



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Instituto de Ingeniería Energética
Departamento Ingeniería Eléctrica

TESIS DOCTORAL

Aplicación de sistemas de autoaprendizaje al diagnóstico de máquinas eléctricas

Autor: Jordi Burriel Valencia

Programa de Doctorado en Ingeniería y Producción Industrial

Director: Dr. Rubén Puche Panadero

Dr. Javier Martínez Román

Valencia, Junio del 2016

Resumen

El mantenimiento de las máquinas eléctricas es esencial para garantizar la continuidad de los procesos industriales. Gran parte de las investigaciones desarrolladas en los últimos años en el campo de la detección de fallos en las máquinas eléctricas rotativas se han enfocado en la búsqueda de características discriminatorias dentro de las magnitudes medibles (mecánicas y eléctricas) de los motores y generadores instalados en la industria tradicional, así como en sectores emergentes como energía eólica, tracción eléctrica, etc. El objetivo de las técnicas de diagnóstico de fallo es el de determinar el estado del motor, identificar, en su caso, el tipo de fallo que se ha producido, y cuantificar su severidad, con el fin de desarrollar acciones preventivas y correctivas, y planes de contingencia que minimicen el impacto económico de la avería. Una de las opciones más empleadas en los últimos años, junto con el análisis de las vibraciones de la máquina, ha sido el análisis de las corrientes del estator del motor, a fin de detectar en las mismas las perturbaciones características producidas por los fallos de la máquina. Además, la corriente, su valor instantáneo y su evolución con el tiempo, se encuentran disponibles muchas veces en las instalaciones, o incluso en los equipos de control de los motores, y, en caso contrario, resultan fácil de medir de forma no invasiva mediante el uso de una pinza amperimétrica adecuada.

En la literatura técnico-científica, se pueden encontrar diferentes técnicas de análisis y procesamiento de señal que permiten discriminar entre una máquina eléctrica en estado de fallo o en estado sano. Generalmente, cada una de estas técnicas o métodos es válida para una magnitud concreta, un régimen concreto de trabajo, para unos fallos determinados, etc... Por tanto, presentan una limitación intrínseca debido a que cada técnica o método presenta unas características diferentes y por tanto se hace necesario un experto

que sea capaz de interpretar los resultados en función de la técnica o método empleado para discernir entre la máquina con fallo o sin él.

Por todo ello, se hace necesaria la automatización de esta labor, mediante el uso de las técnicas de aprendizaje e inteligencia artificial basadas en clasificadores. Por ello, es prioritario investigar y desarrollar nuevos sistemas de detección que maximicen la precisión respecto a las características disponibles. Para mejorar la clasificación de fallos con respecto a los sistemas actuales, se estudian sistemas cada vez más complejos basados en sistemas expertos capaces de aprender a identificar los fallos, como por ejemplo los sistemas de clasificación multidimensionales o los sistemas expertos basados en la simulación de un “cerebro” (red de neuronas).

Dado que mi formación abarca desde el campo de la automática industrial hasta los sistemas expertos e inteligencia artificial, he procurado integrar mis conocimientos adquiridos en estos campos diversos en el desarrollo para esta tesis de investigación de un “suprasistema” automático capaz de generar sistemas óptimos de diagnóstico de fallos. Estos sistemas son adaptables a la tipología de la máquina a la que se les aplica, pudiendo desarrollar mediante este “suprasistema” de forma automática nuevos sistemas óptimos de diagnóstico de fallos cuando se aplican a nuevos tipos de máquinas. Los sistemas de diagnóstico generados son óptimos en cuanto a la maximización de la precisión en la detección del fallo, minimizando los costes de cálculo como objetivo secundario. Además, a partir de las características discriminatorias que se obtienen de los diversos métodos y transformaciones de detección de los fallos, los sistemas expertos de clasificación se han adaptado respecto a la naturaleza de estas características con el fin de obtener el mejor sistema de detección para cada tipo fallo.

El “suprasistema” propuesto se encarga de generar el sistema de diagnóstico óptimo para la detección de fallos en motores asíncronos de inducción. Este sistema de diagnóstico óptimo está compuesto por una técnica de diagnóstico y un sistema experto clasificador. Para ello se han seguido los siguientes pasos:

- **Adquisición de los datos.**

Se ha empleado un banco de ensayos experimental dónde se disponen dos motores acoplados mecánicamente por sus ejes. Uno de ellos, el de estudio, que es un motor asíncrono de inducción (sano o con fallo) y el otro motor acoplado es un motor síncrono de imanes permanentes, que actúa como carga o par resistente en el sistema. Se han realizado diferentes ensayos en los que se miden las magnitudes eléctricas de las 3 corrientes de fase del motor y la velocidad para los diferentes supuestos de condiciones de trabajo:

- Velocidad del motor constante y par resistente constante.
- Velocidad del motor variable y par resistente constante.

- Velocidad del motor constante y par resistente variable.
- Velocidad del motor variable y par resistente variable.

Todos los ensayos se han repetido con la finalidad de comprobar su correcto funcionamiento. Además, se han realizado diferentes ensayos de cada uno de los cuatro bloques antes mencionados, con diferentes velocidades del motor a ensayar, y con diferentes niveles de carga o par resistente. En total, se han obtenido 735 muestras a las cuales se les aplicarán las siguientes técnicas para que sean analizadas con el suprasistema.

▪ **Aplicación de las técnicas y/o métodos de diagnosis.**

Para la extracción de las características relevantes desde el punto de vista del diagnóstico se han aplicado diferentes técnicas y/o métodos de diagnosis según el régimen de funcionamiento. En esta tesis se aportan dos técnicas novedosas, el uso de la ventana Prolate y el empleo de la transformada de Fourier corta en frecuencia (SFFT).

Las señales empleadas para el diagnóstico han sido:

- La corriente de una fase del estator.
- La corriente de una fase del estator multiplicada por una ventana de tipo Hanning (Hann).

Las técnicas empleadas para el régimen estacionario son:

- Análisis de la señal.
- Análisis de la transformada de Hilbert de la señal (módulo de la señal analítica).
- Análisis de la transformada Cepstrum de la señal.
- Análisis de la transformada del módulo de Park de la señal.
- Análisis del orden de armónico (HOTA) de la señal.

La señal se ha analizado en el dominio de la frecuencia mediante:

- Análisis de la transformada de Fourier de la señal.
- Análisis del método de Welch de la señal.

En conclusión, para el régimen estacionario se han explorado 20 métodos diferentes de diagnóstico, resultado de la combinación de las posibilidades anteriormente descritas.

Las técnicas empleadas para el régimen transitorio son:

- Análisis del orden de armónico (HOTA) de la señal mediante el empleo de la Short-Time Fourier Transform y con una ventana Gaussiana.
- Análisis del orden de armónico (HOTA) de la señal mediante el empleo de la Short-Time Fourier Transform y con una ventana Prolate. (*Aportación a la tesis del uso de esta ventana tipo Prolate dentro de la técnica de diagnóstico*).
- Análisis del orden de armónico (HOTA) de la señal mediante el empleo de la Short-Frequency Fourier Transform y con una ventana Gaussiana. (*Aportación a la tesis de la modificación de la STFT mediante el uso de la SFFT*).
- Análisis del orden de armónico (HOTA) de la señal mediante el empleo de la Short-Frequency Fourier Transform y con una ventana de tipo Prolate.

Además, a todas estas técnicas variantes de HOTA también se le han aplicado una serie de mejoras con respecto al algoritmo original presentado recientemente en la literatura técnica. Estas mejoras disminuyen el tiempo de cálculo y el espacio de memoria ocupado, lo que es esencial en el diseño de un sistema experto. Las mejoras propuestas en la presente tesis son:

- *Una nueva etapa de filtrado paso banda para eliminar las altas frecuencias.*
- *La minimización del efecto de borde.*
- *Un nuevo algoritmo más rápido para el ajuste óptimo de la ventana de filtro.*

En conclusión, para el régimen transitorio se exploran 4 métodos diferentes de diagnóstico.

- **Extracción de las características relevantes.**

Para cada una de las técnicas empleadas, se obtienen las características relevantes, en este caso una dupla por cada muestra o señal medida del motor. De tal modo se obtiene una matriz de vectores de características de dimensión igual al número de métodos de diagnóstico empleados (20 estacionario más 4 en transitorio), donde cada vector tiene la dimensión del número de ensayos que se han realizado (más de 700 ensayos).

▪ **Obtención de los Clasificadores.**

Para cada uno de los vectores de características obtenido con cada una de las técnicas de diagnóstico empleadas, se generan diferentes tipos de clasificadores basados en:

- Redes Neuronales Artificiales de 1 capa oculta.
- Redes Neuronales Artificiales de 2 capas ocultas.
- Máquina de vectores de soporte que emplean un kernel de función polinomial.
- Máquina de vectores de soporte que emplean un kernel de función de base radial gaussiana.

Dentro de cada tipo de clasificador se generan diferentes clasificadores en función del algoritmo de optimización que se emplea (sistema de validación empleado), generando multitud de clasificadores de diferentes coeficientes hasta obtener el óptimo.

▪ **Selección del sistema experto óptimo.**

Como paso final, se analizan los resultados obtenidos y se realiza la selección del sistema óptimo de diagnóstico compuesto de un clasificador y una técnica de diagnosis. El suprasistema obtenido aporta las siguientes ventajas:

1. Obtiene un sistema de diagnóstico óptimo final.
2. Obtiene varios sistemas de diagnóstico óptimos locales (para la misma técnica y sistema experto) con todos los tiempos obtenidos del proceso de entrenamiento y los tiempos estimados de diagnóstico. De esta forma, el suprasistema no aporta solamente una solución final óptima, sino que además permite al operador usar una de las soluciones alternativas generadas óptimas locales en el caso de que sea necesario ajustarse a alguna limitación en el tiempo de cálculo.
3. El suprasistema es totalmente autónomo. El operador debe aportar solo las muestras del tipo de motor específico y el suprasistema genera el sistema óptimo de diagnóstico sin necesidad de intervención del operario.
4. El suprasistema es adaptable al tipo de motor de inducción, de forma que puede generar un sistema de diagnóstico específico y óptimo para cada tipo de motor.

La tesis doctoral se ha estructurado en los siguientes capítulos:

- En el capítulo 1 se realiza una breve introducción que sirve de punto de partida y se presentan los objetivos principales de esta tesis.
- En el capítulo 2 se presenta una revisión científico-técnica de los fallos presentes en las máquinas eléctricas rotativas, así como de las principales técnicas de análisis de señal aplicadas al diagnóstico de fallos. Por último, se revisan los principales sistemas de inteligencia artificial empleadas.
- En el capítulo 3 se presenta una descripción de los sistemas de clasificación a emplear en el transcurso de la tesis y sus características.
- En el capítulo 4 se describe el banco de ensayos empleado, se describen los ensayos, así como la gestión de dichas señales.
- En el capítulo 5 se describen las características relevantes en régimen estacionario. También el desarrollo del sistema para aplicarse a este régimen de trabajo y se presentan los resultados obtenidos.
- En el capítulo 6 se describen las características relevantes en régimen transitorio. También el desarrollo del sistema para aplicarse a este régimen de trabajo y se presentan los resultados obtenidos.
- Finalmente, en el capítulo 7 se presentan las conclusiones y principales aportaciones de esta tesis al campo de la diagnosis de máquinas eléctricas. También se presentan las futuras líneas de investigación que se pretenden abordar en un futuro cercano.

Resum

El manteniment de les màquines elèctriques ha estat fonamental per garantir la continuïtat dels processos industrials. Un part important de les investigacions desenvolupades els darrers anys en el camp de la detecció d'averies en màquines elèctriques rotatives s'han centrat en la cerca de característiques discriminatòries entre les magnituds mesurables (mecàniques i elèctriques) dels motors i generadors instal·lats tant en entorns industrials tradicionals com ara bé en sectors emergents com la producció de energia elèctrica a partir de la energia eòlica, el desenvolupament de la tracció elèctrica, etc. Els objectius d'aquestes tècniques de diagnòs són determinar l'estat del motor, identificar el tipus de fallo que s'ha produït i quantificar la severitat del mateix amb la finalitat de desenvolupar accions preventives, correctives i els plans de contingència per tal de minimitzar l'impacte econòmic de l'averia. Una de les opcions més utilitzada per detectar les anomalies degudes a les falles dels propis motors ha estat, conjuntament amb l'anàlisi de les vibracions de la màquina, el anàlisi de les corrents estatòriques del motor. L'ús de la corrent ha estat propiciat a que es una magnitud, el valor instantani de la mateixa i la seua evolució temporal, que està sent mesurada en la instal·lació industrial o, tanmateix, en els equips de control del motor. De no ser així, és una magnitud que pot ser fàcilment mesurada de forma no invasiva utilitzant pinces amperimètriques.

En la literatura científica-tècnica s'hi troben diferents tècniques de anàlisi i processament de la senyal que permeten discriminar l'estat de la màquina elèctrica, és a dir, si es troba en esta sa o en estat d'averia. En termes generals, cadascuna d'aquestes tècniques o mètodes, és apropiat per a l'anàlisi d'una magnitud concreta, un règim de treball específic, per a detectar uns tipus de falles determinades, etc. I cadascuna d'aquestes tècniques presenta els resultats d'una manera específica i amb diferències amb no sols

els tipus de tècnica sinó també dependent del tipus de fallada, de les condicions de funcionament de la màquina, etc. Per tant, aquests mètodes, presenten una clara limitació intrínseca ja que es necessita de personal altament qualificat i amb capacitat d'interpretar els resultats en funció de la tècnica utilitzada, tipus de fallada, règim de treballa, per discernir el estat en el que s'hi troba la màquina.

Per tot això, resulta necessari la tasca d'automatitzar aquesta labor mitjançant l'ús i aplicació de tècniques de aprenentatge e intel·ligència artificial basats en classificadors. Per tant, la investigació encaminada al desenvolupament de nous sistemes de detecció que maximitzen la precisió, tenint en compte les característiques disponibles, és prioritària. Per millorar aquestes característiques de classificació, cada vegada es proposen sistemes més complexos basats en sistemes experts amb capacitat d'autoaprenentatge per identificar falles. Un exemple d'aquests són els sistemes de classificació multidimensionals o sistemes experts basats en la simulació d'un "cervell" o xarxa neuronal.

Tenint en compte que la meua formació abasta diversos camps que van, des de l'automàtica industrial fins els sistemes experts e intel·ligència artificial, he procurat integrar els meus coneixements en aquests camps per a desenvolupar, en aquesta tesi doctoral, un "suprasistema" automàtic capaç de generar sistemes òptims de diagnosi de falles. Els sistemes són adaptables a la tipologia de la màquina, el que permetrà que, mitjançant aquest "suprasistema", es puguin desenvolupar automàticament nous sistemes òptims per a la diagnosi de altres tipus de màquines. Els sistemes de diagnosis generats són òptims. D'una banda maximitzen la precisió en la detecció de la fallada i, d'altra banda, minimitzen els costos computacionals. A més els sistemes experts de classificació s'han adaptat, tenint en compte les característiques discriminatòries que s'obtenen en funció dels mètodes i transformacions utilitzades, amb l'objectiu de obtenir el millor sistema de detecció per a cada tipus de falta.

El "suprasistema" proposat s'encarrega de generar el sistema de diagnosi òptim per a la detecció de fallades en motors asíncrons de inducció. Sistema que estarà compost per una tècnica de diagnosi i per un sistema classificador expert. Per al seu desenvolupament s'han seguit les següents etapes:

- **Adquisició de dades.**

S'ha utilitzat un banc de proves format per dos motors acoblats mecànicament pels seus eixos. El motor a estudiar es un motor asíncron de inducció que pot estar en esta sa o en estat de fallida. En canvi, l'altre motor acoblat es un motor síncron de imants permanents que treballa oferint el par resistent que ha de vèncer el motor analitzat. Durant els assajos s'han enregistrat les tres corrents de fase del motor així com la velocitat per a les següents condicions de funcionament:

- Velocitat i par resistent constants

- Velocitat variable i par resistent constant.
- Velocitat constant i par resistent variable.
- Velocitat i par resistent variables.

Tots els assajos han estat repetits amb la finalitat de comprovar el correcte funcionament. A més, s'han realitzat diferents assajos de cadascun dels quatre blocs comentats anteriorment, amb diferents velocitats i amb distints nivells de par resistent. S'han realitzat un total de 735 assajos als quals se'ls aplicaran les tècniques que seran analitzades amb el suprasistema.

▪ **Aplicació de les tècniques i/o mètodes de diagnosi.**

S'han utilitzat diferents tècniques de diagnosi per obtenir les característiques rellevants des del punt de vista de la diagnosi de la màquina. En aquesta tesi s'aporten dos tècniques novadores, l'ús de la finestra Prolate i l'aplicació de la transformada de Fourier curta en freqüència (SFFT). Així doncs, les senyals utilitzades per a la diagnosi han estat:

- La corrent de fase del estator.
- La corrent de fase del estator multiplicada per una finestra de tipus Hanning (Han).

Les tècniques utilitzades per al anàlisi en règim estacionari són:

- Anàlisi de la senyal.
- Anàlisi de la transformada de Hilbert de la senyal (mòdul de la senyal analítica).
- Anàlisi de la transformada Cepstrum de la senyal.
- Anàlisi de la transformada del mòdul de Park de la senyal.
- Anàlisi de l'ordre d'harmònic (HOTA) de la senyal.

La senyal s'ha convertit al domini de la freqüència mitjançant:

- Transformada de Fourier de la senyal.
- Mètode de Welch de la senyal.

Concloent, s'han explorat fins 20 mètodes distints, resultat de la combinació de les possibilitats descrites, per a la diagnosi en règim estacionari.

Les tècniques aplicades al règim transitori han estat:

- Anàlisi de l'ordre d'harmònic (HOTA) de la senyal mitjançant l'ús de la transformada de Fourier curta en temps (STFT) i la finestra Gaussiana.
- Anàlisi de l'ordre d'harmònic (HOTA) de la senyal mitjançant l'ús de la transformada de Fourier curta en temps (STFT) i la finestra Prolate. *(L'ús d'aquesta finestra en la diagnosi és una aportació d'aquesta tesi).*
- Anàlisi de l'ordre d'harmònic (HOTA) de la senyal mitjançant l'ús de la transformada de Fourier curta en freqüència (SFFT) i la finestra Gaussiana. *(L'ús d'aquesta transformada és una aportació de la tesi).*
- Anàlisi de l'ordre d'harmònic (HOTA) de la senyal mitjançant l'ús de la transformada de Fourier curta en freqüència (SFFT) i la finestra Prolate.

A totes aquestes variants de HOTA se'ls ha aplicat, a més, una sèrie de millores amb respecte a l'algoritme original presentat en la literatura tècnica. Aquestes millores suposen una reducció en el temps de càlcul i en l'espai de memòria necessari. Aquestes millores son essencials per al disseny d'un sistema expert. *Les millores proposades en aquesta tesi són:*

- *Inclusió d'una etapa de filtrat pas banda per eliminar les altes freqüències.*
- *Minimització del efectes de costat.*
- *Algoritme més ràpid per a l'ajust òptim de la finestra de filtrat.*

En conclusió, per al règim transitori s'han analitzat quatre mètodes de diagnosi diferents.

▪ **Obtenció de les característiques rellevants.**

Per cada una de les tècniques utilitzades s'obté una dupla de característiques rellevants per mostra o senyal mesurada del motor. Així doncs s'obté una matriu de vectors de característiques rellevants de dimensions igual al nombre de mètodes de diagnosi utilitzats (20 estacionari més 4 en transitori), on cada vector té la dimensió del nombre d'assajos que s'han realitzat (més de 700).

▪ **Obtenció dels classificadors.**

Es desenvolupen distints tipus de classificadors per cada un dels vectors de característiques obtinguts amb cadascuna de les tècniques de diagnosi utilitzades. Aquests classificadors estan basats en:

- Xarxes neuronals artificials d'una capa oculta.
- Xarxes neuronals artificials de dues capes ocultes.
- Màquina de vectors de suport que empren un kernel de funció polinomial.
- Màquina de vectors de suport que empren un kernel de funció de base radial gaussiana.

Dins de cada tipus de classificador es generen distints classificadors en funció de l'algorisme d'optimització utilitzat (sistema de validació empleat), generant multitud de classificadors de diferents coeficients fins obtenir l'òptim.

▪ **Selecció del sistema expert òptim**

A l'última etapa, s'analitzen els resultats obtinguts i es realitza la selecció del sistema de diagnòstic òptim compostos de un classificador i una tècnica de diagnosi. El suprasistema obtingut aporta els següents avantatges:

1. Obtenir un sistema de diagnosi òptim final.
2. Obté diversos sistemes de diagnosi localment òptims (per a la mateixa tècnica i sistema expert) amb tots els temps obtinguts del procés d'entrament i el temps de diagnosi estimats. Així, el suprasistema aporta, no sols una solució final òptima, sinó que a més permet al operador una de les solucions alternatives generades que son localment òptimes en el cas que fóra necessari ajustar-se a alguna limitació en el temps de càlcul.
3. El suprasistema es totalment autònom. L'operador deu aportar sols les mostres del motor especificat i el suprasistema genera el sistema òptim de diagnosi sense necessitat de que l'operari intervengui.
4. El suprasistema es adaptable al tipus de motor de inducció, d'aquesta forma pot generar un sistema de diagnosi específic i òptim per a cada tipus de motor.

La tesis s'ha estructurat con s'indica a continuació:

- El capítol 1 inclou una breu introducció que serveix com a punt de sortida i es presenten els principals objectius d'aquesta tesi.
- En el capítol 2 es presenta una revisió científico-tècnica dels tipus de fallades que apareixen en les màquines elèctriques rotatives, així com de les principals tècniques d'anàlisi de senyal aplicades al camp de la diagnosi. S'inclou, a més, una revisió dels principals sistemes de intel·ligència artificial utilitzades.
- El capítol 3 es centra en una descripció dels sistemes de classificació a utilitzar en el desenvolupament de la tesis i les seues característiques.
- En el capítol 4 es descriu la bancada de proves utilitzada així com una descripció dels assajos realitzats i de la gestió de les senyals obtingudes.
- En el capítol 5 es presenten les característiques rellevants en règim estacionari. També es presenta el desenvolupament del sistema per a utilitzar-se en aquest règim de treball així com els resultats obtinguts.
- En el capítol 6 es presenten les característiques rellevants en règim transitori. També es presenta el desenvolupament del sistema per a utilitzar-se en aquest règim de treball així com els resultats obtinguts.
- Finalment, en el capítol 7 es presenten les conclusions i principals aportacions d'aquesta tesi en el camp de la diagnosi de màquines elèctriques. També s'inclouen les futures línies d'investigació que es pretenen abordar en el futur.

Summary

Electrical machines maintenance is crucial to guarantee the continuity of industrial process. In recent years, extensive research on the diagnosis of rotating electrical machines has been focused in obtaining discriminatory features from the measurable quantities (mechanical and electrical) of the motors and generators installed in the traditional industry, and also in emerging fields such as wind energy, electric traction, etc. The main goal of the fault diagnosis developed techniques is to determine the condition of the motor by identifying the fault and quantifying its severity in order to develop preventive and corrective maintenance to reduce the economic impact of the damage. Recently, and in order to detect the relevant features produced by failures in the electrical machine, one of the most common quantity being analysed, along with the electrical machine vibration, has been the stator phase current. In addition, the instantaneous value of the stator phase current and its time evolution are often available on the electrical installations cabinets, or even in electronically control equipment and, otherwise, can be measured easily it by means of a non-invasive clamp meter.

In the scientific-technical literature, one can find different signal processing techniques that allow discriminating the faulty or healthy condition of the electrical machine. Usually, each of these techniques or methods is used for one specific magnitude, or for particular working conditions, or for one kind of failure, etc... Therefore, they often have inherent limitations has associated to their different characteristic features and require an expert to recognize the condition of the motor considering the technique used and the fault type.

It is therefore essential to automate this expert task by using artificial intelligence and learning-machine techniques based on automatic classifiers. It is also a priority to research and to develop new detection systems more accurate than the previous ones. Complex systems based on expert systems capable to learn identifying failures (systems such as multidimensional classification systems or expert systems based on simulating a "brain" using neural networks) are used to improve the classification of faults.

My background includes the fields from the industrial automation to artificial intelligence and experts systems and is used to develop an automatic "supra-system" capable to generate optimal fault diagnosis systems. These optimal systems are adaptable to the type of electrical machine applied. This "supra-system" can generate new optimal fault diagnosis systems automatically when applied to new types of electrical machines. Generated optimal diagnostic systems maximize the accuracy in failure detection and minimize the computation costs as a secondary objective. In addition, from the relevant features obtained from the detection failures, expert classification systems have been adapted to the relevant features in order to obtain the optimal diagnostic system for each type of fault.

The proposed "supra-system" is able to generate the optimal diagnostic system for the fault detection induction motors. This optimal diagnostic system is composed by a diagnostic technique and a classifier expert system. This has been achieved following these steps:

- **Data Acquisition.**

An experimental test-bed is used with two motors mechanically coupled directly. One of them, the test case, is an asynchronous induction machine (healthy or faulty) and the other is a synchronous permanent magnet motor, which acts as a load torque on the system. Different tests have been performed, and the three phase stator currents have been measured along with the motor, at different conditions:

- constant load torque and speed of the motor.
- constant load torque and variable speed motor.
- variable load torque and constant speed motor.
- variable load torque and speed motor.

All tests were repeated in order to check the proper condition of each test. Besides, different tests from the four categories above, with different motor speeds and load torque levels have been performed. In total, 735 data sets were obtained and fed to the following diagnosis techniques to be analysed with the "supra-system".

▪ **Diagnosis techniques/methods applied.**

Different diagnosis techniques/methods have been used depending on the test conditions in order to extract relevant diagnosis features. Moreover, two novel techniques have been developed and included within this thesis frame: the use of the Prolate window and the use of the short frequency Fourier transform.

The signals used for diagnosis are:

- The stator phase current.
- The stator phase current multiplied by a Hanning window (Hann).

The techniques used for the steady-state regime are:

- Signal analysis.
- Analysis of the analytical signal modulus (Hilbert transform).
- Analysis of Cepstrum transform.
- Analysis of Park transform modulus.
- Harmonic order tracking analysis (HOTA).

The signal is analysed in the frequency domain by:

- Analysis of the Fourier transform.
- Analysis of Welch method.

In conclusion, for the steady-state regime 20 different diagnostic methods have been explored, resulting from the combination of the possibilities described above.

The techniques used for the transient regime are:

- Harmonic order tracking analysis (HOTA) using the Short-Time Fourier Transform (STFT) with Gaussian window.
- Harmonic order tracking analysis (HOTA) using the Short-Time Fourier Transform (STFT) with Prolate window (***This window is a contribution of this thesis.***).
- Harmonic order tracking analysis (HOTA) using the Short-Frequency Fourier Transform (SFFT) with Gaussian window (***This technique is a contribution of this thesis, and it is modified from the STFT.***).
- Harmonic order tracking analysis (HOTA) using the Short-Frequency Fourier Transform (SFFT) with Prolate window.

Besides, a set of improvements over the original HOTA algorithm recently presented in the technical literature have been implemented to the different HOTA variants above.

These improvements reduce the computation time and the storage memory requirements, which is essential in the design of an expert system. The improvements proposed in this thesis are:

- *A new band-pass filtering stage to attenuate high frequencies.*
- *The minimization of the edge effect.*
- *A new faster algorithm for the optimum setting of the filter window.*

In conclusion, for the transient regime 4 different diagnostic methods are explored.

