

**PRONÓSTICO DE VIDA ÚTIL REMANENTE EN
RODAMIENTOS CON BASE EN LA ESTIMACIÓN DE LA
PROBABILIDAD DE LA DEGRADACIÓN**

**Remaining useful life prognosis of bearings based on the
estimation of probability of degradation**

Mauricio Holguín Londoño

Magister en Ingeniería Eléctrica (Área de Automática)
Ingeniero Electricista

Tesis presentada con el fin de completar el
cumplimiento de requisitos para optar al título de
Doctor en Ingeniería (Área de Automática)

Pereira, Agosto de 2018
UNIVERSIDAD TECNOLÓGICA DE PEREIRA
Doctorado en Ingeniería
Grupo de Investigación en Automática



DEDICATORIA

A Eliana y Sofía, por su amor, apoyo incondicional y sacrificios realizados.

A mis padres, Luciola y Germán, y a mi hermano Germán Andrés por su amor y acompañamiento.

AGRADECIMIENTOS

Quiero expresar mis más sinceros agradecimientos al Profesor Ph.D. Álvaro Ángel Orozco Gutiérrez por su supervisión, acompañamiento y guía durante todo este proceso de investigación, que me permite ahora concluir esta valiosa experiencia. También deseo dar mis agradecimientos al Profesor Ph.D. Andrés Marino Álvarez Meza de la Universidad Tecnológica de Pereira por todo su apoyo y al Ph.D. Henry Medeiros de Marquette University por permitirme la invaluable experiencia de pasantía en su grupo de investigación.

De forma colectiva e individual, le deseo agradecer a los profesores, amigos y colegas del grupo de investigación en Automática por estar siempre dispuestos a colaborar y dar sus soportes.

Especial agradecimiento merece la Universidad Tecnológica de Pereira por haberme dado la oportunidad de realizar esta etapa de formación y brindarme soporte financiero. Igualmente, manifiesto mi gratitud con el Programa de Doctorado en Ingeniería, la Maestría en Ingeniería Eléctrica y el Grupo de Investigación en Automática de la Universidad Tecnológica de Pereira que me brindaron apoyos financieros con el fin de asistir a ponencias, hacer publicaciones y realizar pasantía.

También agradezco al Laboratorio de Vibraciones Mecánicas de la Universidad Tecnológica de Pereira por facilitar una de las bases de datos empleada para el presente estudio.

Contenido

I INTRODUCCIÓN	1
1. INFORMACIÓN GENERAL	2
2. PLANTEAMIENTO DEL PROBLEMA	2
3. JUSTIFICACIÓN	6
3.1. Impacto	6
3.2. Pertinencia	8
3.3. Viabilidad	9
4. OBJETIVOS	10
4.1. Objetivo general	10
4.2. Objetivos específicos	10
5. CONTRIBUCIÓN Y ORIGINALIDAD	10
II ESTADO DEL ARTE Y MARCO TEÓRICO	14
6. MARCO CONCEPTUAL	15
7. ESTADO DEL ARTE	17
8. MARCO TEÓRICO	20
8.1. Rodamientos	20
8.2. Representación de datos de forma de onda	23
8.3. Representación en el dominio del tiempo - TD	23
8.4. Representación en el dominio de la frecuencia - FD	25
8.5. Representación en el dominio tiempo-frecuencia - TFD	26
8.6. Modelo Oculto de Markov (HMM) con observación discreta	30
8.7. Modelo Oculto de Markov (HMM) con observación continua	32
8.8. Modelo Oculto de Markov Jerárquico (HHMM)	33

III	MATERIALES Y MÉTODOS	36
9.	PRONÓSTICO DE VIDA ÚTIL REMANENTE (RUL)	37
10.	MARCO METODOLÓGICO	39
10.1.	RETOS	39
10.2.	MODELO DE PRONÓSTICO	41
10.3.	BASE DE DATOS	42
IV	RESULTADOS, ANÁLISIS Y DISCUSIÓN	45
11.	ANÁLISIS DE DEPENDENCIA, VARIABILIDAD Y CONGLOMERADOS	46
11.1.	Análisis multivariado, de distancia y dependencia	47
11.2.	Análisis de componentes principales	49
11.3.	Biplots	50
11.4.	Algoritmos para conglomerados	51
11.5.	Marco experimental para dependencia, variabilidad y conglomerados	54
11.6.	Resultados para dependencia, variabilidad y conglomerados	54
11.7.	Discusión de resultados para dependencia, variabilidad y conglomerados	71
12.	SELECCIÓN Y FUSIÓN DE DOMINIOS	73
12.1.	Marco experimental para selección y fusión de dominios	73
12.2.	Resultados para selección y fusión de dominios	76
12.3.	Discusión de resultados para selección y fusión de dominios	80
13.	DIAGNÓSTICO DE FALLOS MEDIANTE SEÑALES VIBRO-ACÚSTICAS	84
13.1.	Métodos de banco de filtros	84
13.2.	Caracterización por medidas de similaridad	85
13.3.	Marco experimental para diagnóstico de fallos mediante señales vibro-acústicas	87
13.4.	Resultados para diagnóstico de fallos mediante señales vibro-acústicas	88
13.5.	Discusión de resultados para diagnóstico de fallos mediante señales vibro-acústicas	93

14. METODOLOGÍA DE SINTONIZACIÓN DE MODELO HMM CON OBSERVACIÓN DISCRETA	95
14.1. Marco experimental para metodología de sintonización de modelo HMM con observación discreta	95
14.2. Resultados para metodología de sintonización de modelo HMM con observación discreta	97
14.3. Discusión de resultados para metodología de sintonización de modelo HMM con observación discreta	103
15. METODOLOGÍA DE DIAGNÓSTICO CON CARACTERIZACIÓN MULTI-DOMINIO Y CLASIFICADOR JERÁRQUICO	105
15.1. Marco experimental para metodología de diagnóstico con caracterización multi-dominio y clasificador jerárquico	105
15.2. Resultados para metodología de diagnóstico con caracterización multi-dominio y clasificador jerárquico	108
15.3. Discusión para metodología de diagnóstico con caracterización multi-dominio y clasificador jerárquico	116
16. METODOLOGÍA DE DIAGNÓSTICO MEDIANTE SELECCIÓN ESTOCÁSTICA DE CARACTERÍSTICAS	118
16.1. Estimación multi-dominio de características	118
16.2. Relevancia de características por Relief-F	119
16.3. Marco experimental para metodología de diagnóstico mediante selección estocástica de características	120
16.4. Resultados para metodología de diagnóstico mediante selección estocástica de características	123
16.5. Discusión de resultados para metodología de diagnóstico mediante selección estocástica de características	129
17. CONCLUSIONES Y TRABAJOS FUTUROS	131
17.1. CONCLUSIONES	131
17.2. TRABAJOS FUTUROS	136

Índice de figuras

8.1. Principales partes de un rodamiento, adaptado de [86]	21
8.2. Proceso MFCC	28
10.1. Modelo de pronóstico	43
11.1. Gráfico de dispersión por pares	57
11.2. Caras de Chernoff para el dominio del tiempo	58
11.3. Caras de Chernoff para el dominio de la frecuencia	59
11.4. Caras de Chernoff para el dominio MFCC	59
11.5. Gráfico de valores propios y varianza acumulativa	62
11.6. Gráficos de asociación de componentes principales y variables	64
11.7. Proyección de observaciones sobre las componentes	65
11.8. Biplot sobre componentes 1 y 2 empleando datos de <i>Sr</i>	66
11.9. Gráfico de distorsión	67
11.10. Conglomerados con k-medias para 4, 8, 18 y 24 clusters	68
11.11. Conglomerados con k-armónicas medias de 4, 8, 18 y 24 clusters	69
11.12. Dendrograma para agrupamiento Ward	70
12.1. Relevancia individual de dominios de representación	77
12.2. Análisis Leave-One-Out por dominio	81
13.1. Representación tiempo-frecuencia para señales de vibración y acústicas	89
13.2. Descomposición WPT para señales de vibración y acústicas	90
13.3. Error de clasificación según FBM, similaridad, tipo de señal y SNR	91
14.1. Curvas ROC para HMM entrenados con BD1, con severidad	99
14.2. Curvas ROC para HMM entrenados con BD1, sin severidad	99
14.3. Curvas ROC para HMM entrenados con BD2 UTP	100
14.4. Área bajo curvas ROC según base de datos	100
15.1. Enfoque metodológico HHMM	108
15.2. Modelo HHMM	108
15.3. Análisis intra-dominio de características	112
15.4. Análisis forward para relevancia de características intra-dominio	113
15.5. Análisis backward para relevancia de características intra-dominio	114
16.1. Marco metodológico para diagnóstico y pronóstico con base en SFS	121
16.2. Señales en el dominio del tiempo y de la frecuencia, según tipo de fallo	124
16.3. Selección de características con HMM	124
16.4. Características relevantes que logran la mejor eficiencia	127

Índice de tablas

1.	Características en el dominio del tiempo	24
2.	Características en el dominio de la frecuencia	25
3.	Análisis univariado de datos	56
4.	Distancia de Mahalanobis	60
5.	Valores propios y valores propios corregidos	61
6.	Vectores propios y vectores propios corregidos	63
7.	Características en TD para fusión y selección	74
8.	Características en FD para fusión y selección	74
9.	Relevancia de características intra-dominio	78
10.	Relevancia por fusión de dominios, con degradación	79
11.	Errores de clasificación según FBM y similaridad, con SNR = -20 dB	92
12.	Eficiencias modelos HMM, con BD1 sin severidad	101
13.	Eficiencias modelos HMM, con BD1 y severidad	102
14.	Eficiencias modelos HMM, con BD2 UTP	102
15.	Análisis inter-dominio, o por tipos de fallos	109
16.	Análisis inter-dominio, por severidad de fallos	110
17.	Evaluación y comparación de modelo HHMM	116
18.	Resultados experimentales para $C = 4$	125
19.	Matriz de confusión $C = 4$, Relief-F y 9 características [en %]	125
20.	Resultados experimentales para $C = 10$	125
21.	Matriz de confusión $C = 10$, Relief-F y 7 características [en %]	126
22.	Estudio comparativo de resultados contra el estado del arte	128

NOMENCLATURA

ABREVIACIONES

AI: Inteligencia Artificial, del inglés *Artificial Intelligence*

ALS: Mínimos Cuadrados Alternativos, del inglés *Alternating Least Square*

ANC: Cancelación Adaptativa del Ruido, del inglés *Adaptive Noise Cancellation*

ANFC: Clasificador Adaptativo Neuro Difuso, del inglés *Adaptive Neuro Fuzzy Classifier*

ANFIS: Sistema Adaptativo de Inferencia Difusa Neuronal, del inglés *Adaptive Neural Fuzzy Inference System*

ANN: Redes Neuronales Artificiales, del inglés *Artificial Neural Networks*

AP: Propagación por Afinidad, del inglés *Affinity Propagation*

AR: Modelo Auto Regresivo, del inglés *Autoregressive Model*

ARMA: Auto Regresivo de Media Móvil, del inglés *Autoregressive Moving Average*

BD: Deconvolución Ciega, del inglés *Blind Deconvolution*

CAE: Auto-Encoder Contractivo, del inglés *Contractive Auto-Encoder*

CBM: Mantenimiento con Base en la Condición, del inglés *Condition Based Maintenance*

CeF: Frecuencia Central, del inglés *Central Frequency*

CF: Factor de Cresta, del inglés *Crest Factor*

CL: Levantamiento Cepstrum, del inglés *Cepstrum Liftering*

CIF: Factor de Holgura, del inglés *Clearance Factor*

CSDI: Índice de Densidad Espectral Cumulativa, del inglés *Cumulative Spectral Density Index*

CWT: Transformada Wavelet Continua, del inglés *Continuous Wavelet Transform*

DAE: Auto-Encoder Antiruido, del inglés *Denoising Auto-Encoder*

DBN: Redes Bayesianas Dinámicas, del inglés *Dynamic Bayesian Networks*

DCT: Transformada Discreta del Coseno, del inglés *Discrete Cosine Transform*

DET: Técnica de Evaluación de Distancia, del inglés *Distance Evaluation Technique*

DFT: Transformada Discreta de Fourier, del inglés *Discrete Fourier Transform*

DW: Distancia-Peso, del inglés *Distance-Weight*

DWT: Transformada Wavelet Discreta, del inglés *Discrete Wavelet Transform*

EDM: Mecanizado por Electro-Descarga, el inglés *Electro-Discharge Machining*

EEMD: Descomposición de Modo Empírico en Ensamble, del inglés *Ensemble Empirical Mode Decomposition*

EM: Esperanza-Maximización, del inglés *Expectation Maximization*

EMD: Descomposición de Modo Empírico, del inglés *Empirical Mode Decomposition*

FAA: Administración Federal de Aviación, del inglés *Federal Aviation Administration*

FBM: Métodos de Banco de Filtros, del inglés *Filter Bank Methods*

FD: Dominio de la Frecuencia, del inglés *Frequency Domain*

FFB: Bancos de Filtros de Fourier, del inglés *Fourier Filter Banks*

FFT: Transformada Rápida de Fourier, del inglés *Fast Fourier Transform*

FL: Lógica Difusa, del inglés *Fuzzy Logic*

FRF: Función de Respuesta en Frecuencia, del inglés *Frequency Response Function*

GA: Algoritmos Genéticos, del inglés *Genetic Algorithms*

GCMPE: Entropía de Permutación Multiescala Compositiva Generalizada, del inglés *Generalized Compositive Multiscale Permutation Entropy*

HLB: Valor Inferior de Histograma, del inglés *Histogram Lower Bin*

HHMM: Modelos Ocultos de Markov Jerárquicos, del inglés *Hierarchical Hidden Markov Models*

HMM: Modelos Ocultos de Markov, del inglés *Hidden Markov Models*

HUB: Valor Superior de Histograma, del inglés *Histogram Upper Bin*

ICD: Distancia Inter-clúster, del inglés *Inter-Cluster Distance*

IF: Factor de Impulso, del inglés *Impulse Factor*

IMF: Funciones de Modo Intrínseco, del inglés *Intrinsic Mode Functions*

IQR: Rango Intercuartílico, del inglés *Interquartile Range*

IQRF: Rango Intercuartílico Frecuencial, del inglés *Interquartile Range Frequency*

ISO: Organización Internacional de Estandarización, del inglés *International Standard Organization*

KNN: Vecinos K-Cercanos, del inglés *K-Nearest Neighbors*

KPCA: Análisis de Componentes Principales por Kernel, del inglés *Kernel Principal Component Analysis*

LP: Predicción Lineal, del inglés *Linear Prediction*

LPC: Coeficientes de Predicción Linear, del inglés *Linear Prediction Coefficients*

LPCC: Coeficientes Cepstrales de Predicción Lineal, del inglés *Linear Prediction Cepstral Coefficients*

LPP: Proyección Preservando Localidad, del inglés *Locality Preserving Projection*

LS: Puntuación Laplaciana, del inglés *Laplacian Score*

MF: Frecuencia Media, del inglés *Mean Frequency*

MFCC: Coeficientes Cepstrales en Escala Mel, del inglés *Mel-Frequency Cepstral Coefficients*

MIBS: Índice de Mérito Basado en Asimetría, del inglés *Merit Index Based on Skewness*

ML: Máxima Verosimilitud, del inglés *Maximum-Likelihood*

MPE: Entropía de Permutación Multi-escala, del inglés *Multi-scale Permutation Entropy*

NMF: Factorización de Matriz de No-negación, del inglés *Non-negation Matrix Factorization*

PCA: Análisis de Componentes Principales, del inglés *Principal Component Analysis*

PCC: Coeficiente de Correlación de Pearson, del inglés *Pearson Correlation Coefficient*

PCT: Técnica Cepstrum de Potencia, del inglés *Power Cepstrum Technique*

PDF: Función de Densidad de Probabilidad, del inglés *Probability Density Function*

PE: Entropía de Permutación, del inglés *Permutation Entropy*

PNN: Red Neuronal Probabilística, del inglés *Probabilistic Neural Network*

PSO: Optimización por Enjambre de Partículas, del inglés *Particle Swarm Optimization*

QC: Clasificador Cuadrático, del inglés *Quadratic Classifier*

RMS: Raíz Media Cuadrática, del inglés *Root Mean Square*

RMSF: Raíz Media Cuadrática Frecuencial, del inglés *Root Mean Square Frequency*

ROC: Característica Operativa del Receptor, del inglés *Receiver Operating Characteristic*

RUL: Vida Útil Remanente, del inglés *Remaining Useful Life*

SC: Clasificador Softmax, del inglés *Softmax Classifier*

SCD: Control de Esparcidad y Decorrelación, del inglés *Sparsity Control and Decorrelation*

ScPCA: Gráfico de Dispersión Empleando Análisis de Componentes Principales, del inglés *Scatter plot using Principal Component Analysis*

SDF: Desviación Estándar Frecuencial, del inglés *Standard Deviation Frequency*

SF: Factor de Forma, del inglés *Shape Factor*

SK: Kurtosis Espectral, del inglés *Spectral Kurtosis*

SkF: Asimetría Frecuencial, del inglés *Skewness Frequency*

SM: Matrices de Dispersión, del inglés *Scatter Matrices*

SNR: Relación Señal-Ruido, del inglés *Signal to Noise Ratio*

SOCF: SDF para Valores Frecuenciales sobre Frecuencia Central, del inglés *SDF of Frequency Values Over Center Frequency*

SP: Parámetros Estadísticos, del inglés *Statistical Parameters*

SRCE: Estimación Rápida de Centroide por Enjambre, del inglés *Swarm Rapid Centroid Estimation*

SSA: Análisis Espectral Singular, del inglés *Singular Spectrum Analysis*

STFT: Transformada de Fourier de Tiempo Corto, del inglés *Short-time Fourier Transform*

SV: Valor Singular, del inglés *Singular Value*

SVM: Máquinas de Soporte Vectorial, del inglés *Support Vector Machines*

SVR: Regresión por Vector de Soporte, del inglés *Support Vector Regression*

SW: Auto-Censurado, del inglés *Self-Weighth*

TD: Dominio del Tiempo, del inglés *Time Domain*

TFD: Dominio del Tiempo-Frecuencia, del inglés *Time-Frequency Domain*

TNR, Fracción de Verdaderos Negativos, del inglés *True Negative Rate*

TPR: Fracción de Verdaderos Positivos, del inglés *True Positive Rate*

VF: Varianza Frecuencial, del inglés *Variance Frequency*

VRA: Análisis de Relevancia Basado en Varianza, del inglés *Variance-based Relevance Analysis*

WPT: Transformada de Paquete de Wavelets, del inglés *Wavelet Packet Transform*

WSVM: Máquinas de Soporte Vectorial por Wavelet, del inglés *Wavelet Support Vector Machine*

WT: Transformada Wavelet, del inglés *Wavelet Transform*

ZC: Cruce por Cero, del inglés *Zero-Crossing*

RESUMEN

La demanda de energía eléctrica por parte del sector industrial global solo crece, y se ve afectada fundamentalmente por el hecho que los rodamientos instalados en los motores eléctricos industriales se constituyen en el modo primario de fallo que afecta dicho consumo energético. Por tanto, la demanda de mantenimiento eficiente en motores eléctricos es crítica. Como solución, típicamente se ha empleado el mantenimiento preventivo como filosofía para la gestión de activos donde se busca maximizar la operación mediante inspecciones de rutina con mayor frecuencia cuando se exhiben anomalías, pero esto conlleva a un aumento en la probabilidad de falla debido a la intervención continua y el error humano inherente. Este documento de tesis doctoral presenta un marco integrado de diagnóstico y pronóstico para tratar con la vida útil remanente en rodamientos, con base en la estimación de la probabilidad de la degradación sujeta a modos de fallos definidos y severidades inducidas. Los enfoques metodológicos presentados incorporan análisis de vibración, para apoyar activamente el diagnóstico de fallos de forma no destructiva ni invasiva de máquinas rotativas en etapas tempranas, pero suponen un desafío con respecto a las propiedades de la señal, por ejemplo, su alto componente dinámico y de no estacionariedad. Se trabaja bajo una metodología que supone degradación de rodamientos evidenciada por una serie de estados discretos que representan efectivamente la dinámica y no estacionariedad del proceso de fallo. El conocimiento empírico previo también se incorpora dentro del sistema integrado para la clasificación de fallos y severidades. En definitiva, la metodología propuesta caracteriza diferentes firmas de falla en rodamientos empleando señales de vibración y varios dominios de representación de señales, con el propósito de tratar con la naturaleza estocástica y relaciones complejas en los datos concernientes a fallos y severidades. En la selección de características, se lleva a cabo un estudio sobre fusión y selección de dominios y características para la representación de señales, con el fin de discriminar la información relevante. Específicamente, aquí se presentan esquemas de fusión y selección basados en procedimientos de relevancia forward y backward, así como un enfoque estocástico de selección de características. Estas técnicas están destinadas a resaltar las características relevantes de múltiples dominios de las señales de vibración para las tareas de diagnóstico de fallos y evaluación de severidad, al mismo tiempo que se reduce la dimensionalidad de los datos. Para la etapa de entrenamiento, los enfoques se basan en sistemas estocásticos relacionados con la estimación de la probabilidad de un conjunto de estados discretos, tales como: Mo-

delos Ocultos de Markov con observación discreta, Modelos Ocultos de Markov con observación continua y Modelos Ocultos de Markov Jerárquicos. El marco de diagnóstico y pronóstico integrado también se prueba como una herramienta de análisis de relevancia de características para discriminar múltiples condiciones de salud en rodamientos con caracterización multi-dominio. Los resultados logrados sobre una base de datos pública demuestran que los sistemas propuestos superan los algoritmos del estado-del-arte en cuanto a la cantidad de características seleccionadas y la eficiencia de la clasificación. Además, los resultados de las pruebas experimentales y los procedimientos de validación enseñan que el enfoque propuesto tiene la capacidad de proporcionar una advertencia de condiciones anormales del sistema mediante la identificación de las etapas tempranas de las condiciones de fallo. Las metodologías propuestas y técnicas analíticas desarrolladas en esta investigación para el pronóstico a largo plazo de la vida útil remanente se pueden aplicar en una gran variedad de contextos.

PALABRAS CLAVE

Diagnóstico, Estimación de Probabilidad de la Degradación, Modelos Ocultos de Markov, Pronóstico, Selección de Características, Vida Útil Remanente.

ABSTRACT

Energy demand from global industries only grows, and is affected by the fact that installed bearings in industrial electric motors are constituted as the primary mode of failure affecting energy consumption. Therefore, the demand for efficient maintenance in electric motors is critical. As a solution, preventive maintenance has typically been employed as a philosophy for asset management that aims to maximize operation through routine inspections with increasing frequency when abnormalities are exhibited, but this leads to an increase in the probability of failure due to the repetitive intervention and the inherent human error. This doctoral thesis document presents an integrated diagnostic and prognostic framework for remaining useful life prognosis in bearings, based on the estimation of probability of degradation subject to a defined faults modes and severities induced. The methodological approaches presented incorporate vibration analysis, which actively support the nondestructive and noninvasive fault diagnostics of rotating machines at early stages, but they pose a challenge regarding the signal properties, e.g., highly dynamic and non-stationary. The methodology assumes that bearing degradation consists of a series of discrete degraded states which effectively represent the dynamic and stochastic process of failure process. Prior empirical knowledge is also embedded in the integrated system for classification of detecting faults and severities. In short, the proposed methodology characterizes different failure signatures from bearings using vibration signals and several domains of signal representation with the purpose of deal with stochastic nature and complex relationships in the data concerning failures and severities. For the characteristics selection stage, a study on the fusion and selection of domains and characteristics for signal representation is carried out, in order to discriminate the relevant information. Specifically, here is introduced a fusion and selection schemes based on forward and backward procedures, as well as a stochastic feature selection approach. These techniques are intended to highlight relevant multi-domain features from vibration signals in bearing fault diagnosis and severity evaluation tasks, at the same time that the dimensionality of the data for machine training is reduced. For the training stage, approaches are based on stochastic systems based on the estimation of the probability of a set of discrete states, such as: Hidden Markov Models with discrete observation, Hidden Markov Models with continuous observation and Hierarchical Hidden Markov Models. The integrated diagnostics and prognostics framework is also tested as a feature relevance analysis tool to analyze multiple health conditions in bearings with multi-domain

characterization. Achieved results in a public database demonstrate that the proposed systems outperform state-of-the-art algorithms concerning the both the number of features selected and the classification accuracy. In addition, the results of experimental tests and validation procedures show that the framework has the capability of providing warning of abnormal machine conditions by identifying early stages of machine fault conditions. The proposed methodologies and analytical techniques developed in this research for long term remaining lifetime prognosis can be applied in a variety of contexts.

KEYWORDS

Diagnostic, Feature Selection, Health State Probability Estimation, Hidden Markov Models, Prognostic, Remaining Useful Life.

Parte I
INTRODUCCIÓN

1. INFORMACIÓN GENERAL

Título: PRONÓSTICO DE VIDA ÚTIL REMANENTE EN RODAMIENTOS CON BASE EN LA ESTIMACIÓN DE LA PROBABILIDAD DE LA DEGRADACIÓN

Área de investigación: Automática; Gestión de Sistemas Eléctricos, Electrónicos y Automáticos.

Materias asociadas a la investigación: Procesamiento de señales, procesos estocásticos, aprendizaje de máquina, análisis multivariado, mantenimiento industrial, automatización industrial.

Director: Ph.D. Álvaro Ángel Orozco Gutiérrez.

2. PLANTEAMIENTO DEL PROBLEMA

Sólo en los Estados Unidos de América la industria invirtió cerca de \$600 mil millones de USD con el fin de realizar actividades de mantenimiento en sus sistemas críticos hacia el año 1981, pero estos gastos se doblaron para el inicio de siglo [1], además los servicios de mantenimiento eran de \$186 mil millones de USD para el año 2016, con un crecimiento esperado del 4 % anual [2]. Estas cifras demarcan la importancia y necesidad crítica de realizar gestión de los activos, y aún más cuando entre un tercio y un medio del gasto realizado involucra acciones no efectivas [1]. Sin embargo, y con el fin de invertir en mantenimiento eficaz, el cual se conoce como productivo y está constituido por actuaciones preventivas y predictivas, las empresas norteamericanas han pasado de un 45.4 % en el año de 1991 a un 65.9 % en el año 2008 de gastos en este tipo de acciones [3]. Este último indicador para el resto del mundo industrializado alcanza un 72.4 % [3]. Teniendo en cuenta lo anterior, es evidente que existe una fuerte y constante presión por mejorar la confiabilidad de activos, lo cual se contrapone por un lado a la necesidad de acumular experiencia en la fuerza de trabajo calificada y por otro lado a verla envejecer [4, 5].

Dentro de las aplicaciones industriales, uno de los equipos de mayor importancia y criticidad son los motores eléctricos, los cuales básicamente se seleccionan según criterios relacionados con el costo inicial de adquisición. Además, es de resaltar que el consumo energético de los motores eléctricos representa cerca del 70 % del total en las industrias a nivel mundial, y al considerar que para 2010 las industrias abarcaban cerca del 41.7 % de la demanda mundial, ellos involucran cerca del 30 % de la

demanda energética [6]. Lo más preocupante radica en que el consumo energético industrial gana participación, lo cual se evidencia en el hecho que para el año 2015 ya era el 42.0% [7]. El mal funcionamiento de un motor es tan crítico y significativo, que puede demandar en solo 5 meses el equivalente en energía a su precio de compra [8].

La experiencia y la industria confirman que el anterior criterio de elección repercute directamente en los costos futuros, ya que por cada dólar invertido en la compra de un motor, 100 dólares adicionales se gastarán principalmente en consumos energéticos y también en necesidades de mantenimiento [8]. Por lo tanto, los costos de funcionamiento son mucho mayores que los de compra, mostrando nuevamente la importancia de su buen funcionamiento, e implicando con ello eficiencia energética y disponibilidad [7, 8].

Grandes industrias como las del petróleo, las petro-químicas y las de gas, entre otras, han efectuado estudios para determinar la distribución de los modos de fallo más comunes en motores eléctricos, dando como resultado que los rodamientos son la fuente del 51% de los fallos y solo seguidos de lejos por los devanados con un 16% [9]. Teniendo presente los datos previos, solo los rodamientos representan cerca del 15.3% de los principales modos de fallo afectando el consumo energético mundial actual, y dado la creciente demanda energética industrial, esta proporción crece año tras año.

Con el fin de lograr estrategias eficientes de mantenimiento, se han encarado de forma reciente retos relacionados con la investigación de pronóstico sobre la Vida Útil Remanente (RUL, del inglés *Remaining Useful Life*) en activos, con el fin de gestionar el riesgo residual en equipos antes que fallen [4, 10, 11]. Ahora, el diagnóstico o detección de fallos es un problema de clasificación con varias metodologías presentes en la literatura, en oposición al pronóstico el cual se puede considerar en su etapa infante, aunque en la última década ha sido un área del conocimiento ampliamente tratada [12, 13, 14, 15]. Un diagnóstico eficiente debe entregar información sobre localización y severidad, siguiendo un proceso de detección de ocurrencia y realizando análisis estadístico de las características observadas para llegar a una decisión del estado de salud estructural [16, 17, 18]. A su vez, el pronóstico se debe entender como la estimación del RUL para un activo mediante la predicción de la progresión de una anomalía diagnosticada [11].

En el campo del diagnóstico de fallos, el interés general se ha centrado en técnicas no invasivas, lo cual implica básicamente estrategias basadas en señales [13, 14, 19, 20], buscando características en el dominio del tiempo y de la frecuencia. En

el dominio del tiempo, las características se basan principalmente en valores estadísticos y momentos de las señales así como en factores de análisis de las mismas; mientras que en el dominio de la frecuencia se han empleado técnicas tales como la Transformada Rápida de Fourier (FFT, del inglés *Fast Fourier Transform*), espectros de potencia, coeficientes cepstrales y espectros de orden superior [13, 14, 21, 22]. Sin embargo, estos últimos se aplican normalmente en señales no Gaussianas [23], ya que si las señales de entrada son aleatorias gaussianas, un sistema de fase no mínima será identificado como un sistema de fase mínima. Otras estrategias emplean técnicas en el dominio tiempo-frecuencia, como lo son las Wavelets [13, 14, 24, 25], las cuales se caracterizan por ser sensibles a la selección de parámetros como el tamaño de ventana y el porcentaje de traslape entre ventanas, lo cual afecta la resolución en tiempo y en frecuencia, además las Wavelets adicionan términos cruzados cuando se analizan señales con componentes desde varias frecuencias [26].

En enfoques convencionales de análisis de señales, las características extraídas usualmente recaen sobre un único espacio de representación; llámese tiempo, frecuencia, tiempo-frecuencia, modal, etc. [25, 27, 28, 29, 30]. Estos enfoques tradicionales carecen de la posibilidad de conjunción para crear diccionarios ampliados, donde se evalúe la relevancia de cada diccionario, la contribución de cada característica y la reducción de las mismas con miras a incrementar la exactitud del diagnóstico a la vez que se disminuye la complejidad computacional.

Continuando con el diagnóstico, a partir de señales, también se ha empleado técnicas paramétricas lineales. Entre estas técnicas está el método Auto Regresivo de Media Móvil (ARMA, del inglés *Autoregressive Moving Average*), que trabaja normalmente a corto plazo por la suposición de linealidad del proceso; y la Inteligencia Artificial (AI, del inglés *Artificial Intelligence*) donde se ha empleado principalmente Redes Neuronales Artificiales (ANN, del inglés *Artificial Neural Networks*), que poseen limitaciones conocidas por su dificultad de interpretación y estructura por ser una caja negra, fuera de una lenta convergencia del proceso de formación y ser altamente dependientes de las metodologías de extracción de características [31]. Otros enfoques, como la Lógica Difusa (FL, del inglés *Fuzzy Logic*), tienen dificultades para enfrentar problemas estocásticos ya que la aleatoriedad es inherente en la mayoría de los procesos de degradación [11, 20, 32]. El análisis modal ha mostrado que los modos pueden ser insensibles a la localización del daño y que cambios en sus formas pueden ser no identificables debido a la presencia de ruido y al número restringido de modos distinguibles [16].

Otras estrategias tradicionales para el diagnóstico emplean enfoques basados en

el modelo físico del sistema, las cuales requieren de conocimiento específico del mismo y, generalmente, no reflejan un modelo general para todos los modos de fallo y durante el ciclo completo de vida del activo [11, 28, 33]. De otro lado, los sistemas físicos actuales se basan principalmente en dispositivos con modelos matemáticos no lineales, lo cual implica incertidumbre por suposiciones y simplificaciones debido a la complejidad y naturaleza estocástica involucrada en ellos y en sus propios mecanismos de degradación, así como la dificultad para integrar datos de la operación comportamental como diferentes tipos de carga y regímenes de velocidad [19, 28, 29, 32]. Además, es impráctico modelar un proceso de predicción sobre un ciclo completo de vida útil [34], debido principalmente al tiempo y cantidad de datos que se involucra.

En cuanto al pronóstico, algunas estrategias se basan en observaciones visuales del espectro en dos dimensiones y gráficos de contorno de tiempo-frecuencia, los cuales son incapaces de reunir todos los requerimientos de un modelo de pronóstico [27]. Los modelos para pronóstico normalmente se han entrenado en enfoques fuera de línea, con lo cual solo responden adecuadamente a procesos de fallo con características iguales a los datos de entrenamiento, y por ende, no se adecuan a situaciones donde las características son cambiantes. Además, usualmente el costo computacional de entrenamiento ha hecho que no se puedan usar en línea para aplicaciones en tiempo real [35]. Por tanto, los modelos de pronóstico han sido hasta la fecha teóricos y restringidos a cantidades pequeñas de modelos y modos de fallo [10, 33, 36].

Los Modelos Ocultos de Markov (HMM, del inglés *Hidden Markov Models*) son apropiados para el análisis de un sistema aleatorio dinámico. Principalmente, permiten realizar predicción sobre el diagnóstico del proceso de degradación al tener la ventaja de fácil interpretación y poder actuar en ambientes de aprendizaje competitivo y con fines de caracterizar los diferentes estados de salud de un mecanismo [11, 20, 30]. Sin embargo, cuando el número de modos diferentes de fallo crece se requiere entrenar un gran número de modelos, normalmente uno por cada modo de fallo y otro adicional por el modo de operación normal, lo cual incrementa la complejidad computacional, tanto al entrenar como al diagnosticar y si es del caso, pronosticar.

Entre las variantes de los HMM están los HMM Jerárquicas (HHMM, del inglés *Hierarchical Hidden Markov Models*), que se comportan como un encadenamiento de sub-HMM jerarquizadas, permitiendo analizar múltiples estados de salud [37]. Otras implementaciones de modelos HMM, como las Redes Bayesianas Dinámicas (DBN, del inglés *Dynamic Bayesian Networks*), reducen el número de parámetros al

factorizar y facilitar la flexibilidad adicional respecto a la estructura del modelo [38]. Aunque estos últimos enfoques facilitan el desarrollo sobre varios modos de fallo, tienen el inconveniente de dificultar la implementación de índices adecuados de degradación con el fin de poder implementar la fase de pronóstico [11, 34].

El desarrollo de metodologías y/o aplicaciones prácticas que permitan pronóstico eficiente a lo largo del proceso de degradación con base en la estimación de los diferentes estados de salud de un activo se encuentra en estado embrionario, además, aunque se ha avanzado en los procesos de clasificación de fallos o estimación de estados de salud estructural, estos se han centrado en algunos componentes específicos sin tener aún presente todos los procesos, o niveles de degradación, en un sistema [4, 20, 22, 25, 27, 29, 32]. Aún más limitadas se encuentran las metodologías de pronóstico de RUL [10, 28, 33, 35]. Además, aunque los métodos basados en datos son preferidos, usualmente requieren grandes cantidades de señales para lograr una buena estimación. Asimismo, no siempre es permitido que los mecanismos corran hasta lograr un evento de fallo [28, 36], lo cual se evidencia en las mismas bases de datos disponibles, donde prácticamente ninguna registra el ciclo completo de vida del activo.

Es por todo lo expuesto que, esta propuesta doctoral busca como propósito de investigación determinar la forma de parametrizar adecuadamente modelos de inferencia estocástica con el fin de diagnosticar y realizar predicción de vida útil remanente en rodamientos, que integre de forma efectiva fusión de características sobre varios espacios de representación de señal, involucrando modos de fallo y niveles de severidad.

3. JUSTIFICACIÓN

3.1. Impacto

El pronóstico de daño es un problema de predicción con base en extrapolación, donde una serie de modelos que describen el daño presente y los diferentes perfiles futuros de carga se combinan para perfilar el espacio potencial de condición segura de operación y bajo una evaluación de riesgo probabilístico. Se diferencia del diagnóstico, en que este último se centra en el análisis de las causas raíces, mientras que el pronóstico trata del comportamiento futuro como resultado de un análisis sobre datos relevantes.

El *Damage Prognosis Workshop*, patrocinado por *Los Alamos National Laboratory* entregó como uno de sus resultados la siguiente definición para pronóstico de daño [17]:

Es el acoplamiento de la información relacionada con el ambiente de carga para el cual se diseñó un sistema; el monitoreo de su salud estructural; el monitoreo de su uso; las condiciones ambientales y de operación pasadas, presentes y futuras; componentes instalados, modelamiento numérico y pruebas del sistema con el fin de estimar la vida útil remanente del sistema. El pronóstico de daño mide el estado actual del sistema, estima el ambiente de carga futura y predice mediante simulación y la experiencia pasada la vida útil remanente.

Según el estándar ISO 13381-1 de 2015, la definición de pronóstico de daño es [18]:

El pronóstico es una estimación del tiempo para fallo y del riesgo para uno o más modos incipientes de fallo. El pronóstico es usualmente efectivo para modos de fallo conocidos, relacionados con la edad, o con características progresivas de degradación. Es el análisis de los síntomas de fallo con el fin de predecir condiciones futuras y vida residual dentro de los parámetros de diseño.

Con base en las anteriores definiciones, se justifica el denominado Mantenimiento con Base en la Condición (CBM, del inglés *Condition Based Maintenance*) que se realiza comúnmente con base en la evaluación del estado de salud de un activo, en contraposición del mantenimiento preventivo que se realiza normalmente con base en el tiempo o ciclos de servicio. Para el caso de los motores eléctricos, y específicamente para los rodamientos, poder diagnosticar el estado de salud permite incrementar la disponibilidad, mantener las actividades de producción bajo planeación, reducir los tiempos de paro y mejorar el entorno de seguridad de la operación industrial.

Si además de diagnosticar el estado de salud de los activos, un programa de mantenimiento efectivo se desarrollara bajo un pronóstico en el tiempo de la vida útil remanente, la operación de los activos se maximizaría a la vez que se minimizan las intervenciones por mantenimiento. En el caso particular de los rodamientos, un mantenimiento con base en la condición impactaría directamente sobre el 30 % del consumo mundial de energía, y como ya se explicó previamente los fallos en rodamientos son la fuente del 51 % de los modos de fallo en motores y representan cerca del 15.3 % de los principales modos de fallo afectando el consumo energético mundial.

3.2. Pertinencia

La única aplicación que se puede considerar hasta la fecha de un sistema de inferencia de vida útil remanente está relacionada con la evaluación aplicada a rotores en helicópteros, especialmente el componente de caja de cambios, y cuenta con aprobación por parte de la Administración Federal de Aviación, (FAA, del inglés *Federal Aviation Administration*) [39]. Igualmente ciertos componentes del rotor de helicópteros han extendido su vida en un 15 % con base en aplicaciones que usan exclusivamente datos de vibraciones [27, 40].

Este mismo tipo de aplicaciones en rotores de helicópteros se han desarrollado en el Reino Unido con mejoras que implican reducciones en pérdidas de equipos hasta en un 50 %, éxito que se atribuye al hecho que la velocidad del rotor (sin implicar par motor) se mantiene en un rango que no excede en el 2 % a la nominal [41]. También se destacan aplicaciones en el monitoreo y evaluación de daño acumulado en maquinaria rotativa desarrollados por parte de la Naval de Estados Unidos para sus navíos [42].

A excepción de las aplicaciones en helicópteros, prácticamente todas las otras implementaciones realizadas presentan falencias, o carecen, en lo relacionado con la inferencia o pronóstico de la degradación y se centran fuertemente en el diagnóstico. Todo lo anterior es un indicador que esta tecnología está en una fase incipiente y en desarrollo [33, 36].

Con el fin de lograr objetivos de disponibilidad, producción, eficiencia energética y entornos seguros de operación relacionados con los motores eléctricos y atendiendo el criterio que los rodamientos representan el 51 % de los fallos en motores, se hace latente la necesidad de plantear una metodología que permita diagnosticar el estado estructural de los rodamientos, presentar un indicador de degradación con base en señales y así indicar la progresión de las anomalías mediante el pronóstico de vida útil remanente. Además, es necesario encontrar como integrar el conocimiento cualitativo y la escasa información cuantitativa con el fin de mejorar la exactitud del pronóstico de fallo [29, 35].

Adicionalmente, la implementación de una metodología para el RUL contribuye a los objetivos del CBM, y específicamente a reducir los costos totales del mantenimiento hasta en un 25 %, e involucrando consecuentes mejoras en calidad y productividad [3, 11, 29].

3.3. Viabilidad

Entre los estudios con mayor éxito y despliegue sobre pronóstico de vida útil remanente están los relacionados con la caja de cambios en rotores de helicópteros [43, 44, 45, 46, 47] y uso en el monitoreo y evaluación de daño acumulado en maquinaria rotativa, desarrollados por parte de la Naval de Estados Unidos para sus navíos [42].

Otros estudios de resaltar involucran plataformas marítimas de exploración petrolera, mediante técnicas como la Función de Respuesta en Frecuencia (FRF, del inglés *Frequency Response Function*) o procesos de Wiener, y que permiten inferir la localización de daños [48, 49]. Otro estudio utilizó un modelo ARMA para medir registros en tiempo de la aceleración y estimar cambios en la frecuencia de resonancia y niveles de amortiguamiento modal sobre una estructura real [50].

También se encuentran aplicaciones que enfrentan el pronóstico desde la perspectiva basada en modelo, pero restringidas a fases de desarrollo como en [51, 52], donde se implementan modelos para la suspensión de un vehículo y para el modelado de una turbina de viento, y se hace un estudio de diagnóstico y pronóstico de la vida útil del sistema proyectado sobre diferentes condiciones de operación.

Ya que los métodos generales de inferencia de vida útil se basan en el conocimiento previo, bien sea por modelo o con base en datos, de cómo debe operar un sistema y en la frecuencia relativa de ocurrencia de diferentes tipos de defectos, esta concepción encaja de forma perfecta en la estructura de un doble proceso estocástico el cual debe encontrar probabilísticamente el estado de degradación total del sistema y la posibilidad de transición entre estados [53, 54]. Además, este modelo se asemeja en gran medida al procesamiento de voz, en cuanto a la naturaleza cuasi-estacionaria de las señales y diferencias en caracteres de parámetros iguales de diferentes sistemas, como es el caso de variabilidad en fonemas iguales proviniendo del mismo tracto vocal o en señales de vibración bajo mismas condiciones de operación desde máquinas supuestamente idénticas. Ya que en la mayoría de sistemas de procesamiento de voz se emplean los HMM, se presenta ésta como la técnica a ser empleada primordialmente para realizar la implementación del sistema de diagnóstico y pronóstico de vida útil remanente [11, 20, 30]. Los HMM permiten modelar y simular la dinámica cambiante de datos estocásticos, donde los cambios comportamentales se ven reflejados directamente en las variables latentes del proceso [55, 56].

4. OBJETIVOS

4.1. Objetivo general

Desarrollar una metodología que permita diagnosticar y estimar la salud estructural en rodamientos, con base en estados discretos de salud, señales indicativas de la degradación y severidad de fallos, con el fin de predecir la progresión de las anomalías mediante un indicador de degradación general para el pronóstico de vida útil remanente.

4.2. Objetivos específicos

- Elaborar una metodología para el diagnóstico con base en señales y estados discretos de salud que abarque de una forma efectiva la integración de espacios de representación, la selección de características, análisis de relevancia de las mismas, clasificación y diagnóstico de fallos.
- Proponer una metodología de estimación de la probabilidad del estado actual de salud en rodamientos, cubriendo modos de fallos y niveles de severidad, con base en la información de salida de la etapa de diagnóstico y abarcando de forma efectiva la asignación de probabilidad a cada posible estado discreto y la evaluación global de un índice de fallo para el tiempo actual.
- Determinar una metodología para la predicción de vida útil remanente con base en la salida de la etapa de estimación y datos históricos de vida.
- Validar el método propuesto, integrando de forma efectiva los pasos de diagnóstico, estimación y pronóstico con el fin de proveer un sistema completo de estimación de vida útil remanente en rodamientos. La validación se debe realizar sobre datos que cubran niveles de severidad sobre diferentes modos de fallos en rodamientos.

5. CONTRIBUCIÓN Y ORIGINALIDAD

Con base en los objetivos específicos, este documento presenta una novedosa metodología para el diagnóstico de rodamientos con base en señales y estados discretos de salud abarcando de forma efectiva la integración de espacios de representación,

la selección de características, análisis de relevancia de las mismas, clasificación y diagnóstico de fallos; permite estimar de la probabilidad del estado actual de salud, cubriendo modos de fallos y niveles de severidad. Estas metodologías, permiten en conjunto predecir la vida útil remanente. Se presenta la validación de las metodologías propuestas, integrando de forma efectiva los pasos de diagnóstico, estimación y pronóstico.

Con el fin de respaldar las metodologías propuestas, difundir los resultados y garantizar la originalidad, se lista a continuación el conjunto asociado de trabajos, publicaciones en revistas indexadas y ponencias en eventos:

1. **Mauricio Holguín**, Álvaro Ángel Orozco, Germán A. Holguín, Mauricio Álvarez, 2016, "Optimal state selection and tuning parameters for a degradation model in bearings using Mel-Frequency Cepstral Coefficients and Hidden Markov Chains," en la revista *Ingeniare - Revista Chilena de Ingeniería*, Vol. 24, No. 4, pp. 570-580, ISSN 0718-3305.
2. **Mauricio Holguín-Londoño**, Oscar Cardona-Morales, Edgar F. Sierra-Alonso, Juan D. Mejía-Henao, Álvaro Orozco-Gutiérrez, German Castellanos-Domínguez, 2016, "Machine Fault Detection Based on Filter Bank Similarity Features Using Acoustic and Vibration Analysis," en la revista *Mathematical Problems in Engineering*, Hindawi Publishing Corporation, Vol. 2016, Article ID 7906834.
3. **Mauricio Holguín Londoño**, Álvaro A. Orozco Gutiérrez, Mauricio Álvarez López, 2015, "Fusión y selección de espacios de representación y sus características en tiempo, frecuencia y tiempo-frecuencia, para diagnóstico de modos de fallos y severidad en rodamientos, usando HMM," en el CIBIM XII, XII Congreso Iberoamericano de Ingeniería Mecánica, Nov. 10-13, 2015, Guayaquil, Ecuador.
4. **Mauricio Holguín Londoño**, Álvaro Ángel Orozco Gutiérrez, 2016, "Metodología para el análisis de vida útil remanente en rodamientos, aplicación a los motores eléctricos," en el I Congreso Internacional de Electromecánica y Eléctrica, Junio 15-17, 2016, Santo Domingo, Ecuador.
5. Proyecto de Investigación Colciencias - Universidad Tecnológica de Pereira: "Diseño y desarrollo de un sistema prototipo en línea para el diagnóstico de motores de combustión interna diésel en servicio con base en vibraciones mecánicas. Aplicación a los sistemas de transporte masivo." Código 1110-669-46074. Julio 2015 a Julio 2018. Investigador principal: **Mauricio Holguín Londoño**.

