKEWNESS&s_tid=doc_srchtile. [Accessed: 27-Nov-2017].
- [34] E. Herrera Pérez, Comunicaciones I: señales, modulación y transmisión, 1st ed. Mexico: Limusa, 2012.
- [35] G. Chujit, A. Phinyomark, H. Hu, P. Phukpattaranont, and C. Limsakul, “Evaluation of EMG Feature Extraction for Classification of Exercises in Preventing Falls in the Elderly,” Thailand, 2012.
- [36] Sijiang Du and M. Vuskovic, “Temporal vs. spectral approach to feature extraction from prehensile EMG signals,” IEEE, San Diego, 2004.
- [37] A. Gilat, Matlab: una introducción con ejemplos prácticos. Barcelona: Reverté, 2006.
- [38] MathWorks, “Parallel Computing Toolbox™: User’s Guide R2017a,” MATLAB Manual. pp. 1–729, 2017.
- [39] R. Buyya, C. Vecchiola, and S. T. Selvi, Mastering cloud computing: Foundations and Applications Programming. Waltham: Morgan Kaufmann, 2013.
- [40] T. Mathieu, G. Hernandez, and A. Gupta, “Parallel Computing with MATLAB,” MathWorks, 2011. [Online]. Available: https://www.mir.wustl.edu/Portals/0/Documents/Uploads/CHPC/PCT_Masterclass.pdf. [Accessed: 28-Nov-2017].

- [41] MathWorks, “Execute for-loop iterations in parallel on workers in parallel pool - MATLAB parfor,” 2008, 2017. [Online]. Available: <https://www.mathworks.com/help/distcomp/parfor.html>. [Accessed: 28-Nov-2017].
- [42] NVIDIA, “Procesamiento Paralelo Cuda.” [Online]. Available: <http://www.nvidia.es/object/cuda-parallel-computing-es.html>. [Accessed: 28-Nov-2017].
- [43] R. I. Ramírez-Castro and L. A. Montejo, “Transformada de Hilbert, Descomposición Modal Empírica y sus Aplicaciones en el Análisis de Vibraciones Libres,” Jul. 2011.
- [44] G. Rilling, P. Flandrin, and P. Es, “On empirical mode decomposition and its algorithms,” *IEEE-EURASIP Work. nonlinear signal image Process.*, vol. 3, pp. 8–11, 2003.
- [45] J. Bian, P. Wang, Q. Mei, and M. Lei, “Fault detection of rolling bearings through vibration analysis via the hybrid CEEMD-EMD approach,” *Proc. 2014 Progn. Syst. Heal. Manag. Conf. PHM 2014*, no. 1, pp. 245–250, 2014.
- [46] X. Weihua, L. Junfeng, and P. Haipeng, “Analysis of transformer winding vibration based on modified empirical mode decomposition,” *Intell. Control Autom. (WCICA)*, 2010 8th World Congr., pp. 5906–5909, 2010.
- [47] W. Liao, P. Han, and X. Liu, “Fault Diagnosis for Engine Based on EMD and Wavelet Packet BP Neural Network,” *2009 Third Int. Symp. Intell. Inf. Technol. Appl.*, vol. 2, no. 1, pp. 672–676, 2009.
- [48] S. Luo, L. Zhang, Y. Zhao, Y. Xiao, and S. Fang, “Application of EEMD in fault diagnosis of diesel engine,” *IEEE*, Tianjin, Jul. 2010.
- [49] Z. Xuan, S. Xie, and Q. Sun, “The Empirical Mode Decomposition Process of Non-stationary Signals,” *2010 Int. Conf. Meas. Technol. Mechatronics Autom.*, vol. 2, no. 3, pp. 866–869, 2010.
- [50] V. Mathivet, *Inteligencia Artificial para desarrolladores. Conceptos e implementación en C#, Española*. España: Ediciones ENI, 2015.
- [51] R. G. Keys and D. J. Foster, “1. A Data Set for Evaluating and Comparing Seismic Inversion Methods,” in *Comparison of Seismic Inversion Methods on a Single Real Data Set*, USA: Society of Exploration Geophysicists, 1998, pp. 1–12.
- [52] C. Snijders, U. Matzat, and U.-D. Reips, “Big Data: Big Gaps of Knowledge in