▪ **Extraction of relevant features.**

For each of the techniques used, the relevant features are obtained, in this case a couple for each motor test or measured current. Thus, a matrix is obtained with as many feature vectors column as the number of diagnostic techniques used (20 in steady-state and 4 in transient) and as many rows as the number of tests used in this thesis (over 700 tests).

▪ **Obtaining the classifiers.**

For each of the feature vectors obtained with each of the diagnostic techniques used, different classifiers are generated based on the following types:

- Artificial neural networks with one hidden layer.
- Artificial neural networks with two hidden layers.
- Support Vector Machine with polynomial kernel function.
- Support Vector Machine with Gaussian kernel function.

The optimization algorithm (validation system) is used to generate different classifiers of each type of classifier; a multitude of classifiers is generated with different coefficients to obtain the optimal classifier.

▪ **Selecting the optimal expert system.**

In the final step, the results are analysed and the selection of the optimal system composed by the best classifier and the best diagnosis technique. The “supra-system” obtained has the following advantages:

1. It achieves a final optimal diagnostic system.

2. It produces several local optimal diagnostic systems (for the same technique and expert system) taking into account the training time and the estimated diagnosis time. Thus, the “supra-system” not only provides a final optimal solution, but also allows the operator to choose one of the generated alternative local optimal solutions in case calculation time restrictions must be taken into account
3. It is completely autonomous. The operator must provide the signals from a specific motor type and the “supra-system” generates the optimal diagnostic system without operator intervention.
4. It is adaptable to the induction motor type, as it can generate an optimal diagnostic system specific for each type of motor.

The thesis has been structured into the following chapters:

- Chapter 1 is a brief introduction that it used as starting point and the main objectives of this thesis are presented.
- Chapter 2 is a scientific-technical review on the faults in rotating electrical machines, as well as on the main signal processing techniques applied to fault diagnosis. Finally, the main artificial intelligence systems are reviewed.
- Chapter 3 shows a description of the classification systems to be used in this thesis and its features.
- In chapter 4 the test bed and the tests are described as well as the management of the acquired signals.
- Chapter 5 describes the relevant features in the steady-state regime of the electrical machine as well as the development of the system to be applied to this work regime and the results obtained.
- Chapter 6 describes the relevant features in the transient regime of the electrical machine as well as the development of the system to be applied to this work regime and the results obtained.
- Finally, in Chapter 7 the main conclusions and contributions of this thesis in the field of diagnosis electrical machines are presented. Also the future research topics that should be addressed in the near future are shown.

Índice

Índice general.

ÍNDICE.....	I
ÍNDICE GENERAL.....	I
ÍNDICE DE FIGURAS.....	VI
ÍNDICE DE TABLAS.....	XIV
ABREVIATURAS.....	XVII
CAPITULO 1. INTRODUCCIÓN.....	1
1.1 ANTECEDENTES Y ESTADO ACTUAL.....	1
1.2 DESARROLLO DEL ESTUDIO.....	3
1.3 OBJETIVOS.....	5
CAPITULO 2. ESTADO DEL ARTE.....	7
2.1 INTRODUCCIÓN.....	7
2.2 FALLOS DE UN MOTOR DE INDUCCIÓN.....	8
2.2.1 Fallos de tipo eléctrico.....	11
2.2.1.1 Fallos eléctricos en el estator del motor.....	11
2.2.1.2 Fallos eléctricos en el rotor del motor.....	12
2.2.1.3 Fallos eléctricos de tipo externo.....	13
2.2.2 Fallos de tipo mecánico.....	13

2.2.2.1	Fallo mecánico por desequilibrio.	13
2.2.2.2	Fallo mecánico en los cojinetes.	14
2.2.2.3	Fallo mecánico de excentricidad entre el rotor y el estator.	15
2.2.3	Frecuencias características producidas por los fallos.	16
2.2.3.1	Fallos eléctricos.	19
2.2.3.2	Fallos mecánicos.	20
2.3	TÉCNICAS DE DETECCIÓN DE FALLOS POR MEDIO DE LA CORRIENTE DE FASE.	25
2.3.1	Técnicas de detección de fallos en régimen estacionario.	26
2.3.1.1	Análisis de la corriente estatórica mediante transformada FFT.	26
2.3.1.2	Vector de Park (PVA).	29
2.3.1.3	Vector de Park Extendido (EPVA).	34
2.3.1.4	Cepstrum.	36
2.3.1.5	Componente alterna del módulo de la señal analítica.	39
2.3.1.6	Análisis del Orden Armónico para régimen estacionario (HOTA estacionario)...	44
2.3.1.7	Otras técnicas de detección de fallos en régimen estacionario.	52
2.3.2	Técnicas de detección de fallos para un régimen de trabajo transitorio.	52
2.3.2.1	Transformada de Fourier corta en el tiempo (STFT).	52
2.3.2.1.1	Búsqueda del espacio de mínima entropía.	58
2.3.2.2	Transformada Wavelet Discreta.	60
2.3.2.3	Transformada Continua Wavelet y Transformada Continua Compleja Wavelet.	65
2.3.2.4	Transformada Hilbert Huang.	67
2.3.2.5	Distribución de Wigner-Ville.	69
2.3.2.6	Análisis del Orden Armónico para régimen transitorio (HOTA transitorio).	72
2.3.2.7	Otras técnicas de detección de fallos en régimen transitorio.	77
2.3.3	Filtros de la señal.	78
2.3.3.1	Filtro de ventana de Hann.	78
2.3.3.2	Periodograma de Welch.	81
2.4	SISTEMAS EXPERTOS EN LA DETECCIÓN Y CLASIFICACIÓN DE PATRONES.	83
2.4.1	Sistemas expertos basados en deducción lógica.	84
2.4.1.1	Lenguajes de programación lógica.	85
2.4.1.2	Arboles de decisión.	86
2.4.2	Sistemas expertos basados en espacios de características.	87
2.4.2.1	K-vecinos más cercanos (kNN).	88
2.4.2.2	Máquina de vectores de soporte (SVM).	89
2.4.3	Sistemas expertos basados en emulación biológica (Redes Neuronales Artificiales).	92
CAPITULO 3. SISTEMAS EXPERTOS ORIENTADOS A LA CLASIFICACIÓN		95
3.1	INTRODUCCIÓN.	95
3.2	MÁQUINA DE VECTORES DE SOPORTE (SVM).	96
3.2.1	Método de clasificación del clasificador SVM.	100
3.2.2	Entrenamiento de un clasificador SVM.	102
3.2.3	Kernels del clasificador SVM.	107
3.2.4	Coefficientes de ajuste del clasificador SVM.	112

3.2.5	<i>Ventajas y desventajas del clasificador SVM.</i>	115
3.3	REDES NEURONALES ARTIFICIALES.	116
3.3.1	<i>Unidad básica de una red neuronal.</i>	118
3.3.2	<i>Topología de red neuronal supervisada.</i>	122
3.3.2.1	Datos sin relación secuencial.	122
3.3.2.2	Datos relacionados secuencialmente.	125
3.3.3	<i>Entrenamiento de la red.</i>	126
3.3.3.1	Algoritmos de entrenamiento.	127
3.3.3.1.1	Algoritmos de retropropagación del gradiente descendiente.	127
3.3.3.1.2	Algoritmos de retropropagación del gradiente conjugado.	128
3.3.3.1.3	Algoritmos casi Newtonianos.	129
3.3.3.2	Funciones de coste.	130
3.3.4	<i>Mejora en la generalización del aprendizaje.</i>	135
3.3.4.1	Estructura de la red ajustada al problema.	135
3.3.4.2	Entrenamiento con detención temprana.	138
3.3.4.3	Reentrenamiento de la red neuronal artificial.	139
3.3.5	<i>Ventajas y desventajas de las redes neuronales artificiales.</i>	140
3.4	CONCLUSIONES.	141
CAPITULO 4. BANCO DE ENSAYOS DE LOS MOTORES		143
4.1	INTRODUCCIÓN	143
4.2	EL BANCO DE ENSAYOS.	143
4.3	ELEMENTOS QUE FORMAN EL BANCO DE ENSAYOS.	146
4.3.1	<i>Sistema de potencia.</i>	147
4.3.1.1	Motores de inducción.	147
4.3.1.2	Auto-transformador.	150
4.3.1.3	Variadores de frecuencia.	151
4.3.1.4	Variador de frecuencia marca ABB modelo ACSM1 (Driver del servomotor).	158
4.3.1.5	Motor de imanes permanentes de tipo síncrono.	160
4.3.2	<i>Sistema de control automatizado.</i>	161
4.3.2.1	El autómatas programable.	162
4.3.2.2	Ordenador personal.	164
4.3.3	<i>Subsistema de medida y adquisición de datos.</i>	165
4.3.3.1	El Osciloscopio.	166
4.3.3.2	Sondas de Tensión.	169
4.3.3.3	Pinzas Amperimétricas.	169
4.4	BASE DE DATOS DE MUESTRAS DE ENSAYOS.	170
4.4.1	<i>Ensayos para la generación de sistemas de diagnóstico estacionario.</i>	171
4.4.2	<i>Ensayos para la generación de sistemas de diagnóstico transitorio.</i>	174
4.5	CONCLUSIONES.	183
CAPITULO 5. GENERADOR DE DETECTORES DE FALLO EN RÉGIMEN ESTACIONARIO		185
5.1	INTRODUCCIÓN	185
5.2	CARACTERÍSTICAS RELEVANTES EN RÉGIMEN ESTACIONARIO.	186

5.2.1	Técnica de Análisis de la corriente estatórica mediante FFT.	187
5.2.2	Componente alterna del módulo de la señal analítica	190
5.2.3	Análisis del Orden Armónico. (HOTA estacionario).	194
5.2.4	Cepstrum.	197
5.2.5	Vector de Park Extendido (EPVA).	199
5.3	MODIFICACIONES DE LAS TÉCNICAS DE DIAGNÓSTICO DE FALLOS.	202
5.4	GESTIÓN DE FALLOS MEDIANTE MÁQUINA DE VECTORES DE SOPORTE.	203
5.4.1	Estudio de los coeficientes del sistema.	204
5.4.2	Algoritmo de generación del sistema de diagnóstico.	205
5.4.2.1	Algoritmo de ajuste de coeficientes por búsqueda en malla.	206
5.4.2.2	Entrenamiento y validación del clasificador SVM.	212
5.4.2.3	Sobreentrenamiento por falta de significancia de fallo en las muestras.	216
5.5	GESTIÓN DE FALLOS MEDIANTE REDES NEURONALES ARTIFICIALES.	217
5.5.1	Estudio de los parámetros del sistema.	218
5.5.1.1	Parámetros topológicos de la red.	219
5.5.1.2	Función de entrenamiento.	223
5.5.1.2.1	Algoritmo de entrenamiento por retropropagación del gradiente escalado.	225
5.5.2	Algoritmo de generación del sistema de diagnóstico.	227
5.5.2.1	Algoritmo de búsqueda de la red neuronal óptima.	229
5.5.2.1.1	Búsqueda de la red neuronal óptima de una capa oculta.	230
5.5.2.1.2	Búsqueda de la red neuronal óptima con dos capas ocultas.	233
5.5.2.2	Entrenamiento y validación de la red neuronal para una red estructural.	235
5.5.2.2.1	Validación cruzada de entrenamiento / validación / prueba.	235
5.5.2.2.2	Algoritmo de particionado de entrenamiento, prueba y validación.	239
5.6	ENTORNO DE DESARROLLO DEL SISTEMA GENERADOR DE DIAGNÓSTICO DE FALLOS.	242
5.7	RESULTADOS OBTENIDOS DEL SISTEMA AUTÓNOMO PARA RÉGIMEN ESTACIONARIO.	244
5.7.1	Resultados obtenidos con sistemas expertos SVM.	250
5.7.1.1	Comparación de tiempos entre kernel polinomial y kernel RBF.	250
5.7.1.2	Comparación de resultados entre kernel polinomial y kernel RBF.	255
5.7.2	Resultados obtenidos con sistemas expertos ANN.	261
5.7.2.1	Comparación de tiempos entre ANN de 1 y de 2 capas ocultas.	261
5.7.2.2	Comparación de resultados entre ANN de 1 y 2 capas ocultas.	265
5.8	CONCLUSIONES DEL CAPÍTULO	270
CAPITULO 6. GENERADOR DE DETECTORES DE FALLO EN RÉGIMEN TRANSITORIO.		273
6.1	INTRODUCCIÓN.	273
6.2	OBTENCIÓN DE CARACTERÍSTICAS RELEVANTES EN RÉGIMEN TRANSITORIO.	276
6.2.1	Análisis del Orden Armónico. (HOTA transitorio).	276
6.2.1.1	Etapa de filtrado paso banda.	283
6.2.1.2	Implementación de la función Prolate como nueva ventana de filtro (PSWF).	284
6.2.1.3	Limitación del recorrido de la ventana para minimizar el efecto de borde.	289
6.2.1.4	Método de ajuste de la proporción de la función de ventana.	292
6.2.1.5	Transformada de Fourier corta en frecuencia (SFFT) con bandas adaptativas.	297
6.2.2	Resumen de variantes optimizadas de HOTA transitorio.	301

6.3	GESTIÓN DE FALLOS MEDIANTE SISTEMAS EXPERTOS.....	302
6.4	VALIDACIÓN PRÁCTICA.	303
6.5	RESULTADOS OBTENIDOS DEL SISTEMA AUTÓNOMO PARA RÉGIMEN TRANSITORIO.	303
6.5.1	<i>Resultados obtenidos con sistemas expertos SVM.....</i>	<i>306</i>
6.5.1.1	Comparación de tiempos entre kernel polinomial y kernel RBF.....	306
6.5.1.2	Comparación de resultados entre kernel polinomial y kernel RBF.	310
6.5.2	<i>Resultados obtenidos con sistemas expertos ANN.....</i>	<i>314</i>
6.5.2.1	Comparación de tiempos entre ANN de 1 y de 2 capas ocultas.	314
6.5.2.2	Comparación de resultados entre ANN de 1 y 2 capas ocultas.....	317
6.6	CONCLUSIONES DEL CAPÍTULO.....	320
CAPITULO 7. CONCLUSIONES, APORTACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN		323
7.1	CONCLUSIONES Y APORTACIONES.....	323
7.2	FUTURAS LÍNEAS DE INVESTIGACIÓN.	327
CAPITULO 8. REFERENCIAS.....		329
CAPITULO 9. ANEXOS.....		343
Anexo 1.	Clúster de redes neuronales.	343
Anexo 2.	Algoritmo de entrenamiento por retropropagación del gradiente.....	346
Anexo 3.	Aceleración computacional mediante hardware CUDA.....	350

Índice de Figuras.

FIGURA 2.1 CLASIFICACIÓN DE LOS FALLOS MÁS IMPORTANTES QUE PUEDEN OCURRIR EN UN MOTOR DE INDUCCIÓN DE JAULA DE ARDILLA.	10
FIGURA 2.2 IMAGEN DEL FRONTAL DEL ESTATOR DE UN MOTOR DE INDUCCIÓN.	11
FIGURA 2.3 IMAGEN DE PERFIL DEL ROTOR DE UN MOTOR DE INDUCCIÓN.	12
FIGURA 2.4 EN ESTA IMAGEN SE PRESENTAN UN EJEMPLO DE DISTRIBUCIONES DE MASA EN EL ROTOR DE UN MOTOR, TENIENDO EN CUENTA QUE CADA VALOR DE MASA ESTÁ REPRESENTADO POR UN COLOR DISTINTO. CADA FLECHA ROJA SE CORRESPONDE CON EL VECTOR DE FUERZA. A) DISTRIBUCIÓN HOMOGÉNEA RESPECTO DEL VECTOR DE ROTACIÓN DEL ROTOR. B) CON DISTRIBUCIÓN DE MASA NO HOMOGÉNEA RESPECTO DEL VECTOR DE ROTACIÓN DEL ROTOR. C) DISTRIBUCIÓN HOMOGÉNEA A LO LARGO DEL EJE DE ROTACIÓN. C) DISTRIBUCIÓN NO HOMOGÉNEA CONFORME A LA LONGITUDINAL DEL EJE DE ROTACIÓN.....	14
FIGURA 2.5 REPRESENTACIÓN DE LOS ELEMENTOS DE UN COJINETE DE BOLAS.	15
FIGURA 2.6 REPRESENTACIÓN ESQUEMÁTICA EN TRES CUADRANTES DE LA TRAYECTORIA DE GIRO DE UN MOTOR CON FALLO DE EXCENTRICIDAD ESTÁTICA (A) Y CON FALLO DE EXCENTRICIDAD DINÁMICA (B). TAL COMO PUEDE OBSERVARSE, EN EL CASO DE EXCENTRICIDAD ESTÁTICA LA POSICIÓN DE MÍNIMO ENTREHIERRO SE MANTIENE EN POSICIÓN FIJA, MIENTRAS QUE EN EL CASO DE EXCENTRICIDAD DINÁMICA ÉSTA VA CAMBIANDO CON LA ROTACIÓN DEL EJE.	17
FIGURA 2.7 SECCIÓN DEL ESPECTRO DONDE SE MUESTRA LA FRECUENCIA CARACTERÍSTICA DE FALLO DE COJINETE DE UNA CORRIENTE DE FASE DE UN MOTOR: (A) SANO, (B) CON EL COJINETE DAÑADO.	18
FIGURA 2.8 REPRESENTACIÓN DE UN COJINETE DONDE SE INDICA EL ÁNGULO DE CONTACTO β , EL DIÁMETRO DE LA BOLA (BD) Y EL DIÁMETRO DE PASO (PD).	21
FIGURA 2.9 FFT DE UN MOTOR SANO TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.....	27
FIGURA 2.10 FFT DE UN MOTOR CON BARRA ROTA TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.	27
FIGURA 2.11 EJEMPLO DE APARICIÓN DE LA FRECUENCIA DE FALLO CARACTERÍSTICA A DISTINTAS DISTANCIAS RESPECTO DEL VALOR DE K	28
FIGURA 2.12 REPRESENTACIÓN DEL VECTOR DE PARK EN UNA SITUACIÓN IDEAL.	30
FIGURA 2.13 VECTOR DE LAS CORRIENTES DE UN MOTOR IDEAL (MOTOR SIMULADO): (A) SANO. (B) CON 10 BARRAS ROTAS. (C) CON UN 40% DE EXCENTRICIDAD ESTÁTICA Y 40% DE EXCENTRICIDAD DINÁMICA. (D) CON 6 ESPIRAS EN CORTO.	31
FIGURA 2.14 VECTOR DE GIRO DE LA COMPONENTE CONTINUA (EN AZUL) Y DE LA COMPONENTE ALTERNA (EN AMARILLO) DEL VECTOR DE PARK.....	32
FIGURA 2.15 VECTOR DE PARK DE LAS CORRIENTES DE FASE DE UN MOTOR SANO CON UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ Y CON CONTROL ELECTRÓNICO.	33
FIGURA 2.16 VECTOR DE PARK DE LAS CORRIENTES DE FASE DE UN MOTOR CON BARRAS ROTAS EN EL ROTOR, CON UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ Y CON CONTROL ELECTRÓNICO.....	33
FIGURA 2.17 VECTOR DE PARK EXTENDIDO DE LAS FASES DE CORRIENTE DE UN MOTOR SANO TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.....	35

FIGURA 2.18 VECTOR DE PARK EXTENDIDO DE LAS FASES DE CORRIENTE DE UN MOTOR CON FALLO DE BARRA ROTA TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.	35
FIGURA 2.19 CEPSTRUM DE UNA DE LAS FASES DE CORRIENTE DE UN MOTOR SANO TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.	38
FIGURA 2.20 CEPSTRUM DE UNA DE LAS FASES DE CORRIENTE DE UN MOTOR CON FALLO DE BARRA ROTA TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.	38
FIGURA 2.21 SEÑAL ANALÍTICA OBTENIDA MEDIANTE LA TRANSFORMADA DE HILBERT DE UNA CORRIENTE DE FASE DE UN MOTOR SANO.....	39
FIGURA 2.22 ESPECTRO DE LA COMPONENTE ALTERNA DEL MÓDULO DE LA SEÑAL ANALÍTICA PARA UNA DE LAS FASES DE CORRIENTE DE UN MOTOR SANO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.	43
FIGURA 2.23 ESPECTRO DE LA COMPONENTE ALTERNA DEL MÓDULO DE LA SEÑAL ANALÍTICA PARA UNA DE LAS FASES DE CORRIENTE DE UN MOTOR CON FALLO DE BARRA ROTA A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.	43
FIGURA 2.24 DIAGRAMA DE FLUJO DEL ALGORITMO DE PROCESO REALIZADO POR LA TÉCNICA DE DIAGNÓSTICO HOTA PARA LA OBTENCIÓN DE CARACTERÍSTICAS REPRESENTATIVAS DE FALLO.....	45
FIGURA 2.25 EJEMPLO DE HOTA PARA EL DIAGNÓSTICO EN RÉGIMEN PERMANENTE. (A) FRECUENCIAS DE LA CORRIENTE DE FASE DEL ESTATOR. (B) FRECUENCIAS DE LA CORRIENTE DE FASE DEL ESTATOR TOMANDO COMO REFERENCIA EL ROTOR. (C) HOTA DE LA CORRIENTE DE FASE DEL ESTATOR. (D) HOTA REDUCIDO DE LA CORRIENTE DE FASE DEL ESTATOR.	46
FIGURA 2.26 HOTA DE UNA DE LAS FASES DE CORRIENTE DE UN MOTOR SANO TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.	50
FIGURA 2.27 HOTA DE UNA DE LAS FASES DE CORRIENTE DE UN MOTOR CON FALLO DE BARRA ROTA TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.	50
FIGURA 2.28 HOTA DE UNA DE LAS FASES DE CORRIENTE DE UN MOTOR CON FALLO DE BARRA ROTA TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 25HZ.	51
FIGURA 2.29 REPRESENTACIÓN PARAMÉTRICA DE LA FUNCIÓN DE VENTANA GAUSSIANA EN EL ESPACIO DE TIEMPO-FRECUENCIA.	53
FIGURA 2.30 EJEMPLO DE FUNCIÓN GAUSIANA REPRESENTADA EN (A) 2D Y (B) 3D.	55
FIGURA 2.31 DESPLAZAMIENTO DE LA FUNCIÓN DE VENTANA GAUSSIANA SOBRE LA SEÑAL DE CORRIENTE ORIGINAL OBTENIENDO ASÍ EL ESPACIO DE TIEMPO FRECUENCIA STFT.	56
FIGURA 2.32 ESPACIO TIEMPO FRECUENCIA PARA UNA CORRIENTE DE FASE DE UN MOTOR CON CONTROL ESCALAR CON GANANCIA DEL DESLIZAMIENTO OPERANDO EN RÉGIMEN TRANSITORIO (45-50Hz).	57
FIGURA 2.33 VECTOR DE ENTROPÍAS DE STFT RESPECTO DEL PARÁMETRO A DE RELACIÓN PROPORCIONAL ENTRE LA DIMENSIÓN DEL ESPACIO TIEMPO Y FRECUENCIA DEL FILTRO DE VENTANA GAUSSIANA.	59
FIGURA 2.34 EJEMPLO DE ESQUEMA DE DESCOMPOSICIÓN EN PROFUNDIDAD DE TRES NIVELES DE UNA SEÑAL MEDIANTE WAVELETS.	61
FIGURA 2.35 COMPARACIÓN DE WAVELETS DEL NIVEL 6, 7 Y 8 (8 NIVELES) PARA UNA SEÑAL DE ARRANQUE DE UN MOTOR DE INDUCCIÓN (A) SANO Y (B) CON DOS BARRAS ROTAS EN EL ROTOR.	63
FIGURA 2.36 REPRESENTACIÓN DE UN ESPACIO DE TIEMPO-FRECUENCIA EN DOS DIMENSIONES CORRESPONDIENTE A LA EVOLUCIÓN DE LAS ONDÍCULAS DE MAYOR NIVEL GENERADAS CON LA TRANSFORMADA DWT A PARTIR DE LA CORRIENTE DE ARRANQUE DE UNA DE LAS FASES DE UN MOTOR DE INDUCCIÓN CON UNA BARRA ROTA EN EL ROTOR.	65

FIGURA 2.37 REPRESENTACIÓN DE UN ESPACIO DE TIEMPO-FRECUENCIA EN DOS DIMENSIONES CORRESPONDIENTE A LA EVOLUCIÓN DE LAS ONDÍCULAS DE MAYOR NIVEL GENERADAS CON LA TRANSFORMADA CWT A PARTIR DE LA CORRIENTE DE ARRANQUE DE UNA DE LAS FASES DE UN MOTOR DE INDUCCIÓN CON UNA BARRA ROTA EN EL ROTOR.....	66
FIGURA 2.38 REPRESENTACIÓN DE UN ESPACIO DE TIEMPO-FRECUENCIA EN DOS DIMENSIONES CORRESPONDIENTE A LA EVOLUCIÓN DE LAS ONDÍCULAS DE MAYOR NIVEL GENERADAS CON LA TRANSFORMADA CCWT A PARTIR DE LA CORRIENTE DE ARRANQUE DE UNA DE LAS FASES DE UN MOTOR DE INDUCCIÓN CON UNA BARRA ROTA EN EL ROTOR.....	66
FIGURA 2.39 SEÑAL CON UNA ENVOLVENTE SUPERIOR E INFERIOR Y EL VALOR MEDIO DE LAS DOS ENVOLVENTES.....	68
FIGURA 2.40 REPRESENTACIÓN DE LOS ESPACIOS DE TIEMPO-FRECUENCIA DEL ESPECTRO HILBERT-HUANG DE (B) LA IMF1 Y (C) LA IMF2 OBTENIDAS DE LA SEÑAL ORIGINAL MOSTRADA EN (A) CON LA TRANSFORMADA HHT A PARTIR DE LA CORRIENTE DE ARRANQUE DE UNA DE LAS FASES DE UN MOTOR DE INDUCCIÓN CON FALLO DE BARRA ROTA.	69
FIGURA 2.41 REPRESENTACIÓN DE UN ESPACIO DE TIEMPO-FRECUENCIA CORRESPONDIENTE A LA DENSIDAD DE ENERGÍA OBTENIDA MEDIANTE LA DISTRIBUCIÓN DE WIGNER-VILLE A PARTIR DE LA CORRIENTE DE ARRANQUE DE UNA DE LAS FASES DE UN MOTOR DE INDUCCIÓN SANO.	71
FIGURA 2.42 REPRESENTACIÓN DE UN ESPACIO DE TIEMPO-FRECUENCIA CORRESPONDIENTE A LA DENSIDAD DE ENERGÍA OBTENIDA MEDIANTE LA DISTRIBUCIÓN DE WIGNER-VILLE A PARTIR DE LA CORRIENTE DE ARRANQUE DE UNA DE LAS FASES DE UN MOTOR DE INDUCCIÓN CON FALLO DE BARRA ROTA.....	71
FIGURA 2.43 ESQUEMA DE FLUJO DEL FUNCIONAMIENTO DE LA TÉCNICA HOTA PARA RÉGIMEN DE TRABAJO TRANSITORIO.....	72
FIGURA 2.44 EJEMPLO DE TRANSFORMACIÓN DE LA CORRIENTE DE FASE A HOTA EN RÉGIMEN TRANSITORIO PARA EL DIAGNÓSTICO DE FALLOS. (A) FRECUENCIAS DE LA CORRIENTE DE FASE DEL ESTATOR. (B) ESPACIO TIEMPO FRECUENCIA STFT (C) ESPACIO TIEMPO FRECUENCIA STFT TOMANDO EL SISTEMA DE COORDENADAS DE REFERENCIA DEL ROTOR. (D) HOTA TRANSITORIO DE LA CORRIENTE DE FASE. (E) HOTA TRANSITORIO SIMPLIFICADO. (F) HOTA TRANSITORIO COMPRIMIDO.	75
FIGURA 2.45 RESULTADO COMPRIMIDO DE HOTA TRANSITORIO OBTENIDO A PARTIR DE SU ESPACIO TIEMPO-FRECUENCIA DE K-ORDEN GENERADO.	76
FIGURA 2.46 RESULTADO SIMPLIFICADO DE HOTA TRANSITORIO OBTENIDO A PARTIR DEL RESULTADO COMPRIMIDO MOSTRADO EN LA FIGURA 2.45. EN ESTE CASO LOS VALORES EN NEGRO REPRESENTAN LOS VALORES OBTENIDOS PARA CADA COMPONENTE ARMÓNICA K MIENTRAS QUE LOS VALORES EN ROJO SON VALORES DE MEDIA DEL FONDO PARA CONOCER EL RUIDO DE FONDO APROXIMADO DE LA ONDA ORIGINAL.....	77
FIGURA 2.47 EJEMPLO GRÁFICO DE UN FILTRO DE COSENO ELEVADO (A) EN FUNCIÓN DEL TIEMPO Y (B) EN FUNCIÓN DE LA FRECUENCIA.	79
FIGURA 2.48 FFT DE UN MOTOR CON FALLO DE BARRA ROTA TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ, DONDE SE APRECIA UN PROBLEMA DE FUGA ESPECTRAL.....	80
FIGURA 2.49 FFT DE UN MOTOR CON FALLO DE BARRA ROTA TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ, DONDE SE HA MINIMIZADO LA FUGA ESPECTRAL MEDIANTE UNA VENTANA DE HANN.	80