6. Diego Fernando Muñoz Torres, 2016, "Pronóstico de vida útil restante en rodamientos, con base en datos de vibraciones y sistemas de inferencia estocástica con degradación no lineal." Proyecto de grado para optar al título de Magister en Ingeniería Eléctrica, Universidad Tecnológica de Pereira, Pereira, **Tutor: Mauricio Holguín Londoño.**
7. Andrés Felipe Valencia López, 2018, "Diseñar una metodología para determinar los niveles de degradación mediante señales de vibración en rodamientos, usando fusión de datos y conglomerados, con optimización de parámetros por algoritmos genéticos." Proyecto de grado para optar al título de Ingeniero Electricista, Universidad Tecnológica de Pereira, Pereira, **Tutor: Mauricio Holguín Londoño.**
8. Pablo A. Aristizabal V., 2016, "Metodología para caracterizar señales de vibración en motores diésel, usando representaciones del tiempo, frecuencia y tiempo-frecuencia y diagnóstico mediante HMM." Proyecto de grado para optar al título de Ingeniero Electricista, Universidad Tecnológica de Pereira, Pereira, **Tutor: Mauricio Holguín Londoño.**
9. Juan Camilo Castaño Guzmán, Carlos Hernán Agudelo Gutiérrez, 2015, "Extensibilidad de criterios de decisión para un clasificador de fallos en rodamientos basado en HMM, sobre resultados en un clasificador de menos estados, empleando características en diferentes espacios de representación." Proyecto de grado para optar al título de Ingeniero Electricista, Universidad Tecnológica de Pereira, Pereira, **Tutor: Mauricio Holguín Londoño.**
10. Diego Fernando Medina C., 2014, "Evaluación de metodologías para el diagnóstico automatizado de la severidad de fallos, a partir de vibraciones mecánicas en rodamientos, usando características de tiempo y de frecuencia, con modelos ocultos de Markov." Proyecto de grado para optar al título de Ingeniero Electricista, Universidad Tecnológica de Pereira, Pereira, **Tutor: Mauricio Holguín Londoño.**
11. Henry Alexander Blandón Hurtado, Yamir Fernando Martínez Mendoza, 2013, "Desarrollo de una metodología para la detección de fallos a partir del análisis de vibraciones en rodamientos usando características de tiempo y modelos ocultos de Markov (HMM)." Proyecto de grado para optar al título de Inge-

niero Electricista, Universidad Tecnológica de Pereira, Pereira, **Tutor: Mauricio Holguín Londoño.**

12. **Mauricio Holguín Londoño**, 2018 "Bearing fault diagnosis from vibration signals based on a stochastic feature selection approach." Artículo en proceso de sometimiento, Elsevier Editorial System(tm) for Engineering Applications of Artificial Intelligence.

Parte II

**ESTADO DEL ARTE Y MARCO
TEÓRICO**

6. MARCO CONCEPTUAL

Todo equipo industrial es un activo esencial de una actividad empresarial enmarcada en un mundo altamente competitivo. Para poder estar al paso de los retos de este mundo globalizado, el mantenimiento se presenta como una actividad fundamental basada en la combinación de tecnologías, acciones administrativas y de gestión durante todo el ciclo de vida útil de un equipo, con el fin de lograr que este cumpla con su función requerida bajo las condiciones de demanda establecidas [57]. Las acciones emprendidas de mantenimiento han evolucionado para ser un factor preponderante con impacto en toda la cadena operativa y financiera de las organizaciones, con lo cual ahora el mantenimiento se enfrenta a una gama de desafíos que involucran la calidad, la eficiencia, la productividad, la seguridad ocupacional y la rentabilidad. Lo anterior implica tratar con tecnologías de innovación que no sean invasivas o impliquen paros en la producción. En definitiva, una política de mantenimiento moderna no solo previene los fallos de los activos, sino que conduce a la utilización máxima de la capacidad instalada, a la calidad de producto y a la satisfacción del cliente utilizando la vida útil adecuada del equipo.

Para lograr los anteriores objetivos se plantea el pronóstico de vida útil remanente, el cual involucra la observación periódica muestreada de un sistema en el tiempo de respuesta dinámica y análisis estadístico de las características que determinen el estado actual del sistema. Para un largo período de tiempo, la salida de este sistema es información que se actualiza periódicamente independientemente que el sistema pueda realizar su función. A este último proceso se le conoce como monitoreo de la condición. Para dar solución al problema del pronóstico de vida útil remanente se requiere de grandes mediciones cuantificables que permitan determinar el desarrollo de unas condiciones de daño e igualmente que indiquen su progresión.

Aunque la solución se debe diseñar para una aplicación específica, existen unos componentes mayores plenamente identificables, para lo cual inicialmente se debe tener presente tres aspectos fundamentales: la causa del daño de interés, las técnicas que se requieren para cuantificar el daño y el objetivo del pronóstico. El origen de un daño para un modo potencial de fallo se puede clasificar en tres categorías generales: i) desgaste gradual, donde se acumula lentamente a nivel de material o componente y normalmente a escala microscópica; ii) eventos discretos predecibles, donde se acumula a tasas más rápidas en eventos inesperados pero si caracterizados a priori, aunque conserva una escala microscópica; y iii) eventos discretos impredecibles, en el que un daño desconocido y severo se inyecta al sistema en tiempos impredecibles.

De acuerdo con la ley de fallo que se basa en la función de confiabilidad Bathtub, las dos primeras categorías representan la vida útil de un activo, o la vida que no comprende la parte senil [58].

Para eventos discretos impredecibles, las mediciones se deben tomar en línea y el sistema se debe evaluar en línea, pero para desgaste gradual no siempre es necesaria una técnica en tiempo real o cercana a tiempo real y puede bastar una evaluación a intervalos discretos fuera de línea. El objetivo final es plantear cuanto tiempo resta hasta que se requiera de mantenimiento, hasta que el sistema falle o no se pierda la función requerida demandada. Otro objetivo puede ser conocer la probabilidad de completar la misión presente dada la evaluación del estado actual de daño. También, un objetivo alternativo puede ser estimar cuanto puede trabajar el sistema bajo condiciones seguras en un ambiente anticipado, antes de no tener más confianza en las capacidades predictivas del sistema que se emplea.

El diagnóstico de fallos en máquinas se puede realizar a través de enfoques de sistemas basados en modelo, basados en señales o basados en conocimiento. Las técnicas basadas en modelo detectan los cambios en el comportamiento de la máquina empleando modelos matemáticos del sistema [52]; sin embargo, su éxito depende de modelos determinísticos de procesos industriales que normalmente poseen suposiciones del sistema real. Los algoritmos basados en señales emplean adquisición de datos en lugar de modelos específicos de entrada-salida para el diagnóstico [59]; sin embargo, aunque el diagnóstico con base en señales entrega una forma más realista de evaluar un fallo en comparación con las técnicas basadas en modelo, este depende de la evaluación en comparación con el estado normal, lo cual recae en el campo de un experto. En contraste, los enfoques basados en conocimiento tratan con el problema de diagnóstico desde una perspectiva del reconocimiento de patrones, comprendiendo una etapa de adquisición de señales, un procedimiento de estimación de características y una evaluación de los tipos de fallos a través de algoritmos de aprendizaje [60].

En general, los sensores de vibración y acústicos son clasificados como dispositivos no invasivos de adquisición de datos [61]; y donde las señales de vibración desde rodamientos son de naturaleza cuasi-estacionarias o no-estacionarias, con patrones de diagnóstico de fallo variando incluso cuando las señales provienen desde equipos supuestamente idénticos [62]. Por tanto, la estimación de características desde señales de vibración involucra parámetros basados en representaciones del tiempo, la frecuencia y del tiempo-frecuencia [21].

En este escenario, el mantenimiento predictivo se debe entender como la estima-

ción de la Vida Útil Remanente de un activo mediante la predicción de la progresión de una anomalía diagnosticada [11]. Un diagnóstico adecuado, debe permitir la toma acertada de decisiones sobre el estado de salud estructural de un activo, debe entregar información sobre localización de avería, y severidad de la misma, siguiendo un procedimiento de detección de eventos y análisis estadístico de las características observadas [16].

7. ESTADO DEL ARTE

En el área industrial moderna, los motores eléctricos son equipos críticos que representan cerca del 70 % del consumo energético, con el agravante que ellos principalmente se seleccionan según su costo inicial de adquisición y no por su consumo. Para el año 2015 las industrias abarcaban ya el 42 % de la demanda energética mundial, y en consecuencia los motores consumen cerca del 30 % de esa demanda [7]. Por lo tanto, la eficiencia energética y disponibilidad de los motores son asuntos a ser preponderantes [6, 8]. En este escenario, la industria ha determinado que los rodamientos son la fuente del 51 % de los modos de fallo más comunes en motores eléctricos [9], por lo que se han encarado de forma reciente retos relacionados con la investigación de pronóstico sobre el riesgo residual en estos equipos, antes que fallen [4, 10, 11].

La tarea global de pronóstico inicia con el diagnóstico de fallo, donde algunas de las estrategias corrientes emplean señales buscando características en el dominio del tiempo y de la frecuencia. Estas señales se basan principalmente en acústica y vibración, donde las primeras tienen una mayor vulnerabilidad al ruido ambiental en comparación con las segundas [63], por lo que es necesario emplear técnicas de pre-procesamiento de señales para reducir interferencias y mejorar la Relación Señal-Ruido (SNR, del inglés *Signal to Noise Ratio*). Debido a que una amplia clase de fallos en maquinarias se definen por energías espectralmente localizadas sobre sub-bandas estrechas, los Métodos de Banco de Filtros (FBM, del inglés *Filter Bank Methods*) se emplean como un medio de pre-procesamiento común para señales acústicas [64, 65] y de vibración [66]; por lo tanto, generalmente se incluye una etapa de banco de filtros que divide la señal medida en un conjunto de bandas espectrales estrechas, que concentran la información dentro de un ancho de banda limitado relacionado con el fallo de la máquina en consideración. En particular, la Transformada de Paquete de Wavelets (WPT, del inglés *Wavelet Packet Transform*) y la Descomposición de Modo

Empírico (EMD, del inglés *Empirical Mode Decomposition*) se utilizan con frecuencia [63, 67]. Sin embargo, la selección de los componentes espectrales representativos se basa comúnmente en el conocimiento a priori, asumiendo cada banda de frecuencias en la que se pueden observar las firmas [68]; por ejemplo, es esperado que los fallos en rodamientos y engranajes se presenten a frecuencias altas, pero este conocimiento no es accesible para todo tipo de fallos.

A la fecha, se han desarrollado varios enfoques para seleccionar las bandas más discriminantes, que se basan principalmente en el uso de distancias de similaridad. En particular, la detección del comportamiento impulsivo puede llevarse a cabo utilizando diferentes medidas de la relación entre componentes de banda estrecha. La estructura de los estimadores de correlación propuestos puede variar desde el Coeficiente de Correlación de Pearson (PCC, del inglés *Pearson Correlation Coefficient*) [69], el Índice de Mérito Basado en Asimetría (MIBS, del inglés *Merit Index Based on Skewness*) [70] y la Kurtosis Espectral (SK, del inglés *Spectral Kurtosis*) [68] hasta selectores más elaborados basados en gráficos cuartil-cuartil [71] e índices basados en entropía [13]. No obstante, la mayoría de estos enfoques se basan en suposiciones de estacionariedad para los procesos correlacionados. Este modelo puede no ser adecuado para una amplia variedad de fallos no estacionarios que tienen una dinámica impulsiva o confinada en el dominio del tiempo.

La gran mayoría de las acciones de mantenimiento basadas en señales acústicas se enfocan en inspecciones visuales, evitando incorporar el procesamiento de señales en el diagnóstico y en el pronóstico [72]. En el Dominio del Tiempo (TD, del inglés *Time Domain*), las características se basan principalmente en valores estadísticos, momentos de las señales y factores de análisis. Otros escenarios emplean modelos de series de tiempo, donde la idea principal es extraer características sustentadas en los parámetros del modelo, con base en enfoques como Modelos Autorregresivos (AR, del inglés *Autoregressive Models*) y modelos ARMA [50, 73]. Todos los anteriores enfoques en el dominio del tiempo no reflejan adecuadamente los cambios en componentes frecuenciales, a medida que se desarrolla la progresión de una fallo y por tanto son insuficientes para identificar la progresión de componentes que fallan [13, 14, 23, 74].

En el Dominio de la Frecuencia (FD, del inglés *Frequency Domain*) se han empleado técnicas como la FFT, coeficientes cepstrales y espectros de orden superior [21, 22, 75, 76]. Uno de los problemas presentes en la detección de fallo es la capacidad de detectar señales de pequeña amplitud, las cuales en el dominio de la frecuencia resultan en un componente de energía distribuido en un amplio rango de

frecuencias, dando como consecuencia que las amplitudes armónicas de la frecuencia del defecto pueden estar ocultas en ruido. La aplicación de la Técnica Cepstrum de Potencia (PCT, del inglés *Power Cepstrum Technique*) se basa en la capacidad de esta para detectar la periodicidad en el espectro, al tiempo que es insensible a la ruta de transmisión de la señal desde una fuente interna hacia a una externa de punto de medición. De lo anterior, dada la naturaleza de las señales de vibración, tales técnicas frecuenciales son inadecuadas para identificar eventos no estacionarios, al menos por si solas, lo cual es recurrente para el caso de defectos en sistemas mecánicos.

Otras estrategias emplean técnicas en el Dominio Tiempo-Frecuencia (TFD, del inglés del *Time-Frequency Domain*), como lo son las Wavelets [24, 25, 26, 30], obteniendo resultados promisorios en la tarea de diagnóstico, pero se caracterizan por ser sensibles a la selección de parámetros como el tamaño de ventana y el porcentaje de traslape [26], y a que solo se puede detectar las características de señal que coinciden con la forma de la wavelet, mientras que todos los demás parámetros se enmascaran, o se ignoran [20].

Estrategias tradicionales alternas para el diagnóstico y la evaluación del RUL emplean enfoques con base en modelos físicos de los mecanismos, pero ellos requieren de conocimiento específico del sistema y generalmente no reflejan un modelo general para todos los modos de fallo y para el ciclo completo de vida del mecanismo, normalmente implican incertidumbre por suposición y simplificación debido a la complejidad y naturaleza estocástica involucrada en los sistemas [11, 28, 33].

En todas estas estrategias, el conjunto de datos proveniente desde señales a menudo consta de un gran número de características y el número de muestras es limitado debido a la carga de trabajo en la adquisición de datos, arrojando que el análisis de las características extraídas recaiga sobre un único espacio de representación, careciendo de conjunción de espacios ampliados, donde se evalúe la relevancia de cada uno y la contribución de cada característica con miras a incrementar la exactitud del diagnóstico, a la vez que se disminuye la complejidad computacional [21, 27, 28, 29, 77].

Independiente que el RUL se base en modelo o en datos, los métodos de inferencia confían en el conocimiento previo sobre cómo opera el sistema y la frecuencia relativa de ocurrencia de las diferentes clases de defectos. Esta perspectiva recuerda la estructura de un doble proceso estocástico el cual debe encontrar, mediante probabilidad, el estado total de degradación de un sistema y la probabilidad de transición entre estados [53, 54]. En gran medida, el modelo previamente mencionado es similar a aquellos que se emplean en el procesamiento del habla, ya que ambos poseen

señales con naturaleza cuasi-estacionaria y además existen diferencias en los parámetros de características similares desde sistemas diferentes. Un ejemplo de esto es la variabilidad en fonemas similares generados por el mismo tracto bucal, o la variabilidad en señales de vibración desde máquinas presumiblemente similares bajo condiciones idénticas de operación.

En consecuencia, para el aprendizaje basado en enfoques soportados en conocimiento se han empleado estrategias de agrupamiento supervisado y no supervisado, tales como HMM [20], ANN [78, 79, 80], Algoritmos Genéticos (GA, del inglés *Genetic Algorithms*) [81], Máquinas de Soporte Vectorial (SVM, del inglés *Support Vector Machines*) [82, 83], etc. En los anteriores, se destacan los HMM por ser apropiados para el análisis de un sistema aleatorio dinámico con el fin de realizar predicción sobre el diagnóstico del proceso de degradación, teniendo las ventajas de fácil interpretación y poder actuar en ambientes de aprendizaje competitivo con fines de caracterizar los diferentes estados de salud de un mecanismo [11, 20, 31, 37]. Los HMM permiten modelar y simular la dinámica cambiante de datos estocásticos, donde los cambios comportamentales se ven reflejados directamente en las variables latentes del proceso [55, 56]. Ya que los Coeficientes Cepstrales en Frecuencias Mel (MFCC, del inglés *Mel-Frequency Cepstral Coefficients*) proveen información de una señal tanto desde los dominios del tiempo como de la frecuencia, permiten la extracción dinámica de características, ya sean ellas lineales o no lineales. Estas características son las que hacen que los MFCC sean comúnmente empleados en señales de vibración y en el procesamiento del habla [84, 85].

8. MARCO TEÓRICO

8.1. Rodamientos

Componentes y fallos

Los rodamientos son elementos mecánicos que reducen la fricción entre el eje de una máquina y las piezas conectadas a éste, sirviendo de apoyo y facilitando el desplazamiento. Además, los rodamientos soportan y transmiten esfuerzos generados en el eje, a otro elemento estructural, por lo que son susceptibles de elevados riesgos de fallo, debido a los ciclos de cargas que soportan. Las partes constitutivas de la gran mayoría de los rodamientos son: pista externa de rodamiento, pista interna de rodamiento, jaula retenedora, elementos rodantes y sellos. Estas partes se pueden

observar en la figura 8.1. Los elementos rodantes (como bolas o rodillos) transfieren la carga entre la pista externa y la interna de rodamiento; la jaula retenedora se encarga de separar uniformemente los elementos rodantes para reducir el calor por fricción generado y con el fin de optimizar la distribución de la carga, además guía los elementos rodantes en la zona sin carga; finalmente, los sellos se encargan de dar un sellado integral para prolongar la vida útil del rodamiento, al mantener el lubricante en su lugar [86].

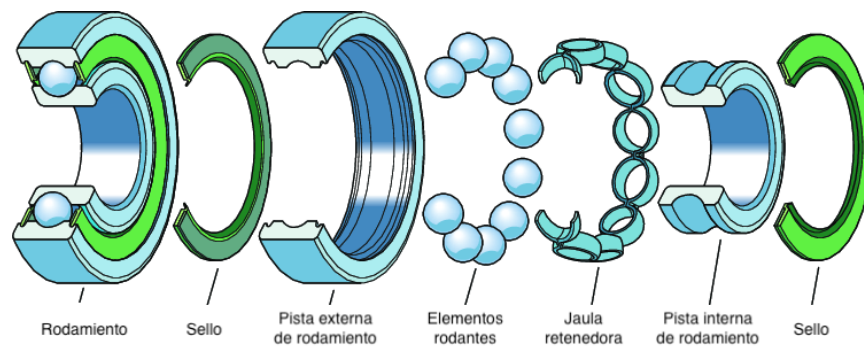


Figura 8.1: Principales partes de un rodamiento, adaptado de [86]

Normalmente, las condiciones de funcionamiento de los rodamientos no son las ideales, por presencia de variaciones en la magnitud de las cargas, discontinuidad en el servicio, malas alineaciones etc., conllevando a que el material se desgaste y provoque grietas que gradualmente se extienden hasta la superficie, que luego desencadenan defectos. Al fallar algún elemento de un rodamiento, en cada vuelta se genera un impacto sobre la parte defectuosa que produce frecuencias periódicas, o frecuencias fundamentales de fallo. En función del tipo de defecto que se encuentre en los rodamientos, se producen diferentes frecuencias características, las cuales dependen de las dimensiones del rodamiento y de la velocidad de giro [87].

Los tipos de defecto que pueden aparecer en un rodamiento se deben al desgaste, que es el daño causado en las superficies en movimiento relativo a causa de la remoción de material de una, o ambas superficies. En el desgaste, las capas superficiales de un sólido se rompen o se desprenden de la superficie. Este desgaste es progresivo, iniciando desde el denominado desgaste por fatiga (cuando existe concentración de tensiones mayores a las que puede soportar el material), pasando por el desgaste abrasivo (daño en las zonas de rozamiento a causa de partículas sólidas), avanzando al desgaste por erosión (que ocurre por corrientes de partículas abrasivas), hasta

el desgaste por frotación (que conjuga todos los tipos de desgaste anteriores) y el desgaste adhesivo (donde se transfiere material de una a otra superficie).

Los tipos de desgaste anteriores permiten identificar situaciones sobre el mantenimiento de los rodamientos. Es así que, el desgaste por fatiga es la condición inicial de desarrollo de una falla, el desgaste abrasivo es consecuencia del desgaste por fatiga y el desgaste por erosión indica una situación de requerimiento de sustitución del lubricante. Los demás tipos de desgaste se consideran situaciones irreversibles del rodamiento, implicando su cambio inminente [88].

Diagnóstico y pronóstico de fallos en rodamientos

Para realizar un proceso de diagnóstico, o pronóstico, sobre cualquier tipo de maquinaria, se debe realizar comparación entre descriptores actuales del comportamiento del sistema y valores de referencia, que normalmente se relacionan con el estado normal de operación. El diagnóstico inicia cuando se detecta la presencia de una anomalía. El proceso de monitoreo de la condición permite conocer el estado actual evolutivo de la anomalía, mientras que el pronóstico se enfoca en determinar la progresión de una anomalía ya diagnosticada y enfocado en la condición futura de falla esperada [89, 90]. Si existe un alto nivel de confianza en el diagnóstico y pronóstico, es posible determinar las acciones de mantenimiento adecuadas a planificar.

El estándar ISO 13381 establece estados genéricos de operación de un sistema, partiendo desde la referencia (que normalmente coincide con el de operación normal), para luego iniciar con un estado de fallo incipiente (o estado de baja criticidad) donde no es necesario realizar acciones inmediatas de mantenimiento, pero, se debe mantener el monitoreo a intervalos normales; un segundo nivel de fallo indica su desarrollo y normalmente es identificable el elemento causante; finalmente, un tercer nivel es indicativo de la presencia de fallo por métodos de análisis de detección. Un nivel adicional de fallo crítico indica una condición severa donde es necesario tomar acciones inminentes [89]. Estos cinco niveles (uno de referencia y cuatro de fallo), desde el punto de vista del pronóstico, se pueden resumir en cuatro: uno de referencia y los tres primeros de progresión; en este contexto el quinto no se tiene en cuenta por ser un nivel observable (por medios convencionales). Estos intervalos también coinciden con otros definidos en estándares como el ISO 17359, donde se habla de intervalos según su probabilidad de detección (Probabilidad Remota, Probabilidad Baja, Probabilidad Moderada, Probabilidad Alta y Certeza) e intervalos según su riesgos (No daño al sistema ni riesgo a la vida, No daño apreciable al sistema ni riesgo

apreciable a la vida, Inicio de pérdida de la función primaria con riesgo despreciable para la vida, Inicio de daño significativo al sistema o riesgo a la vida) [90].

8.2. Representación de datos de forma de onda

Los datos requeridos para realizar la evaluación del diagnóstico y pronóstico de un sistema físico normalmente provienen de datos de monitoreo basados en sistemas no invasivos sobre señales de vibración, acústica, temperatura, presión, lubricantes, etc. Los datos en bruto adquiridos desde sensores se procesan inicialmente (pre-procesamiento) antes de poder ser empleados en análisis subsiguientes (procesamiento), es así como los datos provenientes desde formas de onda requieren de estrategias y técnicas para su análisis e interpretación. En este pre-procesamiento se debe suprimir o atenuar los errores ocasionados por ruido de fondo, factores de intervención humana y fallos de los mismos sensores; lo anterior con el fin de extraer, evaluar y seleccionar las características adecuadas para las etapas de diagnóstico y pronóstico. Para la etapa de pre-procesamiento, en [91] se describe la aplicación de la Cancelación Adaptativa del Ruido (ANC, del inglés *Adaptive Noise Cancellation*) y la técnica de Deconvolución Ciega (BD, del inglés *Blind Deconvolution*) para detectar fallos en rodamientos en presencia de señales contaminadas por ruido, adicionalmente en [92] se demuestra que la solución para errores de datos causados por fallos del sensor es el aislamiento de dichos fallos.

Posterior a obtener datos libres de errores por los diversos tipos de ruidos que los afectan, se dispone de varias técnicas de procesamiento de señales para analizar e interpretar los datos de forma de onda y así extraer información útil con el fin de alimentar las etapas de diagnóstico y pronóstico. Usualmente, estas técnicas de procesamiento se clasifican en tres categorías principales de representación: dominio del tiempo (TD), dominio de la frecuencia (FD) y dominio tiempo-frecuencia (TFD).

8.3. Representación en el dominio del tiempo - TD

Las técnicas de representación basadas en el dominio del tiempo se basan en la obtención de estadísticos distintivos del comportamiento de la forma de onda en el tiempo. Estos estadísticos son características que se fundamentan principalmente en la distribución de las muestras de una señal donde su serie temporal se trata como una variable aleatoria [93]. En este contexto, usualmente la Función de Densidad de Probabilidad (PDF, del inglés *Probability Density Function*) se puede descomponer en

sus momentos constitutivos, en donde se desea que un cambio en la condición de la señal cause un cambio en la función de densidad de probabilidad y, por tanto, en sus estadísticos o momentos. Las características, así extraídas, son indicativos del cambio que monitorea el estado de salud del sistema bajo vigilancia.

Para un segmento de señal $z \in \mathbb{R}^T$ con T muestras, los momentos y características principales se muestran en la tabla 1 [93, 94, 95]. Para la Entropía, $x(i)$ es la señal discreta y $P(x(i))$ las probabilidades en un histograma para la distribución de x .

#	Parámetro	Estimador
1	Media, (μ_z)	$\sum_{i=1}^T \frac{z_i}{T}$
2	Mediana	$\text{median}(z)$
3	Desviación estándar (SD, del inglés <i>Standard Deviation</i>)	$\left(\sum_{i=1}^T \frac{(z_i - \mu_z)^2}{T} \right)^{1/2}$
4	Raíz media cuadrática (RMS, del inglés <i>Root Mean Square</i>), Y_z	$\left(\sum_{i=1}^T \frac{z_i^2}{T} \right)^{1/2}$
5	Valor pico, (β_z)	$\max_i (z_i)$
6	Factor de forma (SF, del inglés <i>Shape Factor</i>)	$\frac{TY_z}{\sum_{i=1}^T z_i }$
7	Factor de cresta (CF, del inglés <i>Crest Factor</i>)	$\frac{\beta_z}{Y_z}$
8	Factor de impulso (IF, del inglés <i>Impulse Factor</i>)	$\frac{T\beta_z}{\sum_{i=1}^T z_i }$
9	Factor de aclaramiento (CIF, del inglés <i>Clearance Factor</i>)	$\frac{T^{1/2}\beta_z}{\sum_{i=1}^T z_i ^{1/2}}$
10	Asimetría (<i>Skewness</i>)	$\frac{\sum_{i=1}^T \frac{z_i^3}{TY_z^3}}$
11	Valor máximo	$\max_i (z_i)$
12	Valor mínimo	$\min_i (z_i)$
13	Rango de la señal	$ \max_i (z_i) - \min_i (z_i) $
14	Límite superior del histograma (HUB, del inglés <i>Histogram Upper Bin</i>)	$\frac{\max_i (z_i)}{\left(\frac{\max_i (z_i) - \min_i (z_i)}{\text{intervalos} - 1} \right)}$
15	Límite inferior del histograma (HLB, del inglés <i>Histogram Lower Bin</i>)	$\frac{\min_i (z_i)}{\left(\frac{\max_i (z_i) - \min_i (z_i)}{\text{intervalos} - 1} \right)}$
16	Rango intercuartílico (IQR, del inglés <i>Interquartile Range</i>)	$\text{iqr}(z)$
17	Kurtosis	$\frac{\sum_{i=1}^T \frac{z_i^4}{TY_z^4}}$
18	Velocidad de la Kurtosis	$\mathcal{K}\{z'\}$
19	Aceleración de la Kurtosis	$\mathcal{K}\{z''\}$
20	Derivada de la aceleración de la Kurtosis	$\mathcal{K}\{z'''\}$
21	Entropía	$-\sum_{i=1}^T P(x(i)) \ln P(x(i))$

Tabla 1: Características en el dominio del tiempo

8.4. Representación en el dominio de la frecuencia - FD

En el dominio de la frecuencia, se confía en el hecho que un defecto en una parte específica de un sistema, genera una señal periódica con frecuencias características únicas, a forma de firma que la identifica [74]. Esta representación permite identificar y aislar componentes de frecuencia, con lo cual un diagnóstico básico se puede ejecutar con la descomposición de una señal en partes simples, y se puede superar la dificultad del dominio del tiempo en cuanto a suposiciones de estacionariedad para los procesos correlacionados con fallos. Lo anterior se logra ya que cambios en características de frecuencia indican la aparición de fallos, gracias a que cada uno posee una firma de espectro diferente [21].

Las características del dominio de la frecuencia usualmente se obtienen de un análisis espectral empleando FFT. Con el fin de mejorar y enfatizar resultados se pueden emplear técnicas de filtros de frecuencia, demodulación, análisis gráficos, etc. También están disponibles diferentes tipos de espectro de frecuencia derivados desde análisis cepstrum, espectros de orden superior y espectros de potencia. En este último se puede visualizar la distribución de potencia con respecto a la variación frecuencial, para lo cual se puede usar la Transformada Discreta de Fourier (DFT, del inglés *Discrete Fourier Transform*) o el método de la entropía máxima. La tabla 2 presenta las características más relevantes empleadas en el dominio de la frecuencia, donde los parámetros son calculados para los vectores de frecuencia y espectro λ , $s \in \mathbb{R}^K$, respectivamente.

#	Parámetro	Estimador
1	Frecuencia media (MF, del inglés <i>Mean Frequency</i>), μ_s	$\sum_{k=1}^K \frac{s_k}{K}$
2	Frecuencia central (CeF, del inglés <i>Central Frequency</i>), Ξ	$\sum_{k=1}^K \frac{\lambda_k s_k}{K \mu_\lambda}$
3	Raíz Media Cuadrática Frecuencial (RMSF, del inglés <i>Root Mean Square Frequency</i>)	$\Xi^{1/2}$
4	Varianza Frecuencial (VF, del inglés <i>Variance Frequency</i>)	$\sum_{k=1}^K \frac{(s_k - \mu_s)^2}{K \mu_s}$
5	Desviación Estándar Frecuencial (SDF, del inglés <i>Standard Deviation Frequency</i>)	$\left(\sum_{k=1}^K \frac{(\lambda_k - \Xi)^2 s_k}{K \mu_s} \right)^{1/2}$
6	SDF para valores sobre frecuencia central (SOCF, del inglés <i>SDF of values over center frequency</i>)	SDF ($\lambda_k \geq \Xi$)
7	Asimetría Frecuencial (SKF, del inglés <i>Skewness Frequency</i>)	$\sum_{k=1}^K \frac{s_k^3}{K \mu_s^{3/2}}$
8	Kurtosis Frecuencial (KF, del inglés <i>Kurtosis Frequency</i>)	$\sum_{k=1}^K \frac{s_k^4}{K \mu_s^2}$
9	El Rango Intercuartílico Frecuencial (IQRF, del inglés <i>Interquartile Range Frequency</i>).	IQRF (λ)

Tabla 2: Características en el dominio de la frecuencia

8.5. Representación en el dominio tiempo-frecuencia - TFD

Las técnicas tiempo-frecuencia encaran los problemas encontrados en los dominios independientes del tiempo y de la frecuencia sobre señales que no son estacionarias. Las técnicas convencionales de tiempo-frecuencia utilizan distribuciones que representan la energía de las señales en funciones bidimensionales, siendo las más empleadas la Transformada Wavelet (WT, del inglés *Wavelet Transform*), la transformada por MFCC y la Transformada de Fourier de Tiempo Corto (STFT, del inglés *Short-time Fourier Transform*). Mientras STFT proporciona una resolución constante en todas las frecuencias, las transformadas WT y MFCC utilizan técnicas de resolución múltiple donde se analizan diferentes frecuencias con diferentes resoluciones.

Transformada Wavelet - WT

La WT descompone una señal en una combinación lineal de unidades de escala de tiempo, analiza las señales originales y las organiza en varios componentes de señal de acuerdo con la traducción de la wavelet madre, la cual cambia la escala y muestra la transición de cada componente de frecuencia [96]. Esta transformada puede producir resolución de alta frecuencia a bajas frecuencias y alta resolución de tiempo a altas frecuencias, así como también puede reducir el ruido en la señal original [97, 98]. La Transformada Wavelet Continua (CWT, del inglés *Continuous Wavelet Transform*) es una integración con respecto al tiempo total del producto de la señal objetivo $f(t)$ y la wavelet madre $\psi_{a,b}$, lo cual se representa por $CWT(a,b) = \int_{-\infty}^{+\infty} f(t)\psi_{a,b}dt$, donde $\psi_{a,b} = \frac{1}{\sqrt{a}}\psi_{a,b}\left(\frac{t-b}{a}\right)$ con a , b y $\psi_{a,b}$ son los parámetros de escala, de traslación y la wavelet madre, respectivamente. CWT proporciona múltiple resolución para caracterizar señales no estacionarias.

La Transformada Wavelet Discreta (DWT, del inglés *Discrete Wavelet Transform*) se basa en la codificación de sub-bandas, proporcionando un cálculo expedito para la transformación wavelet, ya que es simple de implementar y reduce la complejidad computacional desde los dos puntos de vista más críticos: el tiempo de cómputo y los recursos requeridos. Las funciones de base ortogonal que se emplean para este análisis wavelet son familias de función de escalamiento $\varphi(t)$ y la wavelet asociada $\psi(t)$. Este escalamiento se representa por $\varphi_{j,k}(t) = \sum_{k=1}^K H_k\varphi(2^j t - k)$, donde k y j son los índices de traslación y de escala y H_k es el coeficiente de la función de

escalamiento. Con base en lo anterior, la wavelet asociada se crea empleando los coeficientes de la función de escalamiento, así: $\psi_{j,k}(t) = \sum_{k=1}^K (-1)^k \sqrt{2} h_{l-k} \varphi(2^j t - k)$.

Dados coeficientes conocidos, la función de escalamiento y la wavelet asociada se pueden evaluar iterativamente partiendo del hecho que la señal se puede descomponer en coeficientes aproximados $a_{j,k}$, obtenidos del producto interno de la señal a la escala j y la función de escalamiento, y los coeficientes de detalle $d_{j,k}$ que se pueden lograr del producto interno de la señal y el conjugado complejo de la función wavelet. Luego, la señal original se puede descomponer en diferentes escalas así:

$$f(t) = \sum_{j=-\infty}^{+\infty} a_{j_0,k} \varphi_{j_0,k} + \sum_{j=-\infty}^{j_0} \sum_{k=-\infty}^{+\infty} d_{j,k} \psi_{j,k}(t) \quad (8.1)$$

$$f[n] = \sum_{k=0}^{N-1} a_{(j+1),k} \varphi_{(j+1),k}(t) + \sum_{k=0}^{N-1} d_{(j+1),k} \psi_{(j+1),k}(t) \quad (8.2)$$

Para el siguiente nivel de descomposición, los coeficientes se pueden obtener de la siguiente forma, donde $g[k]$ y $h[k]$ son filtros pasa bajo y pasa alto respectivamente.

$$a_{(j+1),k} = \sum_{k=0}^N a_{j,k} \int \varphi_{j,k}(t) \varphi_{(j+1),k}(t) dt = \sum_{k=0}^N a_{j,k} g[k] \quad (8.3)$$

$$d_{(j+1),k} = \sum_{k=0}^N a_{j,k} \int \varphi_{j,k}(t) \psi_{(j+1),k}(t) dt = \sum_{k=0}^N a_{j,k} h[k] \quad (8.4)$$

Finalmente, los coeficientes de descomposición se hallan por la convolución:

$$y[n] = \sum_{k=1}^N h[k] x[n-k] \quad (8.5)$$

Coeficientes cepstrales en escala MEL - MFCC

Los MFCC fueron introducidos por Davis y Mermelstein en 1980 [99] y desde entonces reemplazaron a otros métodos para obtener coeficientes de caracterización en señales cuasi-estacionarias como los Coeficientes de Predicción Linear (LPC, del inglés *Linear Prediction Coefficients*) y los Coeficientes Cepstrales de Predicción Linear (LPCC, del inglés *Linear Prediction Cepstral Coefficients*). Los MFCC son una representación paramétrica basada en el ancho de banda de la percepción del oído

humano, su obtención se basa en dos tipos de filtros los cuales están espaciados linealmente para frecuencias bajo 1 kHz y con espaciamiento logarítmico sobre dicha frecuencia [100]. A este tipo de espaciamiento se le conoce con el nombre de escala tipo Mel. Los MFCC contienen información de la señal tanto del dominio del tiempo como de la frecuencia y permiten extraer características dinámicas tanto lineales como no lineales, que son el punto en común entre las señales de vibración y de voz [84, 85]. El proceso general de implementación de MFCC se muestra en la figura 8.2 y se describe a continuación [101]:

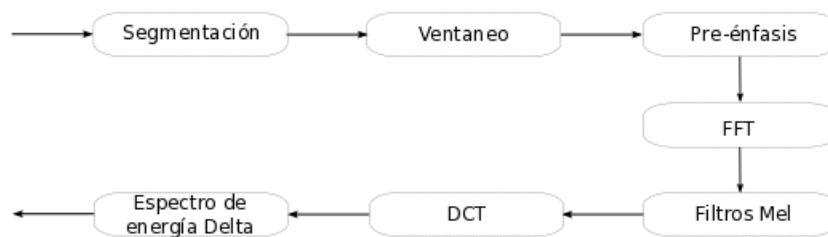


Figura 8.2: Proceso MFCC

- **Segmentación:** divide la señal en marcos de menor tamaño, los cuales se pueden traslapar con su adyacente. Con esta división, se obtienen porciones de señal donde sus características estadísticas sean estacionarias, aunque las características sí pueden cambiar en corto tiempo. Es por esto que, usualmente se toman marcos de 20 a 40 ms.
- **Ventaneo:** usualmente utiliza una ventana tipo Hamming con el fin de ajustar los marcos y permitir integrar todas las líneas cercanas en frecuencia. Una ventana de gran longitud en tiempo entrega baja resolución en tiempo, pero alta resolución en frecuencia; mientras que una ventana de baja longitud en tiempo entrega alta resolución en tiempo, pero baja resolución en frecuencia.
- **Pre-Énfasis:** aplica filtrado que realza las frecuencias altas, incrementando la energía de esas frecuencias.
- **FFT:** Transformada Rápida de Fourier para evaluar el espectro de potencia por segmento, usando $Y = \frac{1}{F} \sum_{n=0}^{F-1} x[n]w[n]e^{-j\frac{2\pi}{F}nm}$, donde F es el número de segmentos, w es la función de ventana Hamming y x es la señal de entrada.

- **Filtros Mel:** la estimación espectral puede contener información no requerida, además de no discernir la diferencia entre dos frecuencias estrechamente espaciadas. Este efecto se vuelve más pronunciado a medida que aumentan las frecuencias. Por esta razón, se toman filtros que poseen una respuesta en frecuencia triangular y se encuentran espaciados de acuerdo con la escala Mel (linealmente separados bajo los 1000 Hz y con espaciamiento logarítmico sobre esta frecuencia) con el fin de tener una idea de cuanta energía existe en diferentes regiones de frecuencia. Con estos filtros, la magnitud de la respuesta es igual a la unidad en la frecuencia central y decrece linealmente a cero en el centro de las dos frecuencias de los filtros adyacentes. $E[m] = \sum_{k=1}^{K-1} \ln [Y[k]^2 H_m[k]]$ es la función para implementar los filtros, donde m representa el número de filtros y H_m es su función de transferencia respectiva [101].
- **DCT:** Transformada Discreta del Coseno (DCT, del inglés *Discrete Cosine Transform*) que convierte el espectro Mel al dominio del tiempo. El resultado de esta conversión es lo que se denomina MFCC. Esto se realiza principalmente debido a que los filtros se superponen y las energías del banco de filtros estarán correlacionadas entre sí. La DCT decorrelaciona estas energías, por lo que las matrices de covarianza diagonal se pueden usar para modelar características en un clasificador. En un punto extremo, se tienen tantos coeficientes cepstrales como número de filtros en el banco, lo cual corresponde a la captura de energía alternando entre cada uno de los bancos de filtros. Producto de esto, la DCT también sirve para reducir el número de coeficientes al juntar aquellos que representan cambios rápidos de energía. Una forma de computar la DCT es $C_i = \sqrt{\frac{2}{M}} \sum_{m=1}^M E_m \cos \left(\frac{i*\pi}{M} \left(m - \frac{1}{2} \right) \right)$.
- **Espectro de Energía Delta:** se adicionan 13 deltas de velocidad y 39 doble deltas, o aceleraciones, ya que los marcos cambian sus características a medida que se traslapan en su evolución, dada la calidad cuasi-estacionaria de las señales que se supone intervienen.

Transformada de Fourier de Tiempo Corto - STFT

La transformada de Fourier de tiempo corto es básicamente la misma transformada de Fourier, pero donde la señal original se divide en pequeños segmentos, o marcos, donde se puede asumir la estacionariedad de la misma. Aplicar una DFT en una larga ventana de observación no permite revelar las transiciones en el contenido

espectral, por lo cual se puede recurrir a aplicar una DFT en periodos cortos de tiempo. Sin embargo, continúa existiendo una relación entre el tiempo y la frecuencia. Al igual que con la transformada MFCC, se requiere implementar una función de ventana con el fin de ajustar los marcos e integrar todas las líneas cercanas en frecuencia. Igualmente, estos marcos ventaneados se pueden traslapar con el fin de incrementar los espectros de frecuencia [102].

8.6. Modelo Oculto de Markov (HMM) con observación discreta

Una cadena de Markov es todo sistema que para un instante de tiempo se puede considerar en uno de un conjunto discreto de estados N , notados como $\{S_1, S_2, \dots, S_N\}$. A tiempos discretos (t), espaciados uniformemente, el sistema realiza transiciones entre estados (q) de acuerdo con un conjunto de probabilidades de transición, estando en un tiempo t en el estado q_t . La cadena es de primer orden si su probabilidad de transición solo depende del estado previo, o: $P\{q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \dots\} = P\{q_t = S_j | q_{t-1} = S_i\} = a_{i,j}$ y donde $a_{i,j} > 0$ y que $\sum a_{i,j} = 1$.

El modelo descrito previamente se conoce como una cadena de Markov observable, ya que en cada instante la salida del proceso es el mismo conjunto de estados. Un HMM es una extensión del modelo observable, donde las salidas son una función probabilística del estado, y por tanto, el modelo es un doble proceso estocástico embebido que no es directamente observable (su estado) pero sí indirectamente a través del conjunto de secuencias de salida [103, 104]. Un HMM se caracteriza por:

- N : número de estados del modelo.
- M : número de diferentes símbolos observables en los estados. El conjunto de símbolos se nota: $V = \{v_1, v_2, \dots, v_M\}$.
- $A = [a_{i,j}]$: distribución de probabilidad de transición entre estados.
- $B = [b_j(k)]$: distribución de probabilidad de observación de emisiones (símbolos) en el estado j , y donde $b_j(k) = P\{v_k \text{ en } t | q_t = S_j\}$.
- $\pi = [\pi_i]$: distribución de probabilidad de estado inicial, con $\pi_i = P\{q_1 = S_i\}$.
- O : secuencia de observación, donde $O = \{O_1, O_2, \dots, O_T\}$, con O_t uno de los símbolos desde V .
- λ : conjunto de parámetros para el modelo HMC, donde $\lambda = \{A, B, \pi\}$.

Para entrenar modelos HMM, los parámetros $\lambda = \{A, B, \pi\}$ se deben ajustar con el fin de maximizar la probabilidad de unas observaciones dados los parámetros del modelo, o $P\{O|\lambda\}$. Para lograr este propósito se emplea el algoritmo Esperanza-Maximización (EM, del inglés *Expectation Maximization*), el cual se describe en detalle en [105, 106], donde la probabilidad de estar en el estado i en el instante t y la probabilidad de estar en el estado j en el instante $t + 1$, o probabilidad de transición, se describe como sigue:

$$\xi_t(i, j) = \frac{P\{q_t = i, q_{t+1} = j | \mathbf{O}, \lambda\}}{P\{\mathbf{O}, \lambda\}} \quad (8.6)$$

$\xi_t(i, j)$ se puede reformular como función de dos nuevas variables $\alpha_t(j)$ y $\beta_t(i)$, así:

$$\xi_t(i, j) = \frac{a_t(i) a_{i,j} b_j(\mathbf{O}_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{i,j} b_j(\mathbf{O}_{t+1}) \beta_{t+1}(j)} \quad (8.7)$$

en donde, la re-estimación de los parámetros del modelo está definida por:

$$\pi'_i = \sum_{i=1}^N \xi_t(i, j) \quad (8.8)$$

$$a'_{ij} = \frac{\sum_{t=1}^{T-1} a_t(i) a_{i,j} b_j(\mathbf{O}_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j=1}^N \alpha_t(i) a_{i,j} b_j(\mathbf{O}_{t+1}) \beta_{t+1}(j)} \quad (8.9)$$

$$b'_j(k) = \frac{\sum_{t=1}^T \sum_{i=1}^N \sum_{t=1}^{T-1} a_t(j) a_{i,j} b_j(\mathbf{O}_{t+1}) (\beta_t = v_k)}{\sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N \alpha_t(j) a_{i,j} b_j(\mathbf{O}_{t+1}) \beta_t \mathbf{O}_t} \quad (8.10)$$

La re-estimación se realiza hasta que se obtenga $P\{O|\lambda\}$, esto es, mientras que un nuevo modelo explique de forma más probable la secuencia de observaciones. También se puede fijar un número máximo de iteraciones, o un proceso de restricción al número de repeticiones donde no se obtenga mejora en $P\{O|\lambda\}$.

8.7. Modelo Oculto de Markov (HMM) con observación continua

Una extensión del modelo HMM ocurre cuando las observaciones pueden tomar valores en escala continua con el fin de representar W segmentos de características en \mathbf{X}_n . En consecuencia, el uso de los estados ocultos permite que el modelo de clasificación sea lo suficientemente genérico con el fin de tratar con clases de fallos y severidades en el diagnóstico y pronóstico de estados de salud en rodamientos [30, 107, 108]. Así, dada una matriz de secuencia de observación $\tilde{\mathbf{X}} \in \mathbb{R}^{W \times M'}$ la cual puede contener todas las características de los dominios de representación ($M' = M$), o ser una matriz donde se han seleccionado características ($M' < M$) según algún criterio definido en un vector de relevancia $\rho \in \mathbb{R}^M$, se puede construir un modelo HMM con observación continua según el conjunto de parámetros $\Psi = \{\Pi, \mathbf{b}, \mathbf{p}_o\}$ y el conjunto de estados $\mathcal{S} = \{\theta_v \in \mathbb{R}^{M'}\}_{v=1}^V$. La matriz $\Pi \in [0, 1]^{V \times V}$ contiene la probabilidad de transición de estados con $\Pi_{vv'} = P(\theta_{j+1} = v' | \theta_j = v)$; donde $v, v', j = \{1, 2, \dots, V\}$ y $P(\cdot | \cdot)$ es una función de distribución de probabilidad condicional, y $\mathbf{B} = \{b_v(\cdot)\}_{v=1}^V$ es una función de probabilidad de salida definida como sigue:

$$b_v(\tilde{x}_w) = \sum_{u=1}^U \omega_{vu} \mathcal{N}(\tilde{x}_w | \vartheta_{vu}, \Sigma_{vu}), \quad (8.11)$$

donde $\tilde{x}_w \in \mathbb{R}^{M'}$ es la w -ésima observación en $\tilde{\mathbf{X}}$, U es el número de mezclas de gaussianas que modelan la distribución de probabilidad, $\omega_{vu} \in [0, 1]$ son los pesos de la mezcla, $\vartheta \in \mathbb{R}^{M'}$ y $\Sigma_{vu} \in \mathbb{R}^{M' \times M'}$ son el vector de medias y la matriz de covarianzas para el estado v y la mezcla u , respectivamente, y $\mathcal{N}(\cdot)$ denota una distribución normal. Además, $\mathbf{p}_o \in [0, 1]^V$ es el vector de probabilidad de estado inicial.

Una vez definido un modelo HMM de observación continua, este se puede entrenar para cada clase de fallo y severidad presente en las señales de vibración en rodamientos. Por tanto, un conjunto de modelos $\Lambda = \{\hat{\Psi}_c : c \in \{1, \dots, C\}\}$ se calculan mediante el estimador de Máxima Verosimilitud (ML, del inglés *Maximum-Likelihood*), como sigue:

$$\hat{\Psi}_c = \arg \max_{\Psi} P(\tilde{\mathbf{X}}_n | \Psi); \quad \forall \{\tilde{\mathbf{X}}_n : y_n = c\}. \quad (8.12)$$

Finalmente, dada una nueva matriz con segmentos de características $\tilde{\mathbf{X}}_* \in \mathbb{R}^{W \times M'}$, la regla de clasificación del modelo HMM con observación continua está dada por:

$$\hat{c}_* = \arg \max_c P(\tilde{\mathbf{X}}_* | \Psi_c). \quad (8.13)$$

En la práctica, el problema de optimización presente en la ecuación 8.12 se puede resolver empleando un enfoque con base en EM [109].