- the Field of Internet Science,” *Int. J. Internet Sci.*, vol. 7, no. 1, pp. 1–5, 2012.
- [53] Y. Baştanlar and M. Özuysal, “Introduction to Machine Learning,” in *miRNomics MicroRNA Biology and Computational Analysis*, vol. 1107, New York: Human Press, 2014, pp. 105–128.
- [54] Y. Liu, T. Zhao, W. Ju, and S. Shi, “Materials Discovery and Design using Machine Learning,” *J. Mater.*, vol. 3, no. 3, pp. 159–177, Sep. 2017.
- [55] D. H. Wolpert and W. G. Macready, “No free lunch theorems for optimization,” *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997.
- [56] S. Shai Shalev and D. Shai Ben, *Understanding Machine Learning: From Theory to Algorithms*. New York: Cambridge University Press, 2014.
- [57] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 53, no. 9. USA: Springer, 2006.
- [58] S. Russell and Norving Peter, *Artificial Intelligence A Modern Approach*, 3rd ed. New Jersey: Pearson Education, 2010.
- [59] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [60] M. Kayri, I. Kayri, and M. Tunay, “The performance comparison of Multiple Linear Regression, Random Forest and Artificial Neural Network by using photovoltaic and atmospheric data,” in *14th International Conference on Engineering of Modern Electric Systems (EMES)*, 2017, pp. 1–4.
- [61] MathWorks, “Create bag of decision trees - MATLAB TreeBagger,” 2009, 2017. [Online]. Available: <https://www.mathworks.com/help/stats/treebagger.html>. [Accessed: 29-Nov-2017].
- [62] MathWorks, “Bag of decision trees - MATLAB TreeBagger class,” 2017. [Online]. Available: <https://www.mathworks.com/help/stats/treebagger-class.html>. [Accessed: 29-Nov-2017].
- [63] A. Villegas, E. Lugo, J. Pacheco, and H. Villegas, “Implementación de una Interfaz Cerebro Computador Basada en la Interpretación del Electroencefalograma para el Control a Distancia de Dispositivos Electrónicos,” in *IV Latin American Congress on Biomedical Engineering 2007, Bioengineering solutions for Latin America health*, September 24th-28th, 2007, Margarita Island, Venezuela, Valencia: Springer, 2008, pp. 220–224.
- [64] A. F. Quiceno, E. Delgado, C. Acosta-Muñoz, and G. Castellanos, “Caracterización de Espectrogramas Usando Análisis de Componentes Principales y Medidas de Energía para Detección de Soplos Cardíacos,” in *IV*

- Latin American Congress on Biomedical Engineering 2007, Bioengineering Solutions for Latin America Health, September 24th-28th, 2007, Margarita Island, Venezuela, Caldas-Colombia: Springer, Berlin, Heidelberg, 2007, pp. 162–166.
- [65] C. E. Metz, “Basic Principles of ROC Analysis,” *Semin. Nucl. Med.*, vol. 8, 1978.
- [66] R. Caruana and A. Niculescu-Mizil, “Data Mining in Metric Space: An Empirical Analysis of Supervised Learning Performance Criteria,” pp. 1–10, 2004.
- [67] RedUsers, *Electricidad - Fundamentos y seguridad: Herramientas para el desarrollo profesional*. Buenos Aires, 2015.
- [68] C. Ortiz and I. Pérez, “Adquisición de señales vibracionales y emisiones acústicas combinando fallos en maquinaria rotativa y elaboración de guías de práctica sobre detección de fallos en engranajes por medio de emisiones acústicas.,” Universidad Politécnica Salesiana, 2016.