FIGURA 2.50 REPRESENTACIÓN DEL PROCESO DE TRANSFORMACIÓN DE UNA CORRIENTE DE FASE UTILIZANDO UN PERIODOGRAMA DE WELCH.	81
FIGURA 2.51 PERIODOGRAMA DE WELCH DE UN MOTOR CON FALLO DE BARRA ROTA TRABAJANDO A UNA FRECUENCIA DE ALIMENTACIÓN DE 50HZ.	82
FIGURA 2.52 ESQUEMA DE UN SISTEMA EXPERTO COMPUESTO POR REGLAS.	85
FIGURA 2.53 TOPOLOGÍA DE UN ARBOL DE DECISION.	87
FIGURA 2.54 ESPACIO DE CARACTERÍSTICAS DE UN CLASIFICADOR k -NN DONDE SE MUESTRAN LOS LÍMITES DE DECISIÓN ENTRE CLASES PARA UN PROBLEMA DE RECONOCIMIENTO DEL HABLA.	88
FIGURA 2.55 ESPACIO DE CARACTERÍSTICAS DE UN CLASIFICADOR SVM DONDE SE OBSERVA EL HIPERPLANO DISCRIMINANTE SEPARANDO LAS MUESTRAS DE CADA CLASE.	90
FIGURA 2.56 ESPACIO DE CARACTERÍSTICAS DE UN CLASIFICADOR SVM DONDE SE OBSERVA EL HIPERPLANO DISCRIMINANTE Y UNA SERIE DE MUESTRAS A LAS QUE SE LES HA ASIGNADO UNA O OTRA CLASE DEPENDIENDO DE LA ZONA EN QUE SE SITUÁN.	91
FIGURA 2.57 ESQUEMA DE UNA RED NEURONAL ARTIFICIAL PERCEPTRON MULTICAPA.	93
FIGURA 3.1 ESPACIO DE CARACTERÍSTICAS DE UN CLASIFICADOR SVM DONDE SE OBSERVA LA FRONTERA DEL HIPERPLANO DISCRIMINANTE Y LAS DOS SUBZONAS DEFINIDAS PARA CADA UNA DE LAS CLASES.	101
FIGURA 3.2 ESPACIO DE CARACTERÍSTICAS DE UN CLASIFICADOR SVM DONDE SE OBSERVA EL HIPERPLANO DISCRIMINANTE Y UNA SERIE DE MUESTRAS A LAS QUE SE LES HA ASIGNADO UNA O OTRA CLASE DEPENDIENDO DE LA ZONA EN QUE SE SITUÁN.	102
FIGURA 3.3 ESPACIO DE CARACTERÍSTICAS DE UN CLASIFICADOR SVM DONDE SE MUESTRAN LAS MUESTRAS REPRESENTATIVAS, LOS VECTORES DE SOPORTE Y EL HIPERPLANO DISCRIMINANTE QUE HA SIDO ENTRENADO.	104
FIGURA 3.4 REPRESENTACIÓN DE UN ESPACIO 2D CON SU CORRESPONDIENTE HIPERPLANO: (A) ESPACIO ORIGINAL CON UN HIPERPLANO LINEAL. (B) ESPACIO DE CARACTERÍSTICAS TRANSFORMADO CON UN KERNEL DONDE SE MUESTRA UN HIPERPLANO LINEAL. (C) ESPACIO ORIGINAL CON UN HIPERPLANO DEFORMADO.	108
FIGURA 3.5 MARGEN DE POSICIONAMIENTO DEL HIPERPLANO. CUANTO MAYOR ES EL VALOR DE C LA SEPARACIÓN ENTRE CLASES ES MÁS ESTRICTA, CUANTO MENOR ES C SE PERMITE UN MAYOR NÚMERO DE CLASIFICACIONES ERRÓNEAS.	113
FIGURA 3.6 REPRESENTACIÓN ILUSTRATIVA DE UNA NEURONA DE TIPO BIOLÓGICO.	118
FIGURA 3.7 MODELO DE UNA NEURONA ARTIFICIAL A) McCULLOCH-PITTS B) ADALINA.	119
FIGURA 3.8 ESQUEMA DE UNA RED NEURONAL ARTIFICIAL PERCEPTRON.	123
FIGURA 3.9 ESQUEMA DE UNA RED NEURONAL ARTIFICIAL PERCEPTRON MULTICAPA.	124
FIGURA 3.10 ESQUEMA DE UNA RED NEURONAL ARTIFICIAL NARX.	125
FIGURA 3.11 ESQUEMA DE UNA RED NEURONAL ARTIFICIAL NAR.	126
FIGURA 3.12 GRAFICA QUE REFLEJA LA EVOLUCIÓN DEL ERROR DE APRENDIZAJE EN CADA CICLO CALCULADO POR LA FUNCIÓN DE COSTE EN CADA ITERACIÓN DEL ENTRENAMIENTO CUANDO EL ERROR ES GRANDE PARA UNA A) FUNCIÓN DE COSTE MSE, B) FUNCIÓN DE COSTE POR ENTROPÍA CRUZADA.	132
FIGURA 3.13 GRAFICA QUE REFLEJA LA EVOLUCIÓN DEL ERROR DE APRENDIZAJE EN CADA ITERACIÓN DEL ENTRENAMIENTO PARA LOS GRUPOS DE ENTRENAMIENTO, VALIDACIÓN Y PRUEBA.	133
FIGURA 3.14 EJEMPLO GRÁFICO DE LA DEFINICIÓN INICIAL DEL NÚMERO DE NEURONAS PARA UNA CAPA OCULTA MEDIANTE LA REGLA DE LA PIRÁMIDE.	137

FIGURA 4.1 ILUSTRACIÓN DEL BANCO DE ENSAYOS.....	145
FIGURA 4.2 SISTEMA DE POTENCIA.	147
FIGURA 4.3 MOTOR SIEMENS MODELO 1LA7090-2AA10.	148
FIGURA 4.4 PLACA DE CARACTERÍSTICAS DEL MOTOR SIEMENS MODELO 1LA7090-2AA10.	149
FIGURA 4.5 FOTOGRAFÍA QUE MUESTRA EL ROTOR SANO Y EL ROTOR CON ROTURA EN UNA DE SUS BARRAS.	149
FIGURA 4.6 AUTOTRANSFORMADOR DE LA MARCA VERILEC MODELO 3ARC 5-2.	150
FIGURA 4.7 PLACA DE CARACTERÍSTICAS DEL AUTOTRANSFORMADOR VERILEC 3ARC 5-2.	151
FIGURA 4.8 VARIADOR DE FRECUENCIA MARCA SIEMENS MODELO MICROMASTER 400.	152
FIGURA 4.9 PLACA DE CARACTERÍSTICAS DEL VARIADOR DE FRECUENCIA MARCA SIEMENS MODELO MICROMASTER 400.	152
FIGURA 4.10 GRÁFICA DE RELACIÓN V/F DE UN CONTROL ESCALAR.....	154
FIGURA 4.11 GRÁFICA DE RELACIÓN V/F DE UN CONTROL ESCALAR CON COMPENSACIÓN DEL DESLIZAMIENTO.	155
FIGURA 4.12 VARIADOR DE FRECUENCIA MARCA ABB MODELO ACS800-01-0005-3+E200+L502.	156
FIGURA 4.13 PLACA DE CARACTERÍSTICAS DEL VARIADOR DE FRECUENCIA MARCA ABB MODELO ACS800- 01-0005-3+E200+L502.	157
FIGURA 4.14 VARIADOR DE FRECUENCIA MARCA ABB MODELO ACSM1-04AS-024A-4+L516.	158
FIGURA 4.15. MÓDULO FEN21 DEL VARIADOR DE FRECUENCIA MARCA ABB MODELO ACSM1-04AS- 024A-4+L516.	159
FIGURA 4.16. MOTOR SÍNCRONO DE IMANES PERMANENTES.....	160
FIGURA 4.17. PLACA DE CARACTERÍSTICAS DEL MOTOR DE IMANES PERMANENTES.	160
FIGURA 4.18. DIAGRAMA DEL SISTEMA DE CONTROL AUTOMATIZADO.	162
FIGURA 4.19. AUTÓMATA PROGRAMABLE DE LA MARCA ABB MODELO PM583ETH.	163
FIGURA 4.20. ORDENADOR PORTÁTIL DE LA MARCA ASUS CON SCADA DE CODESYS.....	164
FIGURA 4.21. SISTEMA DE ADQUISICIÓN DE DATOS.	165
FIGURA 4.22. OSCILOSCOPIO YOKOGAWA MODELO DL750.	166
FIGURA 4.23. TARJETAS MODULARES DEL OSCILOSCOPIO PARA MUESTREO.	168
FIGURA 4.24. CONEXIONES DE LOS PUERTOS DEL OSCILOSCOPIO.	168
FIGURA 4.25. SONDA PARA LA ADQUISICIÓN DE VALORES.....	169
FIGURA 4.26. PINZA AMPERIMÉTRICA UTILIZADA EN EL BANCO DE ENSAYOS.....	170
FIGURA 4.27 . MUESTRA DE CORRIENTE DE FASE CON CONTROL DTC, CARGA AL 100% Y CON TRANSITORIOS DE 10 SEGUNDOS DE DURACIÓN.....	175
FIGURA 4.28 . MUESTRA DE CORRIENTE DE FASE DE UN MOTOR CON FALLO DE BARRA ROTA, CON CONTROL DTC, FRECUENCIA DE ALIMENTACIÓN CONSTANTE A 50Hz Y CARGA EN RAMPA DE 10 SEGUNDOS DE DURACIÓN ENTRE EL 25 Y EL 100% DEL PAR DE CARGA.	176
FIGURA 5.1. ESPACIO DE CARACTERÍSTICAS DE UN SISTEMA EXPERTO SVM DONDE SE HA AMPLIADO LA ZONA DE CONFLICTO (CARACTERÍSTICAS DE CLASES DISTINTAS MUY PRÓXIMAS) ORIGINADA POR UNAS MUESTRAS DE CORRIENTE CON UN DESLIZAMIENTO MUY BAJO Y UNA AMPLITUD MUY CERCANA.	187
FIGURA 5.2. RESULTADO DE LA FFT APLICADO A LA CORRIENTE DE UN MOTOR SANO ALIMENTADO CON UN CONVERTIDOR DE FRECUENCIA CON CONTROL DTC Y REFERENCIA DE 50Hz.	188

FIGURA 5.3. RESULTADO DE LA FFT APLICADO A LA CORRIENTE DE UN MOTOR CON FALLO DE BARRA ROTA ALIMENTADO CON UN CONVERTIDOR DE FRECUENCIA CON CONTROL DTC Y REFERENCIA DE 50Hz. .	189
FIGURA 5.4. SEÑAL ANALÍTICA OBTENIDA MEDIANTE LA TRANSFORMADA DE HILBERT.....	191
FIGURA 5.5 VECTOR DE GIRO DE LA COMPONENTE CONTINUA (EN AZUL) Y DE LA COMPONENTE ALTERNA (EN AMARILLO).	191
FIGURA 5.6. RESULTADO DE LA COMPONENTE ALTERNA DEL MÓDULO DE LA SEÑAL ANALÍTICA CORRESPONDIENTE A UNA MUESTRA DE CORRIENTE DE FASE OBTENIDA DE UN MOTOR SANO CON CONTROL ELECTRÓNICO CONFIGURADO A UNA FRECUENCIA DE TRABAJO DE 50Hz.	192
FIGURA 5.7. RESULTADO DE LA COMPONENTE ALTERNA DEL MÓDULO DE LA SEÑAL ANALÍTICA CORRESPONDIENTE A UNA MUESTRA DE CORRIENTE DE FASE OBTENIDA DE UN MOTOR CON FALLO Y CON CONTROL ELECTRÓNICO CONFIGURADO A UNA FRECUENCIA DE TRABAJO DE 50Hz.	193
FIGURA 5.8. RESULTADO DE HOTA CORRESPONDIENTE A UNA MUESTRA DE CORRIENTE DE FASE OBTENIDA DE UN MOTOR SANO CON CONTROL ELECTRÓNICO CONFIGURADO A UNA FRECUENCIA DE TRABAJO DE 50Hz.	195
FIGURA 5.9 RESULTADO DE HOTA CORRESPONDIENTE A UNA MUESTRA DE CORRIENTE DE FASE OBTENIDA DE UN MOTOR CON FALLO DE BARRA Y CON CONTROL ELECTRÓNICO CONFIGURADO A UNA FRECUENCIA DE TRABAJO DE 50Hz.	196
FIGURA 5.10 . RESULTADO DE CEPSTRUM CORRESPONDIENTE A UNA MUESTRA DE CORRIENTE DE FASE OBTENIDA DE UN MOTOR SANO CON CONTROL ELECTRÓNICO CONFIGURADO A UNA FRECUENCIA DE TRABAJO DE 50Hz.	198
FIGURA 5.11 RESULTADO DE CEPSTRUM CORRESPONDIENTE A UNA MUESTRA DE CORRIENTE DE FASE OBTENIDA DE UN MOTOR CON FALLO DE BARRA ROTA Y CON CONTROL ELECTRÓNICO CONFIGURADO A UNA FRECUENCIA DE TRABAJO DE 50Hz.	198
FIGURA 5.12. RESULTADO DE EPVA CORRESPONDIENTE A LAS TRES FASES DE CORRIENTE DE FASE OBTENIDA DE UN MOTOR SANO TRABAJANDO A 50Hz.	201
FIGURA 5.13 RESULTADO DE EPVA CORRESPONDIENTE A LAS TRES FASES DE CORRIENTE DE FASE OBTENIDA DE UN MOTOR CON FALLO DE BARRA ROTA TRABAJANDO A 50Hz.	201
FIGURA 5.14 ESPACIO DE CARACTERÍSTICAS DE UN CLASIFICADOR SVM CON SU HIPERPLANO DISCRIMINATORIO. LAS CARACTERÍSTICAS REPRESENTATIVAS DE ESTE ESPACIO HAN SIDO OBTENIDAS, A PARTIR DE MUESTRAS DE MOTORES CON DISTINTOS NIVELES DE FALLO DE BARRA ROTA, CON LA TÉCNICA EPVA USANDO UN PERIODOGRAMA DE WELCH Y UNA VENTANA DE HANN.	203
FIGURA 5.15. ALGORITMO PRINCIPAL DE GENERACIÓN DEL MEJOR SISTEMA EXPERTO SVM ENTRENADO PARA LA DETECCIÓN DE FALLOS RESPECTO DE LAS MUESTRAS APORTADAS.....	206
FIGURA 5.16. EJEMPLO DE EXPLORACIÓN EN PROFUNDIDAD PARA EL COEFICIENTE C	207
FIGURA 5.17. BÚSQUEDA EN MALLA PARA $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ Y $r = 2^{-15}, 2^{-13}, \dots, 2^3$	208
FIGURA 5.18 BÚSQUEDA FINA EN MALLA PARA $C = 2^1, 2^{1.25}, \dots, 2^5$ Y $r = 2^{-7}, 2^{-6.75}, \dots, 2^{-3}$	208
FIGURA 5.19. ESQUEMA DEL ALGORITMO DE AJUSTE DE COEFICIENTES POR BÚSQUEDA EN MALLA PARA EL KERNEL POLINOMIAL Y PARA EL KERNEL RBF.....	211
FIGURA 5.20 REPRESENTACIÓN DEL ENTRENAMIENTO Y CLASIFICACIÓN PARA UN CICLO DEL ALGORITMO DE VALIDACIÓN CRUZADA DE UNO A MUCHOS.	213
FIGURA 5.21 REPRESENTACIÓN DEL ENTRENAMIENTO Y CLASIFICACIÓN PARA UN CICLO DEL ALGORITMO DE VALIDACIÓN CRUZADA DE ENTRENAMIENTO/PRUEBA.	214

FIGURA 5.22 ALGORITMO DE ENTRENAMIENTO Y VALIDACIÓN USANDO LA METODOLOGÍA DE VALIDACIÓN CRUZADA DE UNO A MUCHOS.	215
FIGURA 5.23. EJEMPLO DE ESPACIO DE CARACTERÍSTICAS DEL CLASIFICADOR SVM CON HIPERPLANO SOBREENRENADO.....	216
FIGURA 5.24 . ESQUEMA DE UNA RED NEURONAL ARTIFICIAL PERCEPTRON MULTICAPA.	218
FIGURA 5.25. EJEMPLO GRÁFICO DE LA DEFINICIÓN INICIAL DE BÚSQUEDA DEL NÚMERO ÓPTIMO DE NEURONAS PARA UNA CAPA OCULTA MEDIANTE LA REGLA DE LA PIRÁMIDE.....	221
FIGURA 5.26. GRAFICA QUE REFLEJA LA EVOLUCIÓN DEL ERROR DE APRENDIZAJE EN CADA CICLO CALCULADO POR LA FUNCIÓN DE COSTE EN CADA ITERACIÓN DEL ENTRENAMIENTO CUANDO EL ERROR ES GRANDE PARA UNA A) FUNCIÓN DE COSTE MSE, B) FUNCIÓN DE COSTE POR ENTROPÍA CRUZADA.....	225
FIGURA 5.27. ALGORITMO PRINCIPAL DE GENERACIÓN DEL MEJOR SISTEMA EXPERTO ANN ENTRENADO PARA LA DETECCIÓN DE LOS FALLOS CORRESPONDIENTES A LAS MUESTRAS REPRESENTATIVAS DEL PROBLEMA.....	228
FIGURA 5.28. ALGORITMO DE BÚSQUEDA DEL MEJOR CLASIFICADOR ANN PARA UNA ESTRUCTURA DE RED NEURONAL DE UNA SOLA CAPA OCULTA.	232
FIGURA 5.29 ALGORITMO DE BÚSQUEDA DEL MEJOR CLASIFICADOR ANN PARA UNA ESTRUCTURA DE RED NEURONAL DE DOS CAPAS OCULTAS.	234
FIGURA 5.30 REPRESENTACIÓN DEL ENTRENAMIENTO Y CLASIFICACIÓN PARA UN CICLO DEL ALGORITMO DE VALIDACIÓN CRUZADA DE ENTRENAMIENTO/PRUEBA.	237
FIGURA 5.31 ALGORITMO DE APRENDIZAJE POR VALIDACIÓN CRUZADA DE ENTRENAMIENTO, PRUEBA Y VALIDACIÓN.	238
FIGURA 5.32. ALGORITMO DE SELECCIÓN DE MUESTRAS PARA FORMAR UN CONJUNTO DONDE LA REPRESENTATIVIDAD DE LAS MUESTRAS ABARCAN TODO EL RANGO DEL PROBLEMA.....	240
FIGURA 5.33. REPRESENTACIÓN DE UN ESPACIO DE CARACTERÍSTICAS PARA UN MOTOR SANO OBTENIDAS POR LA TÉCNICA EPVA.	241
FIGURA 5.34 REPRESENTACIÓN DE UN EJEMPLO DE DIVISIÓN DEL ESPACIO DE CARACTERÍSTICAS PARA UN MOTOR SANO OBTENIDAS POR LA TÉCNICA EPVA DIVIDIDO EN 23 SUBCONJUNTOS DE 2 ELEMENTOS (LOS SUBCONJUNTOS VACÍOS NO SE TIENEN EN CUENTA).	242
FIGURA 5.35. ENTORNO DE DESARROLLO MATEMÁTICO MATLAB R2014A.	243
FIGURA 6.1. EJEMPLO DE LA GENERACIÓN DE UN ESPACIO TIEMPO/FRECUENCIA PARA UNA SEÑAL DE ARRANQUE DE UN MOTOR UTILIZANDO UNA VENTANA TRIANGULAR CON LAS SIGUIENTES PROPORCIONES: A) LONGITUD MÁXIMA EN TIEMPO, B) LONGITUD MÍNIMA EN TIEMPO Y C) VENTANA AJUSTADA A LA RAMPA DE FALLO DEL ESPACIO TIEMPO/FRECUENCIA.....	278
FIGURA 6.2. EJEMPLO DE LA GENERACIÓN DE UN ESPACIO TIEMPO/FRECUENCIA PARA UNA SEÑAL DE ARRANQUE DE UN MOTOR UTILIZANDO UNA VENTANA GAUSSIANA CON LAS SIGUIENTES PROPORCIONES: A) LONGITUD MÁXIMA EN TIEMPO, B) LONGITUD MÍNIMA EN TIEMPO Y C) VENTANA AJUSTADA A LA RAMPA DE FALLO DEL ESPACIO TIEMPO/FRECUENCIA.....	279
FIGURA 6.3. ESQUEMA DE FLUJO DEL FUNCIONAMIENTO DE LA TÉCNICA HOTA PARA RÉGIMEN DE TRABAJO TRANSITORIO.....	281
FIGURA 6.4 EJEMPLO DE FUNCIÓN DE ONDA ESFEROIDAL PROLATE REPRESENTADA EN (A) 2D Y (B) 3D. .	286
FIGURA 6.5. COMPARATIVA GRÁFICA EN 2D ENTRE UNA VENTANA GAUSSIANA Y UNA VENTANA PROLATE.	287

FIGURA 6.6. DESPLAZAMIENTO DE LA FUNCIÓN DE VENTANA PROLATE SOBRE LA SEÑAL DE CORRIENTE OBTENIENDO ASÍ EL ESPACIO DE TIEMPO FRECUENCIA STFT.....	288
FIGURA 6.7 ESPACIO TIEMPO FRECUENCIA DE UNA CORRIENTE DE FASE GENERADO CON LA TÉCNICA STFT, DONDE SE OBSERVA UN ACUSADO EFECTO DE BORDE.....	289
FIGURA 6.8. RESULTADO SIMPLIFICADO DE HOTA TRANSITORIO OBTENIDO A PARTIR DE UN ESPACIO TIEMPO FRECUENCIA <u>CON</u> EFECTO DE BORDE.	290
FIGURA 6.9. RESULTADO SIMPLIFICADO DE HOTA TRANSITORIO OBTENIDO A PARTIR DE UN ESPACIO TIEMPO FRECUENCIA <u>SIN</u> EFECTO DE BORDE.	290
FIGURA 6.10. DESPLAZAMIENTO DE LA FUNCIÓN DE VENTANA GAUSSIANA SOBRE LA SEÑAL DE CORRIENTE PARA GENERAR EL ESPACIO DE TIEMPO FRECUENCIA STFT, OMITIENDO LA ZONA EN LOS LÍMITES DE LA SEÑAL DONDE SE PRODUCE EL EFECTO DE BORDE.	291
FIGURA 6.11. COMPARATIVA DEL ESPACIO TIEMPO FRECUENCIA PARA UNA MUESTRA DE CORRIENTE DE FASE VARIANDO LA PROPORCIÓN DE VENTANA ENTRE A) 10 UNIDADES EN TIEMPO POR 1 EN FRECUENCIA, B) 3 UNIDADES EN TIEMPO POR 1 EN FRECUENCIA, C) 1 UNIDAD EN TIEMPO POR 3 EN FRECUENCIA Y C) 1 UNIDAD EN TIEMPO POR 10 EN FRECUENCIA.....	294
FIGURA 6.12. COMPARATIVA DEL RESULTADO REDUCIDO PARA HOTA TRANSITORIO PARA UNA MUESTRA DE CORRIENTE DE FASE VARIANDO LA PROPORCIÓN DE VENTANA ENTRE A) 10 UNIDADES EN TIEMPO POR 1 EN FRECUENCIA, B) 3 UNIDADES EN TIEMPO POR 1 EN FRECUENCIA, C) 1 UNIDAD EN TIEMPO POR 3 EN FRECUENCIA Y C) 1 UNIDAD EN TIEMPO POR 10 EN FRECUENCIA.	295
FIGURA 6.13. REPRESENTACIÓN PARAMÉTRICA DE LA FUNCIÓN DE VENTANA EN EL ESPACIO DE TIEMPO- FRECUENCIA.....	297
FIGURA 6.14. DESPLAZAMIENTO DE LA FUNCIÓN DE VENTANA GAUSSIANA SOBRE LA SEÑAL DE CORRIENTE PARA GENERAR EL ESPACIO DE TIEMPO FRECUENCIA MEDIANTE LA TÉCNICA SFFT.	298
FIGURA 6.15. ALGORITMO DE OBTENCIÓN DE LAS FRECUENCIAS REPRESENTATIVAS DE FALLO Y LOS VALORES MEDIOS DEL RUIDO DE FONDO BASADO EN BANDAS ADAPTATIVAS.....	299
FIGURA 6.16. EJEMPLO DE ESPACIO TIEMPO FRECUENCIA, DONDE SE HA MARCADO EN COLOR LAS FRECUENCIAS PROCESADAS POR EL ALGORITMO DE BANDAS ADAPTATIVAS.	301
FIGURA 6.17. ESPACIO DE CARACTERÍSTICAS DEL SISTEMA EXPERTO SVM CON KERNEL POLINOMIAL, Y LA VARIANTE DE HOTA PARA RÉGIMEN TRANSITORIO CON TRANSFORMADA SFFT Y VENTANA GAUSSIANA.	313
FIGURA A3.1. DIAGRAMA DE BLOQUES DE UNA GPU MAXWELL GM107.	31353

Índice de Tablas.