8.8. Modelo Oculto de Markov Jerárquico (HHMM)

Los Modelos Ocultos de Markov Jerárquicos (HHMM, del inglés *Hierarchical Hidden Markov Models*) son procesos estocásticos estructurados en forma de múltiples niveles. Los HHMM generalizan los HMM haciendo que cada uno de los estados ocultos sea por sí mismo un modelo probabilístico independiente, por lo que los HHMM son un HMM donde cada estado es también a su vez un HMM. También se puede ver como un encadenamiento jerarquizado de HMM, donde una capa de nivel mayor HMM posee a su vez un HMM como parte de sus estados ocultos [110, 111].

El encadenamiento jerarquizado ocasiona que en un HHMM se emitan secuencias de símbolos mediante una activación recursiva de uno de los sub-estados contenidos en un estado. Estos sub-estados también pueden estar compuestos por otros sub-estados. Este proceso de activación recursiva termina cuando se llega al estado de menor nivel, o estado de producción, que son los únicos estados que efectivamente emiten símbolos de salida a través de la forma normal de emisión HMM. Los demás estados que no emiten símbolos se llaman estados internos.

La activación de un sub-estado por un estado interno de mayor nivel se denomina transición vertical, la cual puede incluir transiciones verticales adicionales a estados de menor nivel. Al completar una transición vertical el sistema vuelve al estado inicial de mayor nivel donde se originó la activación recursiva. Luego, se realiza una transición de estado dentro del mismo nivel, que se denomina transición horizontal. El conjunto descrito de estados y las transiciones verticales producen una estructura de árbol donde el estado raíz es el nodo en la parte superior de la jerarquía y las hojas son los estados de producción.

Con el fin de simplificar la descripción, se restringe la estructura HHMM a una estructura de árbol subyacente completa, esto es, todas las hojas están a la misma distancia del estado raíz. Un HHMM es una tripleta $HHMC = \{\eta, \zeta, \Sigma\}$ formada por: los parámetros del modelo η , la estructura topológica del modelo ζ y el alfabeto de observaciones Σ [110, 111]. La estructura topológica ζ determina los niveles jerárquicos $D \in \mathbb{N}$ y el número de estados $Q \in \mathbb{N}$ por cada nivel, o $\zeta = \{\forall d_i : \exists Q_i^d \mid d_i \in 1 \dots D\}$. Se debe notar que si el número de estados es igual en todos

los niveles, entonces el subíndice i se puede omitir en \mathcal{Q}_i^d . Ahora, los parámetros del modelo $\eta = \{\mathbb{A}, \mathbb{B}, \Pi\}$ están formados por: la matriz de transición horizontal \mathbb{A} , la distribución de probabilidad de observaciones \mathbb{B} y el vector de transición vertical Π .

La matriz de transición horizontal entrega las probabilidades de transición entre nodos que pertenecen a un mismo nivel y el vector de transición vertical codifica las probabilidades de transición de nodos de mayor nivel a nodos de menor nivel [111]. Las anteriores matrices se definen formalmente como:

$$\mathbb{A} = \left\{ \forall \mathbb{A}^{q_i^d} = \left(a_{ij}^{q_i^d} \right) \mid a_{ij}^{q_i^d} = P \left(q_j^{d+1} \mid q_i^{d+1} \right) \right\} \quad (8.14)$$

$$\mathbb{B} = \left\{ \forall \mathbb{B}^{q_i^D} = \left(b^{q_i^D}(k) \right) \mid b^{q_i^D}(k) = P \left(\sigma_k \mid q^D \right), \sigma_k \in \Sigma \right\} \quad (8.15)$$

$$\Pi = \left\{ \forall \Pi^{q_i^d} = \pi^{q_i^d} \left(q_i^{d+1} \right) \mid \pi^{q_i^d} \left(q_i^{d+1} \right) = P \left(q_i^{d+1} \mid q^d \right) \right\} \quad (8.16)$$

Para entrenar un modelo HHMM se emplea la estimación de parámetros por ML, el cual es una generalización del conocido algoritmo de Baum-Welch, donde también se debe considerar las transiciones verticales que generan las observaciones cuando llegan a un estado de producción. Por lo tanto, además de las variables α y β que corresponden a las transiciones horizontales forward y backward, se agregan variables adicionales para las transiciones verticales 'hacia abajo' y 'hacia arriba'.

Las variables utilizadas para describir una probabilidad de transición se resumen en $\xi^J(t, q_i^d, q_j^d, q^{d-1})$, la cual es la probabilidad de realizar una transición horizontal desde el estado q_i^d hasta el estado q_j^d , donde ambos estados son sub-estados de q^{d-1} en el instante t después de la observación \mathbf{O}_t y antes de la observación \mathbf{O}_{t+1} . Igual que para el entrenamiento HMM, se definen dos variables auxiliares para simplificar la re-estimación: $\gamma_{in}(t, q_i^d, q^{d-1})$ que es la probabilidad de realizar una transición horizontal al estado q_i^d antes de generar \mathbf{O}_t ; y $\gamma_{out}(t, q_i^d, q^{d-1})$ que es la probabilidad de dejar el estado q_i^d mediante una transición horizontal a cualquier estado d del mismo nivel después de generar \mathbf{O}_t . Estas variables auxiliares se pueden escribir así:

$$\gamma_{in}(t, q_i^d, q^{d-1}) = \sum_{k=1}^{|q^{d-1}|} \xi^J(t-1, q_k^d, q_i^d, q^{d-1}) \quad (8.17)$$

$$\gamma_{out}(t, q_i^d, q^{d-1}) = \sum_{k=1}^{|q^{d-1}|} \zeta^J(t-1, q_k^d, q_i^d, q^{d-1}) \quad (8.18)$$

La variable para estimar la probabilidad de una transición vertical es $\chi(t, q_i^d, q^{d-1})$ la cual expresa la probabilidad que el estado q^{d-1} sea accedido en el instante t antes que se genere la observación \mathbf{O}_t , e inicialmente activado el estado q_i^d . En [110, 111] se puede consultar los pasos EM requeridos para la evaluación de las variables, que arroja como resultado el conjunto de nuevos parámetros re-estimados:

$$\hat{\pi}^{q^1}(q_i^2) = \chi(t, q_i^2, q^1) \quad (8.19)$$

$$\hat{\pi}^{q^{d-1}}(q_i^d) = \frac{\sum_{t=1}^T \chi(t, q_i^d, q^{d-1})}{\sum_{t=1}^{|q^{d-1}|} \sum_{t=1}^T \chi(t, q_i^d, q^{d-1})} \quad (2 < d < D) \quad (8.20)$$

$$\hat{a}_{ij}^{q^{d-1}} = \frac{\sum_{t=1}^T \zeta^J(t, q_i^d, q_j^d, q^{d-1})}{\sum_{t=1}^T \gamma_{out}(t, q_i^d, q^{d-1})} \quad (8.21)$$

$$\hat{b}_{q_i^D}^{q^{D-1}}(v_k) = \frac{\sum_{o_t=v_k} \chi(t, q_i^D, q^{D-1}) + \sum_{t>1, o_t=v_k} \gamma_{in}(t, q_i^D, q^{D-1})}{\sum_{t=1}^T \chi(t, q_i^D, q^{D-1}) + \sum_{t=2}^T \gamma_{in}(t, q_i^D, q^{D-1})} \quad (8.22)$$

La re-estimación de parámetros HHMM se realiza en similitud con el procedimiento descrito para la re-estimación HMC, y corresponde a los pasos del algoritmo EM. Por lo tanto, se garantiza que este procedimiento converge a un punto estacionario (probablemente un máximo local) de la función de verosimilitud.

Parte III

MATERIALES Y MÉTODOS

9. PRONÓSTICO DE VIDA ÚTIL REMANENTE (RUL)

La estimación de vida útil remanente es la salida de un modelo predictivo que acopla información proveniente del estado de degradación, del perfil actual de carga y de las condiciones operacionales, aunque también puede incorporar consideraciones ambientales presente, pasadas y futuras. Este modelo también puede considerar información relacionada con información experta sobre el comportamiento de un sistema [112]. Desde esta perspectiva, un modelo de estimación del RUL debe medir el estado actual de degradación, estimar las condiciones operacionales y de carga y predecir mediante simulación basada en la experiencia la vida útil remanente del sistema.

Como punto inicial en el proceso de solución al problema de estimación del RUL está la clasificación para el problema de diagnóstico. Para cada modo distinguible de fallo, su origen puede provenir de tres posibles fuentes: desgaste gradual, eventos discretos predecibles y eventos discretos no predecibles [113]. Para evaluar el problema de clasificación del diagnóstico, se debe tener presente las necesidades de evaluación en línea, en tiempo real o fuera de línea. Estas necesidades afectan los métodos de adquisición de datos y las técnicas y metodologías planeadas para evaluar la degradación, clasificar modos distinguibles de fallos y determinar el RUL. Sin embargo, salvo aplicaciones con altos requerimientos, el objetivo del mantenimiento predictivo es realizar evaluaciones a intervalos de tiempo más cortos en relación con las dinámicas de un sistema. Para el caso particular de evaluar la degradación e inferir el estado futuro de rodamientos con base en vibraciones, es razonable pensar que incluso un procesamiento fuera de línea permite acometer estos objetivos planteados [112]. Por lo tanto, el diagnóstico de fallos en rodamientos aparece como una estrategia para mitigar los desafíos actuales de la industria, donde debe prevalecer la detección de estados anormales en sus primeras etapas con el fin de evitar el desarrollo de fallos mayores [114, 115].

El modelo general para la evaluación del RUL se puede clasificar en técnicas mediante enfoques basados en modelos físicos, basados en señales o basados en conocimiento. Las técnicas basadas en modelos físicos emplean modelos y ecuaciones matemáticas con el fin de modelar, predecir y simular cambios en el comportamiento de un sistema, sin embargo se distinguen por ser intensivas en tiempo, recursos computacionales y por depender de modelos deterministas o probabilísticos de procesos que realmente son estocásticos por naturaleza [52]. Las técnicas basadas en señales utilizan datos adquiridos desde el sistema, logrando indicar cambios de de-

gradación en presencia de variaciones de carga; sin embargo, aunque estos acercamientos proporcionan una forma realista de evaluar fallos en lugar de usar modelos, dependen de un análisis de síntomas de estados normales que es información basada en un experto y además presentan dificultades al clasificar la naturaleza de los cambios [59, 113]. Por el contrario, los enfoques basados en conocimiento abordan el problema del diagnóstico desde una perspectiva de reconocimiento de patrones, que comprende una etapa de adquisición de señal, un procedimiento de estimación de características y una evaluación de fallos a través de algoritmos de aprendizaje [60].

Para los modelos basados en señales o basados en conocimiento, se requieren datos provenientes de sensores con adquisición de datos no invasiva [61], no obstante, las señales de vibración de rodamientos son de naturaleza cuasi-estacionaria o no estacionaria, variando los patrones de diagnóstico de fallos, incluso cuando las señales provienen de supuestamente equipos similares, y hasta del mismo equipo [62]. Además, los sistemas de adquisición de datos que utilizan señales acústicas también han incrementado su uso sobre otras técnicas de medición sin contacto, cuando las ubicaciones de los sensores en la máquina no están disponibles o el procedimiento de medición tiene un alto riesgo para los trabajadores [64]. No obstante, las señales acústicas son más vulnerables al ruido ambiental que las respuestas de vibración [63]. Por lo tanto, el reconocimiento de patrones y la estimación de características a partir de señales debe comprender parámetros basados en los dominios de representación del tiempo, de la frecuencia y del tiempo-frecuencia [21]. Los enfoques clásicos basados en el dominio del tiempo se centran en medidas estadísticas orientadas a procesos estacionarios, que no reflejan adecuadamente el cambio en los componentes de frecuencia a medida que se desarrolla la progresión de una fallo [93]. Las características basadas en frecuencia, como la FFT y el espectro de potencia, permiten superar en parte la restricción del anterior dominio, sin embargo, se ven afectadas en la capacidad de detectar señales de pequeña amplitud, como es el caso de fallos incipientes, y arrojan amplitudes armónicas de la frecuencia del defecto ocultas en ruido. Nuevamente, estas técnicas en frecuencia pueden ser inadecuadas para identificar eventos no estacionarios [21]. Las características de tiempo-frecuencia, como la WT, son conocidas por tratar patrones no estacionarios, pero solo pueden detectar características de señal que coinciden con la forma de la ondícula (*wavelet*), mientras que todos los demás parámetros se enmascaran o se ignoran [20].

Para la etapa de aprendizaje, los enfoques basados en conocimiento conducen diferentes estrategias de agrupamiento no supervisadas / supervisadas, como HMM

[20], ANN [78], GA [81], SVM [82], etc. Además, el análisis de las características extraídas generalmente recae en un solo dominio: tiempo, frecuencia o tiempo-frecuencia, sin considerar la contribución de cada uno de estos dominios al diagnóstico de fallos y al proceso de degradación (gravedad de fallos) [21, 77]. Técnicas de extracción y reducción de características se emplean antes de realizar agrupamiento con el fin de reducir dimensionalidad. El objetivo principal de este tipo de métodos es reducir el número de características para evitar redundancia manteniendo el mayor poder discriminante posible [116]. Se han explorado algunos métodos para reducir dimensionalidad, como el Análisis de Componentes Principales (PCA, del inglés *Principal Component Analysis*) [117] y el Análisis de Componentes Principales por Kernel (KPCA, del inglés *Kernel Principal Component Analysis*) [118]. Sin embargo, el nuevo espacio de características obtenido con este tipo de técnicas generalmente pierde el significado original de ingeniería y muchas veces retira el poder discriminante requerido para la tarea de pronóstico. Por lo tanto, otros métodos de reducción, como selección de características, son más recomendables. En la literatura se encuentran muchas técnicas para seleccionar características relevantes, como Proyección Preservando Localidad (LPP, del inglés *Locality Preserving Projection*) [118], la Técnica de Evaluación de Distancia (DET, del inglés *Distance Evaluation Technique*) [24, 76], Matrices de Dispersión (SM, del inglés *Scatter Matrices*) [25], Auto-Censurado (SW, del inglés *Self-Weigth*) [119] y Puntuación Laplaciana (LS, del inglés *Laplacian Score*) [120]. Sin embargo, durante las tareas de diagnóstico y pronóstico de fallos en rodamientos, la selección de parámetros relevantes para mejorar la precisión de la detección en etapas posteriores de aprendizaje sigue siendo un tema abierto.

10. MARCO METODOLÓGICO

10.1. RETOS

El diagnóstico de fallos cubre la investigación y análisis de la causa o la naturaleza de una avería, mientras que un pronóstico eficiente se centra en información sobre localización y severidad, siguiendo un proceso de detección de ocurrencia, es decir, es la predicción de la progresión de una anomalía diagnosticada, lo cual finalmente se conoce como RUL. Una aplicación efectiva de un pronóstico de fallo en una máquina puede conducir a disminuir o eliminar tiempos de inactividad no programados y aumentar la vida útil y, en consecuencia, entregar una reducción en los costos de mantenimiento y de factores de seguridad humana y medioambiental.

Para establecer un modelo de pronóstico que estime eficazmente el RUL, la revisión de la literatura indica los siguientes retos de investigación:

Modelos y naturaleza de las señales

Se requiere de un modelo para la implementación del pronóstico de vida útil remanente que permita de forma eficaz enfrentar los retos de tratar con sistemas de naturaleza compleja, sin llegar a las suposiciones típicamente existentes de estacionariedad o simplificación estocástica de los sistemas. En este aspecto, además se debe permitir reconocer la naturaleza cuasi-estacionaria de las señales involucradas, donde sus parámetros varían con el tiempo. En estos escenarios es fundamental tratar con señales que impliquen técnicas no invasivas de adquisición de datos.

Deficiencias en diccionarios extendidos

Los procesos normales de diagnóstico y pronóstico encontrados en la literatura se enfocan en un único dominio de representación de las señales, a saber: tiempo, frecuencia y tiempo-frecuencia. Es usual encontrar estudios en la reducción de dimensionalidad enfocados primordialmente en retener el poder discriminante de las características. Sin embargo, se hace necesario considerar el estudio de la contribución de cada uno de estos dominios, de forma individual y conjunta, enfocando el análisis en la selección de características multi-dominio sin sacrificar el poder discriminante restante necesario para afrontar el reto del pronóstico.

Modos de fallos y severidad

Los modelos de diagnóstico normalmente solo responden a procesos con características similares, por lo que hasta ahora se enfocan en sistemas teóricos o restringidos a modos de fallos típicos. Los modelos estrictos para pronóstico están actualmente aún más restringidos, llegándose a confundir con la tarea de diagnóstico y observándose claramente que no enfrentan el reto de distinguir, fuera de los modos de fallos presentes, los niveles de severidad de los mismos, lo cual se constituye en la pieza clave para realizar una tarea de pronóstico.

Aprendizaje de máquina para sistemas aleatorios dinámicos

Los métodos tradicionales para evaluar el RUL recaen en el conocimiento previo y en la frecuencia relativa de ocurrencia de los defectos. Estos enfoques, se parecen

a las estructuras de los dobles procesos estocásticos donde se puede determinar el estado de degradación de un sistema gracias a la probabilidad de transición entre estados. De forma amplia, estos modelos se asemejan a los empleados en el procesamiento de habla, ya que ambos tipos de señales son cuasi-estacionarias por naturaleza y en ambos casos existen variaciones en las características descriptivas de una señal proveniente de sistemas diferentes que describen el mismo fonema, en el caso del habla, o el mismo estado de fallo y degradación, en el caso de sistemas físicos.

Implementación de un índice de degradación extensible

Se hace necesario generalizar la forma en la cual, o desde la cual, se pueden generar diversos índices de degradación aplicables en diversos tipos de sistemas. Para este enfoque, un modelo de pronóstico debe implementar todo el proceso descriptivo del tipo de modo de fallo presente, además de determinar los niveles de degradación con el fin de poder vislumbrar un resultado de pronóstico.

10.2. MODELO DE PRONÓSTICO

El modelo de enfoque metodológico que se propone consiste en la implementación de un sistema integrado de diagnóstico y pronóstico con base en la estimación de probabilidad del estado de degradación, o del tipo de fallo y nivel de severidad, presente. Para obtener la mejor predicción posible sobre la vida útil remanente de una máquina, el modelo que se propone consta de tres subsistemas, a saber: conocimiento histórico experto, diagnóstico y pronóstico. Todo el ciclo incluye monitoreo de condición, clasificación de fallos inminentes, estimación de estado de salud y pronóstico, con el fin de realizar una predicción de largo plazo.

La integración del conocimiento histórico experto se centra en conocer a priori los modelos de fallo más representativos, identificar los principales patrones de fallo que afectan toda la vida útil de la máquina para diagnóstico y pronóstico, además, también determinar las técnicas apropiadas de procesamiento de señales y de extracción de características.

En el módulo de diagnóstico se debe entregar información fiable sobre el patrón, o características, del fallo que se desarrolla desde etapa incipiente. Este módulo se comporta como un sistema de caracterización empleando múltiples dominios de representación, con extracción y selección de características de mayor poder discriminante, no solo desde el punto de vista del diagnóstico, sino también en preparación para la etapa de pronóstico.

La descripción discreta de características predeterminada en la etapa de diagnóstico se emplea luego como insumo para el módulo de pronóstico, donde se debe estimar el estado de la máquina. El resultado final del módulo de pronóstico es una información conjunta de tipo de fallo y nivel de severidad, o degradación, que arroja una perspectiva de la vida útil remanente. Este proceso se realiza mediante técnicas de aprendizaje de máquina, y en particular mediante el uso de Modelos Ocultos de Markov y Modelos Ocultos de Markov Jerárquicos. Bajo estos enfoques de aprendizaje por HMM o HHMM, se emplea entrenamiento hacia adelante (forward) y hacia atrás (backward), con el fin de implementar estrategias de selección de características multi-dominio, y en particular se analizan los dominios con el fin de decodificar desde las características el diagnóstico de fallo (o identificación del modo de fallo) y la severidad (o niveles de degradación).

Con el anterior enfoque, se determina un conjunto de características que incrementen el desempeño discriminante a nivel de diagnóstico y de severidad, lo cual arroja un procedimiento de análisis de relevancia multi-dominio sobre múltiples estados de salud para rodamientos. El proceso de análisis de relevancia se realiza de forma conjunta entre la selección de características, dominios de representación y resultados discriminantes del modelo de aprendizaje. Aunque muchas técnicas de análisis de relevancia se han estudiado, tales como LPP [118], DET [24, 76], SM [25], SW [119] y LS [120], este campo de investigación se considera un área abierta cuando se trata de seleccionar parámetros relevantes que ayuden en la exactitud de las tareas simultáneas de diagnóstico y severidad. Para dar soporte al anterior reto, se presenta una metodología de análisis de relevancia basada en el método Relief-F.

Todo el marco metodológico se prueba sobre una base de datos debidamente soportada y de amplia aceptación, con el fin de realizar evaluación de diagnóstico y severidad (pronóstico), y dando como resultado que la estrategia metodológica propuesta iguala, o supera, desempeños de metodologías reportadas en el estado del arte relacionado con la tarea de pronóstico.

El detalle de los tres módulos descritos se muestra en la Figura 10.1.

10.3. BASE DE DATOS

Con el fin de evaluar la conveniencia del enfoque metodológico propuesto, se emplea una de las base de datos experimentales con mayor reconocimiento y uso frecuente sobre fallos en rodamientos. Esta base de datos es provista por el Bearing Data Center de la Case Western Reserve University [121], donde las señales que con-

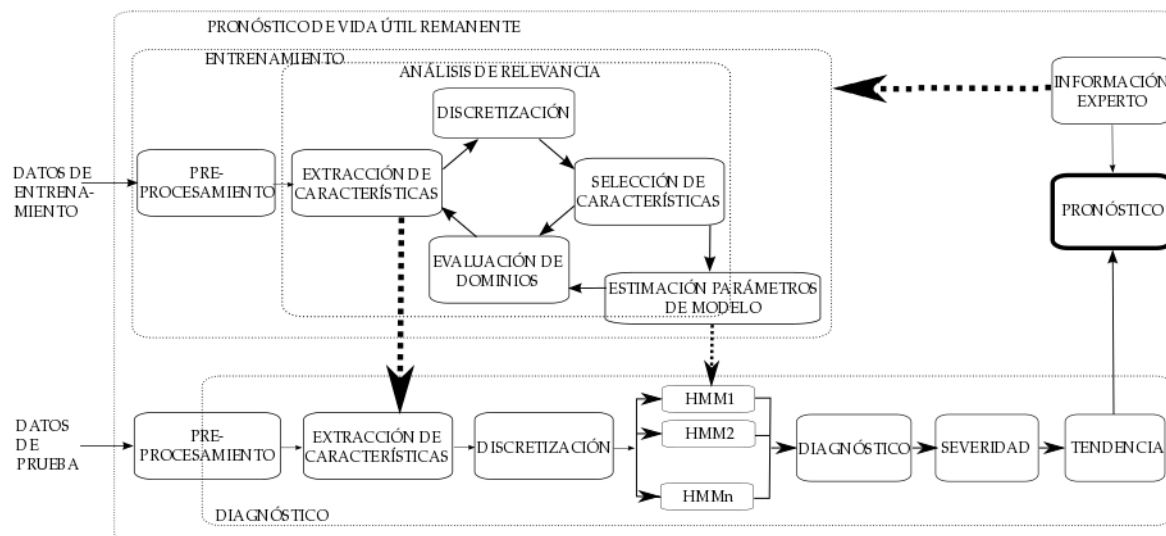


Figura 10.1: Modelo de pronóstico

forman los datos fueron adquiridas usando un motor trifásico de inducción con 2 HP de potencia. El motor se conectó a un dinamómetro y a un sensor de par mediante un acoplamiento de auto alineación. Los acelerómetros se ubicaron en la carcasa con bases magnéticas tanto en el extremo del accionamiento como en del ventilador. A los rodamientos se les indujeron fallos artificiales mediante Mecanizado por Electro-Descarga (EDM, del inglés *Electro-Discharge Machining*) en la pista interior de rodamiento, en la pista exterior de rodamiento y en los elementos rodantes, induciendo tres niveles de severidad, según el diámetro de la EDM así: 0.007" (N1), 0.014" (N2) y 0.021" (N3), todos con una profundidad de 0.011024". los fallos descritos se han etiquetado respectivamente como Inner race (I, del inglés Inner), Outer race (O, del inglés Outer) y Rolling Element (B, del inglés Ball); además se registran datos para el estado base correspondiente a un sistema fin fallo, o estado Normal.

Los rodamientos con fallo inducido se reinstalaron nuevamente en el motor de prueba y se registraron datos de vibración para cargas del motor de 0 a 3 HP, diferenciando cuatro velocidades de operación: 1730 RPM, 1750 RPM, 1730 RPM y 1797 RPM. La carga se aplicó utilizando un dinamómetro y los datos de velocidad y potencia se recolectaron mediante un transductor. Las señales de vibración se adquirieron para frecuencias de muestreo de 12 kHz y 48 kHz. Como consecuencia del proceso de adquisición, se dispone de un total de 56,132 señales con un intervalo de

tiempo de 10 s, aproximadamente, cada una. La anotación de la base datos, fuera de tener presente el tipo de fallo inducida, indica el nivel de severidad presente y la velocidad de operación.

Para indagar a profundidad todos los parámetros de la base de datos, se puede consultar [\[121\]](#).

Parte IV

RESULTADOS, ANÁLISIS Y DISCUSIÓN

11. ANÁLISIS DE DEPENDENCIA, VARIABILIDAD Y CONGLOMERADOS

Dentro del problema general de clasificación, identificación de la degradación y pronóstico en rodamientos, es necesario realizar análisis de la dependencia, variabilidad y distancia entre un conjunto de características en un dominio de representación, con el fin de plantear la fundamentación que permita estudiar si es posible eliminar variables altamente correlacionadas o describir nuevos conjuntos de características mínimas de representación, además de indagar la posibilidad de separación de clases de fallos (diagnóstico) mediante solo este conjunto de características.

Una posible reducción del dominio de características debe permitir la disminución de carga computacional en algoritmos que gestionan el riesgo residual de equipos, además esto está directamente vinculado con la cantidad de información para entrenamiento, las características obtenidas y los modelos de evaluación de la severidad empleados, que normalmente están relacionados con modelos que tratan con la dinámica variante de sistemas bajo observación. Por tanto, dentro de los puntos a considerar inicialmente, es crucial determinar la posibilidad de reducir la dimensionalidad con el fin de lograr tener un nuevo conjunto de características que retenga la mayor cantidad posible de información contenida en las variables iniciales. En este contexto, PCA permite realizar análisis de relevancia y reducción de dimensionalidad sobre las señales y características e indagar la posibilidad de separación de clases con estas nuevas variables.

Las observaciones, o características, derivadas de variables en los dominios del tiempo, la frecuencia y del tiempo-frecuencia, se distinguen por estar definidas en un rango continuo, y por ende pueden tomar cualquier valor en su espacio de representación. Fuera de la necesidad de disminuir el tamaño de las observaciones, su agrupamiento es esencial para el entrenamiento de modelos que así lo requieran. Una metodología de agrupamiento es k -medias, la cual se fundamenta en encontrar k centroides para los datos de forma tal que se minimice la suma del cuadrado de las distancias entre cada observación y su centroide. Otra forma de agrupamiento es el método de Ward, donde los grupos son formados directamente desde los elementos. El número de intervalos, o centroides, también es de mayor importancia ya que un número bajo de ellos le resta representatividad de las muestras y un número alto introduce complejidad innecesaria. Es por ello que, desde el problema general de clasificación e identificación de la severidad de fallos en rodamientos, se

desea analizar la discretización de observaciones y determinar el número óptimo de conglomerados que permiten representar de forma discreta la escala continua de observaciones. Además, se desea verificar si el agrupamiento permite discriminar los fallos (diagnóstico) y sus severidades (pronóstico).

11.1. Análisis multivariado, de distancia y dependencia

Varias medidas para análisis multivariado de datos se pueden encontrar en [122], una de ellas es el vector de medias $\vec{\bar{x}}$, que es la medida de centralidad más empleada en la descripción multivariante de datos $X \in \mathbb{R}^{N \times P}$. El vector de medias es de dimensión P , con componentes iguales a las medias de cada una de las variables sobre N muestras.

$$\vec{\bar{x}} = \frac{1}{n} X' \mathbf{1} \quad (11.1)$$

La covarianza mide la dependencia lineal entre dos variables (s_{jk}). Para datos multivariantes se puede representar esta información en una matriz que se denomina de varianzas y covarianzas (S).

$$s_{jk} = \frac{1}{n} \sum_{i=1}^N (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad S = \frac{1}{n} \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})' \quad (11.2)$$

S se puede obtener desde la matriz de datos centrados \tilde{X} , que es la matriz resultado de restar a cada dato su media. Con este dato se puede encontrar la matriz de varianzas corregidas \tilde{S} .

$$\tilde{X} = X - \mathbf{1} \vec{\bar{x}}' \quad \tilde{S} = \frac{1}{n-1} \tilde{X}' \tilde{X} \quad (11.3)$$

La variabilidad total de los datos (VT) se resume por la traza de S o por su varianza media (\bar{s}^2). Si la dependencia entre variables es alta, intuitivamente la variabilidad conjunta es pequeña. Este aspecto no queda recogido en esta medida, que prescinde de las relaciones de dependencia existentes.

$$VT = tr(S) \quad \bar{s}^2 = \frac{1}{p} \sum_{i=1}^p s_i^2 \quad (11.4)$$

La varianza generalizada (VG) es una mejor medida de variabilidad global y se calcula mediante el determinante de la matriz de varianzas y covarianzas $VG = \det(\mathbf{S})$, siendo una medida del hipervolumen ocupado por los datos. Tiene el inconveniente de no permitir comparar datos con diferente número de variables.

La varianza efectiva es una medida de variabilidad promedio, dada por $VP = |\mathbf{S}|^{1/p}$ y tiene la ventaja que cuando todas las variables tienen las mismas dimensiones, esta medida tiene las unidades de la varianza. La desviación efectiva es: $DE = |\mathbf{S}|^{1/2p}$.

La distancia de Mahalanobis es una medida de distancia ponderada que asume los puntos centralizados. Se define por:

$$d_i = \left[\left(\vec{x}_i - \vec{\bar{x}} \right)' \mathbf{S}^{-1} \left(\vec{x}_i - \vec{\bar{x}} \right) \right]^{1/2} \quad (11.5)$$

Una medida global de variabilidad respecto a la media de una variable vectorial promediando las distancias entre los puntos y la media es la distancia promedio (V_m)

$$V_m = \sum_{i=1}^n tr \left[\frac{1}{n} \left(\vec{x}_i - \vec{\bar{x}} \right)' \left(\vec{x}_i - \vec{\bar{x}} \right) \right] = tr(\mathbf{S}) \quad (11.6)$$

La dependencia por pares entre variables se mide por la matriz de correlación (\mathbf{R}). Es una matriz cuadrada y simétrica que tiene unos en la diagonal principal y fuera de ella los coeficientes de correlación lineal entre pares de variables (r_{jk}). Los coeficientes de correlación lineal tienen un valor absoluto de uno si existe relación lineal exacta entre las variables.

$$r_{jk} = \frac{S_{jk}}{S_j S_k} \quad (11.7)$$

El coeficiente de correlación múltiple (R_j^2) toma el elemento diagonal j en \mathbf{S} , que se denomina S_{jj} , y el elemento diagonal j en la inversa de \mathbf{S} , que se denomina S^{jj} y los relaciona por:

$$R_j^2 = 1 - \frac{1}{S^{jj} S_{jj}} \quad (11.8)$$

R_j^2 entrega un valor grande cuando se puede reconstruir la variable.

La dependencia directa entre dos variables, controlando el efecto de las demás, se mide por el coeficiente de correlación parcial ($r_{jk,12\dots p}$). La colección de estos coeficientes da lugar a la matriz de correlaciones parciales P .

$$r_{jk,12\dots p} = \frac{S^{ij}}{\sqrt{S^{ii}S^{jj}}} \quad (11.9)$$

El coeficiente de dependencia efectiva (*CDE*) entrega una medida de la dependencia conjunta al medir el alejamiento del conjunto de variables de la situación perfecta de dependencia lineal. Este coeficiente se mide por el determinante de la matriz de correlación y es igual a 1 (uno) cuando todas las variables son no correlacionadas, y es 0 (cero) si una variable es combinación del resto.

$$CDE = 1 - |P|^{\frac{1}{p-1}} \quad (11.10)$$

11.2. Análisis de componentes principales

PCA construye una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos, en el cual la varianza de mayor tamaño del conjunto es capturada en el primer eje, la segunda varianza más grande en el segundo eje y así sucesivamente. Este tipo de concepto permiten por un lado, encontrar una representación con reducción de la dimensionalidad, y por otro lado obtener la estructura fundamental de los datos [123].

PCA tiene por objeto reducir mediante transformación la dimensión de un conjunto de variables a otro conjunto de variables no correlacionadas y que permiten retener la mayor información posible desde las originales. La transformación de los datos se realiza mediante el criterio del error cuadrático medio para las distancias entre los puntos originales y sus proyecciones sobre los espacios de menor dimensión. Así, para un punto \mathbf{x}_i y una dirección de norma unidad $\mathbf{a}_i = (a_{11}, \dots, a_{1p})'$, donde p representa la dimensión del espacio original, la proyección del punto es [123, 124]:

$$z_i = a_{11}x_{i1} + \dots + a_{1p}x_{ip} = \mathbf{a}'_i \mathbf{x}_i \quad (11.11)$$

y el vector que representa esta proyección es: $z_i \mathbf{a}_i$. Llamando d_i a la distancia entre el punto y su proyección, el criterio del error cuadrático medio queda como:

$$\underset{\mathbf{a}_1}{\text{minimizar}} \sum_{i=1}^n d_i^2 = \underset{\mathbf{a}_1}{\text{minimizar}} \sum_{i=1}^n |\mathbf{x}_i - z_i \mathbf{a}_1|^2 \quad (11.12)$$

Lo cual equivale a maximizar:

$$\underset{\mathbf{a}_1}{\text{maximizar}} \sum_{i=1}^n z_i^2 = \underset{\mathbf{a}_1}{\text{maximizar}} \sum_{i=1}^n \mathbf{a}_1^T \mathbf{x}_i^T \mathbf{x}_i \mathbf{a}_1 \quad (11.13)$$

La anterior operación se realiza sobre el vector de proyecciones, con lo cual:

$$\mathbf{z}_1 = (z_1, z_2, \dots, z_i, \dots, z_n)^T = \mathbf{X} \mathbf{a}_1 \quad (11.14)$$

$$\underset{\mathbf{a}_1}{\text{maximizar}} (\mathbf{z}_1^T \mathbf{z}_1) = \underset{\mathbf{a}_1}{\text{maximizar}} (\mathbf{a}_1^T \mathbf{X}^T \mathbf{X} \mathbf{a}_1) \quad (11.15)$$

Entre los métodos recomendados para seleccionar el número de componentes principales están [125]:

- En un gráfico de λ_i contra i , seleccionar componentes hasta donde aproximadamente tienen valores iguales.
- Seleccionar componentes hasta cubrir una proporción determinada de varianza, evitando tomar pocas componentes en desventaja de otros.
- No tomar componentes asociados a valores propios con magnitud inferior a una cota dada. Una cota normalmente aceptada es la varianza media: $\sum \lambda_i / p$, que cuando se trabaja con la matriz de correlación conduce a un valor medio de 1 y por ende a despreciar los valores propios inferiores a este valor.

11.3. Biplots

Un Biplot es una representación gráfica de datos multivariantes donde, al igual que un diagrama de dispersión que muestra la distribución conjunta de dos variables, representa tres o más variables. El prefijo "bi" se refiere a la superposición en la misma representación de individuos y variables. El biplot aproxima la distribución de una muestra en un espacio de dimensión reducida y le superpone representaciones de las variables. Las representaciones de las variables son normalmente vectores

y coinciden con las direcciones en las que mejor se muestra el cambio individual de cada variable [126].

Los biplots son útiles para describir gráficamente los datos o para mostrar los resultados proporcionados por modelos más formales. La forma más común del biplot es un diagrama de dispersión en el que los puntos representan a los individuos y los dos ejes a las variables.

11.4. Algoritmos para conglomerados

Método de k -medias

Para una muestra con n elementos y p variables, se desea obtener una división en k grupos. El algoritmo de k -medias consta básicamente de los siguientes pasos [122]:

- Elegir k_0 puntos o centroides iniciales, por alguno de los siguientes criterios:
 - asignación aleatoria de elementos a los grupos.
 - tomar los k_0 puntos más alejados entre sí.
 - con base en información previa.
- Calcular distancias euclídeas desde cada punto al centro y asignarlo al centroide más cercano. Con cada punto nuevo se debe re-evaluar la media del grupo.
- Definir criterio de optimización. Comprobar si reasignar un elemento mejora el criterio.
- Terminar proceso, si no es posible mejorar.

El criterio de homogeneidad emplea la suma de cuadrados dentro de cada grupo ($SCDG$) por variable, que equivale a la suma ponderada de las varianzas por grupo:

$$SCDG = \sum_{g=1}^{k_0} \sum_{j=1}^p \sum_{i=1}^{n_g} (x_{ijg} - \bar{x}_{jg})^2 \quad (11.16)$$

donde x_{ijg} es el valor de la variable j en el elemento i del grupo g . Y como criterio queda:

$$\min SCDG = \min \sum_{g=1}^{k_0} \sum_{j=1}^p n_g S_{jg}^2 \quad (11.17)$$

donde n_g es el número de elementos del grupo g y S_{jg}^2 es la varianza de la variable j . La varianza de cada variable en cada grupo es una medida de heterogeneidad y al minimizarlas para todas las variables se obtienen grupos más homogéneos. Otro criterio de homogeneidad es minimizar las distancias al cuadrado entre los centros de grupos y los puntos de cada grupo. Este criterio es:

$$\min \sum_{g=1}^{k_0} \sum_{i=1}^{n_g} (\mathbf{x}_{ig} - \bar{\mathbf{x}}_g)' (\mathbf{x}_{ig} - \bar{\mathbf{x}}_g) = \sum_{g=1}^{k_0} \sum_{i=1}^{n_g} d^2(i, g) \quad (11.18)$$

donde $d^2(i, g)$ es el cuadrado de la distancia euclídea del elemento con su centro de grupo. Como un escalar es igual a la traza, se puede reescribir:

$$\min \sum_{g=1}^{k_0} \sum_{i=1}^{n_g} \text{tr}[d^2(i, g)] = \min \text{tr} \left[\sum_{g=1}^{k_0} \sum_{i=1}^{n_g} (\mathbf{x}_{ig} - \bar{\mathbf{x}}_g)' (\mathbf{x}_{ig} - \bar{\mathbf{x}}_g) \right] \quad (11.19)$$

y por tanto, se obtiene que

$$\min \text{tr}(\mathbf{W}) = \min \text{SCDG} \quad (11.20)$$

y donde \mathbf{W} es la matriz suma de cuadrados dentro de los grupos.

Maximizar este criterio implica su cálculo para todas las posibles particiones. Por ello, se puede emplear un criterio alternativo denominado el criterio de la curva de distorsión. Fuera de la forma descrita, existen variantes del algoritmo encaminadas a mejorar su eficiencia computacional [127]. La colección de centroides para diferentes procesos de discretización se conoce como Libro de Códigos.

Método de Ward

Método de agrupamiento jerárquico propuesto por Ward y Wishart. En este método se parte de los elementos directamente, en lugar de utilizar la matriz de distancias, y se define una medida global de heterogeneidad de una agrupación de observaciones en clusters. Esta medida es la suma de las distancias euclídeas al cuadrado entre cada elemento y la media de su grupo [128]:

$$\mathbf{W} = \sum_{g=1}^{k_0} \sum_{i=1}^{n_g} (\mathbf{x}_{ig} - \bar{\mathbf{x}}_g)' (\mathbf{x}_{ig} - \bar{\mathbf{x}}_g) \quad (11.21)$$

Se supone inicialmente que cada dato es un grupo, $g = n$, y por tanto $\mathbf{W} = 0$. A continuación se agrupan los elementos que produzcan el incremento mínimo de \mathbf{W} , lo cual implica tomar los más próximos con la distancia euclídea. Para la siguiente iteración se tendrán $n - 1$ grupos, $n - 2$ de un elemento y uno de dos elementos. Se toma nuevamente el criterio de agrupar los elementos para los cuales \mathbf{W} crece lo menos posible. Los valores de \mathbf{W} indican el crecimiento del criterio al formar grupos y pueden utilizarse para decidir cuántos grupos naturales existen.

Para representar gráficamente el anterior procedimiento se emplea un dendrograma, que es un tipo de representación gráfica o diagrama de datos en forma de árbol (de la raíz griega *dendro*) que organiza los datos en subcategorías que se van dividiendo en otras. Este tipo de representación permite apreciar claramente las relaciones de agrupación entre datos e incluso entre grupos de ellos, aunque no las relaciones de similaridad o cercanía entre categorías. Observando las sucesivas subdivisiones, se puede inferir los criterios de agrupación de los mismos, la distancia entre los datos según las relaciones establecidas, etc.

Método de k -armónicas medias

El criterio de homogeneidad para k -armónicas medias es igual al criterio empleado para k -medias, con la diferencia de emplear el promedio armónico (HA) en lugar de la suma de cuadrados dentro de cada grupo. El promedio armónico de M números x_1, x_2, \dots, x_M es el recíproco del promedio aritmético de los recíprocos de los números en el conjunto [129, 130]:

$$HA(x_1, x_2, \dots, x_M) = \frac{M}{\sum_{m=1}^M \frac{1}{x_m}} \quad (11.22)$$

El promedio armónico disminuye cuando uno de los números del conjunto es muy pequeño. De esta forma, HA se comporta más como una función de mínimos que como una de promedios, pero en oposición a una función de mínimos, HA toma en cuenta todos los puntos conduciendo a límites más suaves.

Criterio de la curva de distorsión

La teoría de la tasa de distorsión se aplica para seleccionar el número de conglomerados, o número de agrupaciones que maximizan la eficiencia mientras minimizan el error. La estrategia del algoritmo es generar una curva de distorsión para

los datos de entrada corriendo un algoritmo estándar de agrupamiento (por ejemplo, k -medias) para todos los valores de k entre 1 y un máximo, y luego evaluando la distorsión del agrupamiento en particular. La curva de distorsión se transforma posteriormente por una potencia negativa seleccionada con base en la dimensión de los datos. Los saltos en los valores resultantes significan selecciones razonables de k , con el mayor salto representando la mejor selección [131]. La distorsión de un agrupamiento de datos se define como: Sea un conjunto de datos (X) modelados mediante unas variables aleatorias de dimensión p , que consisten de una mezcla de distribuciones con G componentes de covarianza común. Si se determinan k centros de agrupamiento, con los centros lo más cerca posibles a una muestra dada en X , la distorsión promedio mínima por dimensión cuando se fija el número de centros es:

$$d_k = \frac{1}{p} \min_{c_1, \dots, c_k} E \left[(X - c_x)^T S^{-1} (X - c_x) \right] \quad (11.23)$$

En la práctica, se genera un conjunto de centros empleando un método estándar de agrupamiento y computando la distorsión con el resultado.

11.5. Marco experimental para dependencia, variabilidad y conglomerados

Para el presente análisis, se han seleccionado las señales de la base de datos muestreadas a una frecuencia de 12 kHz, tomando 3 muestras, o señales, aleatorias de cada uno de los 3 tipos de fallos y de los 3 tipos de severidad por fallo, además de 3 señales de referencia para la operación normal. Todas las señales pertenecen a un único régimen de trabajo a una velocidad de 1730 RPM. Además, se recuerda que la base de datos se anota según su tipo de fallo: Normal, Fallo en Bola (B), Fallo Cara Interna (I) y Fallo Cara Externa (O); adicionalmente se emplean etiquetas para indicar la severidad del fallo según no se induzca (N0 para operación normal), o se induzca fallo de 0.007" (N1), de 0.014" (N2) o de 0.021" (N3).

11.6. Resultados para dependencia, variabilidad y conglomerados

A cada una de las señales z_w se les calcula los siguientes parámetros univariados en el dominio del tiempo (TD): media, desviación típica, varianza, coeficiente de variación, coeficiente de asimetría, kurtosis y mediana. Además se realizan gráficos de dispersión y caras de Chernoff. En el dominio de la frecuencia (FD), luego de una

FFT, se calculan los parámetros: Media frecuencial, Varianza frecuencial, Asimetría, Kurtosis, Mediana frecuencial, así como el 3er y 4to momento estadístico. En el dominio del tiempo-frecuencia (TFD) se evalúan 12 coeficientes MFCC. Luego de calcular todos los parámetros descritos para cada una de las señales, se crea un conjunto de características multi-dominio $\{\mathbf{X}_n\}_{n=1}^N$ producto de concatenar las parámetros de los dominios TD, FD y TFD de cada \mathbf{z}_w en el vector fila $\mathbf{x}_{w,n} \in \mathbb{R}^M$ de $\mathbf{X}_n \in \mathbb{R}^{W \times M}$, donde M es el número de características multi-dominio y W el número de señales.

Continuando con el análisis multivariado, se calculan los siguientes parámetros: vector de medias, matriz de covarianzas, matriz de datos centrados, matriz de varianzas corregidas, traza, determinante y valores propios de la matriz de varianzas corregidas. Se calculan los siguientes parámetros de variabilidad: varianza total, varianza media, varianza generalizada, desviación típica generalizada, varianza efectiva y desviación efectiva. Se calculan los siguientes parámetros de variabilidad y distancias: distancia de Mahalanobis y distancia promedio. Finalmente se calculan los siguientes parámetros de medida de dependencia lineal: matriz de correlación, coeficientes de correlación múltiple, matriz de correlación parcial, coeficientes de dependencia efectiva.

Resultado de las 30 señales y 7 características univariadas en el dominio del tiempo, se obtiene la matriz de parámetros univariados que se observa en la tabla 3. Se puede observar valores marcadamente diferentes de todos los parámetros, excepto la kurtosis, para las señales pertenecientes a la operación normal. De los coeficientes de variación, que son una medida independiente de las unidades de las variables, se observa alta variabilidad general, menos para las señales pertenecientes a la operación normal. La kurtosis presenta valores mezclados entre altos y bajos dentro de un mismo tipo de fallo, además solo es constantemente baja en operación normal. En general no se puede realizar una separación de clases entre los tipos de fallos con solo la observación de los parámetros univariados analizados, aunque es de aclarar que si se puede realizar una separación de la clase operación normal contra las operaciones en fallo empleando umbrales, como es el caso cuando la desviación típica es inferior a 0.07. Umbrales similares se pueden obtener para los demás coeficientes, excepto la kurtosis, para identificar la operación normal.