TABLA 2.1 TASA DE APARICIÓN DE FALLOS EN MÁQUINAS ELÉCTRICAS ROTATIVAS DE INDUCCIÓN OBTENIDAS DEL ESTUDIO REALIZADO POR [3].	9
TABLA 2.2 COMPONENTES FRECUENCIALES GENERADAS SEGÚN EL FALLO.	24
TABLA 3.1. DEFINICIÓN DE LOS COEFICIENTES DISPONIBLES EN CADA KERNEL.	114
TABLA 3.2. FUNCIONES DE TRANSFERENCIA ($\gamma(\tau)$) MÁS UTILIZADAS.	121
TABLA 4.1. CARACTERÍSTICAS ELÉCTRICAS DEL MOTOR SIEMENS MODELO 1LA7090-2AA10.	148
TABLA 4.2. CARACTERÍSTICAS TÉCNICAS DEL AUTO-TRANSFORMADOR VERILEC 3ARC 5-2.	151
TABLA 4.3. CARACTERÍSTICAS TÉCNICAS DEL VARIADOR SIEMENS MODELO MICROMASTER 400.	153
TABLA 4.4. CARACTERÍSTICAS TÉCNICAS DEL VARIADOR ABB MODELO ACS800-01-0005-3+E200+L503.	157
TABLA 4.5. CARACTERÍSTICAS TÉCNICAS DEL VARIADOR ABB MODELO ACSM1-04AS-024A-4+L516.	159
TABLA 4.6. CARACTERÍSTICAS TÉCNICAS DEL MOTOR DE IMANES PERMANENTES.	161
TABLA 4.7. CARACTERÍSTICAS TÉCNICAS DE LA PINZA AMPERIMÉTRICA.	170
TABLA 4.8. VARIACIONES DISTINTAS DE PARÁMETROS GENERADAS CON CONEXIÓN POR AUTOTRANSFORMADOR.	172
TABLA 4.9 VARIACIONES DISTINTAS DE PARÁMETROS GENERADAS CON CONEXIÓN MEDIANTE VARIADOR DE LA MARCA SIEMENS.	173
TABLA 4.10. VARIACIONES DISTINTAS DE PARÁMETROS GENERADAS CON CONEXIÓN MEDIANTE VARIADOR DE LA MARCA ABB.	174
TABLA 4.11. VARIACIONES DISTINTAS DE PARÁMETROS GENERADOS CON CONEXIÓN MEDIANTE VARIADOR DE LA MARCA SIEMENS INTRODUCIENDO TRANSITORIOS DE 10 SEGUNDOS DE DURACIÓN.	177
TABLA 4.12. VARIACIONES DISTINTAS DE PARÁMETROS GENERADOS CON CONEXIÓN MEDIANTE VARIADOR DE LA MARCA ABB INTRODUCIENDO TRANSITORIOS DE 10 SEGUNDOS DE DURACIÓN.	178
TABLA 4.13. VARIACIONES DISTINTAS DE PARÁMETROS GENERADAS CON CONEXIÓN POR AUTOTRANSFORMADOR, CON VARIACIONES EN RAMPA DE LA CARGA DE 10 SEGUNDOS DE DURACIÓN POR RAMPA.	179
TABLA 4.14. VARIACIONES DISTINTAS DE PARÁMETROS GENERADAS CON CONEXIÓN MEDIANTE VARIADOR DE LA MARCA SIEMENS, CON VARIACIONES EN RAMPA DE LA CARGA DE 10 SEGUNDOS DE DURACIÓN POR RAMPA.	179
TABLA 4.15. VARIACIONES DISTINTAS DE PARÁMETROS GENERADAS CON CONEXIÓN MEDIANTE VARIADOR DE LA MARCA ABB, CON VARIACIONES EN RAMPA DE LA CARGA DE 10 SEGUNDOS DE DURACIÓN POR RAMPA.	181
TABLA 4.16. VARIACIONES DISTINTAS DE PARÁMETROS GENERADOS CON CONEXIÓN MEDIANTE VARIADOR DE LA MARCA SIEMENS INTRODUCIENDO TRANSITORIOS DE 10 SEGUNDOS DE DURACIÓN Y CON VARIACIONES EN RAMPA DE LA CARGA DE 10 SEGUNDOS DE DURACIÓN POR RAMPA.	182
TABLA 4.17. VARIACIONES DISTINTAS DE PARÁMETROS GENERADOS CON CONEXIÓN MEDIANTE VARIADOR DE LA MARCA ABB INTRODUCIENDO TRANSITORIOS DE 10 SEGUNDOS DE DURACIÓN Y CON VARIACIONES EN RAMPA DE LA CARGA DE 10 SEGUNDOS DE DURACIÓN POR RAMPA.	183

TABLA 5.1 FUNCIONES DE TRANSFERENCIA RECOMENDADAS PARA PROBLEMAS DE CLASIFICACIÓN DE PATRONES.	222
TABLA 5.2. RESULTADOS DE LOS TIEMPOS OBTENIDOS EN LA EJECUCIÓN DE LAS TÉCNICAS DE DIAGNÓSTICO EN DONDE SE HAN OBTENIDO LAS CARACTERÍSTICAS REPRESENTATIVAS DE FALLO A PARTIR DE LAS MUESTRAS.....	248
TABLA 5.3. RESULTADOS DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN SVM CONFIGURADO CON UN <u>KERNEL POLINOMIAL</u> , Y LOS TIEMPOS TOTALES DE LOS DIVERSOS SISTEMAS DE DIAGNÓSTICO.....	252
TABLA 5.4. RESULTADOS DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN SVM CONFIGURADO CON UN <u>KERNEL GAUSSIANO RBF</u> , Y LOS TIEMPOS TOTALES DE LOS DIVERSOS SISTEMAS DE DIAGNÓSTICO.	253
TABLA 5.5. RESULTADOS MÁXIMO, MÍNIMO Y PROMEDIO DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN SVM CON <u>KERNEL POLINOMIAL</u> Y CON <u>KERNEL GAUSSIANO RBF</u>	254
TABLA 5.6. RESULTADOS OBTENIDOS RESPECTO DEL CLASIFICADOR EXPERTO DE TIPO SVM CONFIGURADO CON UN <u>KERNEL POLINOMIAL</u>	257
TABLA 5.7. RESULTADOS OBTENIDOS RESPECTO DEL CLASIFICADOR EXPERTO DE TIPO SVM CONFIGURADO CON UN <u>KERNEL RBF GAUSSIANO</u>	258
TABLA 5.8. RESULTADOS MÁXIMO, MÍNIMO Y PROMEDIO OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN SVM CON <u>KERNEL POLINOMIAL</u> Y CON <u>KERNEL GAUSSIANO RBF</u>	260
TABLA 5.9. RESULTADOS DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN ANN DE <u>UNA SOLA CAPA OCULTA</u> , Y LOS TIEMPOS TOTALES DE LOS SISTEMAS DE DIAGNÓSTICO.	262
TABLA 5.10 RESULTADOS DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN ANN DE DOS CAPAS OCULTAS, Y LOS TIEMPOS TOTALES DE LOS SISTEMAS DE DIAGNÓSTICO.	263
TABLA 5.11. RESULTADOS MÁXIMO, MÍNIMO Y PROMEDIO DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN ANN PARA <u>UNA SOLA CAPA OCULTA</u> Y PARA <u>DOS CAPAS OCULTAS</u>	264
TABLA 5.12. RESULTADOS OBTENIDOS RESPECTO DEL CLASIFICADOR EXPERTO DE TIPO ANN CON UNA SOLA CAPA OCULTA.....	266
TABLA 5.13. RESULTADOS OBTENIDOS RESPECTO DEL CLASIFICADOR EXPERTO DE TIPO ANN CON <u>DOS CAPAS OCULTAS</u>	267
TABLA 6.1. RESULTADOS DE LOS TIEMPOS OBTENIDOS EN LA EJECUCIÓN DE LAS VARIANTES DE LA TÉCNICA DE DIAGNÓSTICO HOTA TRANSITORIO EN DONDE SE HAN OBTENIDO LAS CARACTERÍSTICAS REPRESENTATIVAS DE FALLO A PARTIR DE LAS MUESTRAS.....	305
TABLA 6.2. RESULTADOS DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN SVM CONFIGURADO CON UN <u>KERNEL POLINOMIAL</u> , Y LOS TIEMPOS TOTALES DE LAS VARIANTES DEL SISTEMA DE DIAGNÓSTICO HOTA TRANSITORIO.	308
TABLA 6.3. RESULTADOS DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN SVM CONFIGURADO CON UN <u>KERNEL RBF</u> , Y LOS TIEMPOS TOTALES DE LAS VARIANTES DEL SISTEMA DE DIAGNÓSTICO HOTA TRANSITORIO.	308
TABLA 6.4. RESULTADOS MÁXIMO, MÍNIMO Y PROMEDIO DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN SVM CON <u>KERNEL POLINOMIAL</u> Y CON <u>KERNEL GAUSSIANO RBF</u>	309
TABLA 6.5. RESULTADOS OBTENIDOS RESPECTO DEL CLASIFICADOR EXPERTO DE TIPO SVM CONFIGURADO CON UN <u>KERNEL POLINOMIAL</u>	311

TABLA 6.6. RESULTADOS OBTENIDOS RESPECTO DEL CLASIFICADOR EXPERTO DE TIPO SVM CONFIGURADO CON UN <u>KERNEL RBF GAUSSIANO</u>	311
TABLA 6.7. RESULTADOS MÁXIMO, MÍNIMO Y PROMEDIO OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN SVM CON <u>KERNEL POLINOMIAL</u> Y CON <u>KERNEL GAUSSIANO RBF</u>	312
TABLA 6.8. RESULTADOS DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN ANN DE <u>UNA SOLA CAPA OCULTA</u> , Y LOS TIEMPOS TOTALES DE LOS SISTEMAS DE DIAGNÓSTICO.....	315
TABLA 6.9. RESULTADOS DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN ANN DE DOS CAPAS OCULTAS, Y LOS TIEMPOS TOTALES DE LOS SISTEMAS DE DIAGNÓSTICO.	315
TABLA 6.10. RESULTADOS MÁXIMO, MÍNIMO Y PROMEDIO DE LOS TIEMPOS OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN ANN PARA <u>UNA SOLA CAPA OCULTA</u> Y PARA <u>DOS CAPAS OCULTAS</u>	316
TABLA 6.11. RESULTADOS OBTENIDOS RESPECTO DEL CLASIFICADOR EXPERTO DE TIPO ANN CON <u>UNA SOLA CAPA OCULTA</u>	318
TABLA 6.12. RESULTADOS OBTENIDOS RESPECTO DEL CLASIFICADOR EXPERTO DE TIPO ANN CON <u>DOS CAPAS OCULTAS</u>	318
TABLA 6.13. RESULTADOS MÁXIMO, MÍNIMO Y PROMEDIO OBTENIDOS DEL SISTEMA EXPERTO DE CLASIFICACIÓN ANN PARA <u>UNA SOLA CAPA OCULTA</u> Y PARA <u>DOS CAPAS OCULTAS</u>	319

Abreviaturas

AC	Corriente alterna
ANN	Redes neuronales artificiales
AS	Señal analítica
BFGS	Broyden-Fletcher-Goldfarb-Shanno
CCWT	Transformada Wavelet continua compleja
CE	Función de coste por entropía cruzada
CFC	Diagrama de funciones continuas
CLIPS	Sistema de producción integrado en lenguaje C
CPU	Unidad central de proceso
CUDA	Arquitectura unificada de dispositivos de cómputo
CWT	Transformada Wavelet continúa
DDR3	Memoria con doble tasa de transferencia (versión 3 que mejora la velocidad de transferencia respecto a versiones anteriores).
DHWT	Transformada Wavelet discreta armónica
DSP	Procesador digital de señales
DTC	Control directo del par

DWT	Transformada Wavelet discreta
EMD	Descomposición en modo empírico
EPVA	Aproximación al vector de Park extendido
FFT	Transformada rápida de Fourier
FFT⁻¹	Inversa de la transformada rápida de Fourier
FPGA	Matriz de puertas lógicas programables
FrFT	Transformada fraccional de Fourier
FTP	Protocolo de transferencia de archivos
GPU	Unidad gráfica de proceso
HD	Alta definición
HOTA	Análisis del orden armónico
HT	Transformada de Hilbert
HHT	Transformada de Hilbert Huang
IDE	Entorno de desarrollo integrado
IEC	Comisión electrotécnica internacional
IF	Frecuencia instantánea
IMF	Funciones de modo intrínseco
kNN	K vecino más cercano
LW	Pesos de la capa
MAE	Error absoluto medio
MATLAB	Matrix Laboratory
MPL	Perceptron multicapa
MPVA	Aproximación al vector de Park multicapa
MSE	Error cuadrático medio
NAR	Red autorregresiva
NARX	Red autorregresiva con entradas externas
PC	Ordenador personal.
PLC	Autómata lógico programable

POO	Programación orientada a objetos
PPT	Transformada polinomial de fase
PROLOG	Lenguaje de programación en lógica
PSWF	Función elemental de onda esferoidal Prolate
PVA	Aproximación al vector de Park
PWM	Modulación del ancho del pulso
QP	Programación cuadrática
RAM	Memoria de acceso aleatorio
RBF	Función de base radial
SAE	Suma del error absoluto
SATA	Interfaz de transferencia de datos en serie “Tecnología avanzada Serial”
SATA2	SATA versión 2. Mejora la velocidad de transferencia respecto a SATA .
SATA3	SATA versión 3. Mejora la velocidad de transferencia respecto a SATA2 .
SCADA	Sistema de supervisión del control y adquisición de datos.
SFFT	Transformada de Fourier corta en frecuencia
SOM	Mapa auto-organizado
SSE	Suma del error cuadrático
STFT	Transformada de Fourier corta en el tiempo
SVM	Máquina de vectores de soporte
TDL	Línea de retraso del tiempo
TKEO	Operador de energía Teager-Kaiser
TMm	Tiempo medio para procesar una muestra
TTc	Tiempo total para generar características
UART	Receptor transmisor universal asíncrono
USB	Bus serie universal
VSD	Variador de velocidad / Convertidor de frecuencia
XOR	OR exclusiva
WVD	Distribución Wigner-Ville

Capítulo 1

Introducción

1.1 Antecedentes y estado actual.

A lo largo de los años, muchas de las investigaciones desarrolladas en torno a la detección de fallos se han enfocado en la búsqueda de características discriminatorias dentro de las magnitudes medibles de los motores industriales, de forma que se pueda determinar si existe algún tipo de fallo, así como evaluar su severidad. Principalmente los estudios más recientes están basados en el estudio de la medida de corriente; ya que en esta magnitud utilizando técnicas no invasivas, aprovechando los sensores de corriente ya disponibles en la propia instalación o dentro del control de la máquina, se pueden observar las perturbaciones que provocan los diferentes fallos que se pueden presentar.

De esta forma y gracias a la observación de estas características significativas, se puede deducir si se ha producido algún tipo de fallo en el motor. No obstante, la detección correcta se ve limitada por la capacidad del observador humano o del clasificador utilizado para analizar las características.

Una de las metas más importante para la detección de fallos es poderlo hacer lo más tempranamente posible y además poder extraer la mayor información de este, de forma que se eviten o minimicen los daños que se puedan ocasionar. Cuanto menos significativo es el daño más difícil es la detección del fallo.

Para el desarrollo de un sistema de detección de fallos se buscan y extraen características discriminatoria con las que poder identificar la existencia del fallo, pudiendo discernir sobre la confirmación/negación del fallo mediante esta característica; y en algunos casos se puede llegar a especular incluso sobre la gravedad de este.

Por ejemplo para detectar un fallo de barra rota el método más usado es utilizando la transformada FFT, que trata de detectar la amplitud característica en la posición $2fs$ (s =deslizamiento, f =frecuencia fundamental) respecto de la frecuencia fundamental cuando se transforma al espectro de frecuencias una de las fases de corriente del motor mediante la transformada FFT. Con esta característica, además de conocer si existe fallo o no, también se puede estimar con la amplitud del pico la gravedad de este; pero no se puede decir a ciencia cierta ni el número de barras afectadas ni la posición. Existen otros métodos que explotan otras transformaciones de estas características (Extended Park Vector Approach, Cepstrum, etc.) y que aportan una información similar.

Otro ejemplo es el uso de modelos de descomposición en modo empírico (EMD) para la obtención de componentes ortogonales de funciones de modo intrínseco (IMF) desde donde calcular la oscilación típica del motor para extraer las características representativas respecto de diversos fallos. Con esta técnica se puede conocer si existen múltiples fallos en el estator, rotor o el rodamiento.

Como se puede extrapolar de los ejemplos anteriores, las investigaciones actuales se centran en facilitar al máximo la detección de los fallos, cosa primordial limitar pérdidas económicas en una industria dada por los daños ocasionados en el motor y el posible tiempo fuera de servicio de la cadena de montaje a la se encuentra acoplado.

Con técnicas de diagnóstico donde la detección de fallo sea más precisa, el operario responsable podría valorar mejor los riesgos del uso de la maquinaria sin poner en riesgo el resto de la cadena de montaje, pudiendo reparar este motor alargando su vida útil. Esta información de fallo es esencial sobretodo para motores poco accesibles, donde el coste para acceder a este y sustituirlo es alto.

Por ello, es prioritario investigar y desarrollar nuevos sistemas de detección que maximicen la precisión respecto a las características disponibles.

Cualquier sistema a implementar que maximice la precisión en la detección y minimice el tiempo de detección lógicamente debe ser totalmente automático, pues la precisión y el tiempo de detección que puede llegar a desarrollar un operario humano de forma manual está más limitado que la precisión y el mínimo tiempo de detección que puede llegar a alcanzar actualmente una máquina especializada.

Para mejorar la clasificación de fallos con respecto a los sistemas actuales, se estudian sistemas cada vez más complejos basados en sistemas expertos discriminatorios multidimensionales. Por ejemplo, parte de dicho estudio se enfoca en las máquinas de vectores de soporte (SVM), utilizando núcleos de transformación de espacios polinomiales y Gaussianos para la obtención de diversos hiperplanos discriminatorios; aunque de momento solo se han aplicado sobre espacios bidimensionales.

Los fallos más relevantes dentro del ámbito de los motores de inducción son la rotura de barras en el rotor, daños en los cojinetes, daños en el anillo del rotor, cortocircuito y entre distintas partes del motor. Los sistemas de clasificación más comunes actualmen-

te son los denominados sistemas de razonamiento lógicos y de aprendizaje (inteligencia artificial, sistemas expertos, redes bayesianas, etc.), que ya se aplican con bastante éxito en muchos otros campos (lingüística, industria, medicina, robótica, mecatrónica, videojuegos, etc.). Hasta ahora, las características discriminatorias obtenidas a partir de los métodos que se utilizan para la detección de fallos suelen obtenerse mayormente en el dominio de la frecuencia o tiempo-frecuencia.

1.2 Desarrollo del estudio.

Dado que mi formación abarca desde el campo de la automática industrial hasta los sistemas expertos e inteligencia artificial, he utilizado mi conocimiento adquirido en desarrollar para esta tesis de investigación un suprasistema automático capaz de generar sistemas óptimos de diagnóstico de fallos personalizable a la tipología del motor que se desarrolla, pudiendo desarrollar mediante este suprasistema de forma automática nuevos sistemas óptimos de diagnóstico de fallos conforme se sustituye la tipología del motor. Los sistemas de diagnóstico generados están principalmente orientados a maximizar la precisión en la detección del fallo principalmente, minimizando los costes de cálculo y de almacenamiento de memoria como objetivos secundarios.

Partiendo de las características discriminatorias que se obtienen de los diversos métodos de detección de los fallos; los sistemas expertos de clasificación son adaptados teniendo en cuenta la naturaleza de estas características con el fin de obtener el mejor sistema de detección para cada tipo fallo.

En primer caso (capítulo 2), se ha desarrollado un extenso estado del arte analizando todos los elementos relevantes concernientes a la dinámica de los sistemas de diagnóstico. Se han analizados los diversos tipos de fallos y se han expuesto los diversos métodos de diagnóstico existentes en la actualidad tanto para un régimen de trabajo estacionario como un régimen de trabajo transitorio del motor. Además, también se ha analizado y desarrollado diversos sistemas expertos existentes dentro del campo de la inteligencia artificial, orientándolos en esta tesis a su posible aplicación como sistema de toma de decisiones ante las características representativas de fallo.

Dentro del capítulo 3 se ha explicado con mayor profundidad los sistemas expertos seleccionados como posibles sistemas de toma de decisiones en el diagnóstico dentro del suprasistema. Para poder validar experimentalmente el suprasistema, como se describe posteriormente en el capítulo 4 se ha montado un banco de ensayos del que se han obtenido una base de datos de muestras usada para validar el correcto funcionamiento del suprasistema y para obtener resultados experimentales.

A partir del análisis del estado del arte se ha implementado el suprasistema destinado a generar sistemas óptimos de diagnóstico de fallos tanto en régimen de trabajo estacionario (capítulo 5) como en régimen de trabajo transitorio (capítulo 6), basados en di-

versas técnicas de diagnósticos y sistemas expertos, para una serie de muestras obtenidas de la tipología del motor a estudio.

En el capítulo 7 se exponen las conclusiones de la tesis. En ella se ha tratado el diagnóstico de motores de inducción para el régimen de trabajo estacionario y el régimen de trabajo transitorio, analizando el fallo de barras rotas en el motor, siendo el suprasistema de generación de sistemas de diagnóstico capaz de desarrollar sistemas de diagnóstico óptimos para ambos regímenes de trabajo.

En el capítulo 0 y 6 de la tesis se ha realizado un estudio experimental en laboratorio para validar el correcto funcionamiento del suprasistema automático de generación de sistemas de diagnóstico óptimos. Obteniendo en este caso unos resultados excelentes (sistemas de diagnóstico óptimos).

Además del desarrollo del suprasistema, también se han desarrollado y aplicado nuevas **innovaciones** (sin aparición previa en el estado del arte), tanto para las técnicas de diagnóstico implementadas en el suprasistema como para los sistemas expertos utilizados. Estas innovaciones están orientadas a mejorar la precisión en el diagnóstico de fallos y a mejorar los costes de cálculo asociados al diagnóstico.

Para complementar la tesis, se han añadido 3 anexos para extender con mayor detalle conceptos, métodos y técnicas comentadas en la tesis. En el anexo 1 se desarrolla un algoritmo de toma de decisiones basada en un Clúster de redes neuronales generadas durante el entrenamiento. En el anexo 2 se describe en detalle el algoritmo de entrenamiento por retropropagación del gradiente conjugado escalado. En el último anexo 3 se explica la prueba desarrollada para mejorar el coste de tiempo acelerando los cálculos mediante paralelización con hardware CUDA.

Dentro del grupo de investigación es posible acceder a todo el material y medios necesarios para desarrollar una tesis de este tipo. En el laboratorio de esta unidad están disponibles máquinas eléctricas, con motores sanos y motores con fallos provocados adrede para su estudio. Estos motores con fallo abarcan todos los tipos de fallo que más comúnmente ocurren en la práctica.

Para la observación de los anteriores motores se han utilizado equipos de medición (osciloscopios) de tipo digital, los cuales permiten el almacenamiento y transmisión de la información muestreada para su tratamiento desde un equipo informático.

Los equipos informáticos disponibles son ordenadores de gama alta, necesarios y suficientes para ejecutar los complejos algoritmos de simulación y de clasificación que requiere esta tesis.

1.3 Objetivos.

El objetivo general de este trabajo es el de mejorar las capacidades de la detección de fallos respecto de los actuales sistemas de detección de averías electromecánicas de los sistemas rotatorios.

Con dichas mejoras se espera que dentro de la industria disminuyan los costes al disminuir los problemas provocados por fallos de los sistemas rotatorios, así como evitar un mayor número de accidentes que pueden afectar al personal.

El objetivo es el desarrollo de un suprasistema que es capaz de generar diversos sistemas de diagnóstico óptimo en diferentes regímenes de trabajo y con una alta precisión respecto de la detección de fallos; superando incluso los resultados de detección de fallos obtenidos con las técnicas actuales de diagnóstico.

Para ello son necesarios la consecución de los objetivos parciales que se han perseguido:

- Revisión del estado del arte sobre averías en los motores de inducción.
- Revisión de las técnicas de análisis de fallos en motores, así como las características relevantes utilizadas para la detección de los fallos.
- Estudio de diversos sistemas de clasificación (redes neuronales, máquinas de vectores de soporte, etc.) con posibilidad de adaptación para la clasificación de las características de detección de fallos en máquinas rotatorias.
- Experimentación con diversos sistemas automáticos de detección de fallos desarrollados como resultado de los estudios previos.
- Valoración de los resultados y recopilación de los sistemas automáticos clasificatorios más acordes con los objetivos.

Capítulo 2

Estado del arte

2.1 Introducción

Desde su aparición a finales del siglo XIX, los motores de inducción se han convertido en equipos imprescindibles en la mayor parte de los accionamientos industriales, convirtiéndose en un elemento esencial para cualquier máquina que requiera de algún esfuerzo de tipo mecánico.

Con el paso de los años, las mejoras e innovaciones introducidas en los motores de inducción han ido mejorando su potencia específica. Un motor construido actualmente es aproximadamente varias veces más potente que un motor con el mismo tamaño construido al inicio del siglo XX.

En la actualidad, más del 80% de las máquinas que realizan la función de convertir energía eléctrica en energía mecánica son motores de inducción de corriente alterna [1], llegando a consumir una parte muy importante de la energía requerida en la mayoría de instalaciones.

Dentro del grupo de motores de inducción de corriente alterna destaca un tipo en particular que es alabado por sus diversas cualidades, como son su simplicidad en la fabricación, su bajo precio (causa derivada de la cualidad anterior) y su gran robustez. Estos motores son denominados “de jaula de ardilla” y es actualmente el más extendido y utilizado en muchos tipo distintos de maquinaria.

Aunque los motores de inducción del tipo de jaula de ardilla son los más robustos que existen, como en cualquier tipo de máquina, su utilización conlleva una degradación progresiva a causa de diversos factores. Estos factores pueden ser, por ejemplo, un defecto en su fabricación, estrés dado por su uso, funcionamiento en un ambiente nocivo, etc.

En consecuencia, dada la expansión de este tipo de motores y su importancia relevante dentro de la industria, a lo largo de los años los expertos competentes en el tema (técnicos e ingenieros) han ido desarrollando sistemas cada vez más complejos para el control, protección, mantenimiento y el diagnóstico de estos motores eléctricos. Con estos sistemas de diagnóstico se pretende monitorizar y detectar, de la forma más exacta y rápida posible, todas aquellas anomalías no previstas, que se presentan a lo largo de su vida útil.

Una de las facetas de diagnóstico más estudiada dentro del conjunto de los motores de inducción versa en la detección de fallos. Este hecho es de vital importancia para poder ahorrar costes al evitar averías derivadas y poder controlar de forma óptima la maquinaria de producción (donde se produce el fallo) minimizando los tiempos de parada.

Ha sido en estas últimas dos décadas donde se han concentrado los esfuerzos para investigar sobre los fallos que aparecen en los motores de inducción, así como su posible detección y diagnóstico [2]. Actualmente el principal problema es la optimización de las técnicas de diagnóstico desarrolladas de cara a mejorar su precisión y fiabilidad en aquellos casos en que los datos no reflejan suficiente información para poder discernir sobre la existencia/ausencia de fallo, dándose un estado de incertidumbre no deseado.

En los siguientes apartados de este capítulo se van a describir los fallos más relevantes en un motor de inducción, las diversas técnicas existentes para la detección de fallos y los sistemas expertos más utilizados para mejorar la clasificación de datos.

2.2 Fallos de un motor de inducción.

La complejidad estructural de un motor de inducción de jaula de ardilla (cojinetes, estator, rotor, etc.) hace que puedan surgir fallos de diversas tipologías. Los fallos más comunes por lo que se ven afectados los motores de inducción son de carácter mecánico y eléctrico.

Aunque las tasas de aparición para cada tipo de fallo varían dependiendo del estudio analizado (puesto que son porcentajes estadísticos). Como se puede observar en las tasas de fallo recopiladas por los estudios de Thorsen [3], Bell [4] y Clay [5], todos estos porcentajes de fallo se mantienen dentro de un rango. Por ejemplo, en la siguiente Tabla 2.1 se muestra los resultados obtenidos respecto a las tasas de aparición de fallos para el estudio realizado por Thorsen (1999) [3].

Tabla 2.1 Tasa de aparición de fallos en máquinas eléctricas rotativas de inducción obtenidas del estudio realizado por [3].