Una forma gráfica de ayudar con la identificación de relaciones entre las variables es con gráficos de dispersión de variables por pares, como es el caso de la figura 11.1 que contiene pares de variables en el dominio del tiempo. En este gráfico es fácilmente distinguible la relación existente entre la desviación típica y la varianza, también se puede observar la separación de los datos para operación normal en todas las

Fallo	Severidad	Media	Desv Típica	Varianza	Coef var	Coef asim	Kurtosis	Mediana
Normal	N0	0.0140	0.0648	0.0042	4.6235	-0.1169	2.9804	0.0159
	N0	0.0140	0.0642	0.0041	4.5842	-0.1244	2.9418	0.0161
	N0	0.0115	0.0648	0.0042	5.6521	-0.1182	2.9441	0.0134
B	N1	0.0038	0.1543	0.0238	40.8430	0.0195	2.8191	0.0034
	N1	0.0042	0.1561	0.0244	36.9490	0.0152	2.9424	0.0039
	N1	0.0046	0.1504	0.0226	32.8675	0.0274	2.9053	0.0037
	N2	0.0045	0.1261	0.0159	28.3039	0.0615	15.3007	0.0062
	N2	0.0045	0.1529	0.0234	33.8180	0.2867	15.9628	0.0062
	N2	0.0046	0.1206	0.0145	26.1558	-0.0007	8.2736	0.0071
	N3	0.0046	0.1168	0.0136	25.3974	0.0165	3.1091	0.0045
	N3	0.0045	0.1177	0.0139	26.2453	0.0417	3.0964	0.0037
I	N1	0.0045	0.3117	0.0972	68.6438	-0.0107	5.2328	0.0017
	N1	0.0048	0.3127	0.0978	65.6562	-0.0114	5.2639	0.0027
	N1	0.0048	0.3169	0.1004	65.8153	-0.0179	5.3720	0.0029
	N2	0.0030	0.1811	0.0328	59.8681	0.0244	17.0098	0.0034
	N2	0.0033	0.1820	0.0331	55.8384	0.0248	19.2038	0.0034
	N2	0.0028	0.1789	0.0320	64.6947	0.0490	18.1755	0.0034
	N3	0.0031	0.4450	0.1980	142.8843	0.3070	8.5181	-0.0057
	N3	0.0028	0.4538	0.2059	161.0901	0.2987	7.9115	-0.0069
	N3	0.0026	0.4451	0.1981	168.8393	0.3076	8.6332	-0.0069
O	N1	0.0049	0.5746	0.3302	116.0993	-0.0041	8.0054	0.0041
	N1	0.0044	0.5832	0.3401	132.6871	-0.0023	7.9459	0.0041
	N1	0.0043	0.5826	0.3394	136.5625	0.0006	7.9491	0.0037
	N2	0.0027	0.0953	0.0091	35.3350	-0.0183	3.5648	0.0030
	N2	0.0027	0.0956	0.0091	35.8875	0.0153	4.0721	0.0028
	N2	0.0023	0.0932	0.0087	41.3577	0.0137	3.7628	0.0022
	N3	0.0035	0.5608	0.3145	160.1530	0.1099	22.5734	0.0037
	N3	0.0035	0.5499	0.3024	156.9384	0.1583	23.6102	0.0037
	N3	0.0036	0.5673	0.3219	155.6481	0.1309	24.6299	0.0037

Tabla 3: Análisis univariado de datos

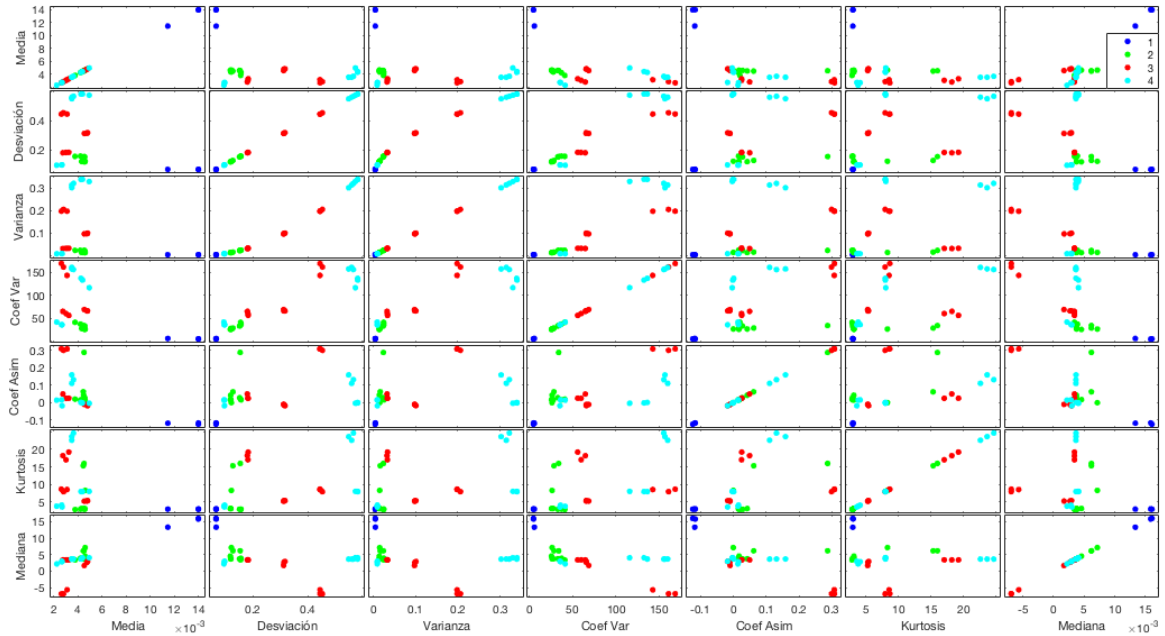


Figura 11.1: Gráfico de dispersión por pares

relaciones que involucran la media, como se ve con las marcas identificadas con la etiqueta 1 (azul). En general, para los demás pares se puede observar relaciones de dependencia no lineal, donde la etiqueta 2 (verde) es para fallo en Bola, la etiqueta 3 (rojo) para fallo en cara interna y la etiqueta 4 (cian) para fallo en cara externa.

La figura 11.2 muestra el gráfico de Caras de Chernoff para el dominio del tiempo, donde la relación entre variables y características de las caras es la siguiente:

- Tamaño de la cara: media
- Longitud del arco relativo entre frente y mandíbula: desviación típica
- Forma de la frente: varianza
- Forma de la mandíbula: coeficiente de variación
- Distancia entre ojos: coeficiente de asimetría
- Posición vertical de los ojos: kurtosis

- Altura de los ojos: mediana

La posición de las caras en el gráfico se determina por la disimilitud de los datos.

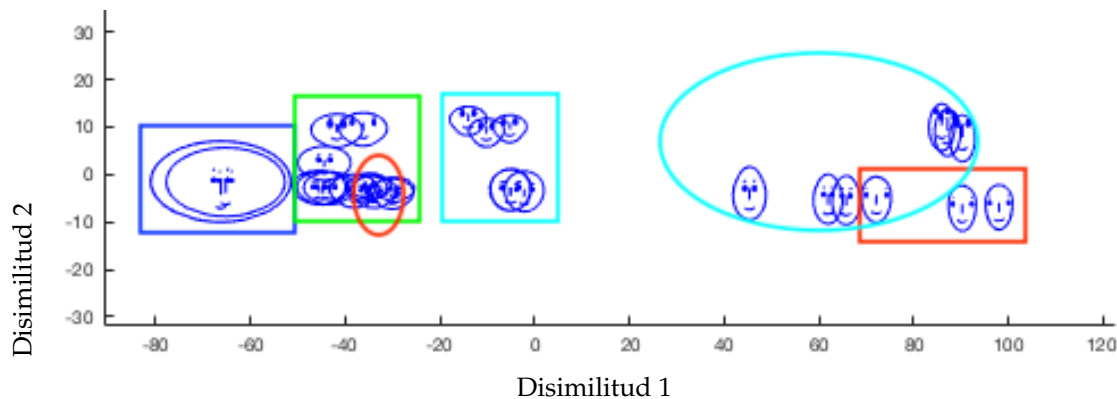


Figura 11.2: Caras de Chernoff para el dominio del tiempo

La figura 11.2 indica la pertenencia de cada cara contra su tipo de fallo. La posición de las caras en el gráfico se determina por la disimilaridad de los datos. Se puede observar la clara separación de la operación normal y la forma como se intersectan los otros tipos de fallos. La figura 11.3 muestra el gráfico de las Caras de Chernoff para el dominio de la frecuencia, donde continúa existiendo el mismo tipo de observaciones dadas para el dominio del tiempo, pero con intersecciones distribuidas diferente. La figura 11.4 muestra el gráfico de las Caras de Chernoff para el dominio tiempo-frecuencia empleado MFCC, donde se observa separabilidad total entre el modo de operación normal y todos los demás, los cuales aparecen como un solo grupo. En este conjunto de gráficas, se emplea la etiqueta color azul para la operación normal, la etiqueta color verde para fallo en Bola, la etiqueta color rojo para fallo en cara interna y la etiqueta color cian para fallo en cara externa.

El vector de medias $\vec{\bar{x}} = 100 * [0,48 \ 26,46 \ 10,50 \ 7056,66 \ 05,00 \ 892,74 \ 0,39]$ en el dominio del tiempo, comprueba una vez más que la kurtosis promedio de los datos es alta, al igual que el coeficiente de variación promedio. Estos valores son indicativos de la alta variabilidad de los datos y la presencia de varias poblaciones (tipos de fallos). El valor promedio bajo del coeficiente de asimetría verifica nuevamente

la simetría existente. Además, se observa que el coeficiente de variación posee covarianza con todas las variables, mientras que la kurtosis posee covarianza solo con la desviación estándar, la varianza y los coeficientes de variación y simetría.

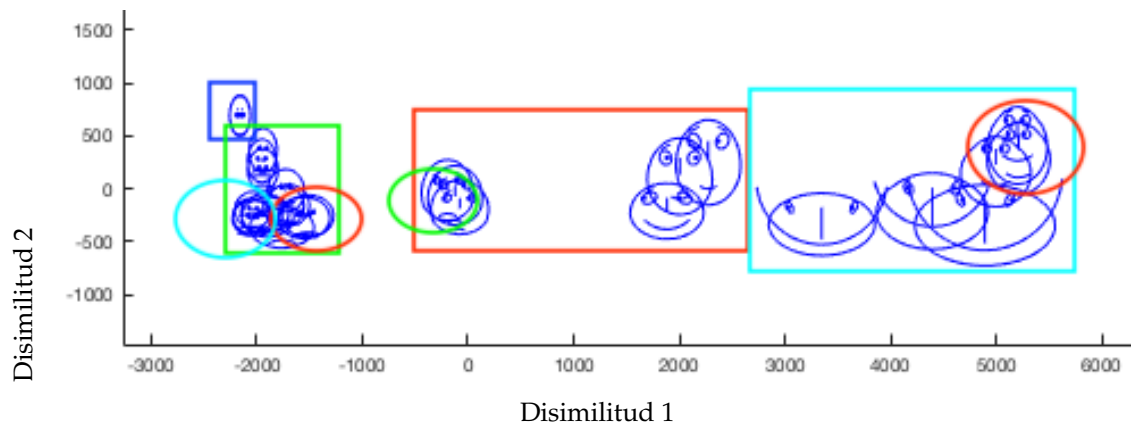


Figura 11.3: Caras de Chernoff para el dominio de la frecuencia

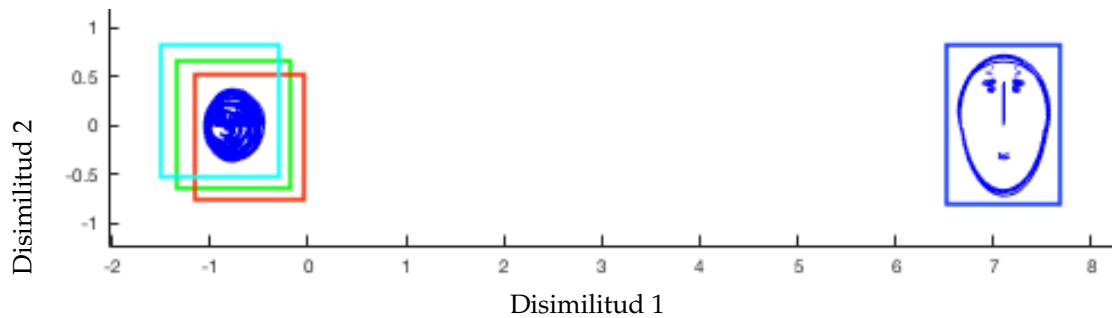


Figura 11.4: Caras de Chernoff para el dominio MFCC

En cuanto a los parámetros de medida de variabilidad, se evaluaron:

- La variabilidad total de los datos: $VT= 3053.8$

- La varianza media: $\bar{s}^2 = 436.2553$
- La varianza generalizada: $VG = 1.9493e-015$
- La varianza efectiva: $VP = 0.0079$
- La desviación efectiva: $DE = 0.0890$

El valor de VT y de \bar{s}^2 es alto, por lo que existe una dependencia conjunta entre las variables relativamente baja. De VG se puede verificar que el hipervolumen ocupado por los datos es pequeño, ya que al ser datos de vibración están perfectamente delimitados. Los valores bajos de VP y DE comprueban una dependencia baja entre las señales de vibración. Entre las medidas de dependencia lineal se tiene la distancia promedio, la cual arroja un valor de $V_m = 2952$, comprobando la independencia entre los datos.

Para las medidas de distancias, inicialmente se muestran resultados para la distancia de Mahalanobis, mostrada en la tabla 4, donde el elemento i, j representa la distancia de la variable i a la j . Solo se muestran los resultados de la diagonal superior, ya que la distancia j a i es igual a la distancia entre i a j , y distancias de los pares de señales con mayor representatividad, debido al tamaño máximo de 30 señales empleadas, según se ha descrito previamente.

Señal	Normal	Normal	Normal	B N0	B N0	B N0	I N0	I N0	I N0	O N0	O N0	O N0
Normal	-	0.26	1.23	4.10	3.87	3.65	4.38	4.42	4.47	5.08	4.47	4.43
Normal	-	-	1.32	4.01	3.82	3.68	4.39	4.40	4.44	5.19	4.45	4.40
Normal	-	-	-	3.54	3.28	2.88	3.91	4.07	4.14	4.29	3.80	3.77
B N0	-	-	-	-	0.53	1.67	2.98	2.69	2.71	4.83	3.21	3.12
B N0	-	-	-	-	-	1.19	2.64	2.40	2.43	4.43	3.05	3.00
B N0	-	-	-	-	-	-	2.45	2.52	2.62	3.64	2.99	3.01
I N0	-	-	-	-	-	-	-	0.81	0.95	4.08	3.73	3.74
I N0	-	-	-	-	-	-	-	-	0.17	4.41	3.70	3.71
I N0	-	-	-	-	-	-	-	-	-	4.51	3.73	3.74
O N0	-	-	-	-	-	-	-	-	-	-	2.64	2.88
O N0	-	-	-	-	-	-	-	-	-	-	-	0.32
O N0	-	-	-	-	-	-	-	-	-	-	-	-

Tabla 4: Distancia de Mahalanobis

Es claro como las señales 1 a 3 (pertenecientes al modo de operación normal) están claramente separadas de las demás señales. Las señales 4 a 6 (pertenecientes al modo de operación en fallo en bola con severidad baja) se pueden diferenciar de los modos de fallo en cara interna y cara externa con severidad igualmente baja. Además, se puede observar que no existe una distancia marcada entre las señales con severidades medias y altas en los tres modos de fallo descritos.

Para el caso del análisis de componentes principales se realiza el cálculo de la matriz de covarianzas (S) y de la matriz de covarianzas corregidas (S_r), con el fin de obtener los valores propios y vectores propios respectivos. Los resultados ordenados de forma descendente son los indicados en la tabla 5. Luego se realizan gráficos comparativos para la selección del número de componentes de representación en el nuevo espacio reducido. Con los datos anteriores se puede graficar el ordinal de los valores propios contra el valor propio y el ordinal contra la varianza acumulada, tanto para los datos desde S como para S_r . Estos gráficos se pueden observar en la figura 11.5, donde los gráficos superiores son los valores propios y varianza acumulada para S , y los gráficos inferiores corresponden a S_r .

	Val propios S	Val propios S_r
1	2917.2	4.1492
2	34.7	1.3648
3	0	0.7445
4	0	0.4204
5	0	0.0688
6	0	0.0161
7	0	0.0030

Tabla 5: Valores propios y valores propios corregidos

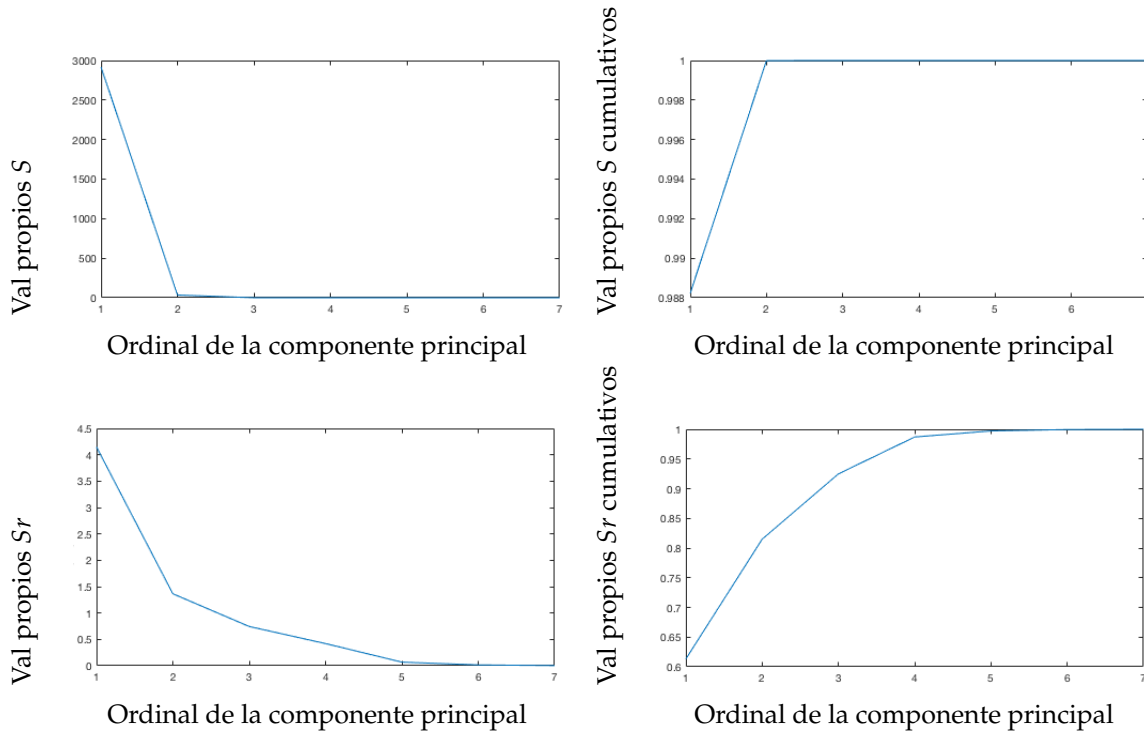


Figura 11.5: Gráfico de valores propios y varianza acumulativa

Con base en los resultados previos, se puede proceder a determinar el número de componentes principales a usar como representación de las variables originales, para ello se realiza el análisis de los siguientes tres criterios de selección:

- En las gráficas de ordinales contra valores propios, se puede observar que justo desde el ordinal 2 en *S* los componentes tienen valores aproximadamente iguales, y para el ordinal 3 en *Sr*.
- La selección de componentes con el fin de cubrir mínimo un 85 % de varianza arroja un ordinal de 2 para los datos desde *S* y de 3 para los datos desde *Sr*.
- La varianza media para el caso de datos desde *S* arroja $\sum \lambda_i / p = 1475.95$, lo que implica despreciar componentes bajo este valor, que en este caso son todas las componentes menos la primera. La varianza media para el caso de datos desde

Sr arroja $\sum \lambda_i / p = 0.967 \approx 1$, lo que implica despreciar componentes bajo este valor, que en este caso son las componentes más allá de la tercera.

En definitiva, se opta por una representación con tres componentes principales, ya que algunos resultados con 1 o 2 componentes están sustentados en cálculos que no emplean la matriz de covarianzas corregidas y enmascaran variabilidad de los datos. Además, pocas componentes eliminan la representatividad de los datos. Los vectores propios asociados a estas tres componentes son los mostrados en la tabla 6.

Variables	Componentes desde S			Componentes desde Sr		
	C1	C2	C3	C1	C2	C3
Media	0.0000	-0.0001	0.0100	0.3163	0.5150	-0.0958
Desv. típica	-0.0033	-0.0011	0.4926	-0.4244	0.3549	-0.2204
Varianza	-0.0021	-0.0006	0.4178	-0.3986	0.4344	-0.2175
Coef. variación	-0.9977	-0.0675	-0.0016	-0.4619	0.1831	-0.1331
Coef. asimetría	-0.0014	0.0021	-0.7629	-0.3641	-0.3313	0.1757
Kurtosis	-0.0675	0.9977	0.0023	-0.2821	0.1950	0.8786
Mediana	0.0001	0.0002	0.0249	0.03681	0.4887	0.2728

Tabla 6: Vectores propios y vectores propios corregidos

Estos resultados muestran que, desde la matriz de varianzas y covarianzas, las 3 componentes están relacionadas con pocas de las variables, así la componente 1 se relaciona casi completamente con el coeficiente de variación y desprecia las demás, la componente 2 se relaciona casi totalmente con la kurtosis y desprecia las demás y la tercera componente recibe aportes ponderados de la desviación típica y varianza contra aportes negativos del coeficiente de asimetría. En este conjunto de componentes, las variables media y mediana no aportan significativamente. Los resultados desde la matriz de varianzas corregidas muestran aportes ponderados de casi todas las variables, así, la primera de las componentes contrapone la media contra las demás y desprecia los aportes de la mediana. La componente 2 es una ponderación de todas las variables contra aportes inversos del coeficiente de asimetría. Finalmente, la componente 3 pondera los aportes de las tres últimas contra los aportes de las cuatro primeras, con baja representatividad de la media y alto aporte de la kurtosis.

En general, las componentes desde la matriz de covarianzas esconden representatividad de las variables, tal como se observa en las componentes desde la matriz de covarianzas corregidas. Estas asociaciones entre variables y componentes principales se pueden observar en la figura 11.6, donde las componentes con igual signo para misma variables se encuentra juntas en el gráfico, por ejemplo desviación típica y varianza para S y desviación típica, varianza y coeficiente de variación en Sr . Componentes con signos diferentes en distintas variables se encuentran más alejadas. En el gráfico de componentes asociadas a S , la media y la mediana aparecen juntas, pero tienen signos diferentes en la componente 2, lo cual ocurre por la magnitud de dicha componente que está cercana a cero. Igualmente es de resaltar el cambio inverso en signos de componentes entre el coeficiente de asimetría con la media.

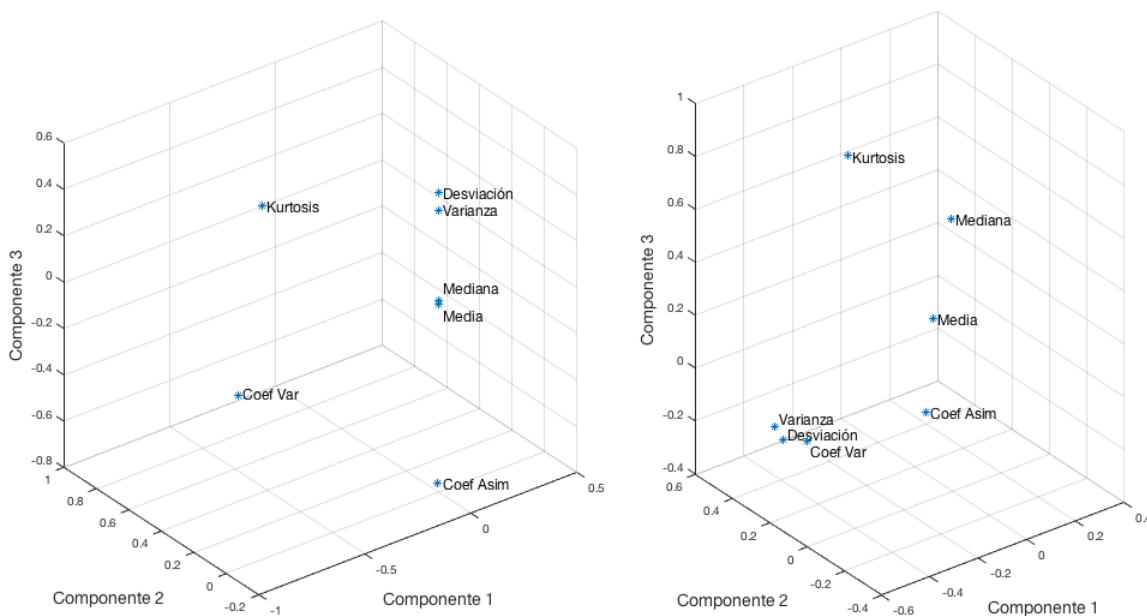


Figura 11.6: Gráficos de asociación de componentes principales y variables

Ahora, en la figura 11.7, se muestra como se proyectan las observaciones sobre el nuevo espacio definido por las 3 componentes principales, separando los grupos de la siguiente forma: las gráficas superiores muestran la proyección de observaciones divididas solo en dos conjuntos (normal vs fallos) según se emplee S o Sr ; las

gráficas inferiores muestran la proyección de observaciones sin operación normal y dividido en tres conjuntos (uno por cada tipo de fallo) y según se emplee S o Sr . Se destaca que los gráficos para normal contra fallos siempre muestran la separación de estas dos clases, mientras que los gráficos que proyectan solo los tres tipos de fallos (sin normal) indican que las clases se mezclan en algunas de las observaciones, sin embargo aparece mejor separación en la proyección desde Sr .

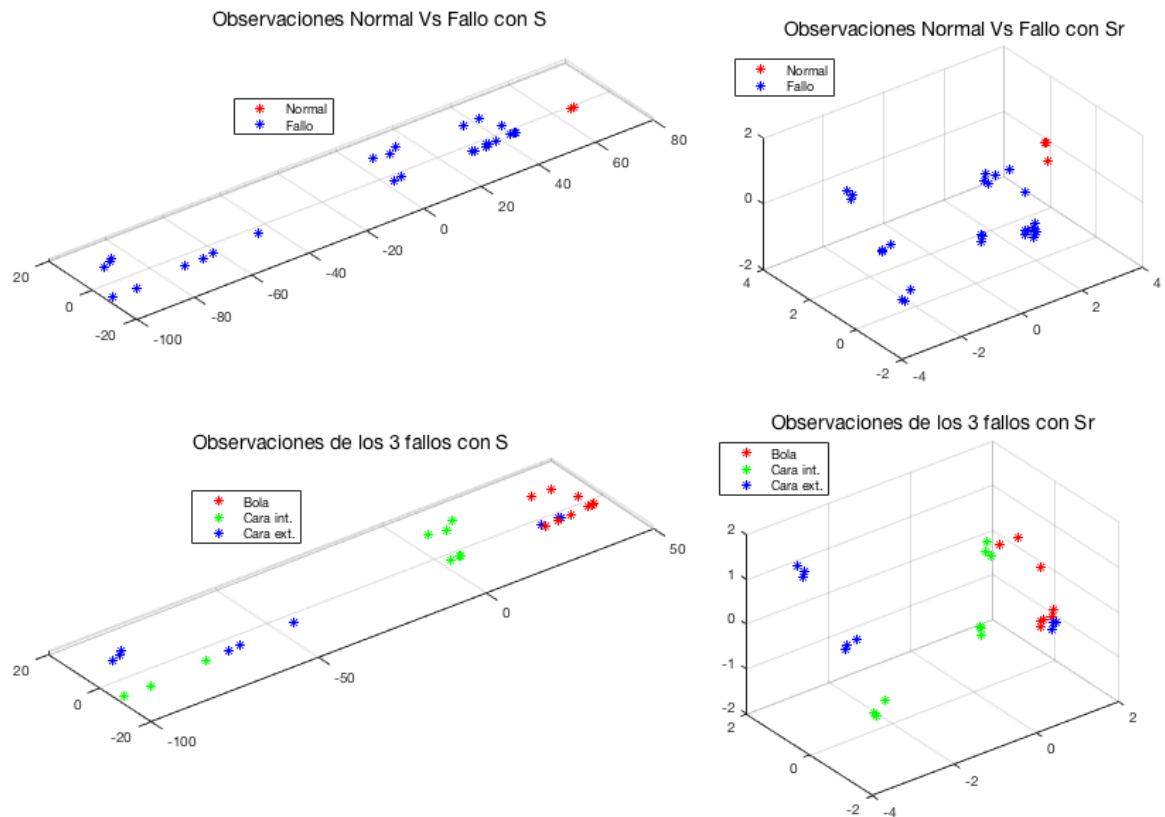


Figura 11.7: Proyección de observaciones sobre las componentes

Seguidamente se muestra el biplot sobre los datos proyectados en las componentes 1 y 2 y empleando datos desde Sr . En la figura 11.8 se muestra este gráfico, donde

los vectores indican las direcciones que mejor representan el cambio individual de variables. El signo de las componentes, visto previamente al analizar los valores propios, también se ven reflejados en el gráfico biplot.

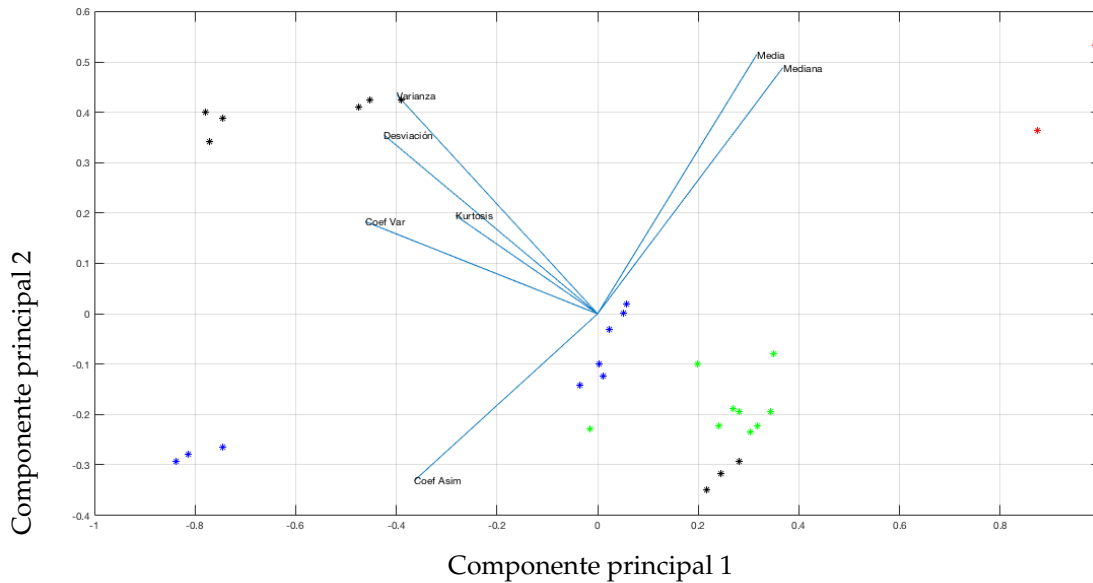


Figura 11.8: Biplot sobre componentes 1 y 2 empleando datos de Sr

Para el análisis de conglomerados, se toman las 30 señales definidas y se efectúa segmentación en marcos de 200 ms y filtrado mediante una ventana tipo Hamming con el propósito de obtener un total de 12 coeficientes MFCC. Los coeficientes así obtenidos son organizados de forma matricial para seguidamente realizar agrupamiento por medio de k -medias desde 2 hasta un máximo de 30 grupos y luego evaluar el criterio de distorsión. Luego de 10 ejecuciones, en 8 de ellas el criterio arrojó entre 23 y 25 grupos. En la figura 11.9 se pueden observar un ejemplo de ejecución del procedimiento, donde el número óptimo de grupos está en 23 grupos. Se debe tener presente que en estos gráficos de curvas de distorsión se representan directamente los valores de salto entre dos clusters, según el criterio de Ward [128].

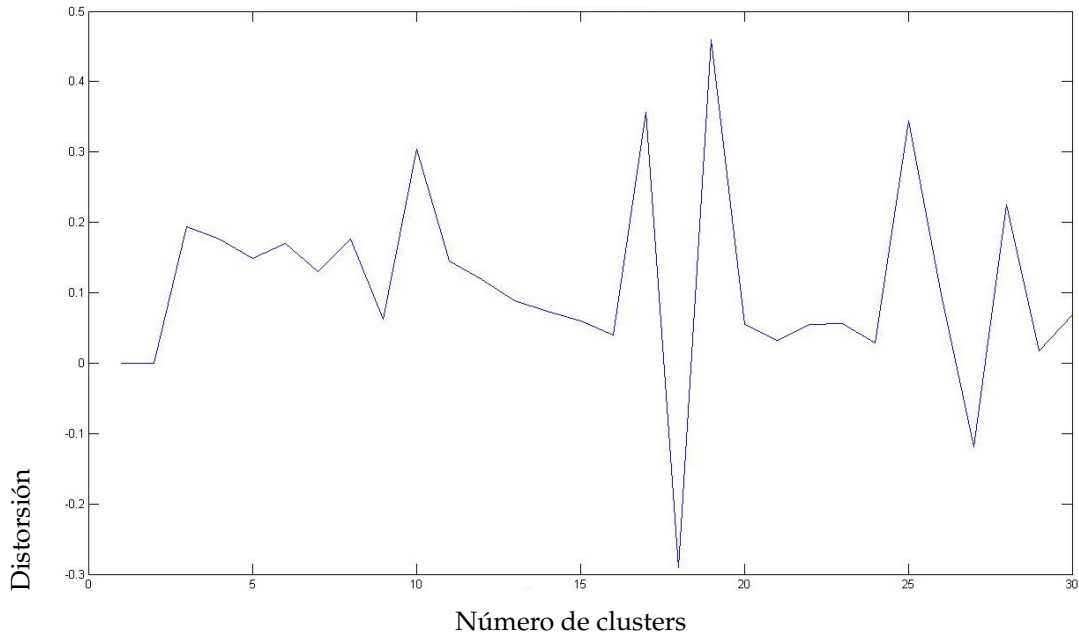


Figura 11.9: Gráfico de distorsión

En la figura 11.10, se muestra conjuntamente un total de 14400 observaciones provenientes de los coeficientes MFCC al ser registradas a medida que se presenta un nuevo marco de ventana (puntos en color azul) y los valores discretos de ubicación de cada uno de los centroides (líneas en color rojo) para una comparación de conglomerados de 4, 8, 18, y 24 grupos respectivamente. En esta figura, se puede observar que al emplear un número bajo de grupos se sacrifica especialmente los valores extremos; sin embargo, cuando el número de centroides crece los valores alejados tienen representación, pero se inicia un aglutinamiento de centros alrededor de zonas más pobladas de datos.

Al hacer agrupamiento por k-armónicas medias, se obtiene la figura 11.11, donde se muestran conglomerados con 4, 8, 18, y 24 clusters respectivamente. Se observa que este método le da mayor importancia a datos extremos cuando el número de grupos es bajo; sin embargo, cuando el número de clusters crece, se evidencia aglutinamiento de centroides en zonas muy pobladas de observaciones, como se puede observar de la figura 11.11 para 18 y 24 clusters.

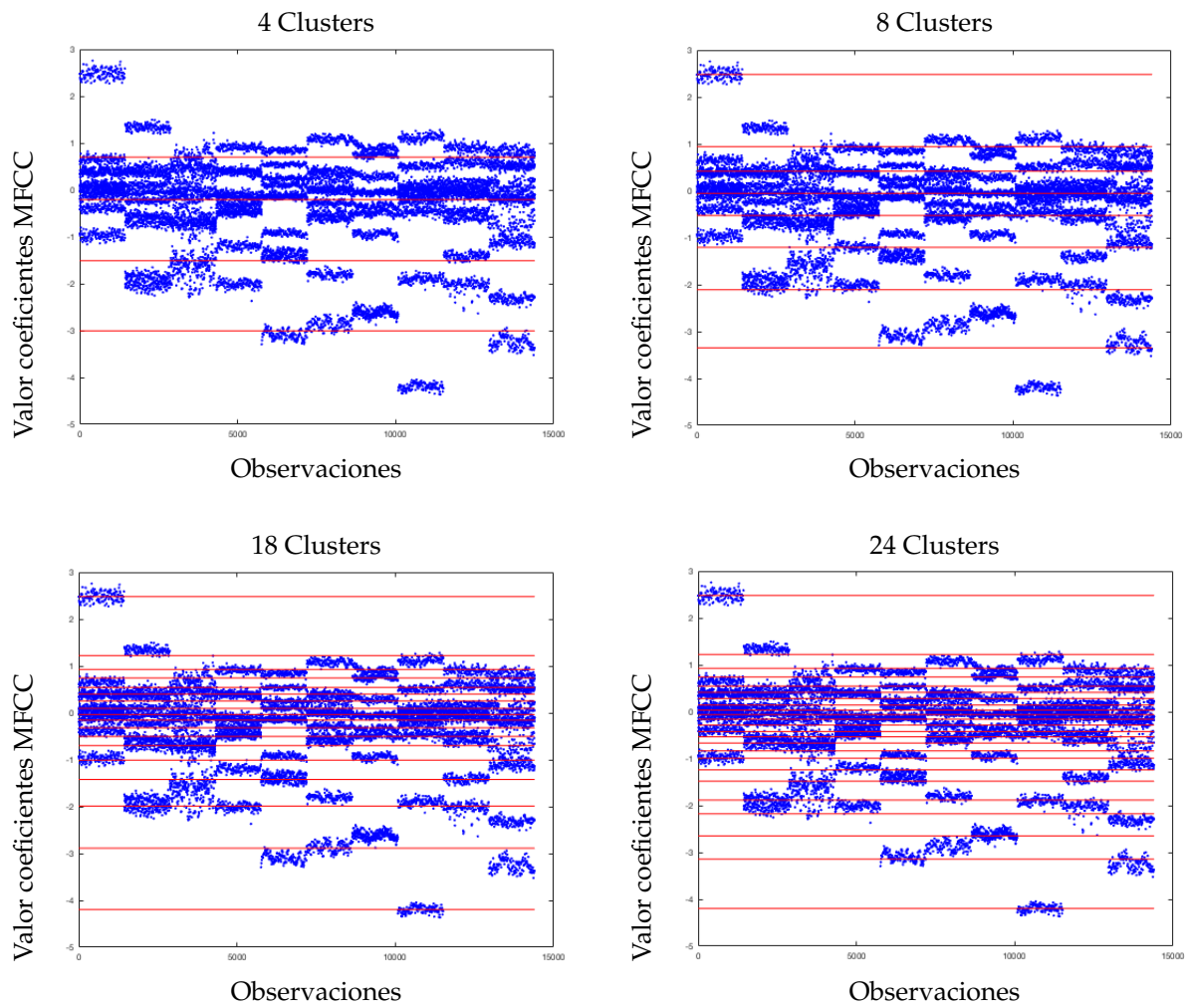


Figura 11.10: Conglomerados con k-medias para 4, 8, 18 y 24 clusters

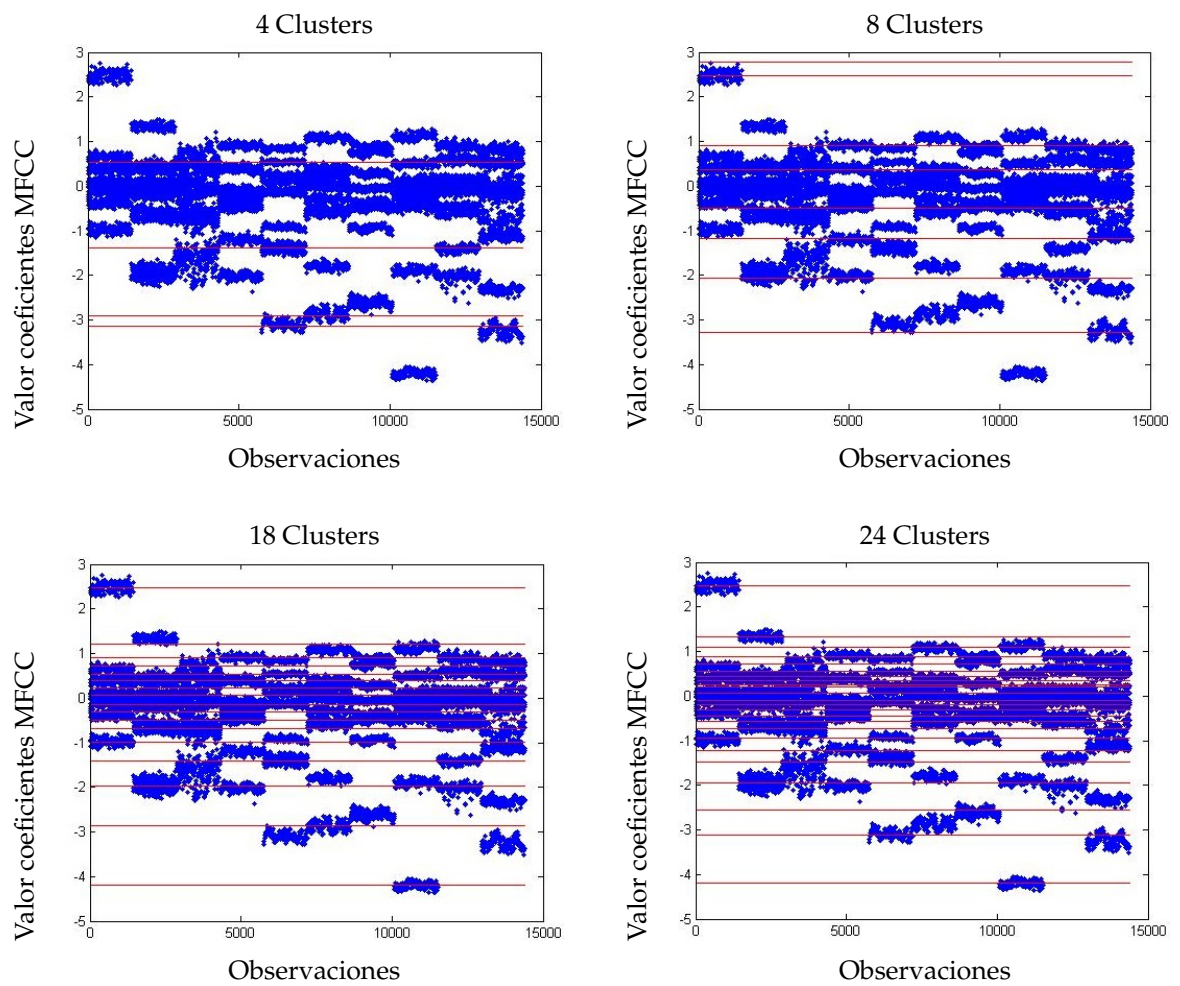


Figura 11.11: Conglomerados con k-armónicas medias de 4, 8, 18 y 24 clusters

Finalmente, empleando el método de agrupamiento jerárquico propuesto por Ward y Wishart [128], con hasta un máximo de 36 conglomerados y graficando los resultados en un dendrograma, se obtiene la figura 11.12. Se debe recordar la organización de la base de datos para poder darle un significado a las agrupaciones, la cual inicia con tres señales de operación normal y luego existen 9 señales por tipo de fallo divididas en grupos de 3, según el nivel de severidad. Con esta descripción del orden de las señales, es claro como en el dendrograma siempre se agrupan correctamente las ternas de señales para un mismo fallo, sin embargo, las agrupaciones de niveles de severidad no son correctas; por ejemplo, se asocian las señales de las ternas (13, 14, 15) con la terna (16, 17, 18) y la terna (28, 29, 30), con lo cual se agrupa severidad 1 y 2 de fallo en cara interna con severidad 3 de fallo en cara externa del rodamiento. Análogamente, las señales 4 a 12 que pertenecientes a un tipo de fallo en bola, se agrupan con las señales de la terna (25, 26, 27) que pertenecen a un fallo en cara externa de severidad 2.

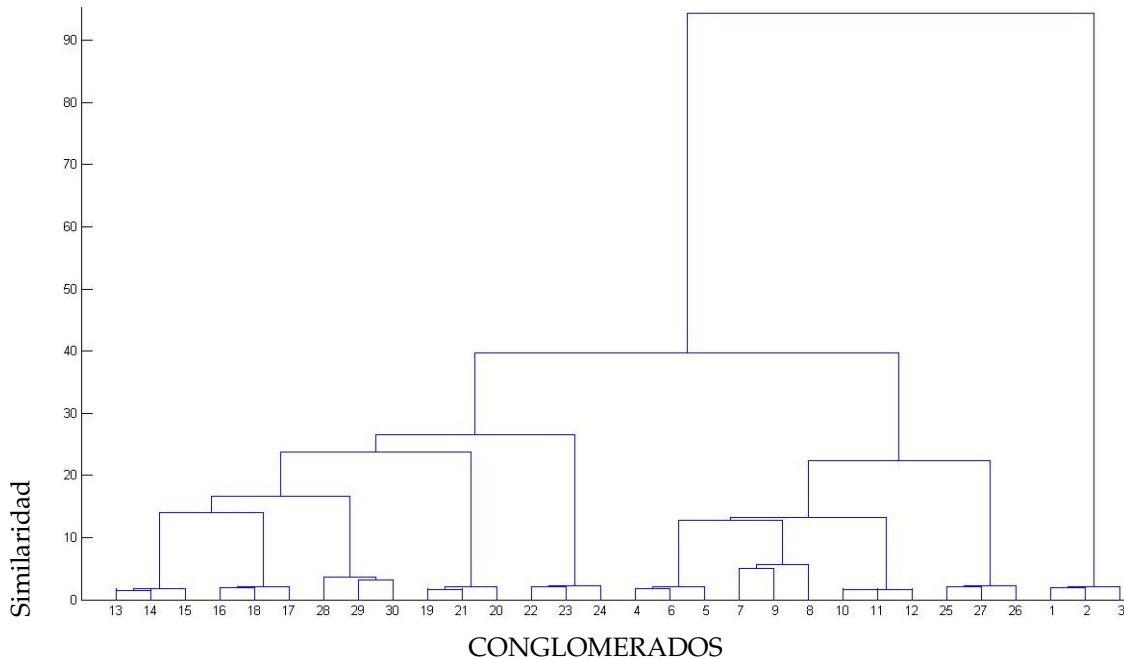


Figura 11.12: Dendrograma para agrupamiento Ward

11.7. Discusión de resultados para dependencia, variabilidad y conglomerados

De los resultados presentados, se puede observar que la separación entre señales cuando se dividen en solo dos clases, a saber: operación normal y operación con fallo, es perfectamente realizable desde varios tipos de análisis, como es el caso de emplear umbrales para la matriz de datos univariados (tabla 3), en los gráficos de dispersión por pares (ver figura 11.1), en los análisis por caras de Chernoff (ver figuras 11.2 , 11.3 y 11.4) y de la distancia de Mahalanobis (ver tabla 4).

Del gráfico de caras de Chernoff se puede deducir la presencia de más de una clase dentro de las observaciones, con datos para las severidades media y alta mezclados en los tipos de fallos. Estas mezclas son recurrentes, y persisten indiferente del tipo de dominio de representación de las señales. Es de resaltar el caso del dominio tiempo-frecuencia, donde se separa perfectamente la operación normal de la operación con fallo, pero todos los tipos de fallos se mezclan (ver figura 11.4). La distancia de Mahalanobis corrobora la separación del modo de operación normal del resto, pero sin embargo, las distancias de los demás modos de operación muestran que no existe una distancia marcada entre las señales con severidades medias y altas en los tres modos de fallo descritos (ver tabla 4).

Se puede observar que existe una alta variabilidad de los datos, lo cual indica que no se puede realizar separación de clases entre tipos de fallos con solo observación de parámetros univariados. Lo anterior lo corrobora el análisis multivariado, con baja dependencia conjunta entre las variables y bajo hipervolumen de los datos.

Los resultados vistos desde la matriz de varianzas y covarianzas enseñan que los valores propios enmascaran información y solo muestran cierta representatividad, aunque, la matriz de covarianzas corregidas indica una mayor representatividad para todas las componentes principales. Los métodos presentados para la selección del número de componentes principales son sensibles a la matriz empleada, especialmente se observa que para la matriz de varianzas y covarianzas el número de componentes es particularmente bajo, lo cual muestra nuevamente como se omite información de variabilidad de las variables originales; sin embargo, cuando se emplea la matriz de covarianzas corregidas, el número de componentes crece, lo cual indica que este método incluye mayor información (ver figura 11.5 y tabla 6).

Los resultados muestran que desde la matriz de varianzas y covarianzas las com-

ponentes principales son individualmente casi una asignación a una variable; pero desde la matriz de covarianzas corregidas se muestran aportes ponderados de casi todas las variables, tal como se puede observar en la figura 11.6. En el contexto de la separación de clases, el análisis de relevancia de las variables originales en el aporte a las componentes principales y el análisis de proyección de observaciones solo sirven para el caso simple de separación de operación normal contra fallo, más no para la separación de fallos y niveles de severidad (ver figura 11.7).

Para el método de conglomerado por k -medias, se observa que al emplear un número bajo de grupos se sacrifica las observaciones con valores extremos (ya sea datos máximos o mínimos), y por tanto su representatividad; sin embargo, cuando el número de centroides crece los valores extremos ganan representación, pero se inicia una concentración de centros alrededor de zonas más pobladas de datos. Al hacer agrupamiento por k -armónicas medias, se da mayor importancia a datos extremos cuando el número de grupos es bajo; sin embargo, cuando el número de clusters crece, k -medias y k -armónicas medias se comportan similar, perdiendo resolución al ubicar varios centroides en zonas pobladas de observaciones (ver figuras 11.10 y 11.11). De acá la importancia de definir un número adecuado de centroides, o clusters.

Del análisis por dendrograma es evidente que las señales pertenecientes a un mismo tipo de fallo poseen las menores distancias entre ellas. Sin embargo, las agrupaciones por niveles de severidad no son agrupadas coincidiendo con el tipo de fallo. En el dendrograma es claro que la máxima distancia entre señales ocurre entre las pertenecientes al modo de operación normal contra todas las demás. Este último resultado ha sido consistente con resultados previos, donde siempre ha sido posible esta misma separación, pero ahora también es posible realizar una identificación de clases de las señales del tipo de fallo en bola, al tener una distancia entre ellas que claramente las distingue de las demás; sin embargo, los otros dos tipos de fallos (en cara interna y externa) siguen presentando mezcla en sus agrupaciones cuando se tiene presente los niveles de severidad, como se aprecia en la figura 11.12.