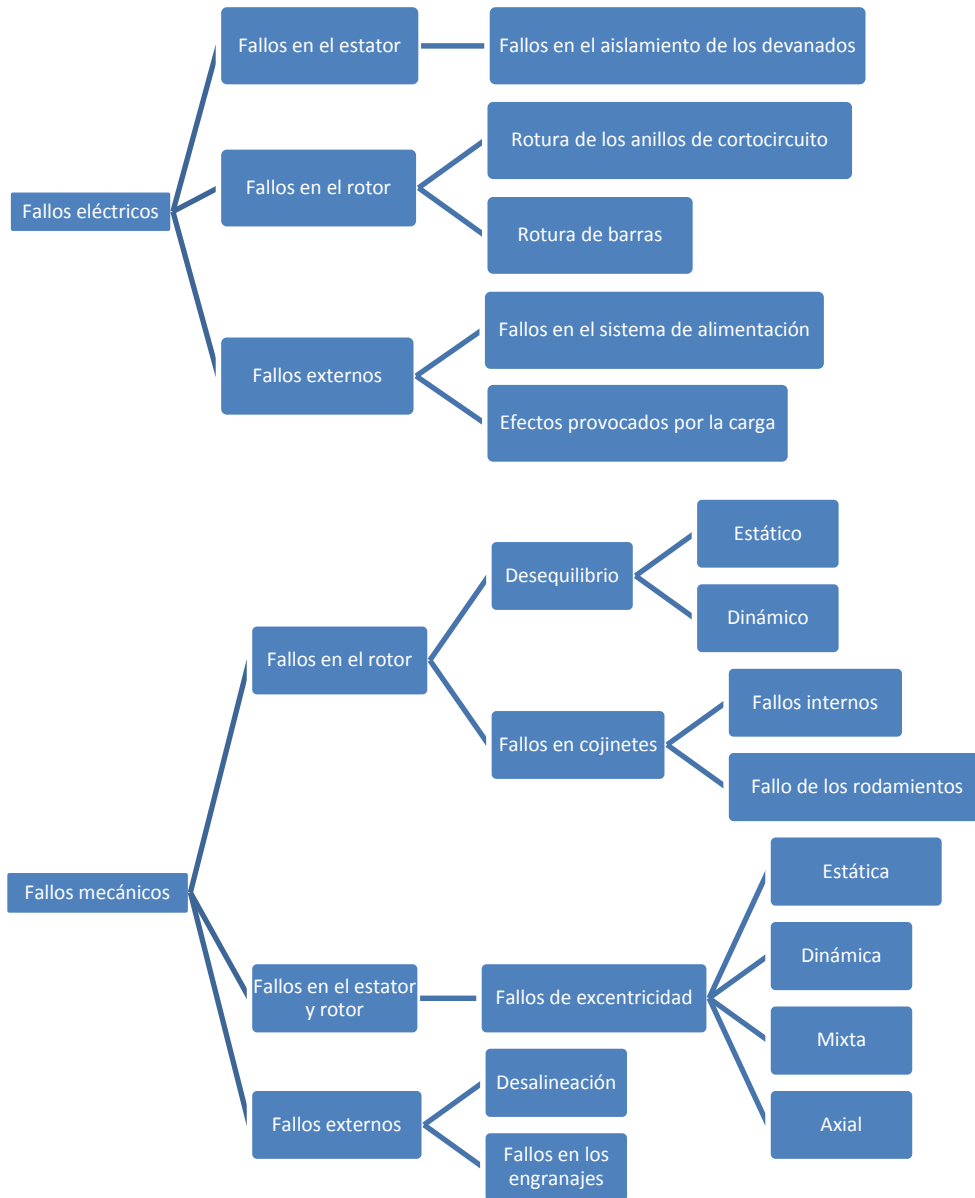
Fallo	Porcentaje (%)
Fallo en los cojinetes	51,6
Fallo en el estator	24,8
Fallo en el rotor	6,0
Fallo en el acoplamiento	3,2
Otros tipos de fallos	14,4

Fuente: Thorsen y Dalva (1999) [3]

Los fallos más importantes de este tipo de maquinaria rotativa se pueden agrupar en dos conjuntos dependiendo de la naturaleza del fallo (fallos eléctricos y fallos mecánicos). La clasificación completa se muestra en el esquema de la Figura 2.1 [6].

Aunque no todos los tipos de fallos que se producen en un motor de inducción dan como resultado el paro total de este, sí se puede dar una degradación paulatina de su funcionamiento que puede acabar generando posteriormente otros fallos de mayor gravedad.

Cuando el fallo es crítico actúan las protecciones y se produce la parada del motor, de forma que puede ser, si la avería lo requiere, prácticamente instantánea. Cuando el fallo no es crítico, entran en juego los sistemas de diagnóstico de fallos orientados a la detección de fallos en los motores, los cuales pueden evitar en un futuro próximo una avería mayor que conlleve una pérdida económica importante para el proceso de producción donde el motor se encuentra funcionando.



Fuente: Puche (2008) [6]

Figura 2.1 Clasificación de los fallos más importantes que pueden ocurrir en un motor de inducción de jaula de ardilla.

2.2.1 Fallos de tipo eléctrico.

En el siguiente apartado se exponen los fallos de tipo eléctricos que pueden darse en cualquier motor de inducción de jaula de ardilla. Es evidente que internamente estos fallos solo se dan en los componentes del motor por los que circulan corrientes eléctricas.

En consecuencia, los fallos de tipo eléctrico se originan internamente en el estator o en el rotor. Estos fallos pueden producirse debido a la construcción del motor con materiales defectuosos o de mala calidad, por un diseño deficiente, por el desgaste habitual del propio uso o por elementos externos que fuerzan los límites eléctricos soportados por el motor; como por ejemplo una sobretensión en el bobinado. Puede ocurrir también que problemas de tipo mecánico ocasionen fallos de tipo eléctrico, como puede ser una sobrecarga mecánica en el eje del rotor.

2.2.1.1 Fallos eléctricos en el estator del motor.

En la Figura 2.2 se muestra una fotografía de la parte frontal del estator de un motor de inducción de jaula de ardilla. En esta imagen se puede observar las bobinas de las fases donde se pueden dar alguno de los diversos tipos de fallo eléctrico.

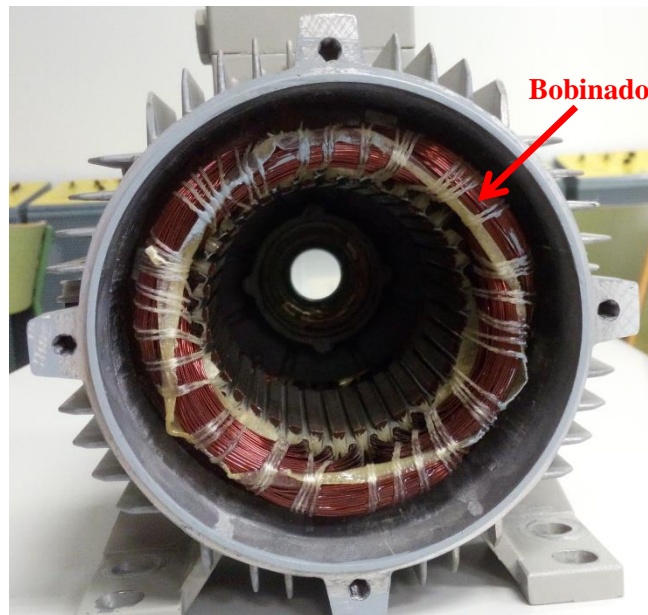


Figura 2.2 Imagen del frontal del estator de un motor de inducción.

A continuación se exponen los fallos que pueden generarse dentro del estator. Estos fallos pueden producirse también en los motores de rotor bobinado, tanto en rotor como en estator. Todos estos fallos se derivan de un defecto en el aislamiento del cableado de la bobina.

- **Cortocircuito entre bobinas de la misma fase.**
- **Cortocircuito entre espiras contiguas.**
- **Cortocircuito entre fases contiguas.**
- **Cortocircuito entre una fase y la masa del motor.**
- **Devanado con aislamiento defectuoso.**

2.2.1.2 Fallos eléctricos en el rotor del motor.

En la siguiente Figura 2.3 se muestra una fotografía de perfil del rotor de un motor de inducción de jaula de ardilla. En esta imagen se muestran las barras y los anillos del rotor donde en caso de rotura de una de estas se genera un fallo de tipo eléctrico.

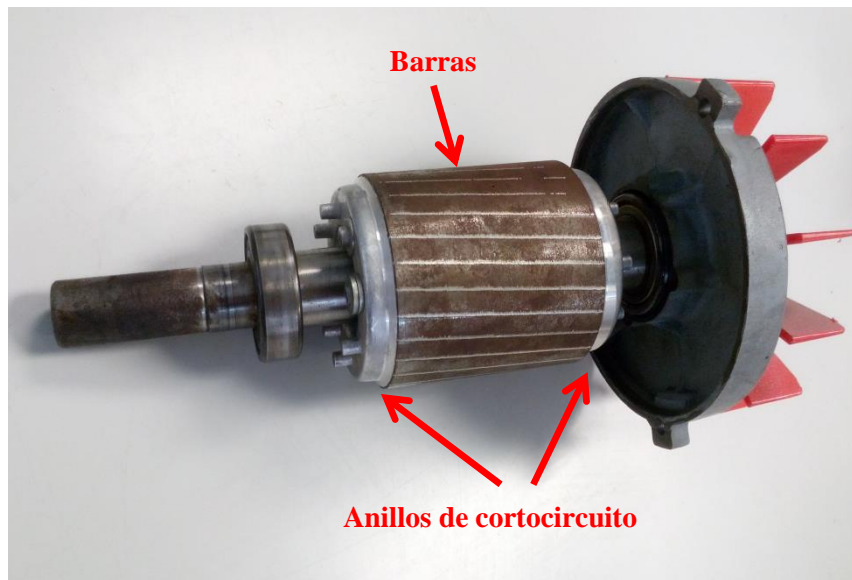


Figura 2.3 Imagen de perfil del rotor de un motor de inducción.

Dentro del rotor de un motor de inducción de jaula de ardilla la corriente circula solo por la estructura con forma de jaula compuesta por las barras y los dos anillos, en consecuencia los fallos de carácter eléctricos se centran en la rotura de alguno de estos elementos:

- **Rotura de una o más barras.** Este es el tipo de rotura más común. La rotura suele producirse cerca de la unión de la barra con el anillo de cortocircuito.
- **Rotura de uno o los dos anillos de cortocircuito del rotor.**

2.2.1.3 Fallos eléctricos de tipo externo.

Como ya se ha comentado anteriormente los fallos eléctricos externos son provocados por equipos mecánicos accionados por el motor, el acoplamiento mecánico entre carga y motor (engranajes, correas, etc.) o los sistemas de alimentación eléctrica del motor. Aunque en ocasiones se trate de un fallo mecánico, éste puede dar lugar a un mal funcionamiento del sistema electromagnético del motor.

En este caso hay dos situaciones capaces de crear daños en el motor de inducción mediante fallos de tipo eléctrico:

- **Mal funcionamiento del sistema de alimentación o de regulación,** como pueden ser sobretensiones.
- **Efectos de enmascaramiento de los armónicos generados por la variación de la carga.** Aunque este punto no es un fallo en sí, el cambio del par de carga genera una serie de armónicos en las corrientes de fase del estator del motor que pueden llegar a enmascarar los armónicos derivados de fallos de tipo eléctrico, dificultando la detección correcta de estos últimos en el espectro.

2.2.2 Fallos de tipo mecánico.

A diferencia de los fallos de tipo eléctrico, los fallos mecánicos pueden producirse en cualquiera de las piezas que componen el motor. No obstante, la mayoría de fallos se centran en las partes móviles como por ejemplo el rotor u otros elementos del eje de rotación.

2.2.2.1 Fallo mecánico por desequilibrio.

Como se comenta en [7] y [8], cualquier tipo de maquinaria rotativa se considera en equilibrio si todas las fuerzas internas o externas del rotor son constantes tanto en su magnitud como en la dirección y el sentido de dichas fuerzas. Sin embargo, dado que el proceso de fabricación de motores no es ideal, siempre existe un desequilibrio de la masa del rotor, por muy pequeño que sea este desequilibrio.

En el caso de que estas fuerzas no estén equilibradas, se pueden dar dos tipos de desequilibrios en el rotor del motor:

- **Desequilibrio estático.** Este tipo de desequilibrio se corresponde con una distribución de masas irregular en el rotor respecto del eje de rotación. En la representación de la Figura 2.4 **a** y **b**, se puede comprobar la diferencia entre equilibrio y desequilibrio respectivamente, donde la distribución de masas se representa con los cambios de color.
- **Desequilibrio dinámico.** Este tipo de desequilibrio se da cuando la distribución de las masas a lo largo del eje de rotación no está distribuida homogéneamente. En la Figura 2.4 **(c)** y **(d)** mostrada a continuación, se expone un ejemplo de la distribución de las masas en equilibrio **(c)** y en desequilibrio dinámico **(d)**.

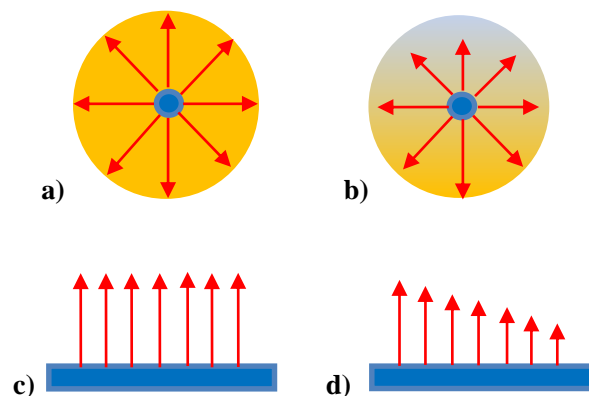
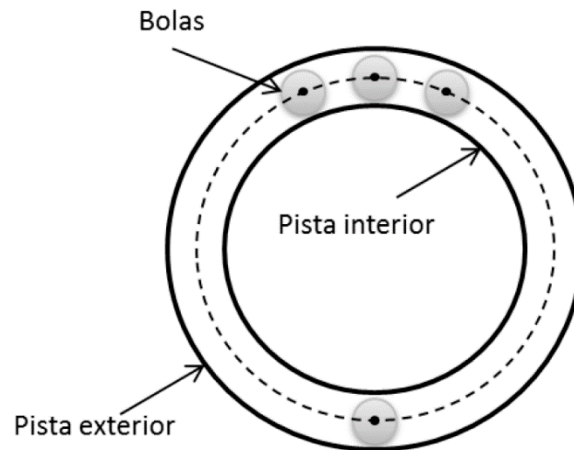


Figura 2.4 En esta imagen se presentan un ejemplo de distribuciones de masa en el rotor de un motor, teniendo en cuenta que cada valor de masa está representado por un color distinto. Cada flecha roja se corresponde con el vector de fuerza. a) Distribución homogénea respecto del vector de rotación del rotor. b) Con distribución de masa no homogénea respecto del vector de rotación del rotor. c) Distribución homogénea a lo largo del eje de rotación. d) Distribución no homogénea conforme a la longitudinal del eje de rotación.

2.2.2.2 Fallo mecánico en los cojinetes.

Los rodamientos o cojinetes se utilizan para minimizar el rozamiento entre la parte móvil (rotor) del motor de inducción y su parte fija (estator).

El sistema de cojinete habitualmente está compuesto por dos pistas (una exterior y una interior) y un conjunto de bolas que separan las dos pistas (Figura 2.5). Estas bolas intermedias con libertad de movimiento rotacional son las encargadas de atenuar el rozamiento y la fricción entre los elementos que están soportados por cada una de las dos pistas.



Fuente: Sapena (2014) [9]

Figura 2.5 Representación de los elementos de un cojinete de bolas.

En general, en los cojinetes del motor se registra la mayoría de posibles fallos de un motor de inducción (Tabla 2.1 y Tabla 2.2). Además del rozamiento y de la fricción generada, el cojinete también absorbe todas las vibraciones producidas por el motor.

2.2.2.3 Fallo mecánico de excentricidad entre el rotor y el estator.

Se considera como excentricidad las desviaciones respecto de la distancia máxima de separación (entrehierro) entre el estator y el rotor para cualquiera de los ángulos de rotación.

Aunque en un motor ideal la distancia de separación entre el rotor y el estator es uniforme para distintos ángulos de rotación, en la realidad bien por fallos del motor o por defectos de funcionamiento o de fabricación nunca se llega a dar esa uniformidad.

En los motores de inducción de tamaño pequeño o mediano el entrehierro que separa el rotor y el estator es tan pequeño (puede ser inferior a 1mm) que, en el caso en que exista un mínimo desalineamiento entre el rotor y el estator es casi inevitable que se produzca algún que otro roce durante el funcionamiento del motor.

Existen los siguientes tipos de excentricidad dependiendo de la deformación o desajuste entre el rotor y el estator:

- **Excentricidad estática.** Este fallo se corresponde a una distorsión del entrehierro causada por un cierto grado de deformación de la circularidad del esta-

tor o por una desviación de la posición central de giro del rotor respecto a su posición idónea en el centro del estator. En este caso la distancia entre las superficies cilíndricas interior del estator y exterior del rotor (obviando las ranuras) no son equidistantes en todos los puntos, pero la posición de giro del rotor en el espacio se mantiene en su eje central por lo tanto se mantiene a lo largo del tiempo la misma distancia no equidistante entre la zona del estator y el rotor.

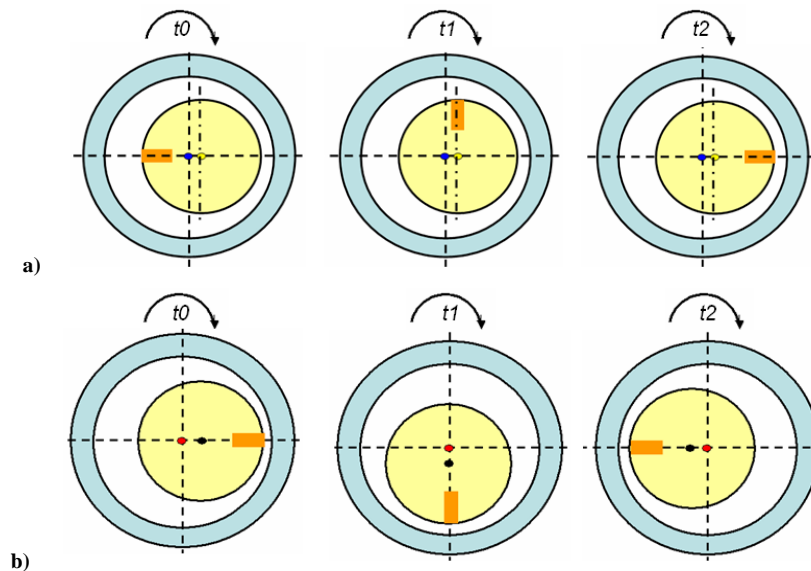
- **Excentricidad dinámica.** Este tipo de fallo se produce al existir un cierto grado de deformación de la circularidad del rotor o al producirse un desplazamiento el centro de giro del rotor. En cualquiera de estos casos, además de que la distancia no es equidistante entre el rotor y el estator, también varía en el tiempo la distancia que separa la superficie del estator con la del rotor en cada punto.
- **Excentricidad Mixta.** Este fallo se corresponde a la aparición tanto de excentricidad estática como excentricidad dinámica al mismo tiempo en el motor.

En la Figura 2.6 se muestra un ejemplo de ciclo de giro entre el rotor y el estator para tres cuadrantes del giro contemplando el fallo de excentricidad estática (**a**) y excentricidad dinámica (**b**). Estas imágenes han sido obtenidas del trabajo de Puche [6].

2.2.3 Frecuencias características producidas por los fallos.

Los tipos de fallos que afectan al estator o al rotor del motor de inducción durante su funcionamiento generan o amplifican una serie de armónicos característicos cuyas frecuencias pueden ser detectadas en las corrientes de fase que circulan por el estator.

Como se puede deducir del esquema de la anterior Figura 2.1, tanto los fallos eléctricos como mecánicos que se muestran en el esquema afectan al estator, al rotor o a ambos. En el caso de los fallos externos los efectos producidos por la carga y los desajustes mecánicos producidos por una desalineación o engranajes defectuosos afectan directamente al rotor y en consecuencia afectan indirectamente al estator.



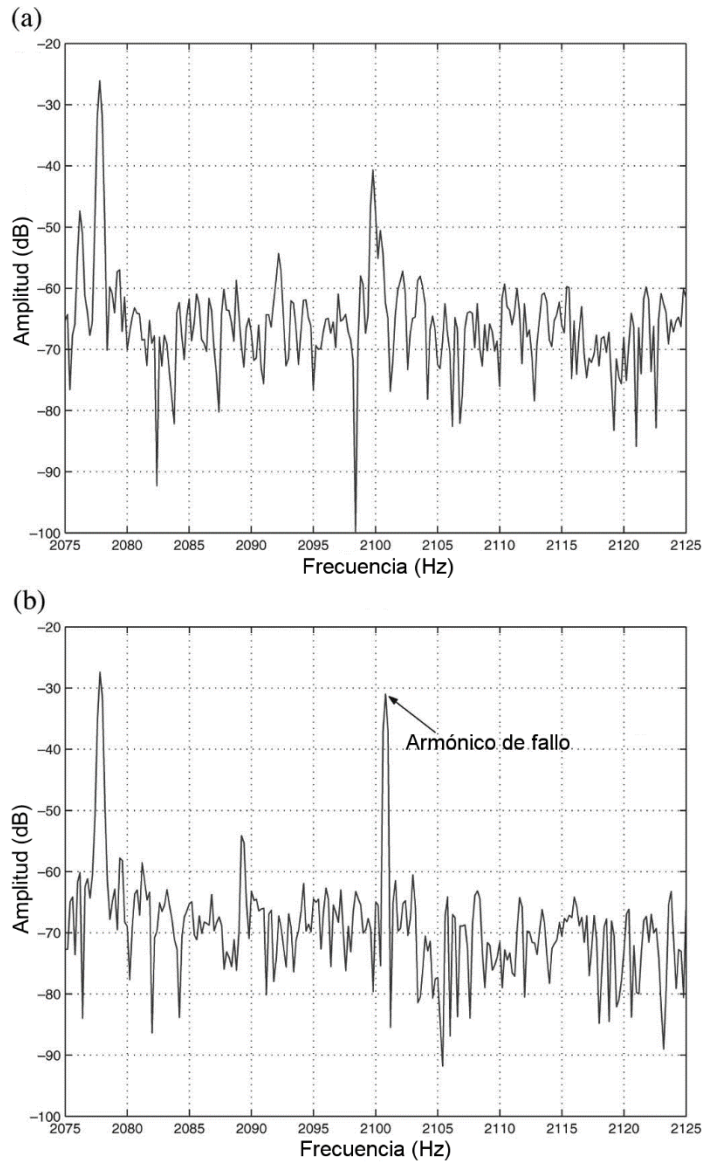
Fuente: Puche (2008) [6]

Figura 2.6 Representación esquemática en tres cuadrantes de la trayectoria de giro de un motor con fallo de excentricidad estática (a) y con fallo de excentricidad dinámica (b). Tal como puede observarse, en el caso de excentricidad estática la posición de mínimo entrehierro se mantiene en posición fija, mientras que en el caso de excentricidad dinámica ésta va cambiando con la rotación del eje.

Las componentes frecuenciales características de los diversos tipos de fallos es un tema estudiado por muchas investigaciones relacionadas con el diagnóstico de motores por medio de las corrientes de fase estáticas.

El análisis del espectro de una corriente de fase de un motor con fallos mostrará una amplitud de la frecuencia de fallo característica dentro del dominio de frecuencias de la señal de fase generado a partir de una función transformada (transformada rápida de Fourier (FFT), Welch,...). Un ejemplo de este pico de fallo se muestra en la Figura 2.7 (b). Esta posición en frecuencia se conoce de forma teórica por la naturaleza del fallo que genera la componente armónica.

Hay que tener en cuenta que también pueden aparecer componentes frecuenciales en las frecuencias de fallo en motores sanos debido a las imprecisiones en la construcción del propio motor, sin embargo estas componentes serán mucho más pequeñas que en el caso de un motor con dicho fallo manifiesto.



Fuente: Jung et al (2006)[9]

Figura 2.7 Sección del espectro donde se muestra la frecuencia característica de fallo de cojinete de una corriente de fase de un motor: (a) sano, (b) con el cojinete dañado.

Como ejemplo visual de la diferencia de amplitud entre la frecuencia característica de un motor con fallo y un motor sin fallo se muestra en la Figura 2.7 extraído del artículo de Jung [7].

En esta Figura 2.7 se muestran los espectros de una de las fases de corriente de un motor sano y un motor con fallo del cojinete. Como se puede observar claramente, el espectro correspondiente al motor con un fallo en el cojinete Figura 2.7 (b) muestra una mayor amplitud en la frecuencia característica de fallo de cojinete, fruto del conjunto de armónicos generados a causa del cojinete defectuoso.

En consecuencia, muchas de las técnicas de diagnóstico de fallos en motores de inducción obtienen la información relevante desde el espectro generado a partir de una o más corrientes de fase del estator.

A continuación se muestran las ecuaciones correspondientes a varios tipos de fallos de los que se puede calcular la frecuencia de fallo de donde obtener la amplitud que caracteriza la presencia o ausencia de dicho fallo (ejemplo en la anterior Figura 2.7).

2.2.3.1 Fallos eléctricos.

Las siguientes ecuaciones definen las posiciones de frecuencias para los fallos de tipo eléctrico comentados en el apartado 2. Estas ecuaciones pueden ser consultadas con mayor detalle en las publicaciones referenciadas en [2], [10]-[14].

Dentro de estas ecuaciones, la variable p corresponde al número de pares de polos, la variable f es la frecuencia fundamental, la variable s corresponde al deslizamiento y k y n son dos secuencias de valores constantes y enteros positivos.

- **Fallos en el aislamiento de los devanados [2].** En la ecuación 2.1 se muestra la fórmula correspondiente para calcular la componente frecuencial de fallo por cortocircuito entre espiras por causa de un devanado defectuoso.

$$f_{CCespiras} = \left[k + \frac{n}{p} \cdot (1 - s) \right] \cdot f \quad \begin{array}{l} k = 1, 3, 5, \dots \\ n = \pm[1, 2, 3, \dots, (2p-1)] \end{array} \quad (2.1)$$

- **Rotura de los anillos de cortocircuito o de las barras del rotor [10]-[13].** En la ecuación 2.2 se muestra la fórmula que permite conocer la componente frecuencial de indicación de fallo en el caso de que alguno de los anillos de cortocircuito o alguna de las barras del rotor sufra algún daño.

$$f_{BarraRota} = \left[1 + 2(k \cdot s) \right] \cdot f \quad k = \pm(1, 2, 3, 4, 5, \dots) \quad (2.2)$$

- **Efectos provocados por el par de la carga** [14]. En la ecuación 2.3 se define las componentes frecuenciales generadas por la carga al variar la posición del rotor.

$$f_{ParCarga} = \left[1 + \frac{k}{p} \cdot (1 - s) \right] \cdot f \quad k = \pm(1, 2, 3, 4, 5, \dots) \quad (2.3)$$

2.2.3.2 Fallos mecánicos.

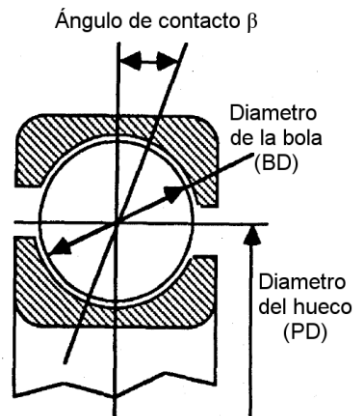
Las ecuaciones que se presentan a continuación calculan las posiciones frecuenciales de fallo para los fallos provocados en el motor de inducción por causas mecánicas, los cuales han sido analizados en el apartado 2.2.2. Estas ecuaciones son desarrolladas con mayor detalle en las publicaciones enunciadas junto a su respectivo fallo.