12. SELECCIÓN Y FUSIÓN DE DOMINIOS

Dentro de la tarea de diagnóstico y pronóstico eficiente de fallo se requiere determinar localización, severidad, salud estructural y entregar pronóstico mediante progresión. En la literatura, el análisis de características para fallos en rodamientos usualmente trata un único dominio de representación, llámese tiempo, frecuencia, tiempo-frecuencia, etc., con enfoques tradicionales que carecen de fusión o conjugación de estos dominios, de evaluación de la relevancia de cada uno, así como de la contribución de información contenida en cada una de sus características, donde la selección/reducción dimensional pueda incrementar la eficiencia del diagnóstico a la vez que disminuye la complejidad computacional.

Algunas estrategias tradicionales para el diagnóstico emplean enfoques basados en un modelo físico del mecanismo, las cuales requieren de conocimiento específico del sistema y generalmente no reflejan un modelo general para todos los modos de fallo y durante el ciclo completo de vida del activo. Además, es recurrente solo realizar detección de modos de fallo, sin severidad, la cual es vital para el pronóstico. De lo anterior, este capítulo presenta una metodología para diagnosticar y pronosticar vida útil en rodamientos, enfocada en conjugar espacios de representación para determinar modos de fallos y niveles de severidad, empleando HMM, donde se evalúa la relevancia de los espacios de representación, y de sus características. Se emplean modelos HMM ya que son apropiados para el análisis de sistemas aleatorios dinámicos, por tener la ventaja de fácil interpretación, poder actuar en ambientes de aprendizaje competitivo y con fines de caracterizar los diferentes estados de salud de un mecanismo y por permitir modelar cambios comportamentales que se ven reflejados directamente en las variables latentes del proceso.

12.1. Marco experimental para selección y fusión de dominios

Los dominios de representación conforman un proceso de modelamiento de la señal bajo una variable espacial dada; este procesamiento de la señal determina los parámetros bases para el desarrollo de estudio de características. El dominio de representación del tiempo (TD) es la representación natural de la señal de vibración, que se basa en el análisis del comportamiento estadístico de la forma de onda. Para un segmento de señal $z \in \mathbb{R}^T$ con T muestras, los momentos y características principales en TD se resumen en la tabla 7 con su respectiva identificación (Id.).

Id.	Característica	Id.	Característica
T1	Valor medio	T10	Factor de holgura
T2	Desviación	T11	Factor de impulso
T3	Asimetría	T12	Límite superior del histograma
T4	Kurtosis	T13	Límite inferior del histograma
T5	Pico positivo máximo	T14	Verosimilitud logarítmica negativa
T6	Raíz media cuadrática	T15	Entropía
T7	Factor de forma	T16	Raíz media cuadrática de picos
T8	Factor de impacto	T17	Desviación sobre umbral del 70 %
T9	Factor de media sin signo		

Tabla 7: Características en TD para fusión y selección

El dominio de representación de la frecuencia (FD) representa una señal con base en sus componentes frecuenciales. Una forma de análisis consiste en aplicar la transformada de Fourier a un segmento de señal $z \in \mathbb{R}^T$ que descompone en funciones sinusoidales las distintas frecuencias relacionadas con la señal. Las principales características empleadas en FD son calculadas para los vectores de frecuencia y espectro (λ y $s \in \mathbb{R}^K$, respectivamente), donde su lista e identificación (Id.) se enseña en la tabla 8.

Id.	Característica	Id.	Característica
F1	Frecuencia media	F9	Inversa del RMS
F2	Varianza	F10	Desviación sobre frecuencia central
F3	Asimetría	F11	Promedio sobre tercer momento
F4	Kurtosis	F12	Promedio sobre cuarto momento
F5	Frecuencia central	F13	Raíz media promedio del primer momento
F6	Desviación	F14	Raíz media geométrica
F7	Raíz media cuadrática	F15	Desviación media absoluta
F8	Media cuadrática	F16	Rango intercuartílico

Tabla 8: Características en FD para fusión y selección

El dominio de representación del tiempo-frecuencia (TFD), es una representación tridimensional de los valores puntuales de tiempo, frecuencia y amplitud de la señal. Con estos puntos se conoce la forma de la señal en el plano TF y se puede obtener el

valor de la concentración de intensidad en cada punto de la distribución global. Las representaciones en TFD que se emplean para este análisis se listan a continuación:

- Coeficientes Cepstrales en Escala Mel (MFCC, o M): se emplea un banco de 24 filtros y 12 coeficientes numerados respectivamente de M1 a M12.
- Transformada Wavelet (WT, o W): se emplean 8 características, donde para las componentes de alta frecuencia se obtiene la Media (W1), el Valor máximo (W2), la Kurtosis (W3) y la Energía de Shannon (W4); análogamente se obtiene para las componentes de baja frecuencia la Media (W5), el Valor máximo (W6), la Kurtosis (W7) y la Energía de Shannon (W8).
- Transformada de Fourier de Tiempo Corto (STFT, o S): se emplean 4 características identificadas como sigue: Kurtosis (S1), Media (S2), Varianza (S3) y Pico máximo (S4).

De la base de datos, para el presente análisis se han seleccionado las señales muestreadas a una frecuencia de 12 kHz, ya que este conjunto de datos presenta registros para todos los modos de fallo y severidades deseables. La base de datos se divide en un conjunto para entrenamiento, con el 60% de los datos, y otro para validación con el restante 40%. Lo anterior conforma $\{\mathbf{Z}_n, y_n\}_{n=1}^N$ un conjunto de N señales de vibración con $\mathbf{Z}_n \in \mathbb{R}^{W \times T}$ conteniendo W segmentos en las filas $\mathbf{z}_{w,n} \in \mathbb{R}^T$ a T instantes de tiempo ($w \in \{1, 2, \dots, W\}$), y con etiquetas de salida $y_n \in \{1, 2, \dots, C\}$ relacionadas con C clases para diagnóstico de fallos o severidades de un rodamiento. Cada w_i segmento, o marco, de señal tiene una duración de 200 ms, al cual se le aplica una ventana Hamming con traslape entre ventanas de $2/3$.

Seguidamente se realiza el proceso de extracción de características según los dominios de representación a saber: TD, FD, MFCC (M), Wavelet (W) y STFT (S), sobre cada uno de los w_i segmentos de señal. Con ellas, se crea un conjunto de características multi-dominio $\{\mathbf{X}_n, y_n\}_{n=1}^N$ producto de concatenar los parámetros en cada dominio para cada $\mathbf{z}_{w,n}$ en el vector fila $\mathbf{x}_{w,n} \in \mathbb{R}^M$ de $\mathbf{X}_n \in \mathbb{R}^{W \times M}$, donde M es el número de características multi-dominio extraídas. Para la obtención del libro de códigos se emplea k -medias con un total de 24 centroides, con lo cual se obtiene la matriz de características multi-dominio discretizada $\{\mathbf{X}_n^*, y_n\}_{n=1}^N$. Finalmente se entrenan modelos HMM por cada estado de degradación bajo un esquema de validación cruzada k -fold y se repite cada entrenamiento un total de 30 iteraciones, donde para cada iteración el conjunto de entrenamiento se selecciona aleatoriamente. Con

el fin de evaluar la eficiencia de los espacios de representación y de sus características, se realizan procedimientos descritos a continuación, corriendo algoritmos en una máquina Intel® Core™ i7-4510, CPU a 2.00 GHz, con 8 GB RAM, S.O. Windows 8.1 de 64 bits.

La anotación de la base de datos (etiquetas $y_n \in \{1, 2, \dots, C\}$) se realiza nuevamente según sean de operación Normal, Fallo en Bola (B), Fallo Cara Interna (I), Fallo Cara Externa (O), y con etiquetas N1, N2 y N3 para indicar la severidad de la fallo según se induzca fallo de 0.007" (N1), de 0.014" (N2) o de 0.021" (N3). Lo anterior arroja un total de $C = 10$ clases. El estado de operación normal, obviamente, no posee niveles de severidad.

12.2. Resultados para selección y fusión de dominios

Análisis de dominios contra estados HMM

Inicialmente se presenta la evaluación de relevancia individual de cada uno de los dominios de representación, verificando la eficiencia y tiempo de entrenamiento, por cada uno de los dominios y empleando todo el conjunto de características. En este análisis no se tiene presente los niveles de severidad. En la figura 12.1 se muestra el tiempo de cómputo para las 30 iteraciones (escala y trazo verde), así como las eficiencias (escala y trazo azul), contra el número de estados HMM recorridos en variaciones desde 1 a 8 estados.

En general, los dominios de representación con todas sus características, muestran una tendencia a entregar una eficiencia de diagnóstico que obtiene los mejores resultados entre 1 y 4 estados HMM, y esta no mejora, o incluso decrece considerablemente, a medida que se incrementan dichos estados hasta llegar al límite de 8, lo cual se relaciona directamente con sobre-entrenamiento del modelo. Lo expuesto no se cumple para el dominio de representación M, el cual presenta mejoras a medida que se incrementa el número de estados. El tiempo de cómputo aumenta con el incremento del número de estados HMM, siendo más crítico en el dominio de representación M, en el cual es de un orden de magnitud superior a los demás. La eficiencia general del proceso de identificación de fallos sin niveles de severidad, respecto al incremento del número de estados HMM, tiene comportamientos variados entre los diferentes dominios, es así como en W permanece alrededor del 80% durante todos los estados; en S inicia en 85% y decrece hasta un 75% aproximadamente; en TD y FD la eficiencia decae ligeramente con el número de estados, pero se mantiene en niveles cercanos al 90%. El dominio M es el único que presenta mejoría a medida

que se incrementa el número de estados HMM y además supera considerablemente a los demás registros.

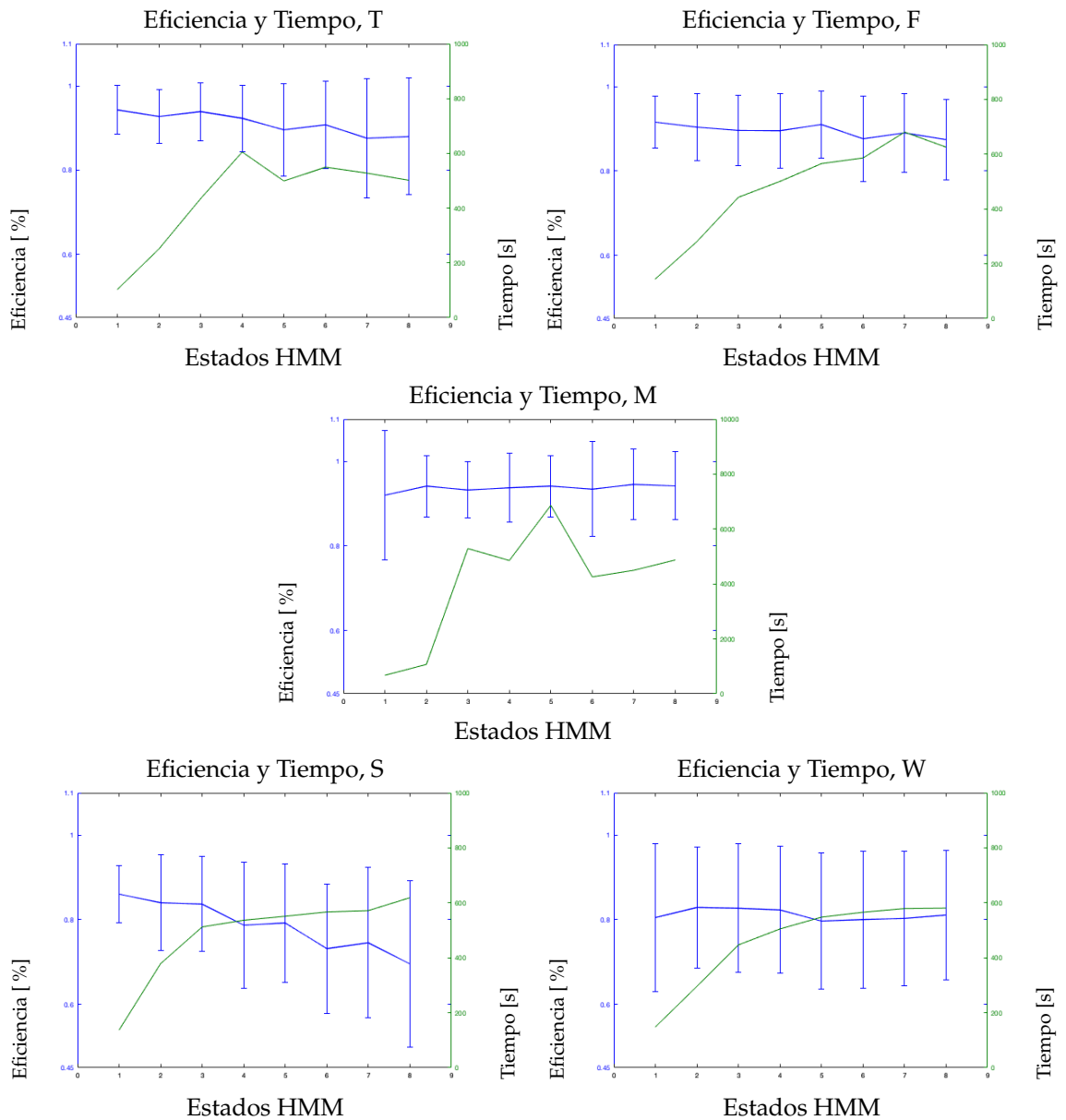


Figura 12.1: Relevancia individual de dominios de representación

Análisis intra-dominio de características

La tabla 9 muestra los valores de eficiencia y desviación como valores decimales, cuando se entrena un modelo HMM de 2 estados empleando únicamente como observación cada una de las características de los dominios de representación. Este tipo de análisis se conoce como análisis intra-dominio, o relevancia individual de cada característica y en él se consideran los modos de fallo sin tener presente la severidad. Ya que el número de características difiere entre espacios, solo se muestran datos para las 17 características de TD, para las 16 de FD, para las 12 de M, para las 8 de W y para las 4 de S.

Dominio	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
TD	0.379 ±.233	0.903 ±0.073	0.508 ±0.173	0.673 ±0.214	0.874 ±0.088	0.891 ±0.093	0.775 ±0.207	0.899 ±0.119	0.844 ±0.121	0.764 ±0.206	0.716 ±0.214	0.865 ±0.089	0.893 ±0.111	0.894 ±0.089	0.806 ±0.187	0.895 ±0.073	0.901 ±0.105
FD	0.803 ±0.239	0.856 ±0.118	0.541 ±0.271	0.574 ±0.297	0.813 ±0.170	0.857 ±0.113	0.883 ±0.112	0.895 ±0.098	0.902 ±0.098	0.651 ±0.181	0.521 ±0.283	0.568 ±0.291	0.103 ±0.083	0.817 ±0.159	0.827 ±0.182	0.836 ±0.133	
M	0.929 ±0.088	0.930 ±0.075	0.928 ±0.118	0.937 ±0.072	0.946 ±0.085	0.920 ±0.095	0.945 ±0.062	0.929 ±0.069	0.929 ±0.112	0.942 ±0.061	0.929 ±0.104	0.935 ±0.081					
W	0.721 ±0.229	0.687 ±0.208	0.720 ±0.130	0.591 ±0.200	0.746 ±0.179	0.697 ±0.218	0.343 ±0.227	0.739 ±0.193									
S	0.425 ±0.090	0.781 ±0.135	0.783 ±0.138	0.649 ±0.165													

Table 9: Relevancia de características intra-dominio

En la tabla 9 se puede apreciar que, en general cada una de las características del dominio M poseen las mejores relevancias individuales con todas sobre el 92 %, mientras que las características de los espacios S y W presentan los desempeños más bajos ya que ninguna supera el 75 %. Las desviaciones estándar de las características del dominio M son las más bajas al no superar el 12 %, mientras que las desviaciones del dominio W siempre están sobre el 17 % para todas las características menos para la número 3. Las desviaciones más altas se presentan para las características F11 y F12 con 28.3 % y 29.1 % respectivamente. Los espacios T y F poseen características con eficiencias individuales que varían desde el 37.9 % al 90.3 % para T y desde el 10.3 % al 90.2 % para F.

Análisis de fusión de dominios

Con el fin de establecer el comportamiento por niveles de degradación, la tabla 10 muestra la eficiencia y desviación como valores decimales, según los diferentes tipos de fallos, niveles de severidad y fusión de dominios de representación, entrenando con 4 estados HMM. La columna $T(s)$ indica el tiempo total en segundos para el conjunto de 30 iteraciones. Los resultados para dominios de representación individuales son los mismos mostrados en la figura 12.1, igualmente solo se muestran fusiones significativas de dominios.

Dominio	$T(s)$	B			I			Normal	O			Total
		N1	N2	N3	N1	N2	N3		N1	N2	N3	
TD+FD	360.413	0.840 ±0.184	1.000 ±0.000	0.840 ±0.246	1.000 ±0.000	0.900 ±0.194	0.940 ±0.190	1.000 ±0.000	0.980 ±0.063	0.920 ±0.140	0.960 ±0.084	0.938 ±0.062
TD+M	321.685	0.860 ±0.212	0.920 ±0.103	0.640 ±0.263	1.000 ±0.000	0.820 ±0.220	0.980 ±0.063	1.000 ±0.000	0.960 ±0.084	0.820 ±0.175	0.980 ±0.063	0.898 ±0.115
TD+W	356.273	0.860 ±0.165	1.000 ±0.000	0.620 ±0.290	1.000 ±0.000	0.940 ±0.135	0.960 ±0.084	1.000 ±0.000	0.940 ±0.135	0.880 ±0.140	0.960 ±0.084	0.916 ±0.115
FD+M	361.258	0.860 ±0.135	0.960 ±0.084	0.840 ±0.158	1.000 ±0.000	0.960 ±0.127	0.940 ±0.135	0.980 ±0.063	0.900 ±0.170	0.860 ±0.190	1.000 ±0.000	0.930 ±0.061
FD+W	414.771	0.840 ±0.207	0.980 ±0.063	0.900 ±0.194	1.000 ±0.000	0.920 ±0.193	0.740 ±0.284	1.000 ±0.000	0.900 ±0.105	0.820 ±0.199	1.000 ±0.000	0.910 ±0.089
M+W	378.088	0.920 ±0.103	1.000 ±0.000	0.660 ±0.232	0.880 ±0.169	0.680 ±0.301	0.840 ±0.227	1.000 ±0.000	0.980 ±0.063	0.760 ±0.227	0.740 ±0.268	0.846 ±0.130
TD+FD+ M	355.367	0.740 ±0.232	0.980 ±0.063	0.780 ±0.290	1.000 ±0.000	0.840 ±0.207	0.980 ±0.063	1.000 ±0.000	0.940 ±0.135	0.880 ±0.140	0.980 ±0.063	0.912 ±0.096
TD+FD+ +W	402.801	0.740 ±0.232	0.960 ±0.084	0.760 ±0.246	1.000 ±0.000	0.760 ±0.158	0.960 ±0.084	1.000 ±0.000	0.960 ±0.127	0.840 ±0.127	0.980 ±0.063	0.896 ±0.108
TD+M+ W	384.791	0.860 ±0.190	0.980 ±0.063	0.740 ±0.250	1.000 ±0.000	0.900 ±0.141	0.980 ±0.063	1.000 ±0.000	1.000 ±0.000	0.840 ±0.158	0.940 ±0.097	0.924 ±0.088
FD+M+ W	399.210	0.820 ±0.382	1.000 ±0.000	0.880 ±0.140	1.000 ±0.000	0.900 ±0.141	0.840 ±0.158	1.000 ±0.000	0.960 ±0.127	0.840 ±0.158	1.000 ±0.000	0.924 ±0.076
TD+FD+ M+W	402.731	0.840 ±0.263	0.980 ±0.063	0.740 ±0.212	1.000 ±0.000	0.820 ±0.175	0.960 ±0.127	1.000 ±0.000	0.940 ±0.135	0.880 ±0.140	0.960 ±0.084	0.912 ±0.088

Table 10: Relevancia por fusión de dominios, con degradación

La tabla 10 muestra resultados de eficiencia de forma globalizada e individualizada según los diferentes tipos de fallos y niveles de severidad. En estos resultados se puede observar que al conjugar los dominios de representación TD+FD y FD+M se logra obtener eficiencias totales del 93.8 % con desviación del 6.2 % y del 93.0 % con

desviación 6.1 %, respectivamente. Es de resaltar, igualmente, los resultados obtenidos con la conjugación de los dominios TD+M+W y FD+M+W, hecho que verifica la necesidad de fusión de dominios de representación donde cada uno puede hacer aportes discriminantes. En contraste existen eficiencias individuales bajas, como es el caso de W. Se resalta además como el estado “Normal” muestra una eficiencia perfecta en todas las conjugaciones de dominios, menos en FD+M, lo cual hace que la decisión binaria de fallo y no fallo sea perfectamente factible, sin embargo, al considerar los niveles de severidad, B N1, B N3, I N2 y O N2 se presentan variaciones marcadas en la eficiencia y desviación. En cuanto al tiempo de entrenamiento, cuando se fusiona M con los otros dominios, el orden del tiempo de entrenamiento decae un orden de magnitud a valores que siempre están bajo los 1000 s, en contraste con los resultados del dominio M, lo cual indica un beneficio de la fusión.

Análisis Leave-One-Out por dominio

Se realiza entrenamiento en cada espacio bajo un esquema Leave-One-Out, iniciando con todas las características y progresivamente retirando la de menor eficiencia, según resultados descritos en la tabla 9. Es así como, por ejemplo, para el dominio TD se inicia con las 17 características, luego se retira T1, luego T3, luego T4, luego T11, y así hasta que queda T2, que es la de mayor relevancia. Estos resultados se muestran en la figura 12.2, donde el eje de las abscisas inicia con todas las características y progresivamente muestra el resultado al retirar la primera característica menos relevante (-1c), la segunda menos relevante (-2c) y así sucesivamente.

La figura 12.2 muestra resultados producto de crear grupos de características de un mismo dominio. Es así como se puede observar que, al retirar las dos características con menor relevancia en W la eficiencia se incrementa, pero en S no hay cambios significativos al retirarlas. En FD y M la eficiencia se mantiene aproximadamente constante, pero en TD la eficiencia es constante hasta retirar la quinta característica con menos eficiencia y decae con las subsiguientes.

12.3. Discusión de resultados para selección y fusión de dominios

En general, las eficiencias de diagnóstico (fallos sin severidad) enseñan que los mejores resultados se presentan para modelos HMM entre 1 y 4 estados, y que para números de estados mayores decrece, lo cual se relaciona directamente con sobreentrenamiento del modelo. Sin embargo, en el dominio de representación M sí se

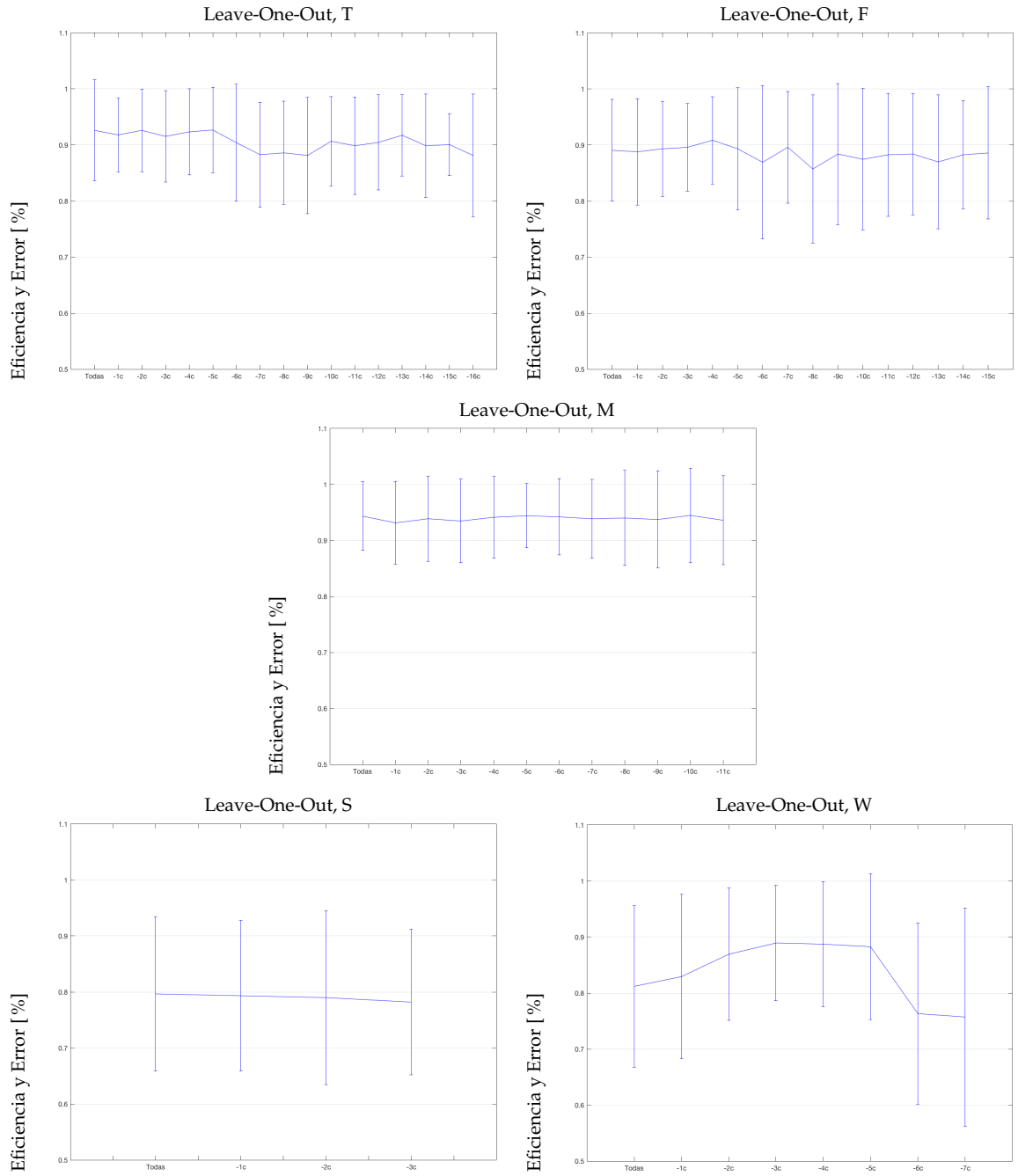


Figura 12.2: Análisis Leave-One-Out por dominio

aprecian mejoras cuando se incrementa el número de estados, pero se observa que esto compromete seriamente el tiempo de cómputo para entrenamiento, llegando a ser de un orden de magnitud superior a los demás dominios (ver figura 12.1). Además, la eficiencia general de identificación de fallos sin severidad, respecto al incremento del número de estados HMM, tiene comportamientos variados entre los diferentes dominios, es así como en W permanece constante, en S decrece, en TD y FD decae ligeramente y M mejora y supera considerablemente a los demás registros. Para todos los casos, se puede generalizar que a menor número de estados HMM menor tiempo de entrenamiento, especialmente para 1, 2 y 3 estados.

Desde el punto de vista de las características al interior de un dominio, en general las características de M poseen las mejores relevancias individuales y menores desviaciones, mientras que las características de S y W presentan las eficiencias más bajas, con este último enseñando las mayores desviaciones. Los dominios TD y FD poseen características con eficiencias en rangos variados (ver tabla 9).

Cuando se tiene presente fallos con sus niveles de severidad y se realiza la fusión de dominios de representación, las fusiones de TD+FD y FD+M logran las eficiencias más altas y enseñando desviaciones bajas (ver tabla 10). Se resalta la necesidad de fusión de dominios de representación, donde cada uno puede hacer aportes discriminantes, en contraste con desempeños individuales bajos, como es el caso del dominio W. Se destaca que, de las eficiencias obtenidas para el estado "Normal" se hace posible la decisión binaria de fallo y no fallo, sin embargo, al considerar niveles de severidad se presentan variaciones marcadas en la eficiencia y desviación. La fusión también permite reducir tiempo de entrenamiento, como fue el caso de fusionar el dominio M con los otros dominios, reduciendo en un orden de magnitud el tiempo en relación con tener este dominio individual. Esto puede implicar, según la aplicación, decidir si se sacrifica eficiencia en el diagnóstico a cambio de lograr mejoras en el tiempo de entrenamiento, lo cual es crucial para aplicaciones en línea.

El análisis Leave-One-Out permite observar que retirar características con bajo valor discriminante individual tiene efecto variado según el dominio de representación, indicando que puede ser posible una pérdida de generalización al tener en cuenta estas características, tal como se puede apreciar en la figura 12.2. Lo anterior demuestra la importancia de combinar espacios, e igualmente de determinar relevancias de grupos de características, ya que el desempeño global bajo de un dominio, o de una característica, no implica que esos dominios y/o características puedan no aportar información de variabilidad del proceso que está ausente en otros dominios y/o características.

Dentro del proceso de mejora, se debe explorar la selección conjunta de dominios y características, con el fin de mantener la mayor información de variabilidad del proceso sin sacrificar eficiencia, además los esquemas de diagnóstico multi-nivel, o jerárquicos, pueden aportar en la separabilidad cuando se tiene presente niveles de severidad. También es importante explorar el uso conjugado de señales no invasivas de naturaleza variada, tal como es el caso de señales de vibración y señales acústicas.

13. DIAGNÓSTICO DE FALLOS MEDIANTE SEÑALES VIBRO-ACÚSTICAS

Las señales de vibración y acústicas son un apoyo activo en el diagnóstico no destructivo y no invasivo de fallos tempranos en máquinas rotativas. No obstante, las señales acústicas son menos utilizadas debido a su vulnerabilidad a las interferencias externas, lo que dificulta un análisis eficiente y robusto para el monitoreo de condiciones. Por tanto, en este capítulo se propone una metodología para caracterizar diferentes tipos de fallos en rodamientos con base en señales acústicas y de vibración. En primer lugar, la señal se descompone en diferentes componentes espectrales de banda estrecha, aplicando diferentes métodos de banco de filtros, como la EMD, WPT y Bancos de Filtros de Fourier (FFB, del inglés *Fourier Filter Banks*). En segundo lugar, un conjunto de características se construye empleando una medida de similaridad denominada en este trabajo como índice de densidad espectral acumulativa, que se utiliza para estimar la dependencia estadística mutua entre cada componente de ancho de banda limitada y la señal en bruto. Finalmente, se lleva a cabo un esquema de clasificación para distinguir los diferentes tipos de fallos presentes en rodamientos, sin severidades. La solidez del enfoque se valida contaminando la señal con varios niveles de ruido gaussiano blanco aditivo.

13.1. Métodos de banco de filtros

Con el fin de separar la información proveniente de sub-bandas espectrales, los FBM descomponen señales en anchos de banda limitados en un conjunto de componentes de banda estrecha. Así, dada una señal $x(t) \in \mathbb{R}(T)$ que posee un ancho de banda finito ΔF (con $F = [0, 1/2\Delta t]$, donde $1/2\Delta t$ es la frecuencia de muestreo) se descompone en $B \in \mathbb{N}$ componentes de banda estrecha $\varphi = \{\varphi_b(t): b \in B\}$ de forma que cada uno posee un ancho de banda ΔF_b tal que $F_b \subseteq F$. Independientemente de sí el tipo de análisis es por señales de vibración o acústicas, se pueden emplear los siguientes enfoques por FBM:

- Descomposición por FFB. La transformada de Fourier provee un enfoque directo para evaluar FBM de forma que varios filtros pasa banda ideales se aplican a la señal de entrada $x(t) \in \mathbb{R}(T)$ con el fin de construir el conjunto de componentes $\varphi = \{\varphi_b(t): b \in B\}$ tal que el soporte de $\varphi_b(f)$ iguala la cerradura del correspondiente ancho de banda, notado como \overline{F}_b . Con el ánimo de obtener

un FBM ortogonal, es necesario que todos los anchos de banda sean disyuntos, esto es, $F_b \cap F_d = \emptyset$, con $(b \neq d) \in B$.

- WPT. La transformada wavelet se puede expresar en términos del producto interno entre cada base $v_{j,k}^n(t) \in \mathbb{R}(T)$ y la respectiva señal $x(t)$. Así, los coeficientes wavelet $w_{j,n,k} \in \mathbb{R}(T)$ son el producto interno $\langle x, w_{j,n,k} \rangle \in \mathbb{R}(T)$ definido como $w_{j,n,k} \int_T x(t) v_{j,k}^n(t) dt$, donde $n \in \mathbb{N} \cup 0$ es el parámetro de oscilación y $j, k \in \mathbb{Z}$ son los índices de la operación de escala y traslación, respectivamente. Por tanto, una función base WPT $v_{j,k}^n(t)$ se define como $v_{j,k}^n(t) = 2^{j/2} v^n(2^j t - k)$, donde la inicialización ($n = 0$) y posterior recursión ($n > 1$) se pueden consultar en [132].
- EMD. Este FBM es un método de análisis de señales tiempo-frecuencia auto-adaptativo con base en escalas temporales locales. EMD puede descomponer una señal compleja e inestable en varias Funciones de Modo Intrínseco (IMF, del inglés *Intrinsic Mode Functions*) casi estables, tal como se muestra en [133].

13.2. Caracterización por medidas de similaridad

Con el objetivo de mejorar la capacidad de discriminación del conjunto de características, se mide la similaridad entre la señal de entrada $x(t) \in \mathbb{R}(T)$ y cada b -ésimo componente de banda estrecha obtenido, $\varphi_b(t) \in \mathbb{R}(T)$, cuantificando su dependencia estadística mutua.

Un enfoque directo para medir la similaridad estadística entre dos procesos es su producto momento, o PCC $\rho_{x,\varphi_b} \in \mathbb{R}[-1, 1]$, que cuantifica la relación lineal de la dependencia de la siguiente manera:

$$\rho_{x,\varphi_b} = \frac{\mathbf{E} \{ x(t) \varphi_b^*(t) \}}{\sigma_x \sigma_{\varphi_b}} \quad (13.1)$$

donde $\sigma_{\xi}^2 = \mathbf{E} \{ |\xi(t)|^2 : t \in T \}$ es la varianza, con $\sigma_{\xi}^2 \in \mathbb{R}^+$. Las anotaciones $\mathbf{E} \{ \cdot \}$ y $(\cdot)^*$ representan los operadores de esperanza y conjugado, respectivamente. Se debe notar que, tanto $x(t)$ como $\varphi_b(t)$ deben ser procesos estocásticos de media cero.

En un enfoque más elaborado, la dependencia estadística entre dos procesos puede estimarse utilizando una medida de información. En particular, se puede emplear la divergencia Kullback-Leibler, Análisis Espectral de Entropía Cruzada (CESA, del

inglés *Cross-Entropy Spectral Analysis*) que mide la diferencia entre un par de PDF, $p_x(u)$ y $p_{\varphi_b}(u)$, de la siguiente manera:

$$r_{x,\varphi_b} = \int_T p_x(u) \ln \left(\frac{p_x(u)}{p_{\varphi_b}(u)} \right) du, \quad r \in \mathbb{R}^+ \quad (13.2)$$

donde u es una muestra i.i.d. de cada una de las distribuciones. Se debe notar que r_{x,φ_b} no es una distancia en sentido estricto ya que no es simétrica y no satisface la desigualdad del triángulo. Aún así, con frecuencia sirve como una estimación de medida de similaridad entre densidades [134]. Sin embargo, cualquiera de las dos medidas presentadas, r_{x,φ_b} o ρ_{x,φ_b} , asumen la estacionariedad de los procesos correlacionados.

En la práctica, la anterior suposición está lejos de ser cierta para la mayoría de los datos de señales de vibración o acústicas. Además, la mayoría de las técnicas actuales para la detección de fallos recaen sobre el análisis espectral de las señales medidas. Con el propósito de proporcionar el soporte analítico para procesos no estacionarios y con significado espectral, se introduce la medida de similaridad denominada Índice de Densidad Espectral Cumulativa (CSDI, del inglés *Cumulative Spectral Density Index*) que se basa en la función de correlación cruzada calculada entre dos procesos no estacionarios de segundo orden. Se define $S_{x\varphi_b}(f)$ como la distribución del contenido de la señal en el dominio de la frecuencia que se representará mediante la densidad espectral cruzada entre $[x(t), \varphi_b(t)]$, por lo que se evalúa la transformada de Fourier de la función en tiempo promedio $R_{x\varphi_b}(t, \tau) \in \mathbb{R}^+$ como sigue:

$$S_{x\varphi_b}(f) = \int_{\mathbb{R}} \bar{R}_{x\varphi_b}(t) \exp^{-j2\pi ft} dt \quad (13.3)$$

donde

$$\bar{R}_{x\varphi_b}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} R_{x\varphi_b}(t, \tau) dt \quad (13.4)$$

Sin embargo, la densidad de valores complejos $S_{x\varphi_b}(f) \in \mathbb{C}$ debe asignarse a un dominio de valores reales. Para este fin, se calcula el valor esperado del módulo al cuadrado $\mathbf{E} \{ |\cdot|^2 \}$ como un operador semidefinido positivo. Por lo tanto, se obtiene una medida de la información espectral compartida entre la señal $x(t)$ y su conjunto de componentes de banda estrecha $\{\varphi_b(t): b \in B\}$. No obstante, la densidad espectral puede contener componentes que se concentran localmente, lo que puede sesgar

fuertemente al estimador del operador de esperanza [135]. Para superar este problema, se introduce el valor esperado sobre la función de energía acumulada en la definición de CSDI de la siguiente manera:

$$\rho_{x,\varphi_b} = \mathbf{E} \left\{ \int_0^f |S_{x\varphi_b}(\bar{f})|^2 d\bar{f} : f \in F_b \right\} \quad (13.5)$$

Se debe notar que, entre más alto sea el valor de similaridad para $\rho_{x,\varphi_b} \in \mathbb{R}^+$, o para $\rho_{x,\varphi_b} \in \mathbb{R}^+$, más alta es la asociación estadística entre variables. En contraste, valores menores de ρ_{x,φ_b} implican una relación cerrada, ya que este puede explicarse como la distancia entre ambas funciones de densidad de probabilidad bajo análisis.

13.3. Marco experimental para diagnóstico de fallos mediante señales vibro-acústicas

La metodología propuesta para la detección de fallos empleando señales de vibración y acústicas comprende las siguientes etapas: (i) realzamiento de señal, aplicando uno de los métodos descritos de descomposición por banco de filtros (EMD, WPT o FFB), (ii) extracción de características, probando cada una de las medidas propuestas de similaridad (PCC, CESA y CSDI) entre los datos medidos y sus componentes de descomposición de banda estrecha, y (iii) validación del desempeño de un clasificador que se alimenta del conjunto de características. Además, se emplea un esquema de clasificación Leave-One-Out para diferentes niveles de SNR, con el fin de validar la robustez y eficiencia de la metodología propuesta. Ya que el objetivo de este experimento se centra en la extracción de características desde diferentes señales de fuente diferente, se emplea el algoritmo de un clasificador básico $k - nn$. Varias razones explican el uso generalizado de este clasificador: es sencillo de implementar, conduce a un rendimiento de reconocimiento confiable (gracias a la no linealidad de sus límites de decisión), y se supone que su complejidad es independiente del número de clases. Por lo tanto, la afinación del clasificador $k - nn$ se lleva a cabo mediante la regla de pulgar (es decir, la raíz cuadrada del número de instancias que tiene cada clase).

La metodología se valida sobre una plataforma de pruebas para rodamientos consistente en un eje impulsado por un motor eléctrico de C.C. de 1.5 HP que puede alcanzar 1720 R.P.M. a través de un acoplamiento rígido. Se consideran tres clases de fallos en rodamientos, así: fallo en cara interna (Inner - I), fallo en cara externa

(Outer - O) fallo en elemento rodante (Ball - B); además se registra la señal de operación sin fallo (Normal). La base de datos posee 40 registros acústicos con duración de tres segundos, a una frecuencia de muestreo de 44.1 kHz y 40 registros de vibración que se adquieren simultáneamente, con una duración de cuatro segundos a una frecuencia de muestreo de 25.6 kHz. Las señales acústicas fueron adquiridas por dos micrófonos ubicados a una distancia de 2 cm delante de cada alojamiento del rodamiento. Simultáneamente, las grabaciones de vibración se recogieron en el plano horizontal, empleando varios acelerómetros montados en soportes del rodamiento ya que el daño simulado se colocó en el extremo del eje. Además, las señales se sub-muestraron ocho veces a 3200 Hz y 5512 Hz para ambos tipos de señales, respectivamente.

Se realizan los procedimientos descritos a continuación, corriendo algoritmos en una máquina Intel® Core™ i7-4510, CPU a 2.00 GHz, con 8 GB RAM, S.O. Windows 8.1 de 64 bits.

13.4. Resultados para diagnóstico de fallos mediante señales vibro-acústicas

Inicialmente, la figura 13.1 muestra algunos ejemplos de representación tiempo-frecuencia para los diferentes tipos de fallos en rodamientos y la señal de operación sin fallo (Normal) para señales de vibración y acústicas.

Desde la base de datos descrita, se implementan los enfoques FBM, fijando $b = 8$ como número considerado de sub-bandas estrechas. La figura 13.2 muestra algunos ejemplos de la descomposición de la señal realizada para la condición normal (parte superior) y fallo en la cara externa (parte inferior). Como se observa, la mejora de la señal proporcionada por WPT permite distinguir visualmente el defecto de la cara externa en los componentes de banda estrecha a alta frecuencia, ya que esos componentes exhiben el comportamiento impulsivo que caracteriza este tipo de daño. No obstante, se debe notar que a menor es el cardinal del ancho de banda estrecho, mayor es el contenido espectral. El clasificador $k - nn$ se emplea con cinco vecinos, usando un esquema de clasificación Leave-One-Out.

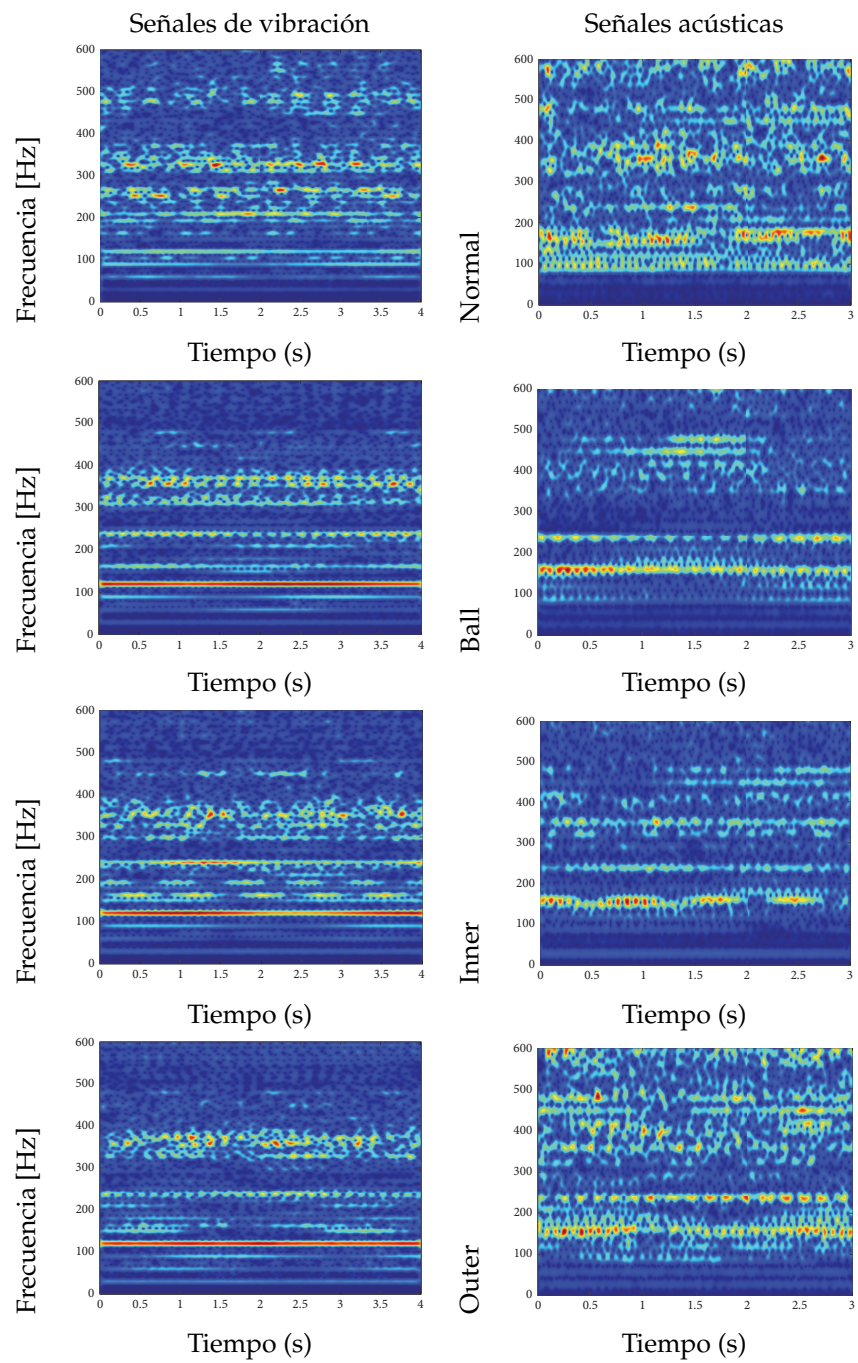


Figura 13.1: Representación tiempo-frecuencia para señales de vibración y acústicas

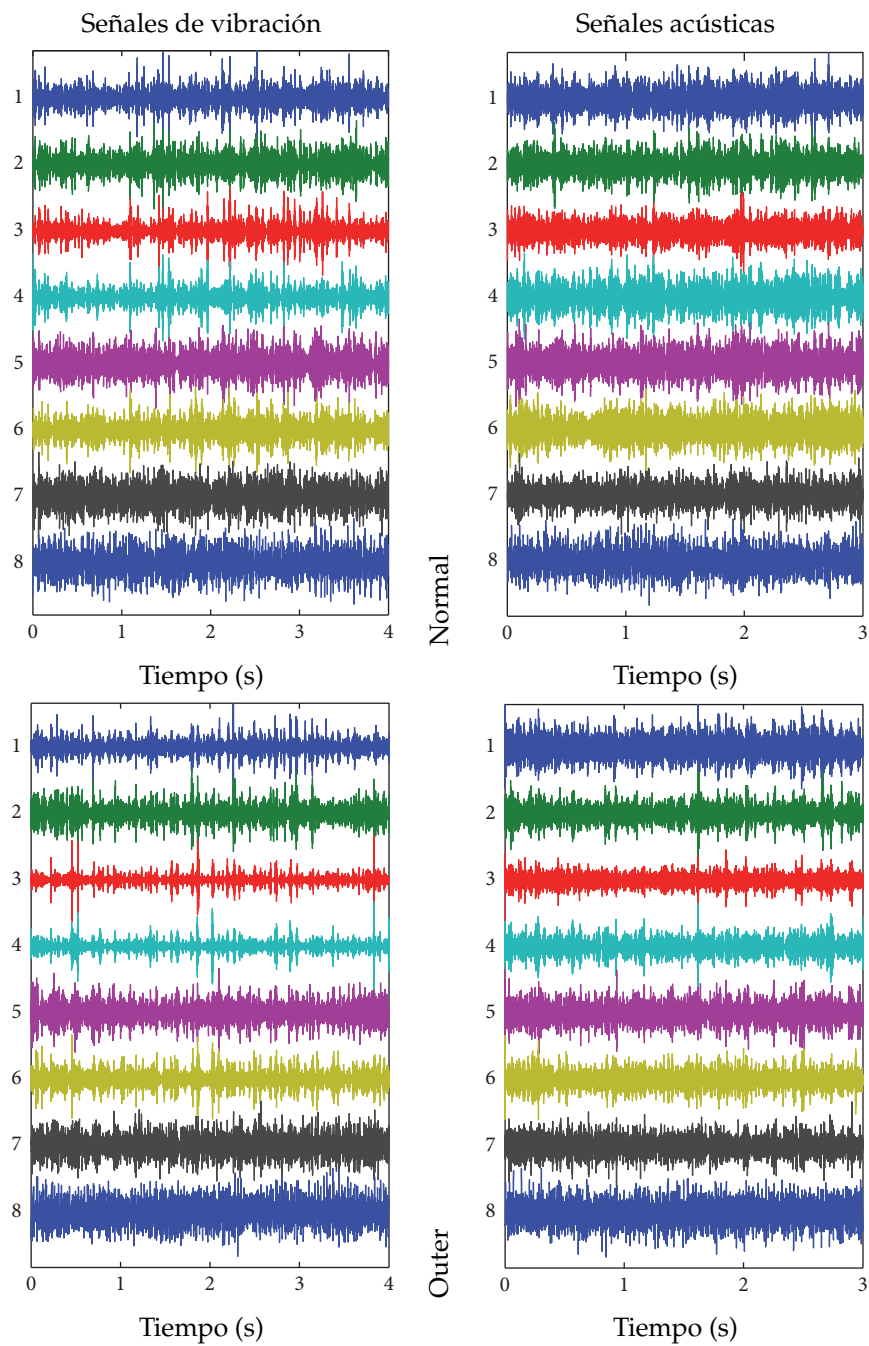
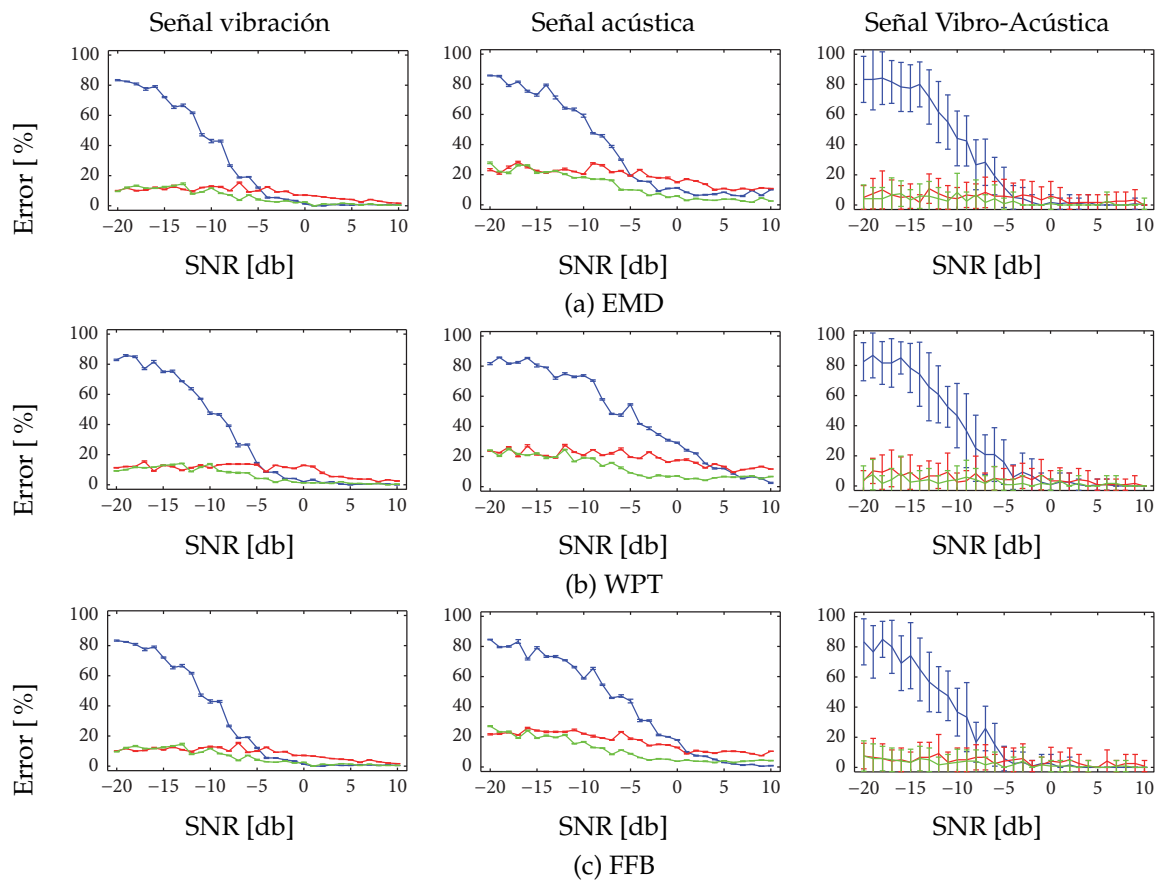


Figura 13.2: Descomposición WPT para señales de vibración y acústicas

Además, la figura 13.3 muestra que cada medida de similitud considerada produce un conjunto de características que proporciona una eficiencia de clasificador equivalente a una SNR muy alta (más allá de 9 dB) independientemente del análisis de señal FBM empleando. Sin embargo, el análisis acústico se comporta peor cuando el nivel de SNR disminuye, de modo que, a valores muy bajos de SNR las características que usan CSDI y entropía cruzada superan claramente la línea base dada por PCC. Además, la combinación de señales acústicas y de vibración disminuye el error de clasificación para valores de SNR por debajo de -5 dB.