En el caso de estas ecuaciones:

- f corresponde a la frecuencia fundamental.
- f_r representa la frecuencia mecánica de rotación del rotor.
- N_b se corresponde con el número de bolas que contiene el cojinete.
- d_B representa el diámetro de cada una de las bolas.
- d_C representa el diámetro medio del cojinete.
- β se corresponde con el ángulo de contacto del cojinete.
- k es una constante entera que abarca todos los enteros positivos.
- q es una constante entera que abarca todos los enteros positivos incluidos el cero.
- N representa el número de ranuras que tiene el estator del motor.
- p representa el número de pares de polos.
- s corresponde al deslizamiento.
- f_c representa a la familia de las frecuencias características dentro del espectro.
- v corresponde al orden de los armónicos generados por este fallo.
- k_p define el signo para calcular la frecuencia de fallo correspondiente a la pista interior (positivo) o a la pista exterior (negativo).

- **Fallos internos en cojinetes [15]-[17].** En la siguiente ecuación 2.4 se define las posiciones frecuenciales donde se observa la amplitud del fallo generado por algún daño en la pista interior o exterior del cojinete. En la Figura 2.8 se muestra un modelo de cojinete donde se representa el ángulo de contacto β , el diámetro de la bola (BD) y el diámetro del hueco (PD).

$$f_{FalloPista} = \left[\frac{N_b}{2} \right] \cdot \left[1 + k_p \frac{d_B \cdot \cos(\beta)}{d_C} \right] \cdot f_r \quad \begin{array}{l} k_p=+1. \text{Pista interior.} \\ k_p=-1. \text{Pista exterior.} \end{array} \quad (2.4)$$



Fuente: Schoen et al. (1995) [15]

Figura 2.8 Representación de un cojinete donde se indica el ángulo de contacto β , el diámetro de la bola (BD) y el diámetro de paso (PD).

- **Daños en las bolas (rodamientos) de los cojinetes [15]-[17].** En la ecuación 2.5 se muestra la ecuación que calcula la componente frecuenciale generada al desarrollarse algún tipo de daño en las bolas de rodamiento que contiene el cojinete

$$f_{Bolas} = \left[\frac{d_C}{2d_B} \right] \cdot \left[1 - \left(\frac{d_B \cdot \cos(\beta)}{d_C} \right)^2 \right] \cdot f_r \quad (2.5)$$

- **Fallo de excentricidad estática [18]-[21].** El fallo de excentricidad estática, comentado en el apartado 2.2.2.3, genera las componentes frecuenciales que se muestran en la frecuencia calculada mediante la ecuación 2.6.

$$f_{ExcentricidadEstática} = \left[\left(\frac{q \cdot N}{p} \cdot (1-s) \right) \pm v \right] \cdot f \quad \begin{array}{l} q = 0, 1, 2, 3, \dots \\ v = 1, 2, 3, 4, \dots \end{array} \quad (2.6)$$

- **Fallo de excentricidad dinámica [17]-[19].** El fallo de excentricidad dinámica, visto también en el apartado 2.2.2.3, genera las componentes frecuenciales cuya posición frecuencial se puede calcular con la ecuación 2.7.

$$f_{ExcentricidadDinámica} = \left[\left(\frac{(q \cdot N) \pm f_c}{p} \cdot (1-s) \right) \pm v \right] \cdot f \quad \begin{array}{l} q = 0, 1, 2, 3, \dots \\ v = 1, 2, 3, 4, \dots \end{array} \quad (2.7)$$

- **Fallo de excentricidad mixta [18]-[20], [22]-[23].** Por supuesto, es factible que el fallo de excentricidad sea mixto (apartado 2.2.2.3). En este caso, además de las frecuencias características que se generan a causa del fallo de excentricidad estática (2.6) y dinámica (2.7), aparece una nueva familia de frecuencias identificativas de un fallo por excentricidad mixta (2.8).

$$f_{ExcentricidadMixta} = |f \pm k \cdot f_r| \quad k = 1, 2, 3, \dots \quad (2.8)$$

- **Fallo de excentricidad axial [24].** Es común que los diversos tipos de excentricidad anteriores sean generados a su vez por una excentricidad dada a lo largo del eje del rotor de forma axial. Una excentricidad de tipo axial genera a su vez una excentricidad estática, dinámica o las dos al mismo tiempo. Por tanto, a priori este tipo de fallo no puede ser detectado por una familia de frecuencias de fallo diferentes a las frecuencias derivadas de los otros tipos de excentricidad del eje, por lo que solo se puede discernir de forma hipotética la excentricidad axial cuando se muestra una frecuencia característica de fallo dada por otro tipo de excentricidad definidas en las formulas (2.6), (2.7) y (2.8).

Aunque existen otros tipos de fallos, dada la naturaleza de algunos de estos fallos no se puede desarrollar en todos estos una formulación específica capaz de indicar una familia de frecuencias características bien definidas que representen el fallo en cuestión.

En resumen, en este apartado 2.2.3 se han explicado algunos de los fallos y perturbaciones que pueden afectar al funcionamiento de un motor de inducción, y que generan nuevas componentes frecuenciales sobre la onda original. Para el caso que nos incumbe en esta tesis, la detección de una u otra componente frecuencial característica de fallo va a permitir la detección de su fallo asociado.

En la Tabla 2.2 se recopilan todo el conjunto de componentes frecuenciales generadas tanto por los fallos de tipo eléctrico como por los fallos de tipo mecánico comentados en el apartado 2.2.3 de la tesis.

Tabla 2.2 Componentes frecuenciales generadas según el fallo.

Fallo	Componente frecuencial	variables
Cortocircuito en el devanado	$f_{CCespiras} = \left[k + \frac{n}{p} \cdot (1-s) \right] \cdot f$	$k = 1, 3, 5, \dots$ $n = \pm[1, 2, 3, \dots, (2p-1)]$
Barras rotas en el rotor	$f_{BarraRota} = [1 + 2(k \cdot s)] \cdot f$	$k = \pm(1, 2, 3, 4, 5, \dots)$
Excentricidad estática	$f_{ExcentricidadEstática} = \left[\left(\frac{q \cdot N}{p} \cdot (1-s) \right) \pm v \right] \cdot f$	$q = 0, 1, 2, 3, \dots$ $v = 1, 2, 3, 4, \dots$
Excentricidad dinámica	$f_{ExcentricidadDinámica} = \left[\left(\frac{(q \cdot N) \pm f_c}{p} \cdot (1-s) \right) \pm v \right] \cdot f$	$q = 0, 1, 2, 3, \dots$ $v = 1, 2, 3, 4, \dots$
Excentricidad mixta	$f_{ExcentricidadMixta} = f \pm k \cdot f_r $	$k = 1, 2, 3, 4, 5, \dots$
Pista del cojinete dañada	$f_{FalloPista} = \left[\frac{N_b}{2} \right] \cdot \left[1 + k_p \frac{d_B \cdot \cos(\beta)}{d_C} \right] \cdot f_r$	$k_p = +1$. Pista interior. $k_p = -1$. Pista exterior.
Rodamientos del cojinete dañados	$f_{Bolas} = \left[\frac{d_C}{2d_B} \right] \cdot \left[1 - \left(\frac{d_B \cdot \cos(\beta)}{d_C} \right)^2 \right] \cdot f_r$	
Efectos del par de carga	$f_{ParCarga} = \left[1 + \frac{k}{p} \cdot (1-s) \right] \cdot f$	$k = \pm(1, 2, 3, 4, 5, \dots)$

2.3 Técnicas de detección de fallos por medio de la corriente de fase.

Gracias a todas las investigaciones que se han realizado hasta ahora, existe un amplio conjunto de técnicas de todo tipo capaces de detectar con cierta precisión uno o varios fallos que se pueden dar dentro de un motor de inducción (apartado 2.2).

Cada una de estas técnicas utiliza al menos alguna magnitud obtenida desde el motor de inducción. Con la correspondiente técnica se puede discernir sobre el fallo en estudio. Los datos medidos de estas magnitudes son transformados y analizados para detectar posibles características que representan al fallo en cuestión. Las principales magnitudes de las que se puede extraer información relevante sobre alguna clase de fallo son:

- Vibraciones del motor.
- Acústica del motor.
- Generación térmica del motor.
- Campo magnético generado por el motor.
- Corrientes de las fases que recorren el estator del motor.
- Potencia del motor.

Actualmente las últimas líneas de investigación se han centrado prácticamente en analizar los datos obtenidos a partir de corrientes de las fases del estator [25]. Esta preferencia por la magnitud de corriente de fase se ha dado por varios factores, entre los que cabe destacar el amplio número de clases distintas de fallos de los que se puede obtener información relevante y el bajo coste que tienen los sensores utilizados para la lectura de datos de estas corrientes de fase. El coste de los sensores es tan bajo que muchas veces ya van instalados de serie dentro del motor.

Esta tesis se centra únicamente en la optimización de los sistemas de diagnóstico mediante sistemas inteligentes para aquellas técnicas de detección de fallos que requieren de la magnitud de corriente de fase, ya que estas son las más destacables actualmente para el diagnóstico de motores. Sin embargo, la clase de optimización que se desarrolla en la tesis puede ser en un futuro extrapolada a otras técnicas de diagnóstico que utilicen otras magnitudes disponibles.

Aunque todas las técnicas de detección de fallos mediante corriente de fase se apoyan en el mismo tipo de magnitud, no todas esas técnicas se pueden aplicar a todas las situaciones posibles. No obstante, estas técnicas se pueden agrupar en varios subconjuntos dependiendo de la operativa del motor.

Como se puede observar en los siguientes apartados 2.3.1 y 2.3.2, la mayoría de las técnicas que se muestran están estrechamente relacionadas entre ellas. Esto se debe a

que algunas de estas técnicas son derivaciones desarrolladas a posteriori, con algún tipo de variación relevante respecto a una técnica anterior.

2.3.1 Técnicas de detección de fallos en régimen estacionario.

Estas técnicas de detección de fallo solo pueden aplicarse cuando el régimen de funcionamiento del motor es estacionario durante todo el tiempo de realización del análisis. Es decir, el motor debe trabajar siempre a una velocidad constante, aunque se pueden dar pulsos de carga en momentos puntuales. Esto se debe a que la característica relevante que determina la aparición de fallo es muy dependiente de la velocidad.

En muchas de las técnicas de detección de fallos para régimen estacionario, además de requerir de una muestra obtenida a partir de una o varias corrientes de fase de donde extraer las características que representan al fallo, también es necesario conocer la velocidad de rotación del motor por lo que se tiene que medir esta por medio de un sensor externo de medición de la velocidad.

Estas técnicas tienen la importante desventaja de necesitar de un régimen de trabajo estacionario, el cual en la práctica es muy difícil de mantener. Con estas técnicas se hace imposible el diagnóstico de fallos en situaciones donde los motores necesitan operar en otros tipos de regímenes de trabajo.

Su principal ventaja deriva en que estas técnicas demandan de un menor coste computacional respecto a otros tipos de técnicas, lo que implica la posibilidad de implementación en hardware menos especializado y de menor precio. En consecuencia, la implementación de estas técnicas tiene preferencia cuando se trata de sistemas de diagnóstico empotrados para diagnóstico online, siempre que se cumpla el requisito de estabilización de la velocidad.

A continuación se exponen las técnicas de detección de fallos para régimen permanente más notorias.

2.3.1.1 Análisis de la corriente estatórica mediante transformada FFT.

Esta técnica de detección de fallos se basa en el estudio de las amplitudes representativas de fallo que surgen cuando se transforma una de las corrientes del estator en el dominio del tiempo al dominio de frecuencias. A la transformada en frecuencia para el análisis se la conoce por sus siglas FFT (Fast Fourier Transform).

Aunque esta técnica se expuso por primera vez por Williamsom (1982)[26], todavía se utiliza en muchos sistemas de diagnóstico.

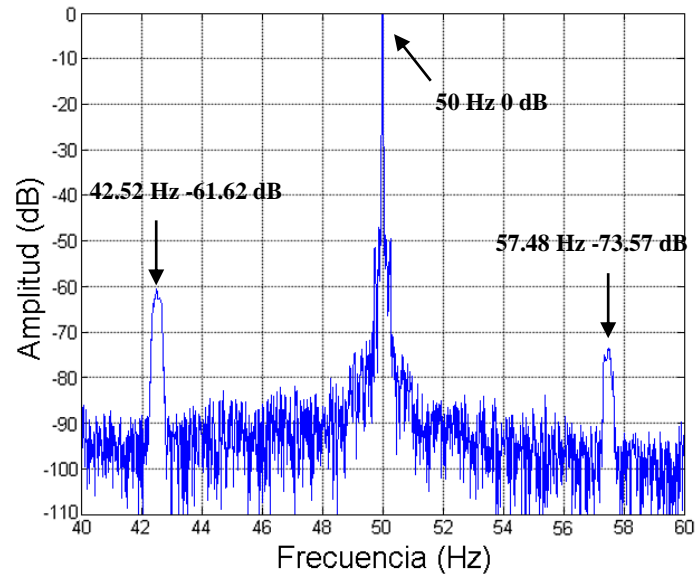


Figura 2.9 FFT de un motor sano trabajando a una frecuencia de alimentación de 50Hz.

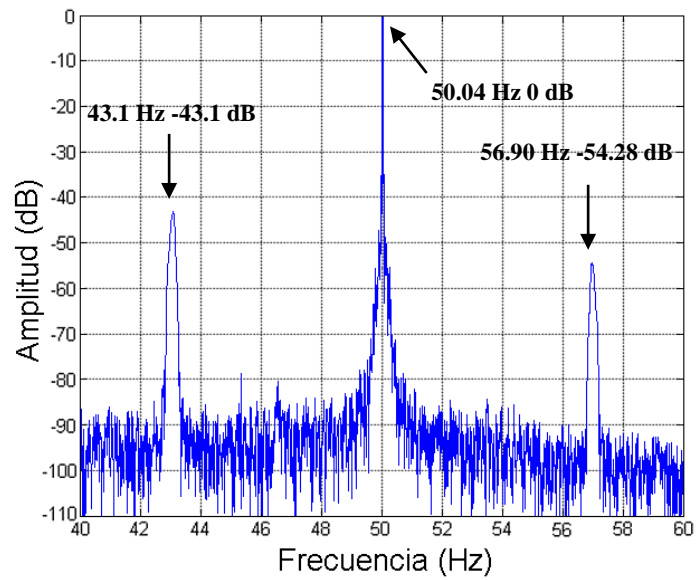


Figura 2.10 FFT de un motor con barra rota trabajando a una frecuencia de alimentación de 50Hz.

Por ejemplo, para un caso en que existen roturas de barras del rotor, donde se rompe al menos una de estas barras, se genera dentro del estator corrientes en varias bandas de frecuencia distintas (ecuación (2.2)) ya que la corriente que idealmente es senoidal a la frecuencia fundamental deja de serlo. De esta forma, la FFT de un motor sano mostrará unos picos poco relevante a las frecuencias características de este tipo de fallo (Figura 2.9), mientras que en un motor con fallo por barra rota estos picos son mucho más destacados (Figura 2.10).

En consecuencia, el diagnóstico de fallos mediante la transformada FFT solo es aplicable si se conocen los parámetros del motor y se tiene una la velocidad mecánica de giro constante, es decir para régimen estacionario.

Como se observa en la Figura 2.11, se genera un grupo de componentes armónicas de orden k relacionadas con el fallo de barras rotas en este caso. Estas componentes armónicas se generan a una misma distancia entre ellas. Conforme aumenta el valor del k -orden para $k > 1$ o disminuye el valor del k -orden para $k < -1$ la amplitud de estas frecuencias características de fallo se va reduciendo.

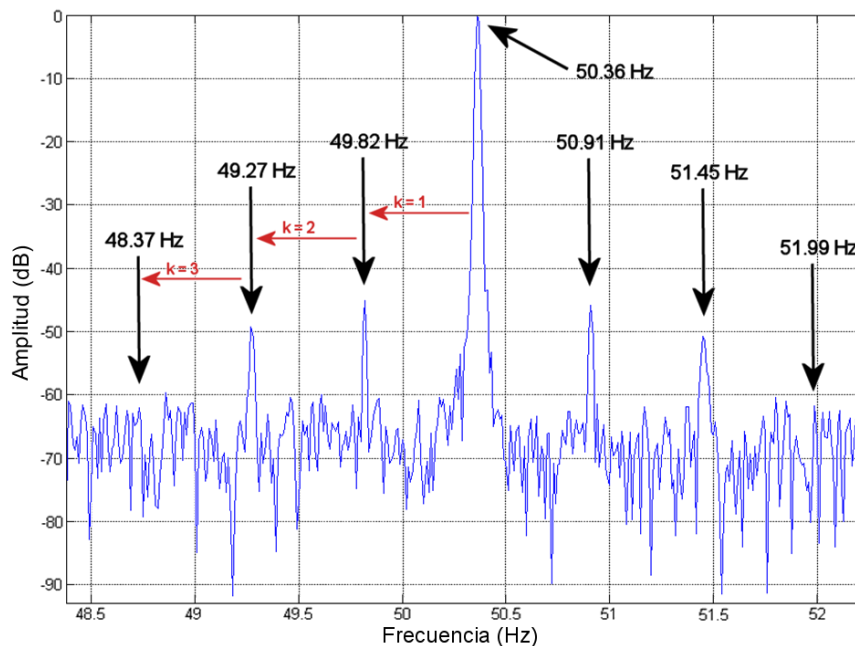


Figura 2.11 Ejemplo de aparición de la frecuencia de fallo característica a distintas distancias respecto del valor de k .

Por tanto, para esta técnica por FFT se suelen descartar todas las repeticiones de k mayores a 1 o menores a -1. Esto es debido a que los picos característicos más relevantes se encuentran en las frecuencias con $k = [1, -1]$.

Uno de los problemas más importantes en este tipo de fallo se da cuando el deslizamiento es tan bajo que la frecuencia característica puede verse enmascarada por la frecuencia fundamental, sobre todo si existe fuga espectral.

La fuga espectral (leakage) se da cuando la frecuencia de control del motor es tan ancha que acaba enmascarando las componentes frecuenciales de fallo más significativas si el deslizamiento es muy bajo. Este efecto no deseado dificulta la detección correcta de fallos.

En el apartado 2.3.3.1 se muestra un ejemplo de minimización de la fuga espectral mediante una ventana de filtro de Hann. En este ejemplo aparece una fuga espectral bastante acusada (Figura 2.48). Sin embargo, en la Figura 2.49 de este mismo apartado se muestra como se ha minimizado enormemente la fuga espectral utilizando la ventana de filtro de Hann.

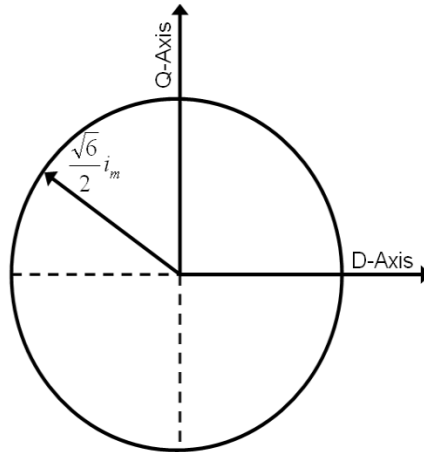
2.3.1.2 Vector de Park (PVA).

La siguiente técnica Vector de Park, denominada en ingles Park Vector Approach (PVA), necesita las tres fases de corrientes del motor para poder realizar el diagnóstico de fallo. Esta técnica se expone con más detalle dentro del artículo de Cruz [27], en la tesis doctoral de Díaz Sánchez [28] y en mi artículo [29].

Su función es la de crear el vector espacial de rotación de las tres corrientes de fase (i_u , i_v , i_w) a partir de la transformada de Park. Con esta transformada las componentes de dicho vector (eje directo D y en cuadratura Q) para cada dimensión del espacio (representación en imagen 2.12) se calculan como.

$$i_D = \sqrt{\frac{2}{3}}i_u - \frac{1}{\sqrt{6}}i_v - \frac{1}{\sqrt{6}}i_w \quad (2.9)$$

$$i_Q = \frac{1}{\sqrt{2}}i_v - \frac{1}{\sqrt{2}}i_w \quad (2.10)$$

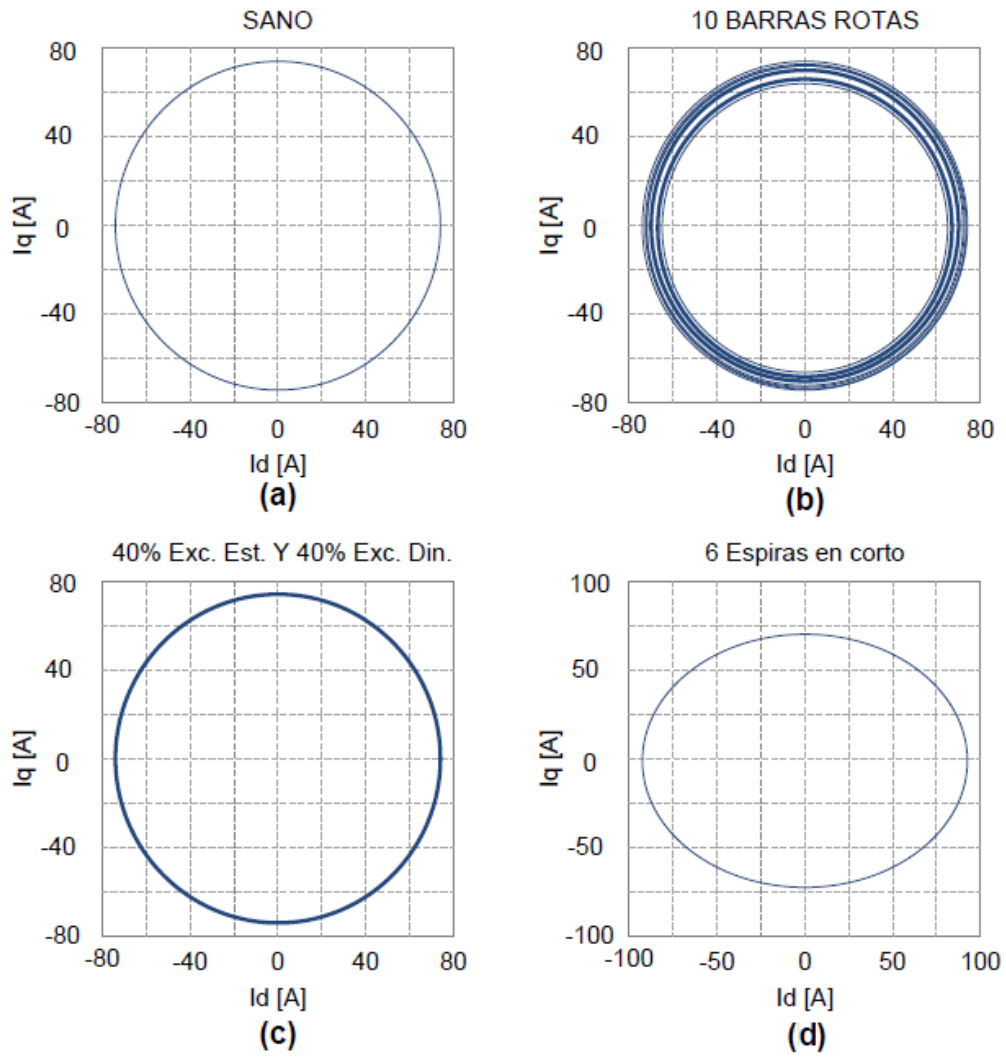


Fuente: Puche (2008) [6]

Figura 2.12 Representación del vector de Park en una situación ideal.

Según la teoría expuesta, se puede comprobar la existencia de barras rotas utilizando un modelo de un motor ideal sano y de un motor con fallo. La diferencia que caracteriza el fallo de barras rotas se observa en el incremento del tamaño del borde que muestra la circunferencia generada por el vector. En la Figura 2.13 se puede observar la representación del vector de un motor idealizado (simulación de un motor) sano, con barras rotas, con excentricidad y con espiras en cortocircuito.

En esta Figura 2.13 se comparan los vectores de Park generados a partir de las corrientes de fase de este un motor simulado sano (a), de un motor simulado con 10 barras rotas (b), de un motor simulado con excentricidad dinámica y estática (c), y de un motor simulado con espiras cortocircuitadas.



Fuente: Díaz-Sánchez (2011) [28]

Figura 2.13 Vector de las corrientes de un motor ideal (motor simulado): (a) Sano. (b) Con 10 barras rotas. (c) Con un 40% de excentricidad estática y 40% de excentricidad dinámica. (d) Con 6 espiras en corto.

Como se observa en la Figura 2.14, existe una componente continua (compuesta por la frecuencia fundamental) y una componente alterna (que refleja las componentes frecuenciales de fallo).

Si el motor está sano el grosor del vector de giro será mínimo, mientras que si aparece una componente alterna en la amplitud del módulo de dicho vector (generada por el fallo de barra rota) añadida a la componente continua del vector de Park, el grosor de la circunferencia resultante de este vector será mayor a causa de la aparición del vector de la componente alterna.

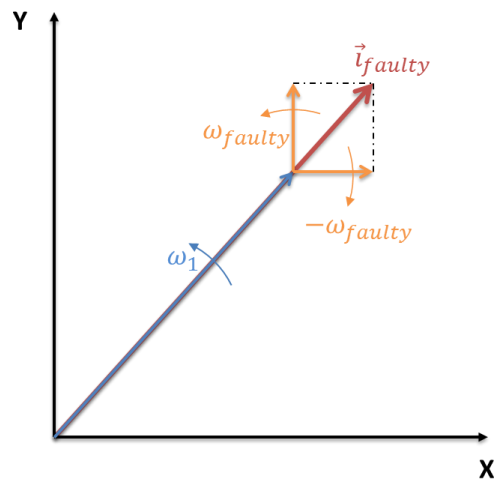


Figura 2.14 Vector de giro de la componente continua (en azul) y de la componente alterna (en amarillo) del vector de Park..

Esta técnica tiene un inconveniente bastante importante. Aunque en el modelo teórico (simulación) la diferenciación y detección de fallos es muy buena al estudiarse las frecuencias implicadas, en un caso real existe un mayor número de componentes no reflejadas en las simulaciones que pueden enmascarar los resultados y dificultan la detección de los fallos.

Un ejemplo descriptivo de este problema se puede observar en la Figura 2.15 donde se representa el vector obtenido de las corrientes de un motor real sano con control electrónico, y en la Figura 2.16 donde se muestra el vector resultante de las corrientes de un motor real con fallo de rotura de barras en el rotor.

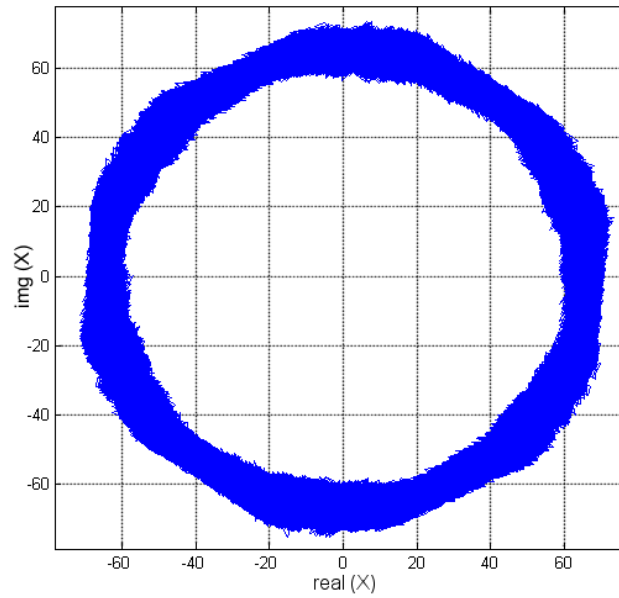


Figura 2.15 Vector de Park de las corrientes de fase de un motor sano con una frecuencia de alimentación de 50Hz y con control electrónico.