Medidas de similitud: Correlación (Azul), Entropía-cruzada (Rojo), CSDI (Verde)

Figura 13.3: Error de clasificación según FBM, similitud, tipo de señal y SNR

Con el propósito de determinar la mejor combinación de FBM y medida de similitud, se consideran los resultados realizados a un nivel de SNR de -20 dB, ya que es el peor escenario probado cuando el ruido contamina altamente la entrada. Por lo tanto, la tabla 11 resume los mejores resultados para cada tipo de combinación de señales. Como se observa, cualquiera de las medidas (entropía-cruzada o CSDI) puede alcanzar la máxima eficiencia usando un solo tipo de datos, vibración o acústica.

Señales	Medida de similitud	EMD	WPT	FFB
Vibración	Correlación	83.3±0.81	82.9±0.45	83.3±0.37
	Entropía-cruzada	14.6±0.37	11.2±0.2	10±0.2
	CSDI	11.2±0	9.2±0.2	9.6±0.2
	Estadísticas	85±0.8 ⁽⁷⁾	85±0.8 ⁽⁸⁾	85±0.2 ⁽⁸⁾
Acústicas	Correlación	85.8±0.29	81.7±0.68	84.6±0.23
	Entropía-cruzada	23.3±0.68	23.7±0.23	24.7±0.42
	CSDI	27.9±0.54	24.2±0.19	27.1±0.35
	Estadísticas	85±0.68 ⁽⁷⁾	83.3±0.29 ⁽⁸⁾	85±0.29 ⁽⁸⁾
Vib.+Acú.	Correlación	83.3±15.3	82.5±12.65	83.3±15.29
	Entropía-cruzada	7.5±7.4	10±6.84	7.5±7.4
	CSDI	4.2±3.7	4.2±3.7	7.5±7.4
	Estadísticas	84.2±13.75 ⁽⁷⁾	82.9±12.65 ⁽⁸⁾	84.2±15.29 ⁽⁸⁾

Tabla 11: Errores de clasificación según FBM y similitud, con SNR = -20 dB

Otro aspecto del entrenamiento a considerar es la estrategia de clasificación para el caso de combinar ambos tipos de datos de forma simultánea. En aras de la simplicidad, se considera una combinación apilada de clasificadores (es decir, combinador de productos), que arroja un aumento notable de la eficiencia, pero solo para valores muy bajos de SNR. De lo contrario, cada medida considerada funciona bien. En aras de la comparación, el rendimiento de la clasificación se calcula mediante 24 características estadísticas tradicionales de la señal en el tiempo, como la media, la desviación estándar y la asimetría [136]. Se estiman las características de cada componente de banda estrecha, y luego, las componentes se agregan sucesivamente hasta que se obtiene la mejor precisión (el superíndice en la tabla 11 indica la cantidad de componentes considerados).

13.5. Discusión de resultados para diagnóstico de fallos mediante señales vibro-acústicas

En este apartado se presenta una metodología para la detección de fallos sobre señales de vibración y acústicas, explorando la combinación de estas señales, su realzamiento y caracterización. Se encuentra que, aunque las señales adquiridas corresponden a una máquina a velocidad constante, es posible encontrar comportamientos estacionarios o no estacionarios, generados por fallos definidos de la máquina. Sin embargo, las vibraciones y las señales acústicas están influenciadas de manera diferente por estos factores externos. Es así que, las señales de vibración muestran un componente estacionario asociado con una perturbación del motor eléctrico que oculta la frecuencia del rodamiento debido a su alta energía, por el contrario, en las señales acústicas esta perturbación externa muestra un comportamiento no estacionario que es similar a los fallos en los rodamientos, como se aprecia en la figura 13.1. Por lo tanto, las señales acústicas son más vulnerables a las interferencias externas que las señales de vibración.

Para el tipo de dispositivo mecánico analizado, pueden existir algunas energías espectralmente localizadas sobre sus sub-bandas estrechas extraídas, incluso si los fallos se comportan de manera no estacionaria como es el caso para rodamientos. En consecuencia, es importante el uso de FBM para llevar a cabo la etapa de pre-procesamiento de señales acústicas y de vibración. Como resultado, los tres FBM considerados (EMD, WPT y FFB) permiten extraer información precisa en el caso de velocidad de eje constante, proporcionando una carga computacional, y de complejidad, comparable en su implementación.

Con el objetivo de mejorar la capacidad de discriminación del conjunto de características, se prueban tres medidas de similaridad entre la señal de entrada y cada componente extraído de banda estrecha. Específicamente, el índice PCC funciona bien en casos de valores de SNR grandes (por encima de 7 dB). Sin embargo, estos casos de datos sin ruido están lejos de ser reales en las condiciones habituales de operación de maquinaria, por lo que la mejor característica de similaridad es el índice CSDI que supera a otros índices en cuanto a los valores de SNR por debajo de -5 dB independientemente del FBM empleado (ver figura 13.3), además, CSDI produce la mejor eficiencia con el beneficio de una carga computacional similar a los otros índices probados.

CSDI permite lograr la mejor eficiencia a niveles bajos de SNR, independientemente del tipo de señal de entrada. Además, las medidas de entropía cruzada y CSDI

logran una eficiencia comparable para las señales acústicas. En cualquier caso, WPT es el mejor conjunto de características, tal como se puede observar en la tabla 11. Sin embargo, otro aspecto importante es que el análisis con señales vibro-acústico alcanza las tasas de eficiencia en clasificación más altas usando medidas de entropía cruzada o CSDI (ver figura 13.3).

14. METODOLOGÍA DE SINTONIZACIÓN DE MODELO HMM CON OBSERVACIÓN DISCRETA

Avanzando en el proceso de diagnóstico y pronóstico, se plantea una metodología para la sintonización de parámetros en el entrenamiento de un modelo HMM con observación discreta, empleando únicamente características desde el dominio de representación MFCC, ya que este dominio ha demostrado que individualmente es el que entrega la mayor eficiencia para las tareas combinadas de diagnóstico (identificación de fallos) y de pronóstico (identificación de fallos y niveles de severidad). Se emplean como señales de estudio datos provenientes de señales de vibración desde rodamientos y desde dos bases de datos diferentes, con cuatro escenarios de operación, a saber: normal, fallo en pista interna, fallo en pista externa y fallo en bola. Una de las bases permite diferenciar tres niveles de severidad para cada uno de los escenarios de operación. Las características son discretizadas empleando agrupamiento por k -medias.

En este análisis no se emplean señales acústicas, ya que han demostrado ser más vulnerables a las interferencias externas que las señales de vibración, además, las perturbaciones externas que afectan a las señales acústicas muestran un comportamiento no estacionario que es similar al de los fallos en los rodamientos, lo cual enmascara su información y dificulta un análisis eficiente y robusto para el monitoreo de condiciones con énfasis en identificación de niveles de severidad.

14.1. Marco experimental para metodología de sintonización de modelo HMM con observación discreta

Con el fin de realizar el entrenamiento y validación del sistema propuesto se emplean dos bases de datos diferentes. La primera base de datos (BD1) es la desarrollada por el Bearing Data Center del Case Western Reserve University [121] y descrita ya previamente. La segunda base de datos (BD2) fue desarrollada por el Laboratorio de Vibraciones Mecánicas de la Universidad Tecnológica de Pereira, con adquisición a una frecuencia de muestreo de 20 kHz y donde se inducen fallos mediante acción mecanizada en la bola, cara interna y cara externa del rodamiento. Cada una de estas bases de datos conforma el conjunto de N señales de vibración $\{\mathbf{Z}_n, y_n\}_{n=1}^N$ con etiquetas de salida $y_n \in \{1, 2, \dots, C\}$ relacionadas con C clases para diagnóstico de fallos o severidades de un rodamiento.

Con el fin de crear bases de datos anotadas, en la BD1 se designan las etiquetas B para fallos en la bola, I para fallos en la cara interna, O para fallos en la cara externa, Normal para la señal base y etiquetas N1, N2 y N3 para los tres niveles de severidad, donde 1 es el nivel bajo y 3 el alto, por lo que la base resultante consta entonces de $C = 10$ carpetas, o clases. En la BD2 solo se cuenta con señales distinguidas por tipos de fallos, sin niveles de severidad, por lo que las etiquetas solo corresponden a cuatro carpetas identificadas como Normal, B, I y O, en igual concordancia con BD1, pero donde $C = 4$ carpetas, o clases.

Sobre las señales presentes, se realiza segmentación en marcos de longitud igual a $T = 200$ ms y con traslape entre marcos de $1/3$. Posteriormente se emplea una ventana Hamming con el fin de ajustar los marcos e integrar las líneas cercanas en frecuencia. Producto de lo anterior, cada señal $\mathbf{Z}_n \in \mathbb{R}^{W \times T}$ posee ahora W segmentos en las filas $\mathbf{z}_{w,n} \in \mathbb{R}^T$, a T instantes de tiempo ($w \in \{1, 2, \dots, W\}$), La subsecuente etapa es específicamente la extracción de los coeficientes MFCC, donde se obtiene un vector que representa una escala continua para cada una de las características, entregando una matriz de características en MFCC $\{\mathbf{X} \in \mathbb{R}^{W \times M}\}_{n=1}^N$, donde M es el número de coeficientes que se extraen. Para entrenar un modelo HMM, con observaciones discretas, se requiere realizar una discretización de las citadas características, para lo cual se emplea el algoritmo de k -medias, y desde el cual se obtiene la matriz de características en el dominio MFCC discretizada $\{\mathbf{X}_n^*, y_n\}_{n=1}^N$.

Con base en las características discretizadas, se procede con el entrenamiento de los modelos HMM, al cual se le desea ajustar sus parámetros. Para evaluar los diferentes modelos HMM que mejor representen a las observaciones, se tomó las curvas de (Característica Operativa del Receptor (ROC, del inglés *Receiver Operating Characteristic*), que son ampliamente aceptadas como método de análisis y comparación de pruebas diagnósticas [137, 138, 139]. Se emplea la sensibilidad $S_E(x_o)$ y la especificidad $S_P(x_o)$ como medidas de la exactitud diagnóstica de una prueba sobre un punto de corte (x_o) a partir del cual el resultado se considera positivo; donde la sensibilidad se define como la fracción de verdaderos positivos (TPR, del inglés *True Positive Rate*) con $TPR = S_E(x_o) = P(X > x_o | C = positiva)$ y representa la fracción de observaciones positivas que efectivamente arrojan un resultado positivo de la prueba. Por otra parte, la especificidad se define como la fracción de verdaderos negativos (TNR, del inglés *True Negative Rate*) con $TNR = S_P(x_o) = P(X \leq x_o | C = negativa)$ y representa la fracción de observaciones negativas que efectivamente arrojan un resultado negativo de la prueba.

Para una población de positivos X_P y otra de negativos X_N con PDF F_P y F_N respectivamente, y donde se fija un punto de corte x_o , la sensibilidad estará dada por $S_E(x_o) = 1 - F_P(x_o) = 1 - P(X_P > x_o)$ y la especificidad $S_P(x_o) = F_N(x_o) = P(X_N \leq x_o)$. La curva ROC estará dada por las coordenadas $(1 - F_N(x_o), 1 - F_P(x_o))$ para $x_o \in \mathbb{R}$, o para una función $\mathfrak{R}(p)$ que a cada $p \in [0, 1]$ le asocie $\mathfrak{R}(p) = P(1 - F_N(X_P) \leq p)$ [140]. Entonces, la curva ROC se obtiene de realizar un muestreo Montecarlo sobre el conjunto solución de los parámetros de sintonización del modelo HMM. Cuando la sintonización de un modelo distinga claramente entre observaciones positivas y negativas la sensibilidad será de 1 y la especificidad 0, (1-especificidad = 1). Cuando la sintonización no sea capaz de discernir las observaciones, entonces la sensibilidad será de 0 y la especificidad de 1 (1-especificidad = 0). Luego, el área bajo la curva ROC se puede emplear como medida de la exactitud diagnóstica y como criterio de comparación entre modelos [137, 139].

En el caso específico de los modelos HMM entrenados, se seleccionó como criterio distinguible de un modelo el número de estados y como parámetros de sintonización el número de coeficientes MFCC, el número de filtros Mel y el número de centroides para discretización por k -medias. Los rangos de cada variable incluyen los siguientes valores:

- Coeficientes: 1, 2, 3, 4, 8, 10, 12, 14, 16, 20.
- Filtros: 1, 2, 3, 4, 8, 10, 12, 14, 16, 20, 24.
- Centroides: 2, 4, 6, 8, 10, 12, 14, 16, 20, 24.

Se realizan los procedimientos descritos a continuación, corriendo algoritmos en una máquina Intel® Core™ i7-4510, CPU a 2.00 GHz, con 8 GB RAM, S.O. Windows 8.1 de 64 bits.

14.2. Resultados para metodología de sintonización de modelo HMM con observación discreta

De los números totales de parámetros para sintonización, se deduce que la curva ROC se construye con un total de 1100 puntos de muestreo Montecarlo, donde cada punto corresponde a valores de sensibilidad y especificidad para un modelo entrenado con valores específicos de parámetros de sintonización. A cada curva se adicionan taxativamente los puntos en coordenadas $(0, 0)$ y $(1, 1)$.

Las figuras 14.1 y 14.2 muestran las curvas ROC para modelos HMM con 1, 2, 3, 4 y 5 estados, empleando la base de datos identificada como BD1 con un 70 % para entrenamiento y el resto para validación. La figura 14.1 enseña los resultados cuando la base de datos presenta separación incluyendo niveles de severidad ($C = 10$), mientras la figura 14.2 muestra los resultados cuando la base de datos presenta separación con solo estados de fallo ($C = 4$), sin tener presente la severidad. La figura 14.3 enseña los resultados para las curvas ROC de los modelos HMM cuando se emplea la base de datos BD2 igualmente con un 70 % de la base empleada para entrenamiento; en esta última base de datos solo existen estados de fallo, sin severidad ($C = 4$).

Para entregar una medida de la exactitud diagnóstica, las curvas ROC son suavizadas mediante el método de media móvil y se calcula el área bajo cada curva, arrojando los resultados que se muestran en la figura 14.4 para cada uno de los conjuntos de curvas en las figuras 14.1 a 14.3, pero ampliando el análisis hasta un máximo de 7 estados en los modelos HMM.

Para determinar el número óptimo de estados para el modelo HMM se busca un compromiso entre el área bajo la curva ROC y el número de estados del modelo HMM empleado. Para el caso de la base de datos BD1 con niveles de severidad, la figura 14.4 muestra que se logran máximos locales de área para 3 y 6 estados. Para el caso de la base de datos BD1 sin severidad, se logran máximos locales de área para 5 y 7 estados, sin embargo, teniendo presente que a más estados mayor será la demanda computacional, en general desde 4 estados se logran valores de área en vecindad del máximo. Para el caso de la base de datos BD2, se aprecia un área creciente y con máximo en el valor extremo del rango de análisis de estados, 7 en este caso.

Es de notar que para las bases de datos que no incluyen severidad, el área ROC crece a medida que se incrementa el número de estados HMM, sin embargo, al incluir niveles de severidad el área ROC tiende a decrecer luego de lograr su máximo. A mayor área bajo la curva ROC, mayor es la exactitud diagnóstica, o capacidad de discernimiento entre poblaciones [141]. Aunque algunos autores [142, 143] recomiendan emplear el estadístico no paramétrico de Wilcoxon para analizar las diferencias bajo las áreas, el concepto de área bajo la curva ROC está estrechamente relacionada con Wilcoxon y que no se afecta significativamente por la forma de distribución de las poblaciones, lo cual implica la ventaja de trabajar correctamente en poblaciones con diferentes formas de distribución de probabilidad.

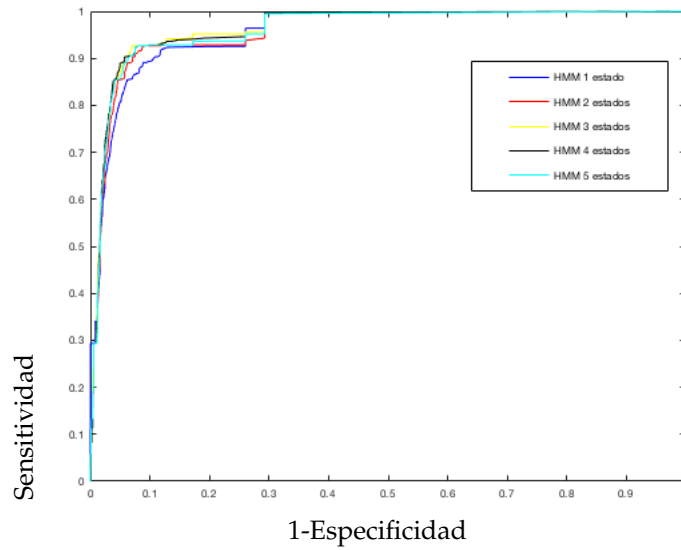


Figura 14.1: Curvas ROC para HMM entrenados con BD1, con severidad

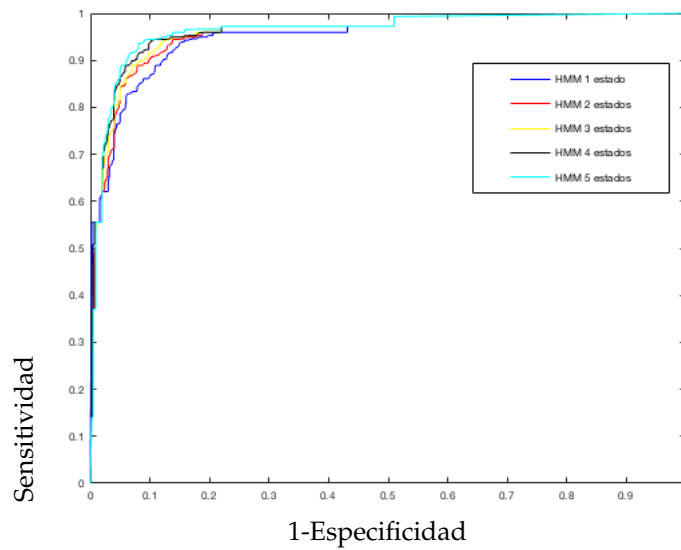


Figura 14.2: Curvas ROC para HMM entrenados con BD1, sin severidad

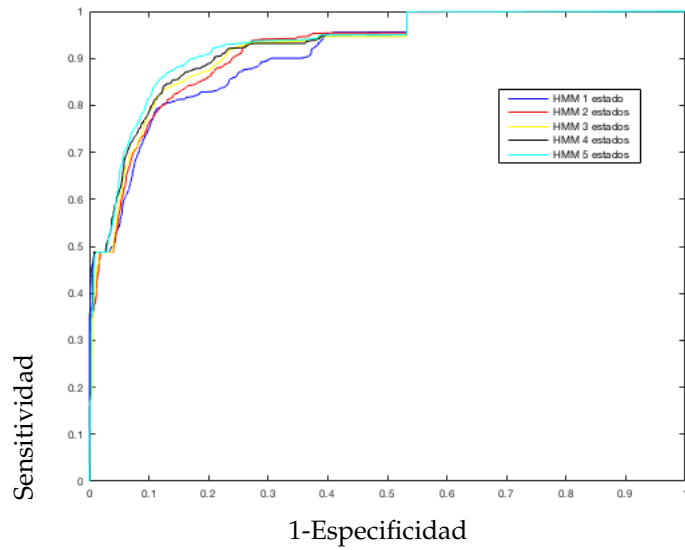


Figura 14.3: Curvas ROC para HMM entrenados con BD2 UTP

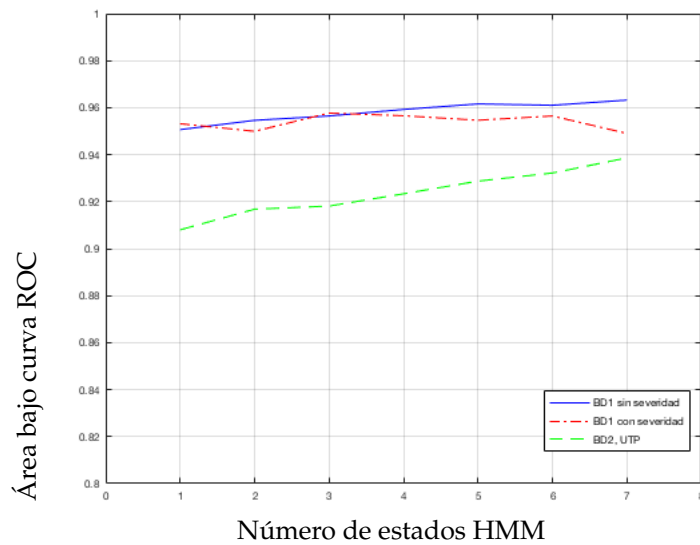


Figura 14.4: Área bajo curvas ROC según base de datos

Teniendo presente los anteriores resultados, y con el objetivo de tener exactitud diagnóstica con menor carga computacional, el número de estados seleccionado en este caso puede ser 3 o 4. Sin embargo, para aplicaciones donde se deba incluir la severidad, incluso 1 o 2 estados puede funcionar.

La evaluación total del modelo HMM, y con el fin de determinar la eficiencia global en la identificación de la degradación y niveles de severidad en rodamientos, se implementa un modelo HMM de 4 estados, con 12 coeficientes MFCC, 24 filtros Mel y 12 centroides para discretización. Cada modelo se entrenó un total de 100 veces y se aseguró independencia entre ejecuciones al realizar, cada vez, la separación de la base de datos para entrenamiento y validación, la extracción de características, la discretización y el entrenamiento del modelo; este tipo de validación se conoce como validación cruzada aleatoria por k-folds, con $k = 100$. Los resultados para cada una de las dos bases de datos descritas se muestran en las tablas 12, 13 y 14, donde se indica el resultado para tres tamaños del conjunto de entrenamiento, 50, 60 y 70 por ciento respectivamente. En estas tablas, cada fila indica el resultado para un tipo específico de fallo, mostrando su confiabilidad y desviación respectiva. La última fila consolida la confiabilidad sobre el proceso total de reconocimiento en los fallos y/o niveles de severidad.

Se destaca nuevamente que la base de datos BD1 se toma desde dos puntos de vista diferentes, una vez como una base con estados de fallo y sin niveles de severidad ($C = 4$ en tabla 12), y otra como una base con estados de fallo y niveles de severidad ($C = 10$ en tabla 13).

	50 %	60 %	70 %
Ball	0.991±0.040	0.974±0.072	0.994±0.039
Inner	0.875±0.172	0.846±0.196	0.808±0.205
Normal	1.000±0	1.000±0	1.000±0
Outer	0.798±0.201	0.829±0.210	0.849±0.201
Total	0.916±0.098	0.912±0.101	0.913±0.107

Tabla 12: Eficiencias modelos HMM, con BD1 sin severidad

		50 %	60 %	70 %
Ball	N1	0.858±0.277	0.906±0.202	0.885±0.223
	N2	0.860±0.307	0.776±0.360	0.798±0.367
	N3	0.962±0.125	0.968±0.120	0.968±0.116
Ball Total		0.893	0.883	0.883
Inner	N1	0.890±0.315	0.940±0.239	0.810±0.394
	N2	0.938±0.192	0.936±0.214	0.930±0.213
	N3	1.000±0	0.998±0.020	1.000±0
Inner Total		0.943	0.958	0.913
Normal		1.000±0	1.000±0	1.000±0
Outer	N1	0.987±0.070	0.986±0.071	0.980±0.116
	N2	1.000±0	1.000±0	1.000±0
	N3	0.870±0.338	0.880±0.327	0.950±0.219
Outer Total		0.952	0.955	0.977
Total		0.937±0.141	0.939±0.133	0.932±0.145

Tabla 13: Eficiencias modelos HMM, con BD1 y severidad

	50 %	60 %	70 %
Ball	0.972±0.055	0.985±0.023	0.981±0.023
Inner	0.935±0.080	0.933±0.070	0.937±0.068
Normal	0.999±0.001	1.000±0	1.000±0
Outer	0.912±0.068	0.905±0.074	0.905±0.074
Total	0.955±0.035	0.956±0.036	0.956±0.036

Tabla 14: Eficiencias modelos HMM, con BD2 UTP

Se resalta de estos resultados que, al variar el porcentaje empleado de las bases de datos para entrenamiento, el resultado total de eficiencia en la identificación de la degradación y severidad no presenta variación significativa, al tener variaciones inferiores a 0.007 en eficiencia y de 0.012 en desviación. Estas variaciones muestran una gran consistencia en los resultados y evidencian el poder de generalización que

se logran sin sufrir de sobre entrenamiento. En las tablas [12](#), [13](#) y [14](#), se puede apreciar que cuando las bases de datos no poseen niveles de severidad, el fallo inducido en los elementos rodantes (B) presenta una eficiencia en su identificación superior al 97 %; cuando la base de datos tiene severidad la eficiencia del proceso total ronda el 93 %, pero con desviaciones totales cercanas al 13 % y en general muy altas desde una perspectiva de eficiencia individual para un estado de severidad, como por ejemplo para fallo tipo I con severidad N1 y 70 % de base para entrenamiento, donde la desviación es del 39.4 %.

Una observación importante sobre la base de datos BD1, ocurre cuando se compara los resultados sin niveles de degradación con los resultados producto de promediar los niveles de degradación de un mismo tipo de fallo. En este caso, el fallo tipo Normal sigue dando 100 %, pero el fallo tipo B pasa de un mínimo de 97.4 % sin degradación a uno de 88.3 % con degradación. Por el contrario, el fallo tipo I pasa de un mínimo de 80.8 % sin degradación a uno de 91.3 % con degradación y el fallo tipo O pasa de un mínimo de 79.8 % sin degradación a uno de 95.2 % con degradación. El anterior resultado indica que la separación de clases e intra-clases se está viendo afectada con la adición de niveles de severidad.

También es de resaltar que la BD1 obtiene mejores resultados cuando se realiza su identificación de fallos incluyendo niveles de severidad. Lo anterior se puede deber primordialmente a que, al no dividir las señales en niveles de severidad, cada conjunto de señales a clasificar posee mayor variabilidad; sin embargo, al dividir las en niveles de severidad existe la posibilidad de afrontar problemas en el número total de modelos a entrenar.

14.3. Discusión de resultados para metodología de sintonización de modelo HMM con observación discreta

Las curvas ROC permiten definir un compromiso entre el área bajo la curva ROC y el número de estados del modelo HMM. A mayor área bajo la curva ROC, mayor es la exactitud diagnóstica, o capacidad de discernimiento entre poblaciones. Para las bases de datos sin severidad, se logra mayor área a medida que se incrementan los estados HMM; pero cuando se tiene presente la severidad, desde 4 estados se logran valores de área en vecindad del máximo. Sin embargo, teniendo presente que a más estados mayor será la demanda computacional, en general para la tarea de determinar degradación y niveles de severidad, se puede trabajar en el rango de 1 a 4 estados (ver figuras [14.1](#), [14.2](#), [14.3](#) y [14.4](#)).

Se observa como gran resultado que los modelos HMM presentan un gran poder de generalización y no sufren de sobre entrenamiento ante variaciones del tamaño de la base para entrenamiento, tal como se observa en la tablas 12, 13 y 14. Ya que este modelo HMM se alimenta con los coeficientes MFCC discretizados, estos se presentan como una poderosa herramienta que permite obtener características dinámicas tanto lineales como no lineales y que involucran información del tiempo y de la frecuencia. Se resalta que, al incorporar niveles de severidad el número total de modelos HMM que se requiere entrenar crece considerablemente, siendo un factor adicional a tener presente desde un punto de vista de uso de recursos computacionales y como reto para mejorar la eficiencia del proceso global.

Nuevamente es de resaltar que la operación normal es perfectamente identificable y separable de los tipos de fallos inducidas, pero ahora se presentan resultado que animan a trabajar en el estudio de clasificación e identificación de la degradación y sus niveles de severidad. Por ejemplo, los resultados descritos indican que la separación de clases e intra-clases se está viendo afectada con la adición de niveles de severidad, abriendo entonces una perspectiva de indagación, al emplear enfoques que permitan afrontar este tipo de problemas.

15. METODOLOGÍA DE DIAGNÓSTICO CON CARACTERIZACIÓN MULTI-DOMINIO Y CLASIFICADOR JERÁRQUICO

Se introduce un enfoque de selección de características con base en Modelos Ocultos de Markov Jerárquicos (HHMM) para resaltar las características relevantes multi-dominio de las señales de vibración, para las tareas de diagnóstico de fallos y evaluación de severidad en rodamientos. El enfoque de trabajo evalúa los dominios del tiempo (TD), frecuencia (FD) y tiempo-frecuencia (TFD), y luego se entrena un HHMM de dos capas utilizando estrategias de selección forward y backward. La primera capa se enfoca en el diagnóstico de fallos en rodamientos, mientras que la segunda capa se refiere a la evaluación de la severidad. Se determinan un subconjunto de características multi-dominio que mejora el rendimiento de discriminación de cada capa HHMM, lo cual permite ahondar en los análisis discriminantes inter-dominio e intra-dominio.

La estrategia metodológica planteada de clasificador jerárquico estocástico basado en HHMM se prueba como una herramienta de análisis de relevancia de características para identificar múltiples condiciones de salud en rodamientos, incluyendo tipos de fallos y severidades. La evaluación de este modelo, sobre una base de datos conocida, demuestra que esta estrategia iguala, o incluso supera, eficiencias de algoritmos reportados en el estado del arte.

15.1. Marco experimental para metodología de diagnóstico con caracterización multi-dominio y clasificador jerárquico

Empleando la base de datos perteneciente al Bearing Data Center del Case Western Reserve University [121] para una frecuencia de muestreo de 12 kHz, se anota la base según etiquetas para el tipo de fallo: B (Ball) para fallos en la bola, I (Inner) para fallos en cara interna, O (Outer) para fallos en cara externa, Normal para la señal base; y etiquetas N1, N2 y N3 para los tres niveles de severidad, donde 1 es el nivel bajo y 3 el alto. La base resultante consta de 10 carpetas, conteniendo 12 registros de 3.3 s cada una, identificadas como: Normal, B N1, B N2, B N3, I N1, I N2, I N3, O N1, O N2 y O N3. La base anotada resultante conforma el conjunto de $N = 120$ señales de vibración $\{\mathbf{Z}_n, y_n\}_{n=1}^N$ con etiquetas de salida $y_n \in \{1, 2, \dots, C\}$ relacionadas con $C = 10$ clases para diagnóstico de fallos o severidades de un rodamiento.

El pre-procesamiento sobre las señales presentes consiste en realizar segmentación en marcos de longitud igual a $T = 200$ ms y con traslape entre marcos de $1/3$; en cada marco se utiliza una ventana Hamming que permite ajustar e integrar las líneas cercanas en frecuencia. Producto de esta segmentación, cada señal $\mathbf{Z}_n \in \mathbb{R}^{W \times T}$ posee ahora W segmentos en las filas de $\mathbf{z}_{w,n} \in \mathbb{R}^T$, a T instantes de tiempo ($w \in \{1, 2, \dots, W\}$),

Para discriminar fallos en rodamientos desde señales de vibración, la subsecuente etapa es específicamente la extracción de características considerando tres dominios diferentes. Primero, dado un segmento de señal $\mathbf{z}_{w,n} \in \mathbb{R}^T$ con T muestras, para su representación natural en el dominio del tiempo (TD) se extraen 17 características temporales de la señal, denominadas: media, desviación estándar, asimetría, kurtosis, máximo, raíz media cuadrática, factor de forma, factor de media sin signo, factor de holgura, factor de impulso, los valores de histograma de límite superior e inferior, valor de verosimilitud, valor de entropía, raíz media cuadrática sobre los picos de señal y desviación estándar de los valores de señal mayores que el percentil 70. En segundo lugar, para el dominio de la frecuencia (FD) se aplica FFT a un segmento de señal $\mathbf{z}_{w,n} \in \mathbb{R}^T$ que descompone en funciones sinusoidales las distintas frecuencias relacionadas y se extraen 16 características calculadas sobre los vectores de frecuencia λ y espectro $s \in \mathbb{R}^K$, de donde obtienen los siguientes valores: media, varianza, asimetría, kurtosis, frecuencia central, desviación estándar, raíz media cuadrática, la media cuadrática, inverso de la raíz media cuadrática, desviación estándar de los valores de frecuencia mayores que la frecuencia central, media sobre la asimetría, media sobre la kurtosis, raíz media cuadrática sobre la media, raíz de la media geométrica, desviación de la media absoluta, y el rango intercuartílico. Finalmente, para características basadas en el dominio del tiempo-frecuencia (TFD), se extraen 12 coeficientes MFCC (M) con el objetivo de resaltar las propiedades lineales y no lineales de la señal.

Las características desde los diferentes dominios de representación genera una matriz multi-dominio $\{\mathbf{X}_n, y_n\}_{n=1}^N$, producto de concatenar los parámetros en cada dominio para cada $\mathbf{z}_{w,n}$ en el vector fila $\mathbf{x}_{w,n} \in \mathbb{R}^M$ de $\mathbf{X}_n \in \mathbb{R}^{W \times M}$, donde M es el número total de características multi-dominio obtenidas (45 en este caso), que representa una escala continua para cada una de las posibles características. Por ello, se requiere una subsecuente etapa de discretización, o cuantización, para lo cual se emplea el algoritmo de k -medias con un total de 24 centroides y desde el cual se obtiene la matriz de características discretizada $\{\mathbf{X}_n^*, y_n\}_{n=1}^N$.

Con el objetivo de descubrir relaciones complejas en los datos para la tarea de diagnóstico y pronóstico de fallos de rodamientos, se emplea un modelo HHMM aplicado a la matriz de representación de características multi-dominio $\{\mathbf{X}_n, y_n\}_{n=1}^N$. En particular, se propone una cadena jerárquica de Markov donde la capa más alta contiene estados HMM como parte de sus estados ocultos. En este sentido, el modelo HHMM requiere definir una estructura topológica ζ , los parámetros del modelo y el alfabeto de observaciones Σ . La estructura topológica ζ determina los niveles jerárquicos $D \in \mathbb{N}$ y el número de estados $Q \in \mathbb{N}$ por cada nivel.

La figura 15.1 muestra el enfoque metodológico propuesto para el diagnóstico y pronóstico de fallos en rodamientos con base en vibraciones, incluyendo las etapas de base de datos, pre-procesamiento, extracción de características, cuantización o discretización, selección de dominios y características, análisis de relevancia de los dominios y sus características, clasificación por HHMM y finalmente evaluación de la severidad. Para el análisis de relevancia de dominios y características, se parte de la matriz $\{\mathbf{X}_n, y_n\}_{n=1}^N$, en la cual sus columnas representan las características, así: $\mathbf{X} = \{F_{T1}, \dots, F_{T17}, F_{F1}, \dots, F_{F16}, F_{M1}, \dots, F_{M12}\} = \{\mathbf{F}_T, \mathbf{F}_F, \mathbf{F}_{MFCC}\}$, donde \mathbf{F}_T , \mathbf{F}_F y \mathbf{F}_{MFCC} representan respectivamente a las 17 características en TD, las 16 en FD y las 12 en M; esta matriz alimenta a su vez a un clasificador (HHMM en este caso) con el fin de discriminar entre las C posibles clases presentes desde una función de asignación de valores $f = \{f_1, \dots, f_M\}$ con $M \in \{1, \dots, 45\}$, donde para cada asignación de f a \mathbf{X} se busca $\{P(C | \mathbf{X} = f)\}_{n=1}^C$. La evaluación de los dominios y sus características se logra al encontrar un nuevo conjunto de características $\{\mathbf{G}_n, y_n\}_{n=1}^N$ donde $\mathbf{G} = \{G_1, \dots, G_L\}$ y $L < M$ cuando se elimina alguna característica F_m en \mathbf{X} , esto es $\mathbf{G} = \mathbf{X} - F_m$ tal que se maximice $\{P(C | \mathbf{G})\}_{n=1}^C$ [144]. Este proceso puede ser iterativo cuando se hace $\mathbf{X} = \mathbf{G}$ y además se puede realizar iniciando con todo el conjunto de características en \mathbf{X} e ir progresivamente buscando retirar la de menor relevancia (análisis backward, o BWD) o se puede iniciar con la característica de mayor relevancia e ir progresivamente adicionando (análisis forward, o FWD). Una forma de evaluar la probabilidad de discriminar clases, dado un conjunto de características, es tomar la eficiencia final del clasificador sobre todas las clases.

Para la etapa del modelo HHMM se entrena una estructura de dos capas ($D = 2$) variando el número de estados Q internos entre 2 y 3 estados. La primera capa tendrá por objetivo la identificación del tipo de fallo y la segunda capa se relaciona con la identificación de los niveles de severidad, tal como se observa en la figura 15.2. Se emplea un esquema de validación cruzada con 30 repeticiones y selección aleatoria

de las señales de la base de datos en un 60 % para configurar el conjunto de datos de entrenamiento. El resto de la base de datos se usa para validación.

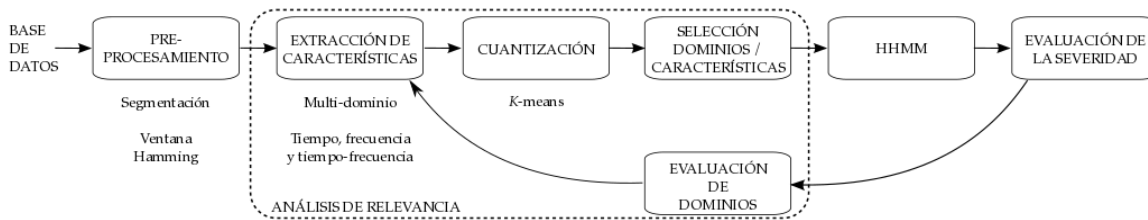


Figura 15.1: Enfoque metodológico HHMM

Se realizan los procedimientos descritos a continuación, corriendo algoritmos en una máquina Intel® Core™ i7-4510, CPU a 2.00 GHz, con 8 GB RAM, S.O. Windows 8.1 de 64 bits.

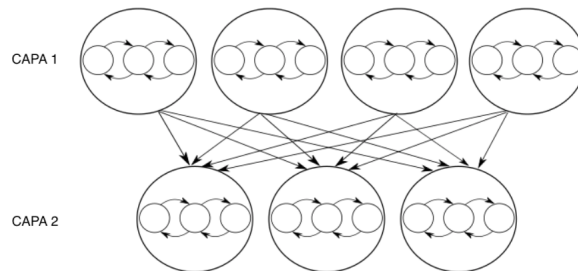


Figura 15.2: Modelo HHMM

15.2. Resultados para metodología de diagnóstico con caracterización multi-dominio y clasificador jerárquico

Análisis inter-dominios

La primera capa del clasificador basado en HHMM se comporta como un sistema de diagnóstico de fallos. En esta capa, se validan los tres dominios de características,

con el fin de determinar su efectividad. La tabla 15 presenta los resultados de validación cruzada, que muestran los valores medios y de desviación estándar para la eficiencia de clasificación de tipos de fallo, según el conjunto de dominios de representación y del número de estados internos del modelo HHMM.

Dominio	Q	B	I	Normal	O
TD	2	1.000±0.000	0.920±0.121	1.000±0.000	0.960±0.047
	3	1.000±0.000	0.973±0.047	1.000±0.000	0.960±0.084
FD	2	0.987±0.042	0.913±0.077	0.960±0.084	0.980±0.032
	3	1.000±0.000	0.940±0.058	0.940±0.190	0.967±0.065
M	2	0.993±0.021	1.000±0.000	1.000±0.000	1.000±0.000
	3	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000
TD+FD	2	1.000±0.000	1.000±0.000	1.000±0.000	0.973±0.047
	3	1.000±0.000	0.980±0.032	1.000±0.000	0.953±0.055
TD+M	2	1.000±0.000	0.960±0.056	1.000±0.000	0.967±0.035
	3	1.000±0.000	0.980±0.063	1.000±0.000	0.973±0.034
FD+M	2	1.000±0.000	0.947±0.053	0.980±0.063	0.987±0.042
	3	1.000±0.000	0.953±0.063	0.980±0.063	0.987±0.028
TD+FD+M	2	0.993±0.021	0.960±0.084	0.980±0.063	0.953±0.045
	3	1.000±0.000	0.967±0.057	1.000±0.000	0.940±0.066

Tabla 15: Análisis inter-dominio, o por tipos de fallos

La tabla 15 enseña resultados de diagnóstico de fallos sobre un 94 % para todos los dominios, pero se destacan ampliamente los resultados obtenidos para el dominio M con 2 y 3 estados internos y para la fusión de dominios TD+FD con 2 estados internos. Además, el dominio M combinado con alguno otro de los dominios siempre mejora el resultado que ese dominio individual puede lograr. Otro resultado a destacar es el diagnóstico del estado normal, el cual es posible para varias de las fusiones de dominios en esta capa, y con menor precisión para cuando se emplea el dominio de la frecuencia, ya sea solo o en conjunción con otros.

La segunda capa del modelo HHMM se entrena pensando en distinguir los niveles de severidad para un tipo de fallo. En la tabla 16 se enseñan los resultados de entrenamiento de esta segunda capa de acuerdo con fusiones de diferentes dominios, y para 2 y 3 estados internos. Se debe tener presente que en este enfoque de dos

capas HHMM, no se requiere clasificar la operación normal por no tener niveles de severidad.

Dominio	Q	Ball (B)			Inner (I)			Outer (O)		
		N1	N2	N3	N1	N2	N3	N1	N2	N3
TD	2	0.820	1.000	0.700	1.000	1.000	1.000	1.000	1.000	1.000
		±0.199	±0.000	±0.236	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000
FD	2	0.860	1.000	0.900	1.000	1.000	1.000	0.940	1.000	1.000
		±0.165	±0.000	±0.105	±0.000	±0.000	±0.000	±0.153	±0.000	±0.000
M	2	1.000	1.000	1.000	1.000	1.000	1.000	0.880	1.000	1.000
		±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.190	±0.000	±0.000
TD+FD	2	0.900	1.000	0.720	1.000	1.000	1.000	1.000	1.000	1.000
		±0.141	±0.000	±0.235	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000
TD+M	2	0.960	0.980	0.660	1.000	1.000	1.000	0.940	1.000	1.000
		±0.084	±0.063	±0.268	±0.000	±0.000	±0.000	±0.190	±0.000	±0.000
FD+M	2	0.660	1.000	0.800	1.000	1.000	1.000	0.920	1.000	1.000
		±0.268	±0.000	±0.133	±0.000	±0.000	±0.000	±0.193	±0.000	±0.000
TD+FD+	2	0.900	1.000	0.780	1.000	1.000	1.000	1.000	1.000	1.000
		±0.141	±0.000	±0.199	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000
M	2	1.000	0.980	0.740	1.000	1.000	1.000	0.940	1.000	1.000
		±0.000	±0.063	±0.268	±0.000	±0.000	±0.000	±0.190	±0.000	±0.000

Tabla 16: Análisis inter-dominio, por severidad de fallos

Los resultados desde la tabla 16 muestran que la eficiencia en la clasificación de los niveles de severidad es menor que las logradas para la primera capa, lo cual se puede explicar por las relaciones complejas existentes al inducir severidades. En particular, discriminar los niveles de severidad para los fallos tipo Ball resulta ser todo un reto. Sin embargo, las características basadas en MFCC permiten evidenciar un dominio de representación que asegura un desempeño discriminante adecuado. Todo lo anterior justifica ampliamente la necesidad de emplear enfoques de modelo basados en estructuras jerárquicas.

Análisis intra-dominios

La figura 15.3 muestra los resultados producto de verificar la capacidad intra-dominio de las características para discriminar fallos y severidades, usando un modelo HHMM de dos capas y 2 estados internos. En esta figura, los resultados se muestran producto de analizar cada clase de fallo, donde el número de características varía con el dominio analizado, por lo tanto: en el dominio TD se analizan 17 características en cada clase (Ball, Inner, Normal y Outer) para un total de 68 características; en el dominio FD se analizan 16 características en cada clase, para un total de 64; y en el dominio M se tienen 12 características en cada clase, para un total de 48. La figura 15.3 enseña los resultados como puntos según los tipos de fallo, o clases, de la siguiente forma: Ball (+), Inner (o), Normal (*) , Outer (x), y el eje de las ordenadas es el valor de la eficiencia de clasificación. Como se observa, las características individuales de los dominios de frecuencia y MFCC muestran los mejores resultados en el diagnóstico de nivel de severidad.

Seguidamente, se realiza un análisis forward y backward para resaltar los subconjuntos de características por dominio. En la figura 15.4 el eje de las abscisas representa el número de características en el subconjunto para el análisis forward por clase, y en la figura 15.5 este mismo eje representa el número de características eliminadas en el subconjunto para un análisis backward por clase. En ambas figuras, el eje de las ordenadas es el valor nominal de eficiencia, y los puntos según tipos de fallo, o clases, se indican de la siguiente forma: Ball (+), Inner (o), Normal (*) y Outer (x). De acuerdo con las figuras 15.4 y 15.5, es posible observar cómo el conjunto de características para cada dominio presenta una capacidad de discriminación diferente entre las dos capas, según se desee determinar tipos de fallos o niveles de severidad.

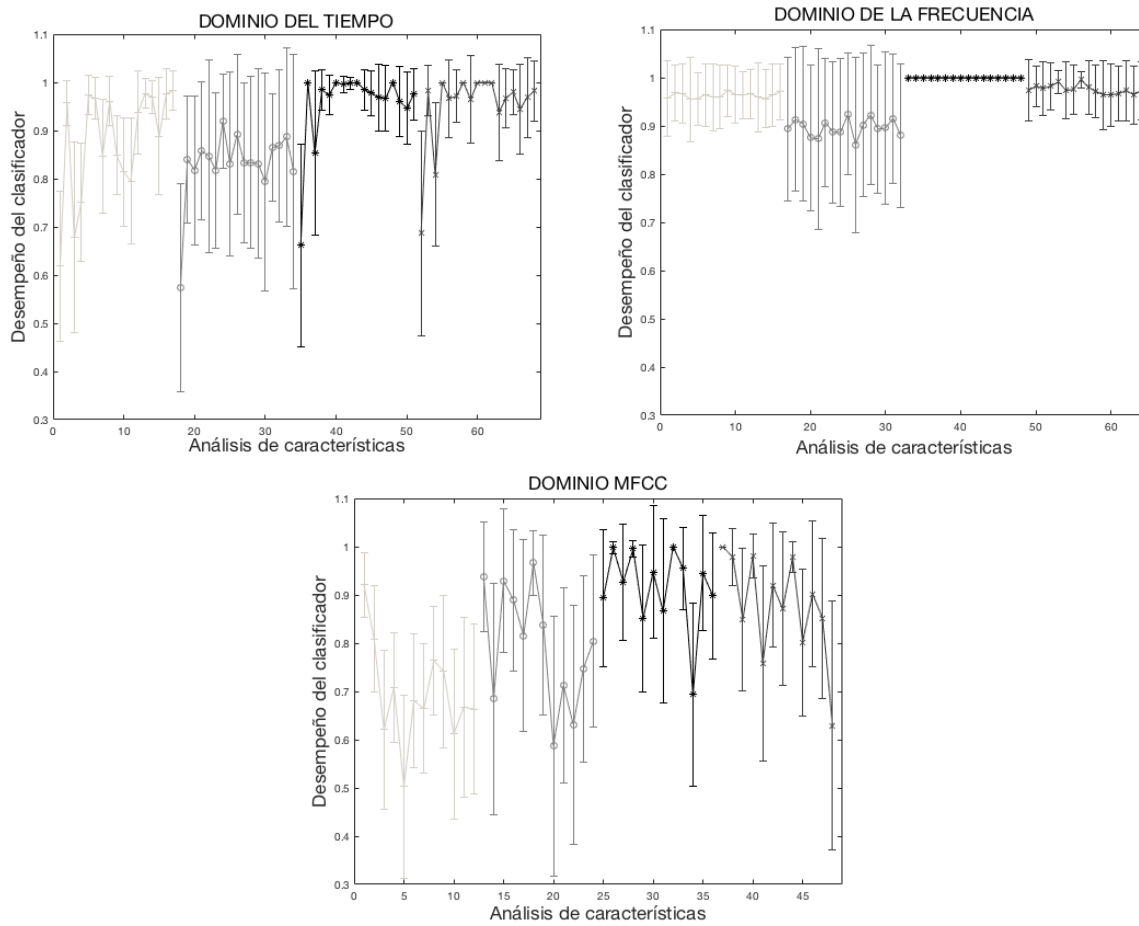


Figura 15.3: Análisis intra-dominio de características

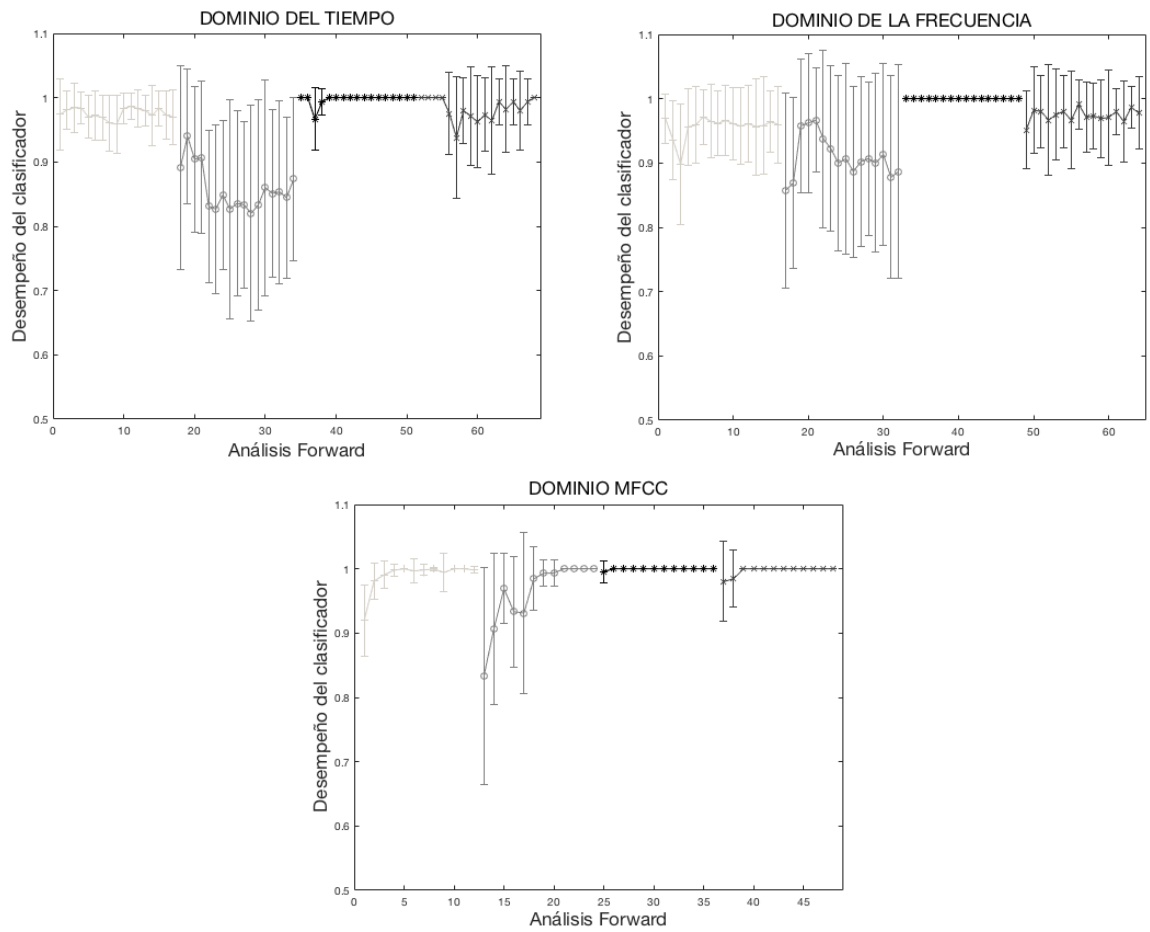


Figura 15.4: Análisis forward para relevancia de características intra-dominio

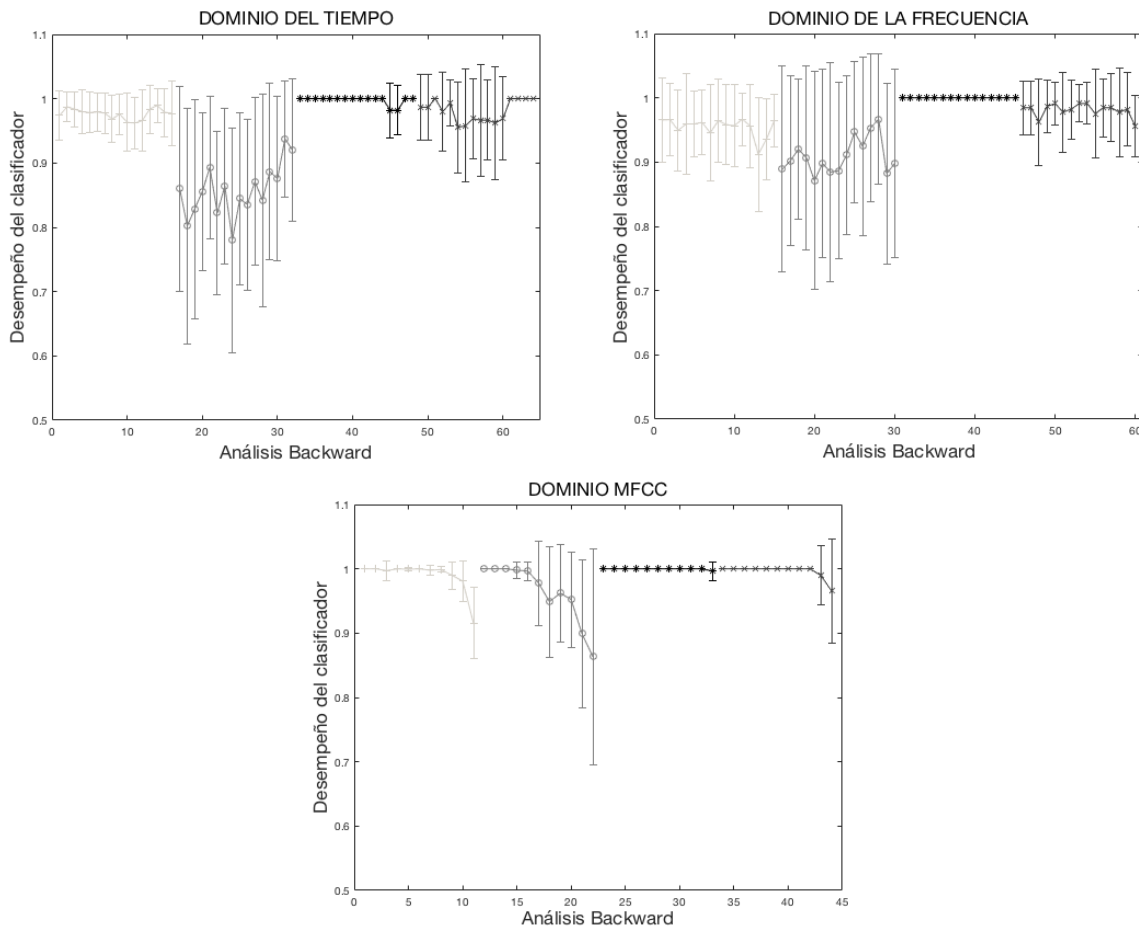


Figura 15.5: Análisis backward para relevancia de características intra-dominio

Con base en los resultados presentados, se propone un modelo notado como RF-HHMM donde la primera capa HHMM solo considere las características 1, 2, 4, 6, 8 y 9 del dominio M. Con respecto a la segunda capa, se propone las características 2, 3, 9, 12 y 16 de FD, para clasificar el nivel de severidad. En aras de la simplicidad, se propone 2 estados internos. Estos resultados están en concordancia con observaciones reportadas en el estado del arte, como en [145], donde los autores informan que con el análisis cepstral se puede detectar un fallo y su naturaleza identificada con

cierta confianza; así como que las técnicas del dominio de la frecuencia son adecuadas para identificar severidades, pero no funcionan correctamente para determinar el componente del rodamiento que es defectuoso.

Evaluación y comparación de modelo HHMM

La tabla 17 presenta los resultados de clasificación, teniendo presente tipos de fallos y severidad, aplicando tanto un modelo HHMM sobre el conjunto relevante de características definidas por el análisis forward y backward (RF-HHMM) y un modelo HHMM sobre el conjunto completo de características multi-dominio (WS-HHMM). Como se puede observar, el modelo RF-HHMM preserva, e incluso mejora, el desempeño en eficiencia logrado por el modelo WS-HHMM. De hecho, el conjunto de características estimado es capaz de descubrir patrones relevantes para identificar clases complejas, por ejemplo, fallos de tipo Ball y Outer en el nivel N1. El peor rendimiento de clasificación se obtiene para el fallo Ball N3 para ambas estrategias estudiadas (87 % y 81 %). Luego, los resultados medios de WS-HHMM y RF-HHMM (97.5 % y 97.7 %, respectivamente) demuestran que la dinámica de fallo en rodamiento se puede derivar de las señales de vibración si se estima un subconjunto de características relevantes tanto a nivel de fallo como de severidad. Además, dado que la probabilidad de detección de un tipo de fallo es del 100 % en ambos esquemas, la probabilidad condicional de detectar un nivel de severidad es independiente del tipo de fallo.

Los resultados logrados están en concordancia con los reportados por los autores del documento [21], donde se propone un enfoque basado en DWT y un Sistema Adaptativo de Inferencia Difusa Neuronal (ANFIS, del inglés *Adaptive Neural Fuzzy Inference System*) para automatizar el proceso de diagnóstico y detección de fallos, logrando una eficiencia de clasificación de 97 % sobre la misma base de datos. Este enfoque ANFIS se basa en un sistema de 5 capas, donde la primera capa implementa la fuzificación, la capa 2 es una operación difusa AND, la capa 3 es la normalización, la capa 4 es la inferencia difusa, y finalmente la capa 5 es la defuzificación. Se resalta que los autores en [21] no reportan resultados por niveles de severidad o por tipos de fallo, solo lo hacen de forma global para el proceso, lo cual genera incertidumbre sobre la eficiencia individual de cada tipo de fallo, de cada tipo de severidad y las desviaciones presentes.

Estrategía	Capa	Ball (B)	Inner (I)	Outer (O)	Normal	Total
WS-HHMM	1	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	
	2, N1	0.91±0.15	1.00±0.00	0.84±0.26		
	2, N2	0.99±0.04	1.00±0.00	1.00±0.00		
	2, N3	0.87±0.15	1.00±0.00	1.00±0.00		
	Total	0.94±0.11	1.00±0.00	0.96±0.09	1.00±0.00	0.975
RF-HHMM	1	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	
	2, N1	0.91±0.15	1.00±0.00	0.93±0.18		
	2, N2	0.98±0.06	1.00±0.00	1.00±0.00		
	2, N3	0.81±0.17	1.00±0.00	0.98±0.06		
	Total	0.93±0.12	1.00±0.00	0.98±0.06	1.00±0.00	0.977
ANFIS [21]						0.970

Tabla 17: Evaluación y comparación de modelo HHMM

15.3. Discusión para metodología de diagnóstico con caracterización multi-dominio y clasificador jerárquico

Producto de los análisis inter-dominio, se puede vislumbrar y justificar la necesidad de emplear un enfoque basado en estructuras jerárquicas para el aprendizaje y entrenamiento de máquina, tal como el que se propone basado en HHMM. En este modelo propuesto, la primera capa del clasificador se comporta como un sistema de diagnóstico de fallos y la segunda capa del modelo se piensa para distinguir los niveles de severidad de un tipo de fallo.

Del análisis inter-dominio, se destaca la capacidad de discriminación de tipos de fallos que presenta el dominio M y la fusión de los dominios TD+FD, además, el dominio M combinado con alguno otro siempre mejora el resultado que ese último dominio individual puede lograr, como se observa en la tabla 15. Otro resultado a destacar es el diagnóstico del estado normal, el cual es posible para varias de las fusiones de dominios en la primera capa HHMM, y con menor precisión para cuando se emplea el dominio de la frecuencia, ya sea solo o en conjunción con otros.

Los resultados del análisis intra-dominio muestran las relaciones complejas existentes al inducir severidades y permiten verificar la capacidad de las características para discriminar fallos y severidades al interior de cada uno de los dominios. En par-

ticular, discriminar los niveles de severidad para los fallos tipo Ball resulta ser todo un reto. Sin embargo, las características basadas en MFCC y FD permiten evidenciar dominios de representación que aseguran un desempeño discriminante adecuado, tal como lo enseña la tabla 16. Todo lo anterior justifica ampliamente la necesidad de emplear enfoques de modelo basados en estructuras jerárquicas.

Los anteriores análisis, inter e intra dominio, permiten proponer un modelo HHMM de dos capas y dos estados internos, donde las características basadas en tiempo-frecuencia son las más relevantes para separar el tipo de fallo (capa 1) y los parámetros basados en frecuencia benefician la discriminación entre los niveles de severidad (capa 2). Además, la estrategia propuesta extrae un subconjunto relevante de características que favorece etapas posteriores de aprendizaje mediante la codificación de relaciones de datos complejos relacionadas con interacciones ocultas entre el tipo de fallo y los niveles de severidad; para ello se realiza un análisis de las eficiencias individuales de las características en cada dominio mediante un análisis forward y backward, donde es posible observar que cada dominio presenta un conjunto de características con capacidad de discriminación diferente entre las dos capas, según se desee determinar tipos de fallos o niveles de severidad (ver figuras 15.3, 15.4 y 15.5).

El modelo HHMM propuesto, empleando un conjunto relevante de características, presenta una eficiencia en la identificación de fallos y severidades del 97.7%, la cual es comparable y ligeramente superior a la que se presenta en un modelo HHMM donde se incluyen todas las características multi-dominios (97.5%), demostrando con ello que la selección de características y enfoque jerárquico permiten mejorar el poder discriminante sobre las señales de vibración. Además estos resultados están en línea con otros reportados en la literatura, como es el caso de los resultados consignados en [21], donde se logra una eficiencia de clasificación de 97% sobre la misma base de datos (ver tabla 17), pero bajo un enfoque que no reporta datos por niveles de severidad ni por tipos de fallo, solo se hace de forma global para el proceso, lo cual genera incertidumbre sobre la eficiencia individual de cada tipo de fallo, de cada tipo de severidad y las desviaciones presentes.

16. METODOLOGÍA DE DIAGNÓSTICO MEDIANTE SELECCIÓN ESTOCÁSTICA DE CARACTERÍSTICAS

El diagnóstico basado en vibraciones plantea un desafío con respecto a las propiedades de las señales, por ejemplo, son altamente dinámicas y no estacionarias. Las técnicas de reducción y selección de características a menudo se usan para reducir el número de características, con el fin de evitar redundancia e incrementar la precisión. Se presenta finalmente un enfoque de selección estocástica de características basado en conocimiento utilizando la técnica supervisada Relief-F para resaltar las características relevantes desde múltiples dominios de representación, con el fin de realizar simultáneamente las tareas de diagnóstico de fallos y evaluación de severidad en rodamientos. En particular, a partir de los dominios del tiempo, frecuencia y tiempo-frecuencia se calculan características de señales de vibración. Relief-F se emplea para calcular un índice vectorial de relevancia multi-dominios, con el fin de medir la contribución de cada dominio en las tareas de diagnóstico de rodamientos. Luego, un modelo HMM de observación continua se entrena agregando en un enfoque dos-por-dos las características seleccionadas para evaluar la exactitud.

La metodología propuesta se valida en comparación con publicaciones previas reportadas en el estado del arte para el diagnóstico y pronóstico de fallos que emplean la misma base de datos y la eficiencia del clasificador como métrica de evaluación.

16.1. Estimación multi-dominio de características

Debido a las complejidades ocultas y dinámicas no-estacionarias de las señales de vibración, se emplean características desde los dominios del tiempo, la frecuencia y del tiempo-frecuencia a extraer desde cada segmento $\mathbf{z} \in \mathbf{Z}$. Para el dominio del tiempo (TD) se consideran 18 características que son la representación natural de la señal, de acuerdo con los siguientes descriptores: Media, mediana, desviación estándar, raíz media cuadrática, valor pico, factor de forma, factor de cresta, factor de impulso, factor de holgura, simetría, valor máximo, valor mínimo, rango, rango intercuartílico, kurtosis, velocidad de la kurtosis, aceleración de la kurtosis, derivada de la aceleración de la kurtosis. Las características para el dominio de la frecuencia (FD) se obtienen desde la aplicación inicial de la FFT con el fin de codificar los pa-

trones armónicos de la señal. Se calculan las 5 características siguientes: frecuencia media, frecuencia centra, raíz media cuadrática frecuencial, desviación estándar frecuencial, kurtosis; además se evalúan desde los datos de componentes frecuenciales las mismas 18 características del dominio del tiempo. Un total de 12 coeficientes MFCC (M) se emplean para resaltar tanto los patrones lineales como no lineales de las señales de vibración para una representación TFD.

Finalmente, luego de calcular todos los parámetros previamente descritos para cada uno de los segmentos de señal, se crea un conjunto de características multi-dominio $\{\mathbf{X}_n, y_n\}_{n=1}^N$ producto de concatenar las parámetros de los dominios TD, FD y TFD de cada $\mathbf{z}_{w,n}$ en el vector fila $\mathbf{x}_{w,n} \in \mathbb{R}^M$ de $\mathbf{X}_n \in \mathbb{R}^{W \times M}$, donde M es el número de características multi-dominio extraídas.

16.2. Relevancia de características por Relief-F

En aplicaciones prácticas, el conjunto de características multi-dominio $\{\mathbf{X}_n\}_{n=1}^N$ posee un gran número de parámetros. Este hecho introduce ruido y complejidad en la tarea de detección y diagnóstico. Por tanto, es necesario identificar las características más discriminantes con el fin de encontrar un balance entre la complejidad del sistema y la eficiencia [22]. Para lograr este propósito, se evalúa la contribución de cada parámetro de los dominios TD, FD y M en términos de la información supervisada empleando vectores de etiqueta $\mathbf{y} = \{y_n\}_{n=1}^N$. Inicialmente, se define a $\Gamma \in \mathbb{R}^{N \times Q}$ como una matriz de características que posee vectores de características fila $\gamma = \text{vec}(\mathbf{X}_n)$, donde $\text{vec}(\cdot)$ representa la función de vectorización. Se calcula el vector de relevancia v como sigue:

$$v_q = \frac{1}{N} \sum_{n=1}^N \left(-\frac{1}{\varphi} \sum_{\gamma_{n'} \in \Omega_n^{y_n}} d(\gamma_n, \gamma_{n'}) + \frac{1}{\varphi} \sum_{c \neq y_n} \frac{p(y=c)}{1-p(y=y_n)} \sum d(\gamma_n, \gamma_{n'}) \right) \quad (16.1)$$

donde $d : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ es una función de distancia definida, $\Omega_n^C = \{\gamma_{n'} : n' = 1, 2, \dots, \varphi\}$ posee los φ vecinos más cercanos de γ_n de acuerdo con d y $p(\mathbf{y} = c) \in [0, 1]$ es la probabilidad que una muestra pertenezca a la clase c ($c \in \{1, 2, \dots, C\}$); $q \in \{1, 2, \dots, Q\}$. En este contexto, un valor mayor de v_q corresponde a la mejor q -ésima característica que discrimina categorías de fallo en la tarea de diagnóstico en rodamientos [146, 147].

Ahora, para estimar la relevancia de las M características multi-dominio se ordena el vector ν en una matriz de relevancia de características $\Delta \in \mathbb{R}^{W \times M}$ concerniente con los W segmentos, como sigue:

$$\Delta = \begin{bmatrix} \nu_1 & \nu_2 & \cdots & \nu_M \\ \nu_{M+1} & \nu_{M+2} & \cdots & \nu_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{W(M-1)+1} & \nu_{W(M-1)+2} & \cdots & \nu_{WM} \end{bmatrix} \quad (16.2)$$

En consecuencia, las M características multi-dominio se pueden seleccionar de acuerdo con el valor del vector de relevancia $\rho \in \mathbb{R}^M$:

$$\rho_m = \frac{1}{W} \sum_{w=1}^W |\Delta_{wm}| \quad (16.3)$$

donde $\Delta_{wm} \in \Delta$ y $m \in \{1, 2, \dots, M\}$. Como resultado, la selección estocástica de características permite explicar la capacidad de preservar la localidad medida por cada característica a lo largo de los W segmentos de las señales de vibración de entrada, ya que el vector de relevancia obtenido ρ preserva las relaciones uno a uno de las características multi-dominio.

16.3. Marco experimental para metodología de diagnóstico mediante selección estocástica de características

Sea $\{\mathbf{Z}_n, y_n\}_{n=1}^N$ un conjunto de N señales de vibración $\mathbf{Z}_n \in \mathbb{R}^{W \times T}$ con W segmentos en las filas $\mathbf{z}_{w,n} \in \mathbb{R}^T$, a T instantes de tiempo ($w \in \{1, 2, \dots, W\}$), y con etiquetas de salida $y_n \in \{1, 2, \dots, C\}$ relacionadas con C clases para diagnóstico de fallos o severidades de un rodamiento. El enfoque de Selección Estocástica de Características (SFS, del inglés *Stochastic Feature Selection*) que se propone, consta de las siguientes etapas: i) Estimación multi-dominio de características, ii) Selección de características por relevancia basada en gráfico, y iii) Clasificación basada en sistema estocástico.

La figura 16.1 resume el proceso del marco metodológico propuesto para un sistema de diagnóstico de fallos mediante enfoque de selección estocástica de características. A continuación, se describen estas tres etapas en mayor detalle.

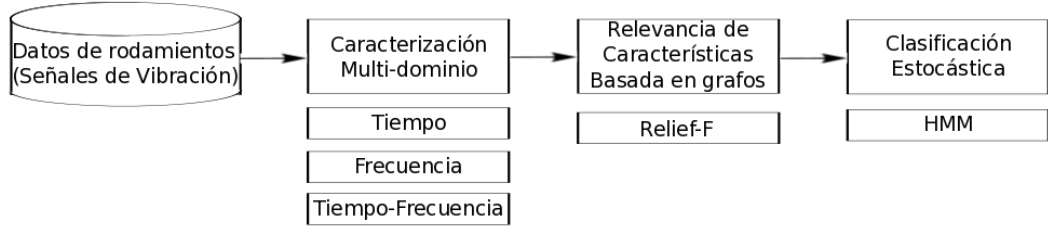


Figure 16.1: Marco metodológico para diagnóstico y pronóstico con base en SFS

Con base en el marco metodológico propuesto, y con el fin de realizar el entrenamiento y validación del sistema SFS, se emplea la base de datos desarrollada por el Bearing Data Center del Case Western Reserve University [121] y descrita previamente. Se emplean las señales pertenecientes a una frecuencia de muestreo de $F = 48$ kHz. Con el fin de anotar la base de datos, se designan las etiquetas Ball (B) para fallos en la bola, Inner (I) para fallos en cara interna, Outer (O) para fallos en cara externa, Normal para la señal base y etiquetas N1, N2 y N3 para los tres niveles de severidad, donde 1 es el nivel bajo y 3 el alto. La base resultante consta de 10 carpetas: Normal, B N1, B N2, B N3, I N1, I N2, I N3, O N1, O N2 y O N3; con un total de 56 señales adquiridas, con una duración de 10 segundos, aproximadamente [121].

Sobre las señales presentes, se realiza una pre-segmentación empleando una ventana Rectangular en marcos de tiempo igual a $\varrho = 85.33$ ms y sin traslape. Los anteriores valores se seleccionan de acuerdo con consideraciones definidas en [148]. El proceso genera una matriz $\mathbf{J} \in \mathbb{R}^{N \times \varrho}$ que posee un total de $N = 5610$ muestras. Con el fin de realizar análisis y comparación de resultados, se consideran dos posibles escenarios diferentes variando el vector de etiquetas $\{y_c\}_{c=1}^C$: i) Problema de 4 clases ($C = 4$), para solo aprendizaje relacionado con los estados de fallo y ii) Problema de 10 clases ($C = 10$), para diagnóstico y severidad. Producto de lo anterior, para el caso $C = 4$ la base de datos se agrupa solo según señales de fallo (Ball (B), Inner (I), Normal, Outer (O)) y para el caso $C = 10$ la base se divide en señales según fallos y severidades. Para ajustar los datos de acuerdo con la metodología propuesta, cada pre-segmento \mathbf{j} se segmenta nuevamente empleando una ventana Hamming con una duración de $T = 21.33$ ms y un traslape de $1/2$. Los anteriores valores se escogen con el fin de generar un total de $W = 7$ ventanas de análisis. Así, finalmente se obtiene el conjunto de análisis $\{\mathbf{Z}_n, y_n\}_{n=1}^N$.

Luego, se evalúa un total de 53 características multi-dominio: 18, 23 y 12 para los dominios TD, FD y M respectivamente. Entonces, se construye un conjunto de características multi-dominio $\{\mathbf{X} \in \mathbb{R}^{W \times M}\}_{n=1}^N$ que posee un total de $N = 5610$ muestras analizadas con $W = 7$ ventanas y $M = 53$ características.

Con el fin de determinar la relevancia de los parámetros multi-dominio bajo el enfoque de selección estocástica, se construye una matriz de características $\Gamma \in \mathbb{R}^{N \times Q}$ con $Q = W * M = 371$ parámetros. En este caso, se fijan dos parámetros con el fin de calcular el vector de relevancia según la ecuación 16.1: tamaño de vecindad $\theta = 1$ y la función de distancia $d(\gamma_n, \gamma_{n'}) = |\gamma_n - \gamma_{n'}| / (\max_n(\gamma_n) - \min_n(\gamma_n))$ [147, 149]. De este proceso se obtiene el vector de relevancia de características v . Ahora, para estimar el nivel de relevancia de las M características multi-dominio se construye la matriz de relevancia Δ según la ecuación 16.2, y seguidamente se computa el vector de relevancia ρ como se indica en la ecuación 16.3.

Debido al gran número de muestras, se considera el subconjunto de datos más representativos de cada clase producto de emplear el algoritmo de agrupamiento k -medias con distancia euclidiana como métrica. Tomando en cuenta es último proceso, se emplea el conjunto de características multi-dominio $\{X_n\}_{n=1}^N$ con $N = 3413$ para el problema $C = 4$ y $N = 4141$ para el problema $C = 10$, para realizar el entrenamiento de un clasificador HMM con observación continua, para representar los W segmentos de características en \mathbf{X} .

La validación del sistema se evalúa mediante el desempeño del clasificador. El número de estados se fija en $V = 2$, según se recomienda en [85] y el número de mezclas de gaussianas en $U = 2$ (valor que se fija de forma empírica). Entonces, la eficiencia del clasificador se evalúa empleando un esquema de validación cruzada de 10 carpetas y se adicionan las características de a dos, según su relevancia en ρ . Los datos originales se dividen aleatoriamente en 10 subgrupos iguales, y se asigna un solo subconjunto para probar el modelo y los nueve restantes para el entrenamiento. Este proceso se repite un total de 10 veces para tomar finalmente el promedio de la eficiencia y desviación estándar.

Con el fin de realizar la comparación del método propuesto, se consideran los siguientes métodos de selección de características, supervisados y no supervisados:

- **Análisis de Relevancia basado en Varianza (VRA, del inglés *Variance-based Relevance Analysis*):** Método no supervisado, el cual genera su clasificación de relevancia sobre las características de entrada con base en un criterio de variabilidad [150].

- Auto-Censurado (SW, del inglés *Self-Weight Ranking*): Método no supervisado, el cual codifica la relevancia de características en términos de una medida de auto-similaridad [119].
- Puntuación Laplaciana (LS, del inglés *Laplacian Score*): Método no supervisado, que evalúa su clasificación de relevancia con el fin de resaltar la importancia de cada característica desde gráficos de bordes [120].
- Distancia-Peso (DW, del inglés *Distance-Weight*): Método supervisado, el cual cuantifica la relevancia de la distancia entre muestras desde diferentes agrupamientos empleando información supervisada [151].

Se realizan los procedimientos descritos a continuación, corriendo algoritmos en una máquina Intel® Core™ i7-4510, CPU a 2.00 GHz, con 8 GB RAM, S.O. Windows 8.1 de 64 bits.

16.4. Resultados para metodología de diagnóstico mediante selección estocástica de características

Inicialmente, se introduce un ejemplo ilustrativo de los parámetros en el dominio del tiempo y de la frecuencia concerniente con el problema de clasificación de 4 clases ($C = 4$). La figura 16.2 muestra la forma de onda en el dominio del tiempo perteneciente a un segmento de señal de vibración con su respectivo espectro en frecuencia. Como se puede observar, el espectro de frecuencia varía en cada estado, notando que la mayor cantidad de información está contenida hasta los 9 kHz.

Con el fin de establecer los mejores resultados experimentales, se construye una selección por relevancia para cada uno de los dos problemas de clasificación mencionados. Así, en la figura 16.3 se muestran los resultados obtenidos para cada uno de los métodos de selección de características empleados, de la siguiente forma: primero, se identifica la mejor eficiencia obtenida con cada método y subsecuentemente, se busca un resultado estadísticamente equivalente con el menor número de características posibles. La tabla 18 muestra el mejor resultado obtenido para cada método de selección de características para el problema $C = 4$, la tabla 19 muestra la matriz de confusión cuando se emplea Relief-F con 9 características; de forma similar, la tabla 20 muestra los resultados respectivos para el problema $C = 10$ y la tabla 21 muestra la matriz de confusión cuando se emplea Relief-F con 7 características.

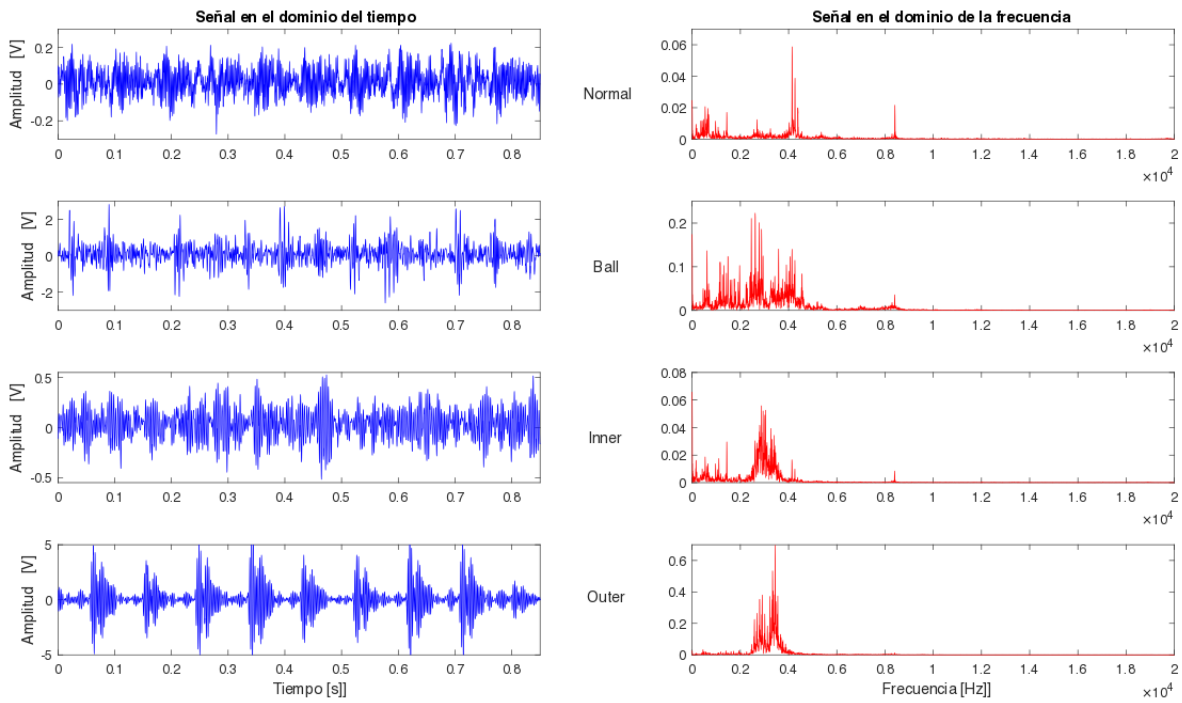
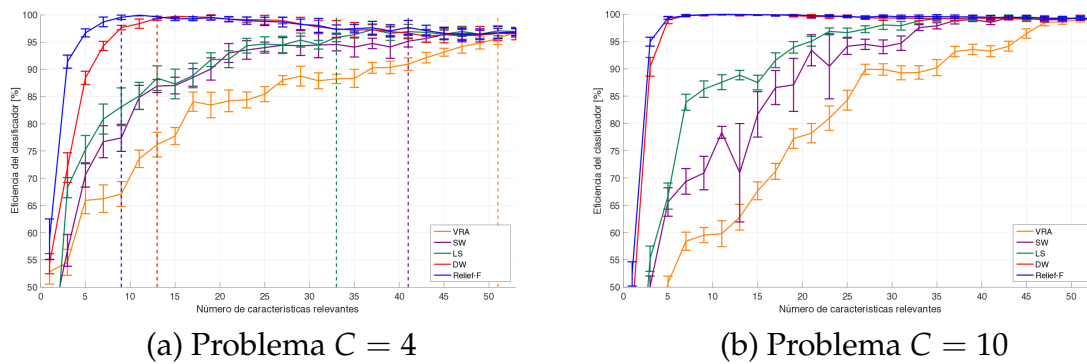


Figure 16.2: Señales en el dominio del tiempo y de la frecuencia, según tipo de fallo



(a) Problema $C = 4$

(b) Problema $C = 10$

Figure 16.3: Selección de características con HMM

Método	# Caract.	Mejor Efic. (%)	Mejor # Caract.	Eficiencia (%)
Sin selección	53	96.97±0.83	-	-
VRA	53	96.92±0.84	51	95.63±1.07
SW	53	96.69±0.76	41	95.17±1.06
LS	37	97.28±0.92	33	95.87±0.79
DW	15	99.74±0.29	13	99.36±0.39
Relief-F	11	99.94±0.12	9	99.56±0.40

Tabla 18: Resultados experimentales para C = 4

	Normal	Ball	Inner	Outer
Normal	100.000	0.000	0.000	0.000
Ball	0.000	99.850	0.150	0.000
Inner	0.000	0.100	99.900	0.000
Outer	0.000	0.000	1.500	98.500

Tabla 19: Mariz de confusión C = 4, Relief-F y 9 características [en %]

Método	# Caract.	Mejor Efic. (%)	Mejor # Caract.	Eficiencia (%)
Sin selección	53	98.97±0.46	-	-
VRA	51	99.01±0.42	47	98.47±0.58
SW	43	99.39±0.43	37	98.84±0.62
LS	43	99.53±0.33	33	98.80±0.49
DW	11	99.95±0.07	7	99.76±0.11
Relief-F	11	99.91±0.16	7	99.74±0.21

Tabla 20: Resultados experimentales para C = 10

	Normal	B N1	B N2	B N3	I N1	I N2	I N3	O N1	O N1	O N3
Normal	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B N1	0.000	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B N2	0.871	0.000	98.545	0.000	0.000	0.000	0.292	0.292	0.000	0.000
B N3	0.000	0.000	0.000	100	0.000	0.000	0.000	0.000	0.000	0.000
I N1	0.000	0.000	0.000	0.000	100	0.000	0.000	0.000	0.000	0.000
I N2	0.000	0.000	0.000	0.000	0.000	99.763	0.237	0.000	0.000	0.000
I N3	0.000	0.000	0.482	0.000	0.000	0.000	99.518	0.000	0.000	0.000
O N1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	100	0.000	0.000
O N2	0.000	0.000	0.000	0.000	0.000	0.471	0.000	0.000	99.529	0.000
O N3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	100

Tabla 21: Matriz de confusión $C = 10$, Relief-F y 7 características [en %]

Con base en los anteriores resultados, la menor eficiencia para el problema $C = 4$ se obtiene con VRA (99.63 ± 1.07 % y 51 características), en contraste, la mejor eficiencia se obtiene con Relief-F (99.56 ± 0.40 % y 9 características). Para el problema $C = 10$, nuevamente la eficiencia más baja se obtiene con VRA (98.47 ± 0.58 % y 47 características), aunque se puede considerar resultados estadísticamente equivalentes para la mejor eficiencia entre DW y Relief-F (99.76 ± 0.11 % y 99.74 ± 0.21 % respectivamente, ambos con 7 características). Al analizar las curvas en la figura 16.3, se nota que los métodos no supervisados no pueden eliminar la información redundante en comparación con las otras técnicas de selección de características, lo cual se debe al alto nivel de de ineficiencia para disminuir el ruido que introducen las características redundantes en el sistema. Así, se puede concluir que esta clase de enfoques no son adecuados para la tarea de identificar características relevantes en estados de fallo en rodamientos. Por otro lado, el método Relief-F resalta sobre las otras técnicas ya que fue siempre responsable de la mejor eficiencia para los dos problemas ($C = 4$ y $C = 10$), además, posee baja complejidad matemática. Además, en las matrices de confusión de las tablas 19 y 21 se puede observar que Relief-F entrega clasificaciones inter-clase con confiabilidades superiores al 98.500% e intra-clase con confiabilidades superiores al 98.545%.

La figura 16.4 muestra la clasificación de relevancia que se obtiene con Relief-F para los escenarios $C = 4$ y $C = 10$. De acuerdo con este método, la mejor eficiencia se obtiene empleando solo el subconjunto de características del dominio M. Especialmente, es necesario evaluar los coeficientes número 4, 5, 7, 8, 9 y 12 de MFCC sobre

cada señal de vibración, para ambos escenarios. Además, es necesario complementar con los coeficientes 3, 6 y 11 de MFCC para el problema $C = 4$ y con el coeficiente 2 de MFCC para el problema $C = 10$. Claramente se puede notar que, comparado con otras técnicas de extracción de características, MFCC es el mejor método ya que codifica tanto información del dominio del tiempo como de la frecuencia de una señal de vibración.

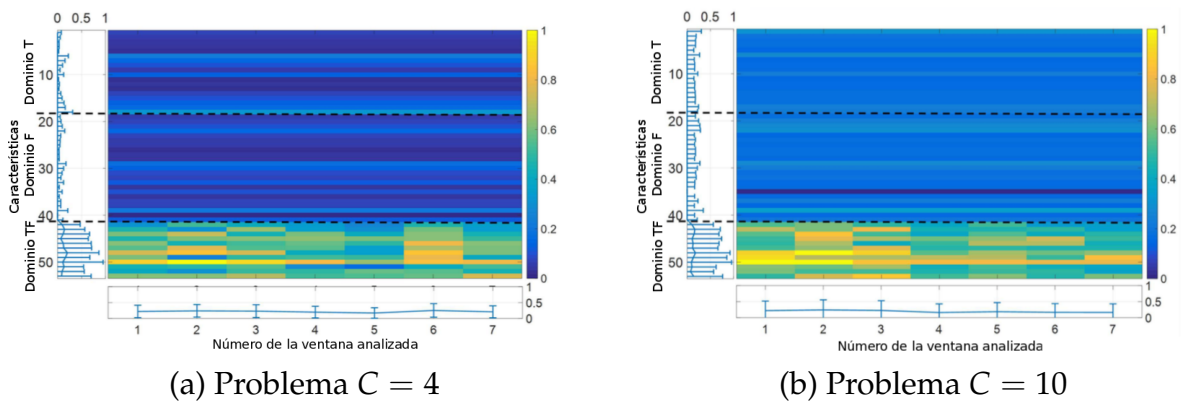


Figure 16.4: Características relevantes que logran la mejor eficiencia

El estudio comparativo entre el presente trabajo y publicaciones previas reportadas en el estado del arte para el diagnóstico y pronóstico de fallos se muestran en la tabla 22. Esta comparación se presenta teniendo en cuenta los siguientes aspectos: i) la metodología propuesta en el documento de interés debe ser verificada sobre la base de datos de vibración de Western Case University [121]; ii) la métrica empleada para evaluar el desempeño del método propuesto debe ser la eficiencia del clasificador.

Referencia	C	Etiquetas	Extracción de características	Selección de características	Número de características	Clasificador	Eficiencia (%)
William y Hoffman [79]	4	I, O, B, N	ZC	-	10	ANN	97.13
Tiwari et al. [152]	4	I, O, B, N	MPE	-	16	ANFC	97.50
Liu et al. [153]	4	I, O, B, N	TD+SP, FD+SP, EMD+SP	-	81	PSO-WSVM	97.50
Muruganatham et al. [154]	4	I, O, B, N	SSA	-	10	ANN	96.53
Zhang et al. [83]	4	I, O, B, N	PE+EEMD	-	12	SVM+ICD	97.75
Ocak and Loparo [31]	3	I, O, B	LPM	-	30	HMM	99.67
Yuwono et al. [30]	3	I, O, B	WT+CL	-	12	HMM+SRCE	95.08
Muruganatham et al. [154]	4	I, O, B, N	SSA	SV	4	ANN	95.14
Shen et al. [24]	4	I-1, O-1, B, N	WPT+SP	DET	30	SVR	100.00
Shen et al. [24]	4	I-2, O-2, B, N	WPT+SP	DET	30	SVR	100.00
Van et al. [76]	7	I-1, I-2, O-1, O-2, B-1, B-2, N	TD+SP, FD+SP, EMD+SP	DET+PSO-KNN	20	KNN	98.58
Van et al. [76]	7	I-1, I-2, O-1, O-2, B-1, B-2, N	TD+SP, FD+SP, EMD+SP	DET+PSO-KNN	54	PNN	97.24
Van et al. [76]	7	I-1, I-2, O-1, O-2, B-1, B-2, N	TD+SP, FD+SP, EMD+SP	DET+PSO-KNN	64	SVM	97.71
Brkovic et al. [25]	4	I, O, B, N	WT+SP	SM	12	QC	100.00
Liang et al. [22]	4	I, O, B, N	TD+SP, FD+SP	NMF+ALS+SCD	3	KNN	92.86
Shao et al. [155]	16		CAE+DAE	LPP	19	SC	100.00
Zheng et al. [120]	6	I-1, I-2, O-1, O-2, B-1, B-2, N	GCMPE	LS	2	PSO-SVM	98.89
Attoui, et al. [21]	10	I-1, I-2, I-3, O-1, O-2, O-3, B-1, B-2, B-3, N	DWT+FFT	-	10	ANFIS	97.0
Este trabajo, WS-HHMM	10	I-1, I-2, I-3, O-1, O-2, O-3, B-1, B-2, B-3, N	TD+SP, FD+SP, TFD+SP	FWD+BWD	45	HHMM	97.5
Este trabajo, RF-HHMM	10	I-1, I-2, I-3, O-1, O-2, O-3, B-1, B-2, B-3, N	TD+SP, FD+SP, TFD+SP	FWD+BWD	11	HHMM	97.7
Este trabajo, SFS	4	I, O, B, N	TD+SP, FD+SP, TFD+SP	Relief-F	9	HMM	99.56
Este trabajo, SFS	10	I-1, I-2, I-3, O-1, O-2, O-3, B-1, B-2, B-3, N	TD+SP, FD+SP, TFD+SP	Relief-F	7	HMM	99.74

Table 22: Estudio comparativo de resultados contra el estado del arte

Se puede observar que la base de métodos propuestos para la evaluación de características son parámetros estadísticos en los dominios del tiempo, de la frecuencia y del tiempo-frecuencia, dominando ampliamente sobre metodologías basadas en métricas de información. Además, se puede observar que las metodologías sin selección de características [79, 83, 152, 153, 154] poseen normalmente las eficiencias menores. Especialmente se destaca como los enfoques basados en HMM [30, 31] requieren de una gran cantidad de características para la clasificación en escenarios con tres clases. Por otro lados, las metodologías con selección de características poseen mejor desempeño [22, 24, 25, 76, 154, 155] y con dos metodologías resaltando sobre las demás: el enfoque que se propone en este documento y el presentado por Zheng et al. en [120]. Aunque el segundo enfoque emplea un menor número de características, los resultados experimentales que se muestran se computan emplean-

do un método de validación que enseña resultados solo sobre el conjunto de datos de entrenamiento (no en datos de prueba), fuera de emplear la técnica de selección no-supervisada LS, la cual en nuestros resultados mostró no eliminar información redundante. Por tanto, el método que se propone basado en SFS es más general, ya que puede trabajar en forma específica sobre todos los niveles de severidad. Con relación a este último aspecto, se resalta que ninguno de los enfoques en comparación con el reportado en este documento emplea los 3 niveles de severidad presentes en las señales, estando restringidos en el mejor de los casos de 2 niveles de severidad. En este sentido, el método propuesto desempeña satisfactoriamente la tarea de diagnóstico/pronóstico equilibrando el número de características y la eficiencia, con el fin de poder entregar información discriminante de todos los niveles de severidad.