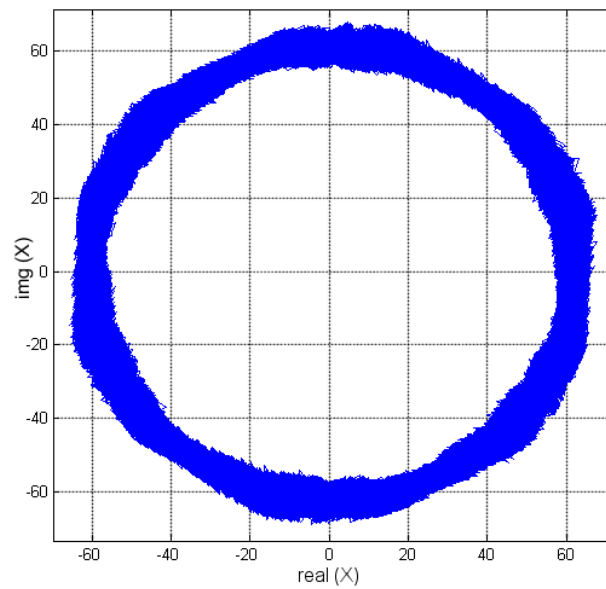


Figura 2.16 Vector de Park de las corrientes de fase de un motor con barras rotas en el rotor, con una frecuencia de alimentación de 50Hz y con control electrónico.

2.3.1.3 Vector de Park Extendido (EPVA).

Esta técnica es una extensión de la técnica PVA comentado en el apartado anterior 2.3.1.2. Esta técnica es conocida en inglés como Extended Park Vector Approach (EPVA) y se utiliza para la obtención de características relevantes de fallo.

Al igual que con la técnica PVA, la técnica EPVA necesita las tres fases de corriente para obtener la característica representativa de fallo.

En este caso se aplica la transformada FFT sobre el vector PVA para transformar el vector al dominio de frecuencias. Además se elimina la frecuencia fundamental del espectro por lo que tampoco se da el problema de la fuga espectral.

Como se expone en los artículos de Zarei [30] y Cruz [31], la técnica del vector de Park traslada el sistema de referencia al espacio resultante del vector de giro de las corrientes de fase. En consecuencia, al obtener el espectro de frecuencias de este vector no se muestra la frecuencia fundamental ya que se elimina al restarle su valor medio, y la componente frecuencial característica de fallo se obtiene a la frecuencia de deslizamiento respecto a 0. Además, como se comenta en [30], la amplitud de las componentes características de fallo se ven multiplicadas por un factor de $3*f$, siendo f la frecuencia de alimentación.

En la Figura 2.17 se muestra una representación de la frecuencia de aparición de fallo para las corrientes de fase de un motor sano, mientras que en la Figura 2.18 se muestra la representación homónima para las corrientes de fase de un motor de las mismas características con fallo de barras rotas en el rotor. Directamente, se puede observar la presencia del fallo.

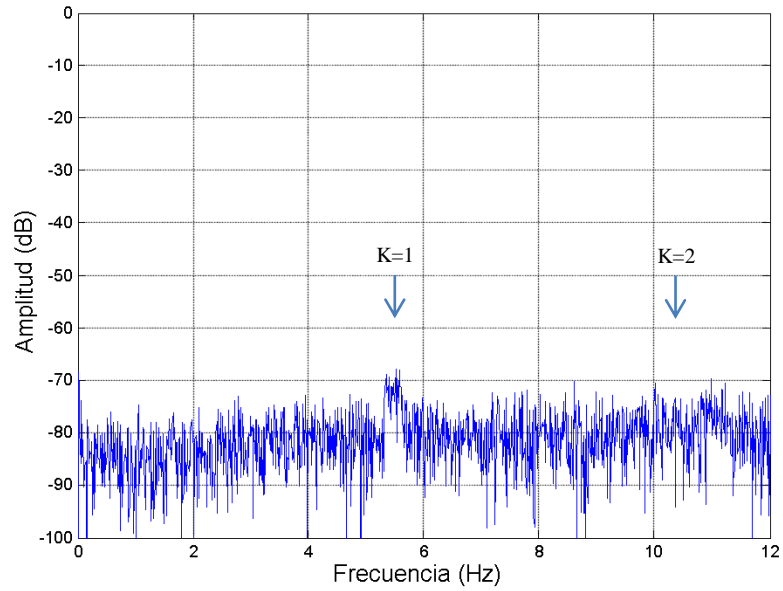


Figura 2.17 Vector de Park Extendido de las fases de corriente de un motor SANO trabajando a una frecuencia de alimentación de 50Hz.

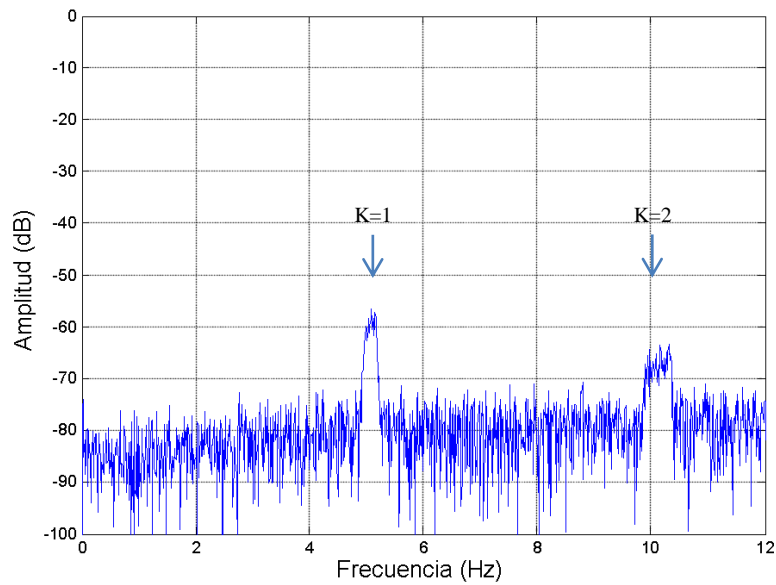


Figura 2.18 Vector de Park Extendido de las fases de corriente de un motor con FALLO DE BARRA ROTA trabajando a una frecuencia de alimentación de 50Hz.

2.3.1.4 Cepstrum.

Esta técnica de diagnóstico explora la característica de repetición existente en las frecuencias de fallo que aparecen en el espectro de las corrientes de fase del motor con fallo para cada valor de k (Figura 2.11). Esta técnica se expone, dentro de los artículos de El Hachemi [32] y Mehrjou [33], como alternativa de diagnóstico para diversos fallos cuando el fallo genera una familia de componentes armónicas.

Según el análisis realizado en los artículos mencionados previamente, esta técnica en teoría mejora la detección de fallos al realizar una doble transformación de Fourier donde la característica resultante de fallo no se ve afectada por el fenómeno de la fuga espectral. Además, a diferencia de otras técnicas en las que se obtiene un conjunto de componentes armónicas de fallo, en este caso el fallo se representa con un solo valor de amplitud dentro del tiempo.

No obstante, el número de artículos en el estado del arte que analizan esta técnica de diagnóstico de fallos es reducido. Es decir, al parecer no se observa un gran interés por el uso de esta técnica para el diagnóstico de fallos. Como se mostrará posteriormente en los resultados obtenidos con esta técnica de diagnóstico, tampoco existe una mejoría real de la técnica de diagnóstico Cepstrum con respecto a las otras técnicas de diagnóstico, mostrándose en este trabajo incluso como la peor técnica de diagnóstico de fallos.

Al realizarse una doble transformación de Fourier el vector de datos resultante se encuentra nuevamente dentro del dominio del tiempo. La posición en tiempo de la característica de fallo se encuentran a la frecuencia de repetición que se daba en el dominio de frecuencias para los k armónicos de fallo, por tanto al desarrollar una transformación FFT y una transformación inversa de la FFT se obtiene una amplitud de la onda en tiempo como características de fallo en la posición obtenida con la fórmula (2.12). Se puede observar esta distancia periódica indicada con una flecha roja para cada k en la anterior Figura 2.11.

De aquí el nombre de dicha técnica, donde la primera parte “ceps” de la palabra “cepstrum” es una inversión de las letras “spec” de “spectrum”, dándolo a entender como su inverso. Es decir, se realiza una FFT^{-1} al FFT adaptado de la muestra.

En la fórmula (2.11) se describe esta doble transformación de frecuencias a partir de la corriente de fase del motor. Como se observa en esta fórmula, no se realiza simplemente una doble conversión de la corriente de fase entre dominios, sino que, antes de volver a realizar la transformada inversa (FFT^{-1}) de Fourier, se aplica una serie de operaciones (normalización respecto de la frecuencia de control, elevación al cuadrado, conversión a escala logarítmica) para mejorar la amplitud de las componentes frecuenciales de fallo.

La posición en tiempo de la característica en caso de fallo aparecerá dentro de $I_{Cepstrum}$ en la posición de tiempo $t_{K-orden}$ definida por la ecuación (2.12), donde f corresponde a la frecuencia de alimentación y s al deslizamiento.

$$I_{Cepstrum} = \left| FFT^{-1} \left(\log_{10} \left(\left[\frac{|FFT(i_{fase}(t))|}{\max(|FFT(i_{fase}(t))|)} \right]^2 \right) \right) \right)^2 \quad (2.11)$$

$$t_{K-orden} = \frac{1}{2 \cdot f \cdot s} \quad (2.12)$$

En el caso de que el motor sea un motor sano, la amplitud en el tiempo de la posición $t_{K-orden}$ que representa al fallo será muy baja (motor sano), mientras que en el caso de un motor con fallo de daños en las barras del rotor la amplitud resultante en el tiempo para la posición $t_{K-orden}$ será mucho mayor.

En la siguiente Figura 2.19 se muestra un ejemplo del Cepstrum de una de las corrientes de un motor sano, mientras que en la Figura 2.20 se muestra el Cepstrum de una de las corrientes de un motor (de las mismas características) con fallo de barra rota. Se puede observar la diferencia de amplitud en el tiempo de la repetición frecuencial de las componentes características entre las dos imágenes. Aunque aparece una pequeña amplitud para la corriente de fase del motor sano, este es mucho menor que su homónimo obtenido de la corriente de fase del motor con fallo de barra rota.

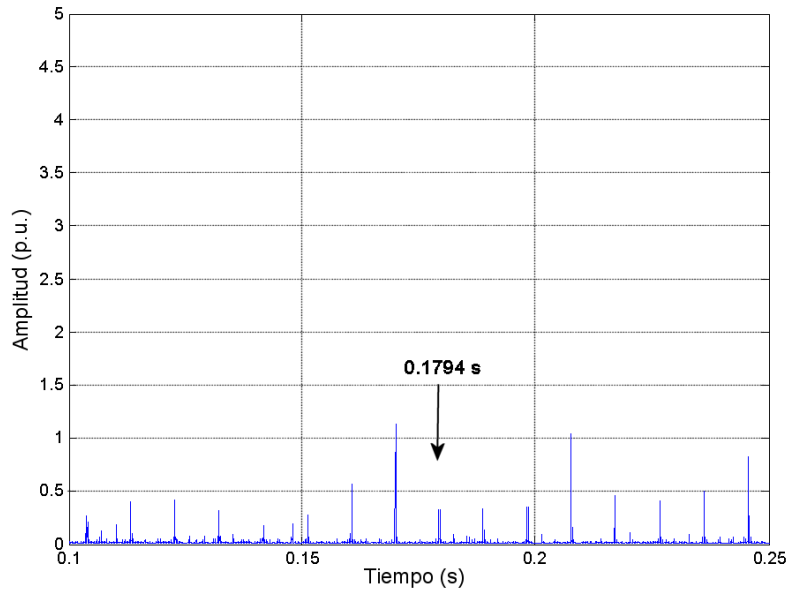


Figura 2.19 Cepstrum de una de las fases de corriente de un motor sano trabajando a una frecuencia de alimentación de 50Hz.

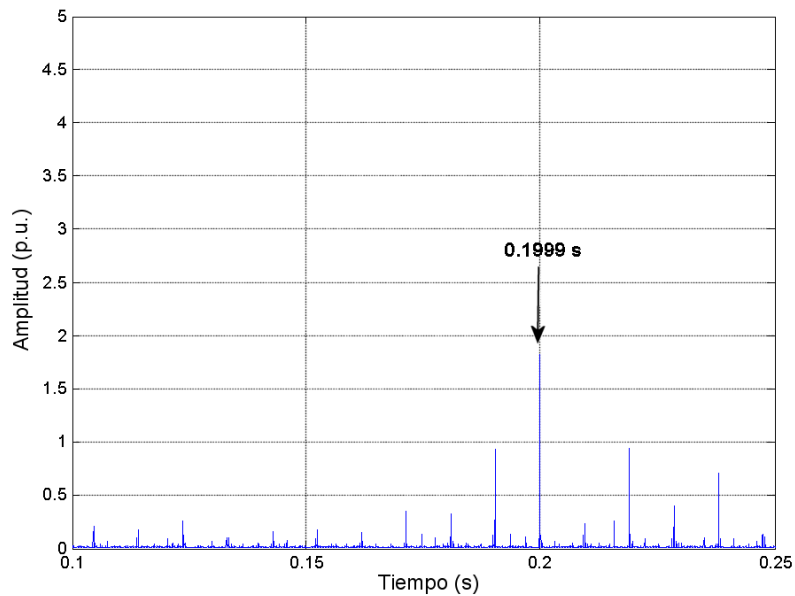


Figura 2.20 Cepstrum de una de las fases de corriente de un motor con fallo de barra rota trabajando a una frecuencia de alimentación de 50Hz.

2.3.1.5 Componente alterna del módulo de la señal analítica obtenida mediante la transformada de Hilbert.

La siguiente técnica emplea la transformada FFT, donde se utiliza la función de transformación de Hilbert para obtener la componente alterna del módulo de la señal analítica (AS) respecto a una de las fases del motor.

Respecto de la técnica FFT, esta técnica elimina casi por completo el problema de la fuga espectral y mejora así la detección de los fallos en los casos afectados por este problema.

Respecto de la técnica del apartado 2.3.1.3 del vector de Park extendido, ambas técnicas obtienen resultados muy parecidos, compartiendo como ya se ha indicado, la cualidad de minimizar el problema de la fuga espectral.

Sin embargo, a diferencia de la técnica EPVA que requiere de las tres corrientes de faso, esta técnica tiene como ventaja que requiere **solo** de una fase de corriente para obtener las características representativas de fallo.

Con solo una corriente de fase, la transformada de Hilbert reconstruye una aproximación del fasor de corriente de la máquina (ejemplo en Figura 2.21).

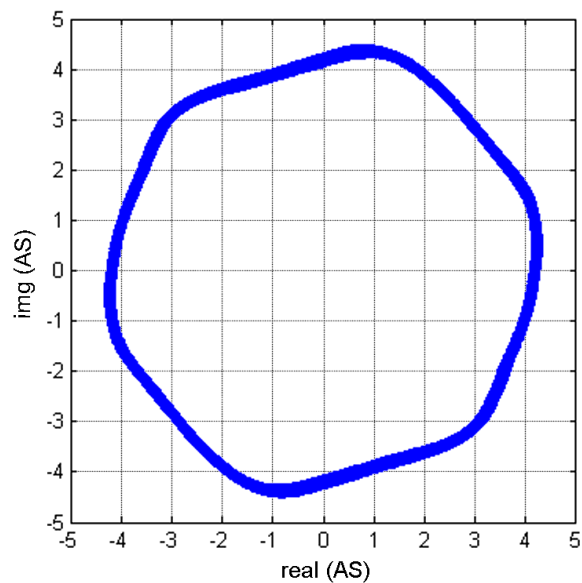


Figura 2.21 Señal analítica obtenida mediante la transformada de Hilbert de una corriente de fase de un motor sano.

Al igual que con la FFT, las características relevantes de fallo se corresponden con una serie de frecuencias dentro del espectro; que en este caso está formado solo por componentes alternas. El módulo de la señal analítica es representado por $|I_{SA}(t)|$.

Como se comenta en el artículo de Puche [34], la transformada de Hilbert es un tipo de transformación muy utilizado en un amplio abanico de campos tecnológicos, como por ejemplo en geofísica, en sistema de diagnóstico de motores, en sistemas de transmisión de señales, etc.

Dentro de este mismo artículo [34], se expone y desarrolla la teoría para poder aplicar la transformada de Hilbert al diagnóstico de motores, validando de forma matemática la obtención de posibles características representativas de fallos. En concreto, este artículo se centra en un fallo por barras rotas en el rotor, aunque puede ser extrapolable a cualquier otro tipo de fallo detectable.

Para su desarrollo se expone un modelo matemático ideal de un motor de inducción. La corriente de fase de un motor sano se representa en la fórmula (2.13) y la formula equivalente que se corresponde con la fase de un motor con barra rota es (2.14).

$$i(t) = I_M \cos(2\pi ft) \quad (2.13)$$

$$i_{BarraRota}(t) = I_M \cos(2\pi ft) [1 + \beta \cos(2\pi f_{BR}t)] \quad (2.14)$$

Dentro de la fórmula (2.13) la variable $i(t)$, como ya se ha comentado, es la corriente del estator del motor sano. La variable I_M representa la amplitud máxima de la componente senoidal a la frecuencia fundamental.

En la fórmula (2.14) la variable $i_{BarraRota}(t)$ se corresponde con la corriente de fase del motor cuando se da una rotura en alguna de las barras del rotor. La variable f_{BR} se corresponde con la posición de frecuencia característica del fallo de barra rota. La variable β representa el índice de modulación o índice de profundidad. Según se expone en el artículo de Puche [34], en el caso de una máquina con fallo de barras rotas, este índice de modulación (β) puede ser expresado de forma aproximada como la división del número de barras rotas consecutivas n_{br} por el número de barras del rotor N (2.15).

$$\beta \approx \frac{n_{br}}{N} \quad (2.15)$$

En dicho motor idealizado, en el caso de tener un fallo de barra rota la frecuencia de los armónicos debidos a éste serán de frecuencias:

$$\omega_{BR} = 2\pi f_{BR} = 2\pi(2ksf) \quad \text{siendo } k = \pm 1, \pm 2, \pm 3, \text{ etc.} \quad (2.16)$$

Para poder desarrollar el análisis de la corriente de fase, en el artículo [34] se exponen una serie de indicaciones a tener en cuenta:

- a) La transformada de Hilbert (HT) de una función trigonométrica es igual a dicha función con un desfase de 90° (el seno se transforma en coseno y viceversa). Extrapolándolo al dominio de frecuencias el espectro resultante de la transformada de Hilbert para esa función tendrá la misma amplitud y frecuencia respecto de la función original, pero manteniendo ese desfase de 90° que se da con esta transformada.
- b) La señal analítica resultante (vector $x(t)$) (2.18) de aplicar HT solo contiene las frecuencias positivas de la señal original, se han eliminado todas las frecuencias negativas y la componente frecuencial continua se ha duplicada.
- c) Todas las frecuencias bajas de la señal original se encuentran en la parte correspondiente a la amplitud $a(t)$ de la señal analítica.
- d) Todas las frecuencias altas de la señal original se encuentran en la fase $\theta(t)$ de la señal analítica.

$$\bar{x}(t) = x(t) + jHT(x(t)) \quad (2.18)$$

Teniendo en cuenta las indicaciones anteriores, se puede obtener la señal analítica para este tipo de fallo del motor con barra rota en el rotor (2.19) expresada como módulo-argumento, combinando la parte real (2.14) y la parte imaginaria (2.20) generada por la transformada de Hilbert tal como se muestra en la fórmula de la señal analítica (2.18). En esta fórmula se ha aplicado la propiedad del desfase de 90° sobre la fórmula de la fase de corriente para motor ideal con fallo de barra rota (2.14).

$$\vec{i}_{BarraRota}(t) = [1 + \beta \cos(2\pi f_{BR}t)] \cdot I_M (\cos(2\pi ft) + j \text{sen}(2\pi ft)) \quad (2.19)$$

$$HT(i_{\text{BarraRota}}(t)) = I_m \text{sen}(2\pi ft) [1 + \beta \cos(2\pi f_{BR}t)] \quad (2.20)$$

Esta señal analítica (2.19) puede separarse en dos vectores correspondientes a la amplitud $a(t)$ y la fase $\theta(t)$. También puede simplificarse como se muestra en (2.21) y sustituyendo el índice de modulación β (2.15) sobre (2.21) se obtiene la señal analítica expresada finalmente en (2.22).

$$\vec{i}_{\text{BarraRota}}(t) = I_M e^{j\omega t} \cdot (1 + \beta \cos(2\pi f_{BR}t)) \quad (2.21)$$

$$\vec{i}_{\text{BarraRota}}(t) = I_M e^{j\omega t} \cdot \left(1 + \frac{n_{br}}{N} \cos(2\pi(2ksf)t)\right) \quad (2.22)$$

La representación en frecuencia de la componente alterna del módulo de la señal analítica de la corriente de fase permite obtener las características representativas para el correspondiente fallo. Para obtener esa componente alterna de la señal analítica (2.22), a esta señal analítica se le resta el valor medio de la señal (filtro de continua).

De este modo se facilita también la detección de las frecuencias características de fallo cuando se da un deslizamiento pequeño sin que la frecuencia fundamental pueda influenciarla, es decir, se elimina el problema de la fuga espectral.

De esta componente alterna del módulo de la señal analítica (2.22) se puede obtener finalmente la amplitud de la frecuencia característica de fallo (en el caso de que exista fallo) a bajas frecuencias teniendo en cuenta el deslizamiento.

En la Figura 2.22 y Figura 2.23 se muestran, respectivamente, el espectro resultante de la componente alterna del módulo de la señal analítica para una corriente de fase de un motor sano y para la fase de un motor con un fallo de barra rota. Como se puede observar, se distingue perfectamente la amplitud característica de fallo de la señal sin fallo, y está claro que ya no puede ser enmascarada por la frecuencia fundamental al no aparecer esta última.

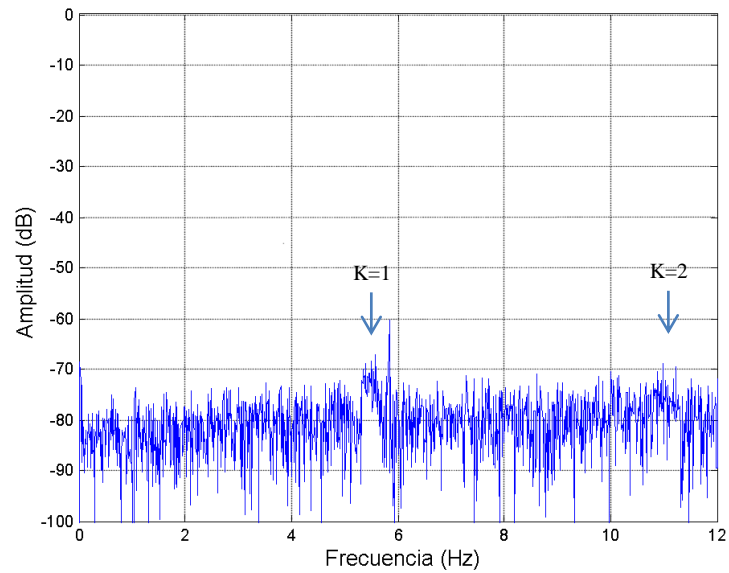


Figura 2.22 Espectro de la componente alterna del módulo de la señal analítica para una de las fases de corriente de un motor SANO a una frecuencia de alimentación de 50Hz.

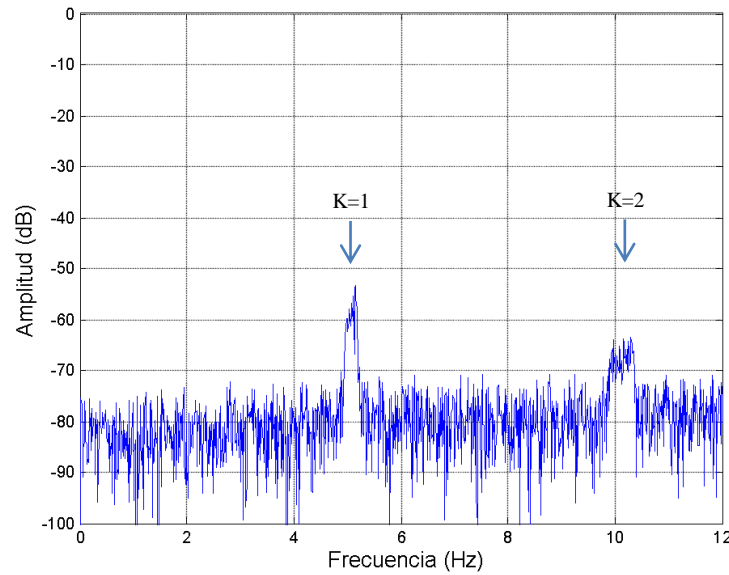


Figura 2.23 Espectro de la componente alterna del módulo de la señal analítica para una de las fases de corriente de un motor con FALLO DE BARRA ROTA a una frecuencia de alimentación de 50Hz.

2.3.1.6 Análisis del Orden Armónico para régimen estacionario (HOTA estacionario).

La siguiente técnica tiene una historia muy reciente, ya que su primera aparición para el diagnóstico de motores de inducción se presenta en el artículo de Sapena (2015)[36]. La nomenclatura de esta técnica se corresponde con las iniciales en inglés HOTA (Harmonic Order Tracking Analysis).

Para este método se emplea la transformada FFT para hacer la transformación en frecuencia. Se mejora en este caso la detección del fallo haciéndolo independiente de la velocidad de trabajo del motor. Además, a diferencia de todas las técnicas anteriores, esta técnica tiene normalizada la escala de los armónicos de fallo a una escala de k -orden armónico (independencia del deslizamiento) y simplifica la información final de los resultados dejando visibles únicamente los datos relevantes (características de fallo).

Dada su naturaleza, esta técnica se puede utilizar tanto en régimen permanente como en régimen transitorio (apartado 2.3.2.6), utilizando en este caso una versión adaptada para cada uno de los regímenes de trabajo.

Como en este apartado solo se comentan técnicas de diagnóstico basadas en régimen permanente se va a exponer únicamente el HOTA desarrollado para dicho régimen de funcionamiento. El HOTA desarrollado para régimen transitorio se puede consultar en el apartado 2.3.2.6.

En la Figura 2.24 se desarrolla el diagrama de flujo del algoritmo HOTA. Como se muestra en este diagrama, como primer paso se genera el vector de giro de la corriente mediante la transformada Hilbert o Park, seguidamente se transforma el eje de coordenadas del vector desde el estator al rotor (proyección de las corrientes). Del vector de corrientes proyectadas se genera su espectro de frecuencias mediante la transformada FFT. Como último paso de este método se transforma el dominio de frecuencias del espectro al dominio del k -orden armónico. Donde el k -orden armónico representa al dominio formado por los k componentes armónicos generados por el fallo.

Finalmente se obtiene el “Resultado HOTA” que se corresponde con el vector de amplitudes del espectro respecto del k -orden armónico, donde cada componente de fallo se muestra en cada posición entera del índice k . En esta tesis, se aplica al fallo de barra rota pero se podría extender a otro tipo de fallos (como por ejemplo excentricidad, cortocircuito entre espiras, etc.) cuyas frecuencias características dependan de un orden de armónico.