16.5. Discusión de resultados para metodología de diagnóstico mediante selección estocástica de características

Los métodos no supervisados para la clasificación de relevancia, con fines de selección de características, no pueden eliminar la información redundante en comparación con las otras técnicas de selección de características supervisadas, lo cual se debe al alto nivel de ineficiencia para disminuir el ruido que introducen las características redundantes en el sistema. Así, se puede concluir que esta clase de enfoques no son adecuados para la tarea de identificar características relevantes en estados de fallo en rodamientos. Se destaca por su eficiencia el método Relief-F sobre las otras técnicas de selección (ver figura 16.3 y tablas 18 y 20), además posee baja complejidad matemática.

En la clasificación de relevancia por Relief-F para los dos escenarios en estudio, la mejor eficiencia se obtuvo empleando solo el subconjunto de características del dominio TFD, y específicamente con base en coeficientes MFCC. En este contexto, y en comparación con otros dominios de representación, MFCC es el mejor método que permite codificar de forma simultánea tanto información del dominio del tiempo como de la frecuencia de una señal de vibración, y permitiendo discriminar entre niveles de severidad.

Del estudio comparativo entre el presente trabajo y publicaciones previas, reportadas en el estado del arte para el diagnóstico y pronóstico de fallos, se puede observar que la base de métodos propuestos para la evaluación de características son parámetros estadísticos en los dominios del tiempo, de la frecuencia y del tiempo-frecuencia, dominando ampliamente sobre metodologías basadas en métricas de in-

formación. Además, las metodologías sin selección de características poseen usualmente las eficiencias menores, en contraste con las metodologías con selección de características que poseen mejor desempeño. Se resalta que ninguno de los enfoques, en comparación con los reportados en este documento, emplea los 3 niveles de severidad presentes en las señales de vibración disponibles desde la base de datos, estando restringidos en el mejor de los casos a 2 niveles de severidad. En este sentido, los métodos propuestos desempeñan satisfactoriamente la tarea de diagnóstico/pronóstico equilibrando el número de características y la eficiencia, con el fin de poder entregar información discriminante de todos los niveles de severidad (ver tabla 22).

El enfoque de selección de características basado en Relief-F y aprendizaje de máquina por HMM logró resultados en concordancia con el estado del arte relacionado con el diagnóstico y pronóstico de fallos, sin embargo, la estrategia presentada permite extraer subconjuntos relevantes de características los cuales favorecen la subsecuente etapa de aprendizaje codificando las relaciones complejas de los datos relacionadas con interacciones ocultas entre los tipos de fallos y los niveles de severidad.

17. CONCLUSIONES Y TRABAJOS FUTUROS

17.1. CONCLUSIONES

Las señales de vibración y acústicas se caracterizan por ser señales cuasi estacionarias o no estacionarias que se encuentran influenciadas de manera diferente por factores externos, tales como perturbaciones derivadas desde el mismo equipo eléctrico. En estos escenarios, las señales de vibración presentan un componente estacionario de alta energía que oculta la frecuencia de fallo de los rodamientos, mientras que, en las señales acústicas esta perturbación enseña un comportamiento no estacionario que es similar a los fallos. Por lo tanto, las señales acústicas son más vulnerables a las interferencias externas que las señales de vibración, por lo han sido menos utilizadas por dificultar la realización de un análisis eficiente y robusto para el monitoreo de condiciones.

Producto de aplicar análisis univariados y multivariados, sobre representaciones en dominios del tiempo y de la frecuencia, las señales de vibración presentan características que se distinguen por su baja dependencia conjunta entre variables (características). Además, para el problema general de diagnóstico, es perfectamente identificable la separación binaria de operación normal y fallo desde varios tipos de análisis: como definir umbrales en parámetros univariados, emplear gráficos de dispersión por pares, usar análisis por caras de Chernoff y definir puntos de corte en la distancia de Mahalanobis. Sin embargo, ninguna de estas técnicas de análisis univariado o multivariado permiten realizar la adecuada clasificación sobre varios tipos de fallos (identificación del origen del fallo, o diagnóstico), y mucho menos sobre fallos y niveles de severidad (identificación del origen del fallo y su nivel de degradación, o pronóstico). Estos análisis univariados y multivariados indican la presencia de clases de fallo que se mezclan, especialmente cuando se trata de severidades de nivel medio y alto, y que persisten indiferente del tipo de dominio de representación de las señales.

La reducción del espacio original de características de representación entrega un conjunto de nuevas variables que permiten contener la mayor representatividad en información contenida desde las variables originales, sin embargo, en el contexto del diagnóstico de fallos, análisis como el de componentes principales, proyección de observaciones y análisis por dendrograma evidencian que esta pérdida de información afecta el contenido de variabilidad proyectado desde las variables originales, lo cual dificulta el proceso de separación de clases, llegando a solo permitirse

la separación binaria de operación normal contra todos los demás modos de fallo, y mostrando nuevamente que no es posible una separación de fallos y niveles de severidad.

Para la implementación de un modelo para el diagnóstico y pronóstico de vida útil remanente en rodamientos, se comprueba la existencia de cuasi-estacionariedad, o no estacionariedad, en las señales y características obtenidas desde diferentes dominios de representación, así como la naturaleza estocástica de los sistemas, con parámetros que varían en el tiempo, incluso en señales que provienen de equipos supuestamente idénticos. En estos escenarios es fundamental tratar con señales que impliquen técnicas no invasivas de adquisición de datos, como son las señales de vibración y acústicas, pero se resalta que estas últimas son más vulnerables a las interferencias externas que las señales de vibración.

Dada la naturaleza cuasi-estacionaria, o no estacionaria, de las señales de vibración, se hace necesario recurrir a representaciones multi-dominio para realizar la caracterización de las señales con el fin de poder entrenar un clasificador para la tarea de diagnóstico y pronóstico. Se destacan las representaciones en el dominio del tiempo, como la representación natural de la señal y basada en estadísticos de la misma; la representación en el dominio de la frecuencia, que busca codificar los patrones armónicos de la señal luego de aplicar la FFT; y representaciones en el dominio tiempo-frecuencia, con el fin de resaltar patrones lineales y no lineales. Estas características presentan comportamientos discriminantes variados según su tipo y dominio de representación, siendo importante identificar y retirar aquellas con bajo valor discriminante individual, ya que conllevan a una pérdida de generalización del sistema.

Con el fin de acometer la tarea de identificar niveles de severidad (diagnóstico), es preciso implementar enfoques de caracterización basados en fusión de dominios de representación, donde cada uno aporte capacidades discriminantes que en conjunto mejoran la eficiencia de la clasificación. En particular se demostró que el espacio de representación tiempo-frecuencia basado en MFCC presenta los mejores resultados como dominio individual y en relación al poder discriminante de cada una de sus características, sin embargo, presenta un tiempo de entrenamiento que es un orden de magnitud superior a los otros dominios. Al fusionar dominios de representación, tales como TD+FD y FD+M, se logran eficiencias mayores con desviaciones bajas en relación al uso de estos dominios de forma individual. Esto puede implicar, dependiendo de la aplicación, decir si se sacrifica eficiencia en el proceso de diagnóstico a cambio de lograr mejoras en el tiempo de entrenamiento, lo cual es crucial para

aplicaciones en línea.

Se demuestra la importancia de fusionar dominios de representación, e igualmente determinar relevancias de grupos de características, ya que el desempeño global bajo de un dominio, o de una característica, no implica que ellos puedan no aportar información de variabilidad del proceso que está ausente en otros dominios y/o características. Además, también es posible que ciertos dominios y características afectan el tiempo de entrenamiento, el poder de generalización e incluso que conlleven a sobre-entrenamiento del clasificador. En este aspecto se destaca al dominio MFCC, el cual reduce un orden de magnitud su tiempo de entrenamiento al ser fusionado con otros dominios, a la vez que mejora su eficiencia a medida que se incrementa la complejidad del clasificador, tal como fue el caso del modelo HMM a medida que se incrementa su número de estados. En este punto, la decisión binaria de fallo y no fallo sigue siendo posible, sin embargo, al considerar niveles de severidad ahora se logran eficiencias que alientan a se presentan variaciones marcadas en la eficiencia y desviación.

Al afrontar la tarea de diagnóstico y pronóstico, se hace importante discriminar la existencia de energías espectralmente localizadas sobre sub-bandas estrechas en las señales adquiridas, lo cual es preponderante con fallos en rodamientos que se comportan de manera cuasi-estacionaria, o no estacionaria. Para esta tarea es importante realizar etapas de pre-procesamiento de las señales, incluyendo realzamiento de características mediante FBM, o por segmentaciones en marcos para procesamiento de caracterización multi-dominio que permita discriminar y analizar relevancia de las representaciones de señal. Para la selección de características se puede emplear diversas técnicas, algunas basadas en medidas de similaridad, en selección por relevancia forward o backward, e incluso por métodos de selección estocástica. Se debe tener presente que en condiciones reales de operación las señales pueden estar contaminadas, por lo que una fusión de dominios ayuda en discriminación de información relevante.

Los modelos HMM se presentan como una gran herramienta para la clasificación estocástica de estados de severidad en rodamientos (pronóstico), al demostrar que en tareas de solo diagnóstico la eficiencia crece a la par del número de estados del modelo, pero en tareas de pronóstico (identificación de fallos y niveles de severidad) se logran buenos resultados en eficiencia desde un número bajo de estados, normalmente en el rango de 1 a 4 estados es suficiente. Estos modelos HMM, además, presentan un gran poder de generalización y no sufren de sobre entrenamiento ante variaciones del tamaño de la base para entrenamiento, aunque al incorporar ni-

veles de severidad el número total de modelos HMM que se requiere entrenar crece. También se destaca, que producto del análisis de modelos HMM empleando solo caracterización MFCC, se puede encontrar que la separación de clases e intra-clases se está viendo afectada con la adición de niveles de severidad, lo cual conlleva a emplear enfoques que permitan afrontar este tipo de problemas, como son los modelos HHMM o una selección estocástica de las características.

Producto del análisis inter-dominio e intra-dominio con el fin de seleccionar tipos de dominios y sus características para clasificación mediante un enfoque de aprendizaje de máquina jerárquico, se propone un modelo HHMM de dos capas y dos estados internos, donde las características basadas en tiempo-frecuencia (específicamente el dominio MFCC) son las más relevantes para separar el tipo de fallo (capa 1) y los parámetros basados en frecuencia benefician la discriminación entre los niveles de severidad (capa 2). Además, la estrategia propuesta se complementa con un subconjunto relevante de características que favorece etapas posteriores de aprendizaje mediante la codificación de relaciones de datos complejos relacionadas con interacciones ocultas entre el tipo de fallo y los niveles de severidad; para ello se realiza un análisis de las eficiencias individuales de las características en cada dominio mediante un análisis forward y backward, arrojando un conjunto reducido de características con poder discriminante diferente entre las dos capas propuestas.

El modelo propuesto de aprendizaje de máquina basado en HHMM, empleando un conjunto relevante de características, presenta una eficiencia total en la identificación de fallos y severidades (pronóstico) del 97.7 %, la cual es comparable y ligeramente superior a la que se presenta en un modelo HHMM donde se incluyen todas las características multi-dominios (97.5 %), demostrando con ello que la selección de características y enfoque jerárquico permiten mejorar el poder discriminante sobre las señales de vibración. Además estos resultados están en línea con otros reportados en la literatura, como es el caso de los resultados consignados en [21], donde se logra una eficiencia de clasificación de 97 % sobre la misma base de datos, pero bajo un enfoque que no reporta datos por niveles de severidad ni por tipos de fallo, solo se hace de forma global para el proceso, lo cual genera incertidumbre sobre la eficiencia individual de cada tipo de fallo, de cada tipo de severidad y las desviaciones presentes.

Los métodos no supervisados para la clasificación de relevancia, con fines de selección de características, no pueden eliminar la información redundante en comparación con otras técnicas de selección de características supervisadas, lo cual se debe al alto nivel de ineficiencia para disminuir el ruido que introducen las caracte-

terísticas redundantes en el sistema. En contraste, se destaca por su eficiencia el método Relief-F sobre otras técnicas de selección, además posee baja complejidad matemática. Relief-F presenta su mejor eficiencia empleando solo características del dominio TFD, y específicamente con base en coeficientes MFCC. En este contexto, y en comparación con otros dominios de representación, MFCC es el mejor método que permite codificar de forma simultánea tanto información del dominio del tiempo como de la frecuencia de una señal de vibración, y además permite discriminar entre niveles de severidad.

El enfoque de selección de características basado en Relief-F y aprendizaje de máquina por HMM (modelo SFS) logró resultados en concordancia con el estado del arte relacionado con el diagnóstico y pronóstico de fallos, sin embargo, la estrategia presentada permite extraer subconjuntos relevantes de características los cuales favorecen la subsecuente etapa de aprendizaje codificando las relaciones complejas de los datos relacionadas con interacciones ocultas entre los tipos de fallos y los niveles de severidad.

Finalmente, del estudio comparativo entre las metodologías propuestas en este trabajo (basadas en modelos HHMM y SFS) y publicaciones previas, reportadas en el estado del arte para el diagnóstico y pronóstico de fallos, se puede observar que la base de métodos propuestos para la evaluación de características son parámetros estadísticos en los dominios del tiempo, de la frecuencia y del tiempo-frecuencia, dominando ampliamente sobre metodologías basadas en métricas de información. Además, las metodologías sin selección de características poseen usualmente las eficiencias menores, en contraste con las metodologías con selección de características que poseen mejor desempeño. Se resalta que ninguno de los enfoques, en comparación con los reportados en este documento, emplea los 3 niveles de severidad presentes en las señales de vibración disponibles desde la base de datos, estando restringidos en el mejor de los casos a 2 niveles de severidad. En este sentido, los métodos propuestos desempeñan satisfactoriamente la tarea de diagnóstico/pronóstico equilibrando el número de características y la eficiencia, con el fin de poder entregar información discriminante de todos los niveles de severidad.

Ya que las metodologías propuestas en este trabajo, HHMM y SFS, permiten eficiencias en el pronóstico de estados de salud en rodamientos que incluyen completamente niveles definidos de severidad, es posible emplear estos mismos niveles como un índice de degradación aplicable en rodamientos, toda vez que los niveles son un indicativo del tipo de severidad presente. Además, este tipo de definición de índice de degradación permite ser extensible a otros tipos de aplicaciones, donde las meto-

dologías propuestas pueden ser ajustadas en concordancia con las particularidades de cada sistema con el fin de entregar pronóstico de vida útil remanente con base en la progresión de la degradación, o nivel de severidad, presente.

17.2. TRABAJOS FUTUROS

Aunque los modelos de pronóstico propuestos demostraron su eficiencia a través de su aplicación en varias bases de datos, y en especialmente sobre una de conocida referencia en el estado del arte, se requieren de mayores validaciones y extensiones con el fin de poder afrontar diferentes tipos de fallos en sistemas de variada naturaleza como sistemas eléctricos, sistemas mecánicos, sistemas electro-mecánicos, sistemas de infraestructura, etc. Para cada una de estas aplicaciones se debe estudiar cual es el conjunto de señales, características, análisis de relevancia, sistema de aprendizaje de máquina y forma de entregar degradación, con el fin de tener un sistema de pronóstico particularizado.

También se debe trabajar sobre la forma de entregar información de predicción, basada en la degradación, de tal forma que sea exacta durante largos periodos de tiempo y ajustable a las necesidades particulares de cada aplicación. En este escenario, el número de estados discretos posibles de degradación se deben estudiar, además de profundizar en modelos de entrega de información sobre la vida útil remanente según la aplicación, tales como tiempo para fallo, reconfiguración de operación para confiabilidad, confiabilidad esperada en un horizonte de tiempo, etc.

Según el tipo de aplicación, se deben ajustar los modelos para operación fuera de línea u operación en línea. Esto implica trabajar fuertemente en el estudio de la complejidad de los algoritmos de extracción de características, de reducción de dimensionalidad, de análisis de relevancia y de aprendizaje de máquina que se ajustan a los requerimientos. De forma genérica, estas aplicaciones tratarán con grandes cantidades de datos, cada vez más disponibles, que se deben integrar dentro del proceso de pronóstico.

Será un objetivo de trabajo a largo plazo poder desarrollar sistemas de software desplegados en campo, que se ajusten a las señales de entrada específicas, que afronten las particularidades del proceso y que entreguen información precisa, confiable y en términos demandados por el usuario final.

Referencias

- [1] A. Heng, S. Zhang, A.C.C. Tan and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities." *Mechanical Systems and Signal Processing*, Vol. 23, Elsevier, pp. 724–739, 2009.
- [2] GEP, Insight Drives Innovation, Market Intelligence News Desk, "Industrial Maintenance Services in North America," Available URL: <https://www.gep.com/mind/blog/industrial-maintenance-services-north-america>, october 03, 2016.
- [3] K. M. Blache, Benchmarks Shed Light on Maintenance & Reliability Perceptions. Reliabilityweb.com, 2015. Available URL: http://reliabilityweb.com/index.php/articles/benchmarking_a_better_understanding en Mayo 25 de 2015.
- [4] S. J. Z. Sikorska, M. Hodkiewicz and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry." *Mechanical Systems and Signal Processing*, Vol. 25, Elsevier, pp. 1803–1836, 2011.
- [5] Revista Business Chile, Adultos mayores: La fuerza laboral del futuro. Cámara Chilena Norteamericana de Comercio - Amcham Chile, Available on: <https://www.amchamchile.cl/2017/09/adultos-mayores-la-fuerza-laboral-del-futuro>, septiembre 15 de 2017.
- [6] The International Energy Agency. Key World Energy Statistics, 2010. International Energy Agency, Cedex, Paris, 2010.
- [7] The International Energy Agency. Key World Energy Statistics, 2016. International Energy Agency, Cedex, Paris, 2017.
- [8] H. De Keulenaer, R. Belmans, E. Blaustein, D. Chapman, A. De Almeida, B. De Wachter, P. Radgen and F. Isi. *Energy Efficient Motor Driven Systems*. Published by European Copper Institute, Tervurenlaan, Brussels, Belgium, 2004.
- [9] O.V. Thorsen and M. Dalva. "A survey of faults on induction motors in offshore oil industry, petrochemical industry, gas terminals and oil refineries." IEEE Paper No. PCIC-94-01, 1994.
- [10] J. Lee, F. Wu, W. Zhao, M. Ghaffari and L. Liao, "Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications." *Mechanical Systems and Signal Processing*, Vol. 42, Elsevier, pp. 314–334, 2014.

- [11] J. Yan, C. Guo and X. Wang, "A dynamic multi-scale Markov model based methodology for remaining life prediction". *Mechanical Systems and Signal Processing*, Vol. 2, Elsevier, pp. 1364–1376, 2011.
- [12] D. A. Tobon-Mejia, K. Medjaher and N. Zerhouni. "The ISO 13381-1 Standard's Failure Prognostics Process Through an Example." *Proceedings of IEEE Prognostics & System Health Management Conference, PHM2010, Macau, China, January 2010, CD ROM*, 11 pp., 2010.
- [13] R. Yan, R. Gao and X. Chen, "Wavelets for fault diagnosis of rotatory machines: a review with applications." *Signal Processing*, Vol. 96, Elsevier, pp. 1–15, 2014.
- [14] Q. He and X. Wang. "Time–frequency manifold correlation matching for periodic fault identification in rotating machines." *Journal of Sound and Vibration*, Vol. 332, pp. 2611–2626, 2013.
- [15] L.Cui, J. Wang and S. Lee. "Matching pursuit of an adaptive impulse dictionary for bearing fault diagnosis." *Journal of Sound and Vibration*, Vol. 333, pp. 2840–2862, 2014.
- [16] R. Rammohan and M.R. Taha. "Exploratory investigations for intelligent damage prognosis using Hidden Markov Models." *Systems, Man and Cybernetics, 2005 IEEE International Conference on, 12-12 Oct., Vol. 2*, pp. 1524–1529, 2005.
- [17] Los Alamos National Laboratory. *Damage Prognosis Workshop*. Phoenix, Arizona, March 27–29, 2001.
- [18] International Standard ISO 13381-1. *Condition monitoring and diagnostics of machines – Prognostics – Part 1: General guidelines*. Second edition 2015-09. Geneva, Switzerland, 2015.
- [19] N. S. Jammu and P. K. Kankar, "A review on prognosis of rolling element bearings." *International Journal of Engineering Science & Technology* 2011, Vol. 3, Issue 10, pp. 7497–7503, 2011.
- [20] S. Sajjad, H. Zaidi, S. Aviyente, M. Salman, S. Kwang-Kuen and E. G. Strangas, "Prognosis of gear failures in DC starter motors using Hidden Markov Models." *IEEE Transactions on Industrial Electronics*, Vol. 58, No. 5, pp. 1695–1705, 2011.
- [21] I. Attoui, N. Boutasseta, N. Fergani, B. Oudjani and A. Deliou, "Vibration-based bearing fault diagnosis by an integrated DWT-FFT approach and an adaptive neuro-fuzzy inference system." In *CEIT, 2015 3rd International Conference on*, pp. 1–6, 2015.

- [22] L. Liang, F. Liu, M. Li, K. He and G. Xu, "Feature selection for machine fault diagnosis using clustering of non-negation matrix factorization." *Measurement*, Vol. 94, pp. 295–305, 2016.
- [23] Y. Li, J. Shiroishi, S. Danyluk, T. Kurfess and S.Y. Liang, "Bearing fault detection via high frequency resonance technique and adaptive line enhancer." in *Biennial Conference on Reliability, Stress Analysis and Failure Prevention (RSAFP)*, 1997.
- [24] C. Shen, D. Wang, F. Kong and W. T. Peter, "Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier." *Measurement*, Vol. 46, No. 4, pp. 1551–1564, 2013.
- [25] A. Brkovic, D. Gajic, J. Gligorijevic, I. Savic-Gajic, O. Georgieva and S. Di Gennaro, "Early fault detection and diagnosis in bearings for more efficient operation of rotating machinery," *Energy*, Vol. 136, pp. 63–71, 2017.
- [26] S. Darley and P. Robson, "Application of wavelet transform to detect faults in rotating machinery." *ABCM Symposium Series in Mechatronics*, Vol. 1, pp. 616–624, 2004.
- [27] H. Teng, J. Zhao, X. Jia, Y. Jia, X. Zhang and L. Cai, "Experimental study on gearbox prognosis using total life vibration analysis." *2011 Prognostics & System Health Management Conference (PHM2011 Shenzhen)*, IEEE, pp. 1–6, 2011.
- [28] W. Achmad and Y. Bo-Suk, "Machine health prognostics using survival probability and support vector machine." *Expert Systems with Applications*, Vol. 38, Elsevier, pp. 8430–8437, 2011.
- [29] Z. Zhi-Jie, H. Chang-Hua, W. Wen-Bin, Z. Bang-Cheng, X. Dong-Ling and Z. Jian-Fei, "Condition-based maintenance of dynamic systems using online failure prognosis and belief rule base." *Expert Systems with Applications* 39, Elsevier, pp. 6140–6149, 2012.
- [30] M. Yuwono, Y. Qin, J. Zhou, Y. Guo, B. G. Celler and S. W. Su, "Automatic bearing fault diagnosis using particle swarm clustering and hidden markov model." *Engineering Applications of Artificial Intelligence*, Vol. 47, pp. 88–100, 2016.
- [31] H. Ocak and K. Loparo, "HMM-Based fault detection and diagnosis scheme for rolling element bearings." *Journal of Vibration and Acoustics*, Vol. 127, pp. 299–306, 2005.
- [32] E. Zio and F. Di Maio. "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system." *Reliability Engineering and System Safety* 95, Elsevier, pp. 49–57, 2010.

- [33] K. Hack-Eun, A.C.C. Tan, J. Mathew and C. Byeong-Keun, "Bearing fault prognosis based on health state probability estimation." *Expert Systems with Applications*, pp. 5200–5213, 2012.
- [34] R. Huang, L. Xi, X. Li, C.R. Liu, H. Qiu and J. Lee. "Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods." *Mechanical Systems and Signal Processing*, Vol. 21, pp. 193–207, 2007.
- [35] Z. Zhi-Jie, H. Chang-Hua, X. Dong-Ling, C. Mao-Yin and Z. Dong-Hua, "A model for real-time failure prognosis based on Hidden Markov Model and belief rule base." *European Journal of Operational Research* 207, Elsevier, pp. 269–283, 2010.
- [36] S. Xiao-Sheng, W. Wenbin, H. Chang-Hua and Z. Dong-Hua, "Remaining useful life estimation - A review on the statistical data driven approaches." *European Journal of Operational Research* 212, Elsevier, pp. 1–14, 2011.
- [37] F. Camcia and R.B. Chinnam. "Dynamic Bayesian Networks for Machine Diagnostics: Hierarchical Hidden Markov Models vs. Competitive Learning." *Proceedings of IEEE International Joint Conference on Neural Networks*, Montreal, Canada, July 31 - August 4, pp. 1752–1757, 2005.
- [38] Z. Xing-hui and K. Jian-she, "Hidden Markov Models in bearing fault diagnosis and prognosis." *2010 Second International Conference on Computational Intelligence and Natural Computing*, CINC2010, IEEE, pp. 364–367, 2010.
- [39] D. White, "Helicopter Usage Monitoring Using the MaxLife System." *DSTO Workshop on Helicopter Health and Usage Monitoring System*, Melbourne, Australia, 1999.
- [40] H. Silverman, "T-HUMS–AH64 Lead the fleet (LTF) summary and glimpse at hermes 450 MT–HUMS." *AIAC Conference*, Melbourne, Australia, Paper 151, 2005.
- [41] J. McColl, "HUMS in the era of CAA, JAA, EASA and ICAO." *AIAC Conference*, Melbourne, Australia, Paper 112, 2005.
- [42] M. DiUlio, C. Savage, B. Finley and E. Schneider, "Taking the integrated condition assessment system to the year 2010." *13th International Ship Control Systems Symposium*, Orlando, FL, 2003.
- [43] D. Siegel, C. Ly and J. Lee, "Methodology and Framework for Predicting Helicopter Rolling Element Bearing Failure." in *IEEE Transactions on Reliability*, Vol. 61, No. 4, pp. 846–857, Dec. 2012.

- [44] Hongzheng Fang, Yi Xiong, Kai Luo and Liming Han, "The study of the helicopter rotor prognostic method and health management system based-on FMA." Proceedings of the IEEE 2012 Prognostics and System Health Management Conference (PHM-2012 Beijing), Beijing, pp. 1–8, 2012.
- [45] L. Zhou et al., "Helicopter gearbox bearing fault detection using separation techniques and envelope analysis." 2016 Prognostics and System Health Management Conference (PHM-Chengdu), Chengdu, pp. 1–5, 2016.
- [46] X. Liqing, Z. Cheng, H. Yan, G. Yunan and F. Haiyan, "A frame design of helicopter health monitoring and diagnosis system based on testability." 2017 Prognostics and System Health Management Conference (PHM-Harbin), Harbin, pp. 1–6, 2017.
- [47] F. Duan, M. Corsar, L. Zhou and D. Mba, "Using independent component analysis scheme for helicopter main gearbox bearing defect identification." 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, pp. 252–259, 2017.
- [48] A. Swamidias and Y. Chen, "Damage detection in a Tripod Tower Platform (TTP) Using Modal Analysis." ASME Offshore Technology, pp. 577–583, 1992.
- [49] L. Xiu, L. Juan, C. Xiu-Rong, S. Xiao-Sheng, Y. Dan-Song and Y. Xue, "Lifetime prognosis of blowout preventer valve for offshore platform based on Wiener process." The 27th Chinese Control and Decision Conference (2015 CCDC), Qingdao, pp. 2432–2436, 2015.
- [50] W. Zhen and Z. Zhigao, "Damage Detection of Offshore Platform Structures Using Time Domain Response Data." 2010 International Conference on Intelligent Computation Technology and Automation, Changsha, pp. 1079–1084, 2010.
- [51] J. Luo, F. Tu, M. Azam, K. Pattipati, L. Qiao and M. Kawamoto, "Intelligent model-based diagnostics for vehicle health management." Proc. SPIE Conference, Orlando, FL., 2003.
- [52] Y. Zhu and Z. Gao, "Robust observer-based fault detection via evolutionary optimization with applications to wind turbine systems." In Industrial Electronics and Applications (ICIEA 2014), IEEE 9th Conference on, pp. 1627–1632, 2014.
- [53] C. Bunks, D. McCarthy and T. Al-Ani, "Condition-Based Maintenance of machines using Hidden Markov Models." Mechanical Systems and Signal Processing, 14(4), pp. 597–612, 2000.

- [54] L. Liporace, "Maximum likelihood estimation for multivariate observations of markov sources." Proc. IEEE Transactions on Information Theory IT-28, pp. 729–734, 1982.
- [55] B. Wu, S. Yu, Y. Luo and C. Feng, "Rolling bearing fault diagnosis method based on SVM-HMM." ICMTMA, International Conference on Measuring Technology and Mechatronics Automation 2010, Changsha City, pp. 295–298, 2010.
- [56] H. Ocak and K.A. Loparo, "A new bearing fault detection and diagnosis scheme based on hidden Markov modeling of vibration signals." Acoustics, Speech and Signal Processing 2001, Proceedings ICASSP 2001, IEEE International Conference on, Vol.5, pp. 3141–3144, 2001.
- [57] Estándar UNE-EN 13306:2011 Mantenimiento - Terminología del mantenimiento. AENOR, Edición 2011-03-09.
- [58] G. Yang, *Life Cycle Reliability Engineering*. John Wiley & Sons, ISBN: 978-0-471-71529-0, 544 pp., 2007.
- [59] Z. Gao, C. Cecati and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques – Part I: Fault diagnosis with model-based and signal-based approaches." IEEE Transactions on Industrial Electronics, Vol. 62, No. 6, pp. 3757–3767, 2015.
- [60] Y. Shatnawi and M. Al-Khassaweneh, "Fault diagnosis in internal combustion engines using extension neural network." IEEE transaction on Industrial Electronics, Vol. 61, No. 3, pp. 1434–1443, 2014.
- [61] Y. Lei, Z. He and Y. Zi, "Application of the EEMD method to rotor fault diagnosis of rotating machinery." Mechanical Systems and Signal Processing, Vol. 23, No. 4, pp. 1327–1338, 2009.
- [62] S. Janjarasjitt, H. Ocak and K. Loparo, "Bearing condition diagnosis and prognosis using applied nonlinear dynamical analysis of machine vibration signal." Journal of Sound and Vibration Vol. 317, pp. 112–126, 2008.
- [63] D. P. Jena and S. N. Panigrahi, "Automatic gear and bearing fault localization using vibration and acoustic signals." Applied Acoustics, Vol. 98, pp. 20–33, 2015.
- [64] N. Baydar and A. Ball, "Detection of gear failures via vibration and acoustic signals using wavelet transform." Mechanical Systems and Signal Processing, Vol. 17, No. 4, pp. 787–804, 2003.

- [65] J. Lin, "Feature extraction of machine sound using wavelet and its application in fault diagnosis." *NDT & E International*, Vol. 34, No. 1, pp. 25–30, 2001.
- [66] Y. Lei, J. Lin, Z. He and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery." *Mechanical Systems and Signal Processing*, Vol. 35, pp. 108–126, 2013.
- [67] M. Amarnath and I. R. Praveen Krishna, "Local fault detection in helical gears via vibration and acoustic signals using EMD based statistical parameter analysis." *Measurement: Journal of the International Measurement Confederation*, Vol. 58, pp. 154–164, 2014.
- [68] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial." *Mechanical Systems and Signal Processing*, Vol. 25, No. 2, pp. 485–520, 2011.
- [69] Z. K. Peng, P. W. Tse and F. L. Chu, "A comparison study of improved Hilbert-Huang transform and wavelet transform: application to fault diagnosis for rolling bearing." *Mechanical Systems and Signal Processing*, Vol. 19, No. 5, pp. 974–988, 2005.
- [70] R. Ricci and P. Pennacchi, "Diagnostics of gear faults based on EMD and automatic selection of intrinsic mode functions." *Mechanical Systems and Signal Processing*, Vol. 25, No. 3, pp. 821–838, 2011.
- [71] J. Obuchowski, A. Wyłomanska and R. Zimroz, "Selection of informative frequency band in local damage detection in rotating machinery." *Mechanical Systems and Signal Processing*, Vol. 48, No. 12, pp. 138–152, 2014.
- [72] M. Amarnath and I. R. Praveen Krishna, "Empirical mode decomposition of acoustic signals for diagnosis of faults in gears and rolling element bearings." *IET Science, Measurement & Technology*, Vol. 6, No. 4, pp. 279–287, 2012.
- [73] M. J. E. Salami and S. N. Sidek, "Parameter estimation of multicomponent transient signals using deconvolution and arma modelling techniques." *Mechanical Systems and Signal Processing*, Vol. 17, pp. 1201–1218, 2003.
- [74] Y. Li, S. Billington, C. Zhang, S. Danyluk and S. Liang, "Adaptive prognostics for rolling element bearing condition." *Mechanical System and Signal Processing*, Vol. 13, pp. 103–113, 1999.
- [75] J.T. Kim and R.H. Lyon, "Cepstral analysis as a tool for robust processing, deconvolution and detection of transients." *Mechanical Systems and Signal Processing*, Vol. 6, pp. 1–15, 1992.

- [76] M. Van and H. J. Kang, "Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection." *IET Science, Measurement & Technology*, Vol. 9, No. 6, pp. 671–680, 2015.
- [77] M. Cococcioni, B. Lazzerini and S. L. Volpi, "Robust diagnosis of rolling element bearings based on classification techniques." *IEEE Transactions* Vol. 9, No. 4, pp. 2256–2263, 2013.
- [78] M. A. Rizvi, A. I. Bhatti and Q. R. Butt, "Hybrid model of the gasoline engine for misfire detection." *IEEE Transactions on industrial Electronics*, Vol. 58, No. 8, pp. 3680–3692, 2011.
- [79] P. E. William and M. W. Hoffman, "Identification of bearing faults using time domain zero-crossings." *Mechanical systems and signal processing*, Vol. 25, No. 8, pp. 3078–3088, 2011.
- [80] R. Tiwari, V. K. Gupta and P. Kankar, "Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier." *Journal of Vibration and Control*, Vol. 21, No. 3, pp. 461–467, 2015.
- [81] Q. Hu, Z. He, Z. Zhang and Y. Zi, "Fault diagnosis of rotating machinery based on improved wavelet package transform and svms ensemble." *Mechanical Systems and Signal Processing*, Vol. 21, No. 2, pp. 688–705, 2007.
- [82] Y. Wang, Q. Ma, Q. Zhu, X. Liu and L. Zhao, "An intelligent approach for engine fault diagnosis based on hilbert–huang transform and support vector machine." *Applied acoustics* 75, pp. 1–9, 2014.
- [83] X. Zhang, Y. Liang and J. Zhou, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized svm." *Measurement* 69, pp. 164–179, 2015.
- [84] H. Ocak and K. A. Loparo, "Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data." *Mechanical Systems and Signal Processing*, Vol. 18, pp. 515–533, 2004.
- [85] F. V. Nelwamondo and T. Marwala, "Faults detection using Gaussian mixture models, Mel-Frequency Cepstral Coefficients and Kurtosis." *IEEE International Conference on Systems, Man and Cybernetics*, Taipei, Taiwan. pp. 290–295, October 8–11, 2006.

- [86] Componentes y materiales de un rodamiento, SKF. Available on: <http://www.skf.com/co/products/bearings-units-housings/principles/general-bearing-knowledge/bearing-basics/components-and-materials/index.html>, 21 de mayo de 2018.
- [87] H. Li, L. Fu and H. Zheng, "Bearing fault diagnosis based on amplitude and phase map of Hermitian wavelet transform." *Journal of Mechanical Science and Technology*, Vol. 25, No. 11, pp. 2731–2740, 2011.
- [88] FAG Sales Europe GmbH, *Averías de los rodamientos Reconocimiento de daños e inspección de rodamientos*. No. publicación WL 82 102/2 SB, FAG Sales Europe GmbH - España, una filial del Grupo FAG Kugelfischer.
- [89] ISO TC 108/SC 5, ISO 13381-1:2015 Condition monitoring and diagnostics of machines -- Prognostics -- Part 1: General guidelines. 2015-09.
- [90] ISO TC 108/SC 5, ISO 17359:2018 Condition monitoring and diagnostics of machines -- General guidelines. 2018-01.
- [91] A. C. C. Tan and J. Mathew, "The adaptive noise cancellation and blind deconvolution techniques for detection of rolling elements bearing faults - A comparison." in *ACSIM Proceedings*, 2002.
- [92] R. XU and C. Kwan, "Robust Isolation of Sensor Failures." *Asian Journal of Control*, Vol. 5, pp. 12–23, 2003.
- [93] A. K. S. Jardine, D. Lin and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance." *MSSP 2006*, pp. 1483–1510, 2006.
- [94] C. M. Moncada, "Estudio de fallas comunes en rodamientos rígidos de una hilera de bolas mediante análisis en el dominio tiempo y de la frecuencia." Tesis presentada para optar al grado de Licenciado en Acústica y al título de profesional de Ingeniero Acústico, Universidad Austral de Chile, 2002.
- [95] A. Barkov, N. Barkova and J. Mitchell, "Condition Assessment and Life Prediction of Rolling Element Bearing-Part 1." *Sound and Vibration*, 10–17, June 1995.
- [96] W. J. Wang and P. D. McFadden, "Application of wavelets to gearbox vibration signals for fault detection." *Journal of Sound and Vibration*, Vol. 192, pp. 927–939, 1996.

- [97] G. Y. Luo, D. Osypiw and M. Irle, "On-line vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing." *Journal of Vibration and Control*, Vol. 9, pp. 931–947, 2003.
- [98] G . O. Chandroth and W. J. Staszewski, "Fault detection in internal combustion engines using wavelet analysis." in *COMADEM 1999*, Chipping Norton, pp. 7–15, 1999.
- [99] S. Davis and P. Mermelstein, "Comparison of parametric representation for monosyllable word recognition in continuously spoken sentences." *IEEE Transactions on Acoustic, Speech and Signal Processing*, Vol. 28, No. 9, pp. 357–366, 1980.
- [100] L. Muda, M. Begam and I. Elamvazuthi, "Voice recognition algorithms using Mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques." *Journal of computing*, Vol. 2, Issue 3, ISSN 2151–9617, 2010.
- [101] X. Huang, A. Acero and H. Hon, *Spoken language processing - A guide to theory, algorithm and system development*. Prentice Hall PTR Upper Saddle River, New Jersey, USA, ISBN: 0130226165, 2001.
- [102] S. Okumura, "The short time fourier transform and local signals." A dissertation submitted to the graduate school in partial fulfillment of the requirements for the degree Doctor of Philosophy in Statistics, Department of Statistics, Carnegie Mellon University Pittsburgh, Pennsylvania, June, 2011.
- [103] L.R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition." *Proc. IEEE*, Vol. 77, No. 2, pp. 257–285, 1989.
- [104] E. Yariv, "Hidden Markov Processes." *Proc. IEEE Transactions on Information Theory*, Vol, 48, No. 6, pp. 1518–1569, June 2002.
- [105] M. Shashanka, "A fast algorithm for discrete HMM training using observed transitions." United Technologies Research Center, East Hartford, CT 06108, Boston University, Available on: <http://cns.bu.edu/~mvss/stuff/ShashankaICASSP2011.pdf>
- [106] J. R. Movellan, "Tutorial on Hidden Markov Models." Machine Perception Laboratory, UC San Diego, Available on: <http://mplab.ucsd.edu/tutorials/hmm.pdf>.
- [107] O. Cappé, E. Moulines and T. Rydén, "Inference in hidden Markov models." In *Proceedings of EUSFLAT Conference*, pp. 14–16, 2009.

- [108] H. Zhou, J. Chen, G. Dong and R. Wang, "Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden markov model." *Mechanical systems and signal processing*, Vol. 72, pp. 65–79, 2016.
- [109] K. Murphy, Hidden markov model (hmm) toolbox for matlab. Available on: <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>, 1998.
- [110] S. Fine, Y. Singer and N. Tishby, "The hierarchical hidden markov model: Analysis and applications." *Machine Learning*, Vol. 32, pp. 41–62, Kluwer Academic Publishers, Boston, 1998.
- [111] O. Samko, A. D. Marshall and P. L. Rosin, "Automatic construction of hierarchical hidden Markov model structure for discovering semantic patterns in motion data." *VISAPP*, 2010.
- [112] S. Doebbling, C. Farrar, M. Prime and D. Shevitz, "Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review." Los Alamos National Laboratory report LA-13070-MS, 1996.
- [113] C. Farrar, H. Sohn, F. Hemez, M. Anderson, M. Bement, P. Cornwell, S. Doebbling, J. Schultzer, N. Lieven and A. Robertson, "Damage prognosis: Current status and future needs." Los Alamos National Laboratory report LA-14051-MS, 2003.
- [114] J. Flett and G. Bone, "Fault detection and diagnosis of diesel engine valve trains." *Mechanical Systems and Signal Processing*, Vol. 72–73, pp. 316–327, 2016.
- [115] D. Martínez-Rego, O. Fontela-Romero, A. Alonso-Betanzos and J. C. Principe, "Fault detection via recurrence time statistics and one-class classification." *Pattern Recognition Letters*, Vol. 84, pp. 8–14, 2016.
- [116] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks." *IEEE Transactions on Industrial Electronics*, Vol. 60, No. 8, pp. 3398–3407, 2013.
- [117] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification." *IEEE Transactions on Instrumentation and Measurement*, Vol. 53, No. 6, pp. 1517–1525, 2004.

- [118] R. Shao, W. Hu, Y. Wang and X. Qi, "The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform." *Measurement*, Vol. 54, pp. 118–132, 2014.
- [119] Z. Wei, Y. Wang, S. He and J. Bao, "A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection." *Knowledge-Based Systems*, Vol. 116, pp. 1–12, 2017.
- [120] J. Zheng, H. Pan, S. Yang and J. Cheng, "Generalized composite multiscale permutation entropy and laplacian score based rolling bearing fault diagnosis." *Mechanical Systems and Signal Processing*, Vol. 99, pp. 229–243, 2018.
- [121] K. A. Loparo, Seeded Fault Test Data, Bearing Data Center, Case Western Reserve University. Available on URL: <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.
- [122] D. Peña, *Análisis de Datos Multivariantes*. McGRAW HILL, ISBN: 978-84-481-3610-9, 2002.
- [123] D. H. Jeong, C. Ziemkiewicz, W. Ribarsky and R. Chang, "Understanding principal component analysis using a visual analytics tool." Charlotte Visualization Center, UNC Charlotte, 2009.
- [124] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications." *Neural Networks Research Centre*, Helsinki University of Technology, Finland *Neural Networks*, Vol. 13, pp. 411–430, 2000.
- [125] S. Valle, W. Li and S. J. Qin, "Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods." Department of Chemical Engineering, The University of Texas at Austin, Austin, Texas, 1999.
- [126] F. Udina, "Interactive biplot construction." Universitat Pompeu Fabra, *Journal of Statistical Software*, Vol. 13, Issue 5, 2005.
- [127] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, "An efficient k-Means clustering algorithm: Analysis and implementation." *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 24, No. 7, July 2002.
- [128] P. Cichosz, *Data Mining Algorithms: Explained Using R*. John Wiley & Sons, 716 pp., ISBN: 9781118332580, 2015.

- [129] B. Zhang, M. Hsu and U. Dayal, "K-Harmonic means – A data clustering algorithm." Software Technology Laboratory, HP Laboratories, Palo Alto HPL, October, 1999.
- [130] H. Dolfing. "K-Harmonic means clustering algorithm." University of Konstanz, Seminar Actual Developments in Visual Data Mining, 2004.
- [131] A. Kolesnikov and Elena Trichina, "Determining the number of clusters with rate-distortion curve modeling." Image Analysis and Recognition Lecture Notes in Computer Science, Vol. 7324, pp. 43–50, 2012.
- [132] G. G. Yen and K.-C. Lin, "Wavelet packet feature extraction for vibration monitoring." IEEE Transactions on Industrial Electronics, Vol. 47, No. 3, pp. 650–667, 2000.
- [133] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, Vol. 454, No. 1971, pp. 903–995, 1998.
- [134] J. Ye, "Fault diagnosis of turbine based on fuzzy cross entropy of vague sets." Expert Systems with Applications, Vol. 36, No. 4, pp. 8103–8106, 2009.
- [135] E. F. Sierra-Alonso, O. Cardona-Morales, C. D. Acosta-Medina and G. Castellanos-Dominguez, "Spectral correlation measure for selecting intrinsic mode functions." In Progress in Pattern Recognition, Image Analysis, Computer Vision and Applications, pp. 231–238, Springer, Berlin, Germany, 2014.
- [136] Y. Lei, Z. He and Y. Zi, "A new approach to intelligent fault diagnosis of rotating machinery." Expert Systems with Applications, Vol. 35, No. 4, pp. 1593–1600, 2008.
- [137] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve." Radiology, Vol. 143, pp. 29–36, 1982.
- [138] N. B. Mohamad, "Optimization of decision threshold for speaker recognition." Faculty of Electrical Engineering, University Teknologi Malaysia, October 2003.
- [139] A. R. van Erkel and P. Pattynama, "Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology." European Journal of Radiology, Vol. 27, pp. 88–94, 1998.
- [140] S. Pérez, "Estimación de la curva ROC acumulativa / dinámica." Máster universitario en modelización e investigación matemática, estadística y computacion, Universidad de Oviedo, Julio de 2015.

- [141] A. J. Dwyer, "In pursuit of a piece of the ROC." *Radiology*, Vol. 202, pp. 621–625, 1997.
- [142] S. Vida, "A computer program for non-parametric receiver operating characteristic analysis." *Comput Methods Programs Biomed*, Vol. 40, pp. 95–101, 1993.
- [143] E. R. DeLong, D. M. DeLong and C. P. DL, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach." *Biometrics*, Vol. 44, pp. 837–845, 1988.
- [144] K. Daphne and S. Mehran, "Toward optimal feature selection." Technical Report, Stanford InfoLab, Deposited On: 28 Oct 2001, Last Modified: 09 Dec 2008, URL: <http://ilpubs.stanford.edu:8090/208/>.
- [145] A. Boudiaf, A. Moussaoui, A. Dahane and I. Atoui, "A comparative study of various methods of bearing faults diagnosis using the Case Western Reserve University data." *J Fail, Anal and Preven*. Vol. 16, pp. 271–284, Doi: 10.1007/s11668-016-0080-7, 2016.
- [146] I. Kononenko, "Estimating attributes: analysis and extensions of relief." in: *European conference on machine learning*, Springer, pp. 171–182, 1994.
- [147] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of relief and rrelief." *Machine learning*, Vol. 53, pp. 23–69, 2003.
- [148] B. Nayana and P. Geethanjali, "Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal." *IEEE Sensors Journal* Vol. 17, No. 17, pp. 5618–5625, 2017.
- [149] J. A. Hernández-Muriel, A. M. Álvarez-Meza, J. D. Echeverry-Correa, A. A. Orozco-Gutierrez and M. A. Álvarez-López, "Feature relevance estimation for vibration-based condition monitoring of an internal combustion engine." *Tecno-Lógicas*, Vol. 20, No. 39, pp. 159–174, 2017.
- [150] G. Daza-Santacoloma, J. D. Arias-Londono, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruíz and G. Castellanos-Dominguez, "Dynamic feature extraction: an application to voice pathology detection." *Intelligent Automation & Soft Computing*, Vol. 15, No. 4, pp. 667–682, 2009.
- [151] B.-S. Yang, T. Han and J. L. An, "Art-kohonen neural network for fault diagnosis of rotating machinery." *Mechanical Systems and Signal Processing*, Vol. 18, No. 3, pp. 645–657, 2004.

- [152] R. Tiwari, V. K. Gupta and P. Kankar, "Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier." *Journal of Vibration and Control*, Vol. 21, No. 3, pp. 461–467, 2015.
- [153] Z. Liu, H. Cao, X. Chen, Z. He and Z. Shen, "Multi-fault classification based on wavelet svm with pso algorithm to analyze vibration signals from rolling element bearings." *Neurocomputing*, Vol. 99, pp. 399–410, 2013.
- [154] B. Muruganatham, M. Sanjith, B. Krishnakumar and S. S. Murty, "Roller element bearing fault diagnosis using singular spectrum analysis." *Mechanical systems and signal processing*, Vol. 35, No. 1, pp. 150–166, 2013.
- [155] H. Shao, H. Jiang, F. Wang and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis." *Knowledge-Based Systems*, Vol. 119, pp. 200–220, 2017.