De esta misma Figura 2.24 se observa un paso adicional sobre el resultado de HOTA de forma que minimiza la información del resultado para contener solo los datos relevantes (resultado reducido HOTA).

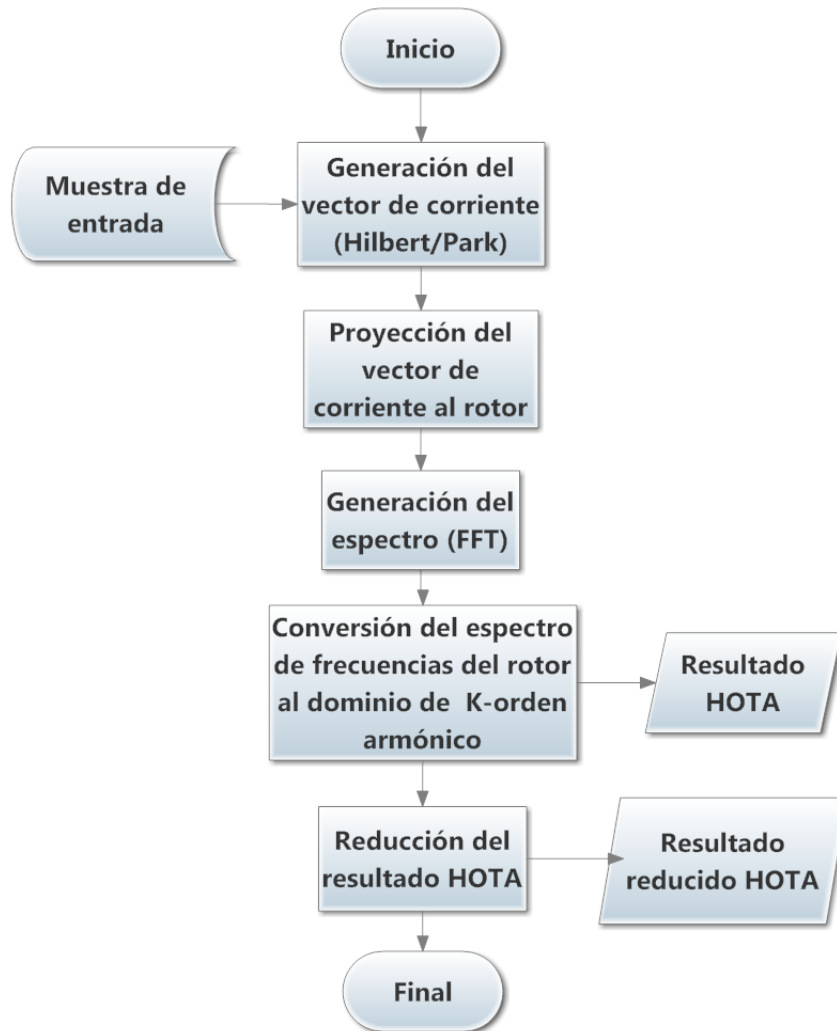
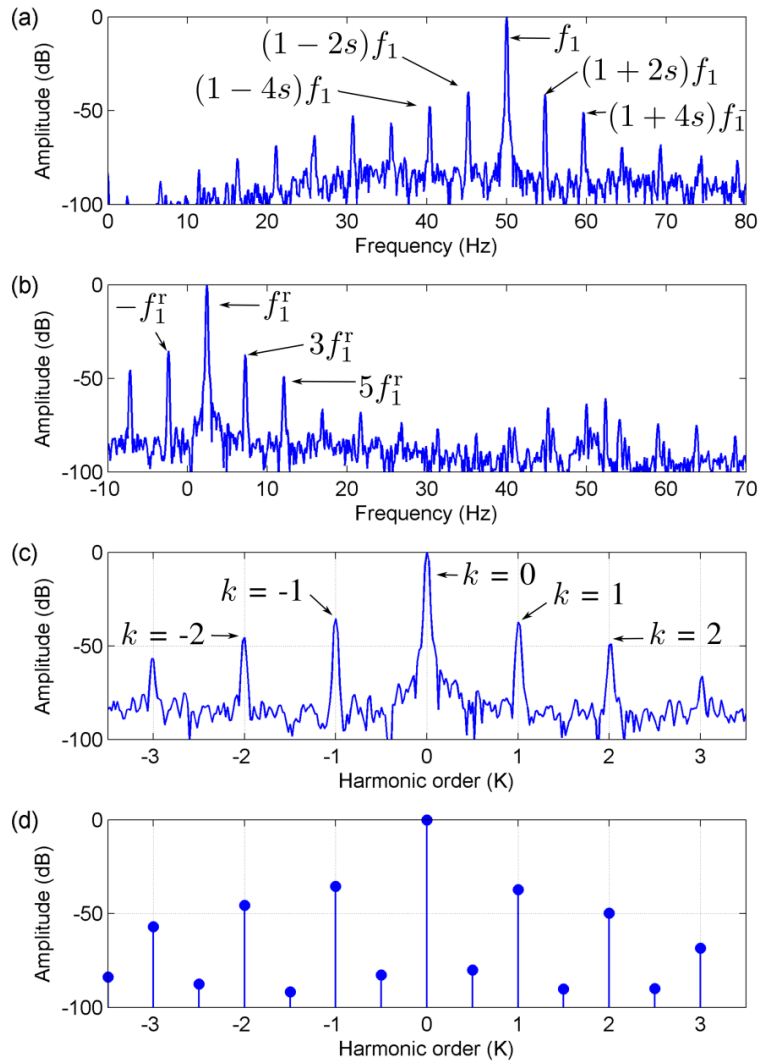


Figura 2.24 Diagrama de flujo del algoritmo de proceso realizado por la técnica de diagnóstico HOTA para la obtención de características representativas de fallo.

En la Figura 2.25 se representa gráficamente de forma aproximada como sería las diversas transformaciones desarrolladas. Partiendo del espectro de frecuencias de una corriente de fase de un motor (a), se traslada el punto de referencia desde el estator al rotor (b) y se reajustan las frecuencias respecto al k -orden armónico (c). En consecuencia, en vez de una escala de frecuencias lo que se obtiene como resultado es una escala de orden k , donde cada valor absoluto de k se corresponde con una componente frecuencial característica del fallo, en este caso de barra rota (d).



Fuente: Sapena Baño (2015) [35]

Figura 2.25 Ejemplo de HOTA para el diagnóstico en régimen permanente. (a) Frecuencias de la corriente de fase del estator. (b) Frecuencias de la corriente de fase del estator tomando como referencia el rotor. (c) HOTA de la corriente de fase del estator. (d) HOTA reducido de la corriente de fase del estator.

Aunque en la imagen obtenida de la investigación original parece que el espectro se desarrolla inicialmente desde Figura 2.25 (a) de esta representación. En realidad, como se ha observado anteriormente en el diagrama de flujo del algoritmo HOTA (Figura 2.24), la transformación de la corriente de fase al espectro de frecuencias se desarrollaría entre (b) y (c) de la representación gráfica y no antes.

Como la única información importante son esas frecuencias características, se puede eliminar del HOTA resultante la información no relevante minimizando notablemente el número de datos imprescindibles, lo que es importante tanto para el almacenamiento como para la transmisión de los datos (resultado reducido HOTA).

Para validar de forma matemática esta técnica de diagnóstico también se hace uso de modelos teóricos de motores sanos y de barra rota. Las fórmulas que generan corrientes de fase ideales de motor sano y de motor de barra rota son las que ya se han expuesto en la técnica de análisis del apartado 2.3.1.5. Estas fórmulas son la (2.13) para corriente de fase de motor sano y (2.14) para corriente de fase de motor con fallo.

Como primer paso, se parte de la corriente de fase del motor para construir el vector correspondiente a su rotación en el espacio. Se puede construir el vector con una única corriente generando la señal analítica (2.22) por medio de la transformación de Hilbert.

El vector resultante se corresponde con la rotación espacial de la corriente en el estator. Se cambia el sistema de referencia del estator al rotor.

La traslación del sistema de referencia se realiza con la fórmula mostrada en (2.23) donde se obtiene la nueva señal analítica con el sistema de referencia desde el rotor.

$$\vec{i}_{BarraRota}^R(t) = \vec{i}_{BarraRota}(t) \cdot \vec{\Phi}_{Rotacion}(t)^* \quad (2.23)$$

La variable $\vec{\Phi}_{Rotacion}(t)^*$ es el complejo conjugado de la posición angular mecánica del rotor, la cual se puede obtener midiendo directamente la posición del rotor con un codificador de posiciones, o con un tacómetro integrando las velocidades obtenidas.

En el modelo ideal, se pueden obtener estas posiciones angulares mecánicas mediante la fórmula (2.24), la cual se puede simplificar para el caso específico de régimen permanente.

$$\vec{\Phi}_{Rotacion}(t) = e^{jP\phi_{Rotacion}(t)} = e^{jP\omega_{Rotor}t} \quad (2.24)$$

En este caso, la variable ω_{Rotor} es la velocidad angular de giro del rotor y la variable p se corresponde con el número de pares de polos específico del motor.

Con la fórmula (2.25) se obtiene el deslizamiento s para régimen permanente, el cual ya ha sido nombrado varias veces en otras técnicas de este tipo de régimen. La variable ω representa la velocidad de sincronismo que en el dominio de frecuencias se convierte en la variable f de la frecuencia de alimentación, mientras que la variable f_r se corresponde con la frecuencia mecánica de rotación en coordenadas eléctricas.

$$s = \frac{\omega - p\omega_{Rotor}}{\omega} = \frac{f - f_r}{f} \quad (2.25)$$

Al reemplazar dentro de la fórmula (2.23) la fórmula de la señal analítica respecto del estator en un motor con fallo (2.22), la posición angular mecánica del rotor para régimen permanente (2.24) y la fórmula del deslizamiento vista en (2.25), se obtiene la siguiente fórmula:

$$\vec{i}_{BarraRota}^R(t) = \left(1 + \frac{n_{br}}{N} \cos(2k(\omega - p\omega_{Rotor})t) \right) \cdot I_M e^{j(\omega - p\omega_{Rotor})t} \quad (2.26)$$

Si se aplica en la ecuación (2.26) en el cambio de notación propuesto en la ecuación (2.27) se obtiene la fórmula (2.28).

$$\omega^r = \omega - p\omega_{Rotor} \quad (2.27)$$

$$\vec{i}_{BarraRota}^R(t) = \left(1 + \frac{n_{br}}{N} \cos(2k\omega^r t) \right) \cdot I_M e^{j\omega^r t} \quad (2.28)$$

De esta última fórmula (2.28), al aplicar la fórmula de expansión del coseno (2.29) queda finalmente la fórmula de la señal analítica de fallo respecto del rotor de la siguiente forma (2.30).

$$\cos(\Phi) = \frac{e^{j\Phi} + e^{-j\Phi}}{2} \quad (2.29)$$

$$\vec{i}_{BarraRota}^R(t) = I_M e^{j\omega^r t} + \frac{n_{br}}{N} \cdot I_M e^{j\omega^r BarraRota t} \quad (2.30)$$

Dónde:

$$\omega_{BarraRota}^r = \omega^r 2k + \omega^r \quad \text{siendo} \quad \begin{matrix} k = \pm(1, 2, 3, 4...) \\ \omega^r = s \cdot f \end{matrix} \quad (2.31)$$

Al realizar el cambio de escala k de (2.32) en los ejes de frecuencia del espectro de la señal analítica respecto del rotor ω^r , obtenida mediante el espectro de la señal obtenida en la ecuación (2.30); todas las frecuencias representativas de fallo recaen dentro de cada una de las posiciones enteras de k de esta nueva escala (Figura 2.27).

$$k = \frac{\omega^{r'} - \omega^r}{2\omega^r} \quad (2.32)$$

En las Figura 2.26 y Figura 2.27 se muestran los resultados de aplicar la técnica de HOTA sobre una corriente de fase de motor sano y de una corriente de fase de motor con fallo de barra rota.

En este ejemplo la diferencia entre la existencia de características representativas de fallo (Figura 2.27) y la ausencia de estas en un motor sano (Figura 2.26) es bastante evidente.

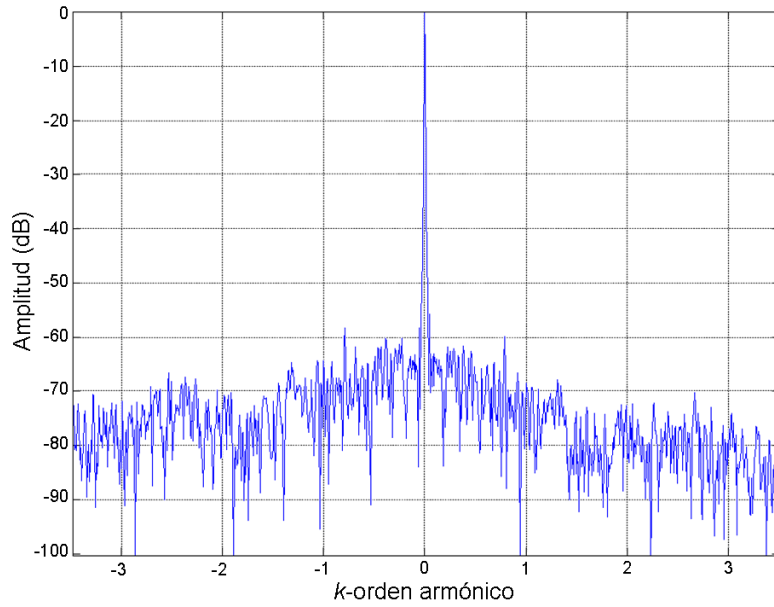


Figura 2.26 HOTA de una de las fases de corriente de un motor sano trabajando a una frecuencia de alimentación de 50Hz.

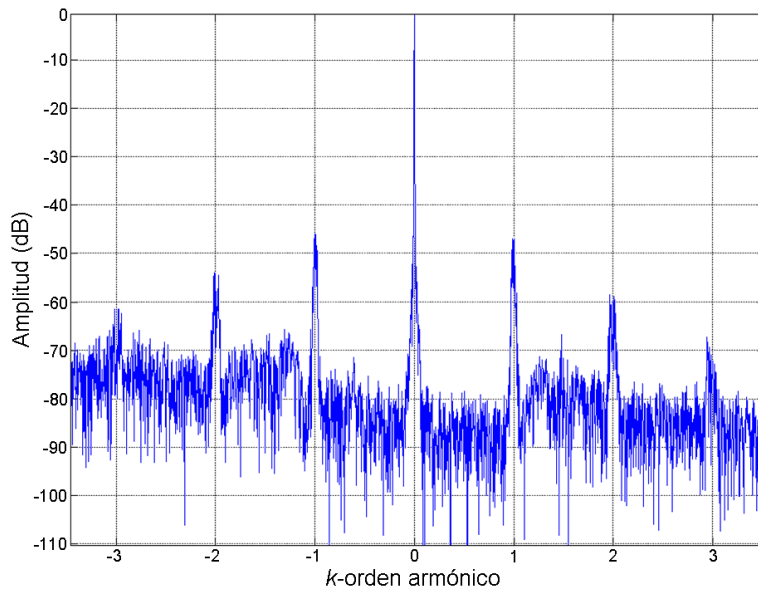


Figura 2.27 HOTA de una de las fases de corriente de un motor con fallo de barra rota trabajando a una frecuencia de alimentación de 50Hz.

Como se puede observar en la Figura 2.27 y como ya fue expuesto en el apartado anterior 2.3.1.1, cuando el valor absoluto de k aumenta, la amplitud considerada como característica de fallo destaca menos.

Gracias al reajuste de las escalas de frecuencias en el orden k , las frecuencias características de fallo siempre se muestran en la misma posición. En consecuencia, esta técnica tiene la ventaja de que es totalmente independiente respecto de las condiciones de funcionamiento del motor, es decir frecuencia de alimentación y velocidad del motor. Facilitándose la comparativa de diferentes muestras obtenidas bajo diferentes condiciones de funcionamiento del motor.

Observando la Figura 2.28 se ve claramente que las frecuencias características de fallo se muestran en el mismo k -orden aunque la frecuencia de alimentación sea distinta a la utilizada en la Figura 2.27 (25Hz en vez de 50Hz).

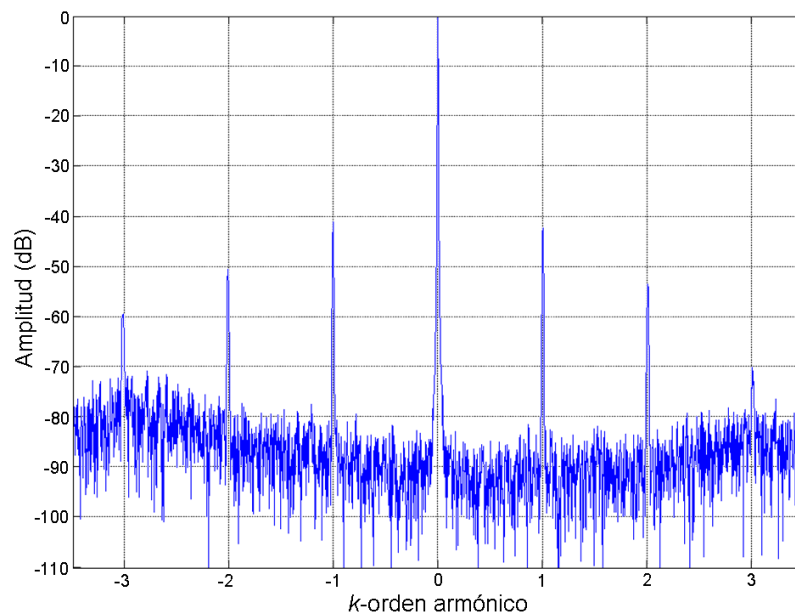


Figura 2.28 HOTA de una de las fases de corriente de un motor con fallo de barra rota trabajando a una frecuencia de alimentación de 25Hz.

2.3.1.7 Otras técnicas de detección de fallos en régimen estacionario.

Además de todas estas técnicas de detección de fallos existentes en el estado del arte (comentadas en el apartado 2.3.1), existen otros tipos de técnicas para el diagnóstico de fallos en régimen estacionario, que también permiten realizar un análisis sobre diversos fallos en los motores de inducción. Algunas de estas técnicas son:

- **Análisis de la potencia instantánea [32], [37].**
- **Análisis espectral de alta resolución (MUSIC) [32].**
- **Análisis de la tensión de neutro [38].**
- **Zero Crossing Transform (ZCT) [39] [40]**

2.3.2 Técnicas de detección de fallos para un régimen de trabajo transitorio.

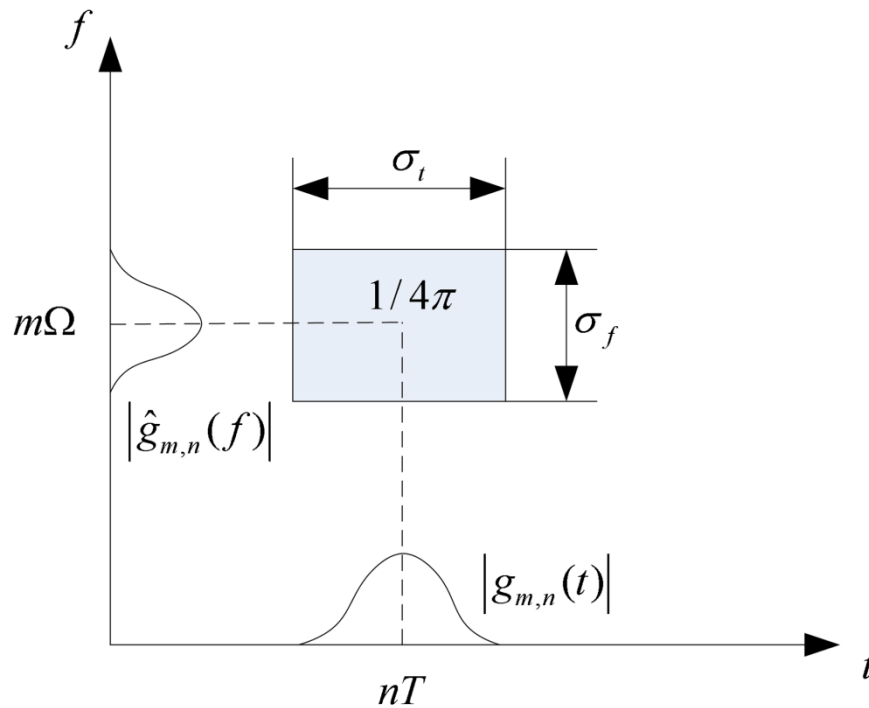
Dada la dificultad de mantener el motor trabajando en régimen estacionario o en el caso de que no se pueda trabajar en dicho régimen, las técnicas de diagnóstico para régimen estacionario no son válidas, ya que la velocidad constante de trabajo es esencial para la detección satisfactoria de las características relevantes. En consecuencia, es necesario el uso de otras técnicas de diagnóstico capaces de obtener ciertas características representativas de los diversos fallos que se dan en el motor de inducción mientras éste se encuentra en un régimen de trabajo con transitorios.

Aunque parezca que estas técnicas de detección de fallos tienen la ventaja de poder analizar señales tanto en régimen transitorio como en régimen estacionario, su mayor coste computacional y menor resolución de detección respecto de las técnicas de detección de fallos en régimen estacionario hace que actualmente estas clases de técnicas sean desplazadas por las técnicas de diagnóstico en régimen estacionario para dicho régimen de trabajo.

En los siguientes subapartados se exponen las técnicas de detección de fallos para régimen transitorio más notorias.

2.3.2.1 Transformada de Fourier corta en el tiempo (STFT).

La transformada STFT es un tipo de transformación en espacio tiempo-frecuencia [41] donde se utiliza una función de ventana (Figura 2.29) a ajustar a su mínima expresión para obtener un espacio tiempo/frecuencia con la mayor resolución posible. Esta ventana se desplaza en el tiempo para enfatizar las componentes frecuenciales de la señal que se está analizando en un intervalo de tiempo determinado.



Fuente: Riera-Guasp et al (2012)[42]

Figura 2.29 Representación paramétrica de la función de ventana Gaussiana en el espacio de tiempo-frecuencia.

Como se comenta en Riera Guasp [42], existen bastantes tipos de funciones distintas usadas como funciones elementales de ventana, como por ejemplo las funciones rectangulares, funciones exponenciales [43], función chirp [44], funciones hiperbólicas [45] y la función Szyper [46] entre otras.

Sin embargo, en este mismo artículo de Riera-Guasp [42] se recomienda la función Gaussiana (2.33) con traslaciones (T) y modulaciones (Ω) para la dimensión de tiempo y la dimensión de frecuencia respectivamente. La función elemental de ventana Gaussiana es comentada por Gabor en su desarrollo en [47] y como se muestra en la ecuación (2.33), esta se expresa en función del tiempo.

$$g_{m,n}(t) = \left(\frac{\alpha}{\pi}\right)^{0.25} \cdot \exp\left(\frac{-\alpha(t-nT)^2}{2}\right) \cdot \exp(j2\pi m\Omega t) \text{ donde } T \cdot \Omega = 1 \quad (2.33)$$

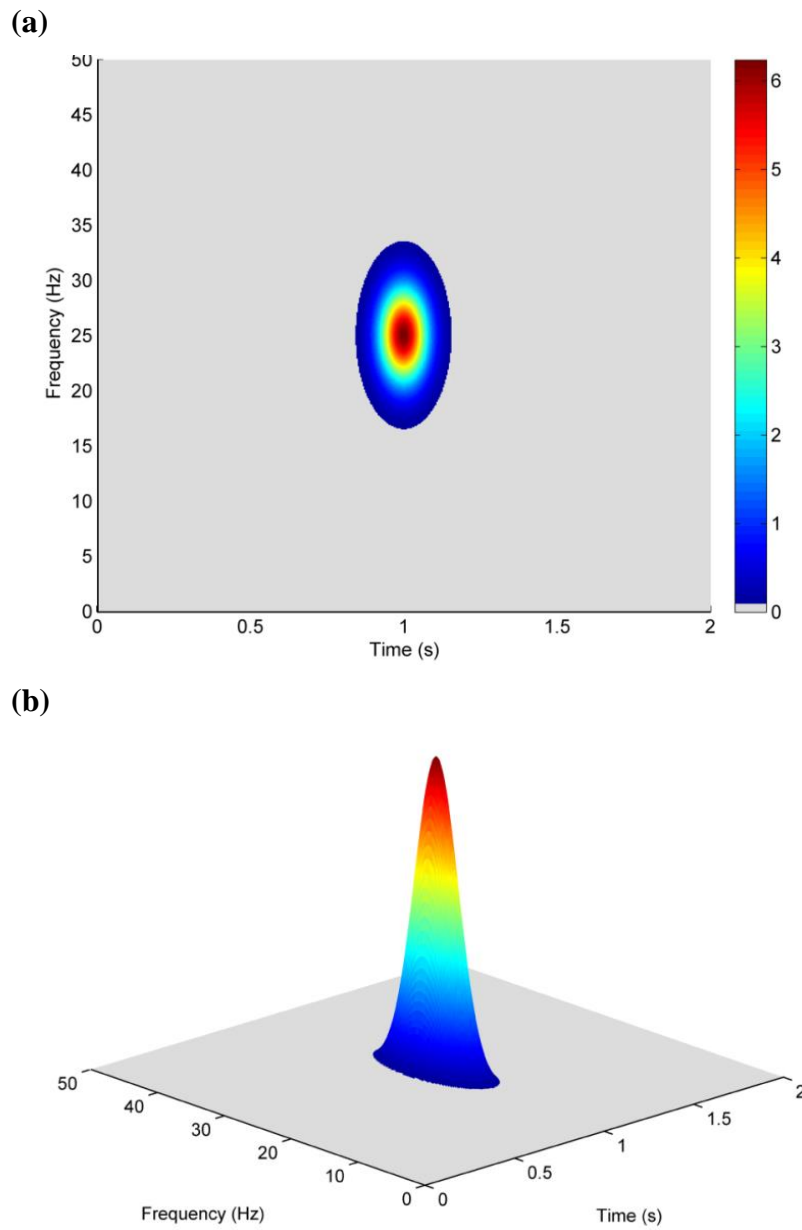
En esta fórmula (2.33) las duplas $\langle m\Omega, nT \rangle$ se corresponden con la posición local en el espacio tiempo-frecuencia para cada valor dentro del rango. Como la función de ventana es mínima para cada dupla, se puede decir que ésta contiene la señal elemental más pequeña en que se puede dividir el espacio de dos dimensiones $\langle f, t \rangle$.

Esta función elemental es óptima en concentrar la mayor energía posible cuando la longitud de la señal es infinita, configurándose con la menor área del filtro de ventana. Como se comenta en Jones [48], esta función elemental funciona de forma similar a un impulso.

En la Figura 2.30 se muestra un ejemplo de la función de ventana Gaussiana optimizada para la detección de fallo de barra rota en 2D (a) y 3D (b), extraído de la investigación de Riera Guasp [42].

Para obtener la mejor definición de las características representativas de fallo dentro del espacio tiempo frecuencia, generalmente se propone que la relación de aspecto T/Ω de la ventana mantenga el mayor solapamiento posible con la trayectoria de dicha característica de fallo.

Como se muestra en la Figura 2.31, la ventana de filtro se va desplazando en el tiempo de forma que al aplicar este filtro en cada unidad de tiempo sobre la señal original se origina el vector de frecuencias para ese punto temporal, construyéndose así el espacio final de tiempo frecuencia.



Fuente: Riera-Guasp (2012)[42]

Figura 2.30 Ejemplo de función Gaussiana representada en (a) 2D y (b) 3D.

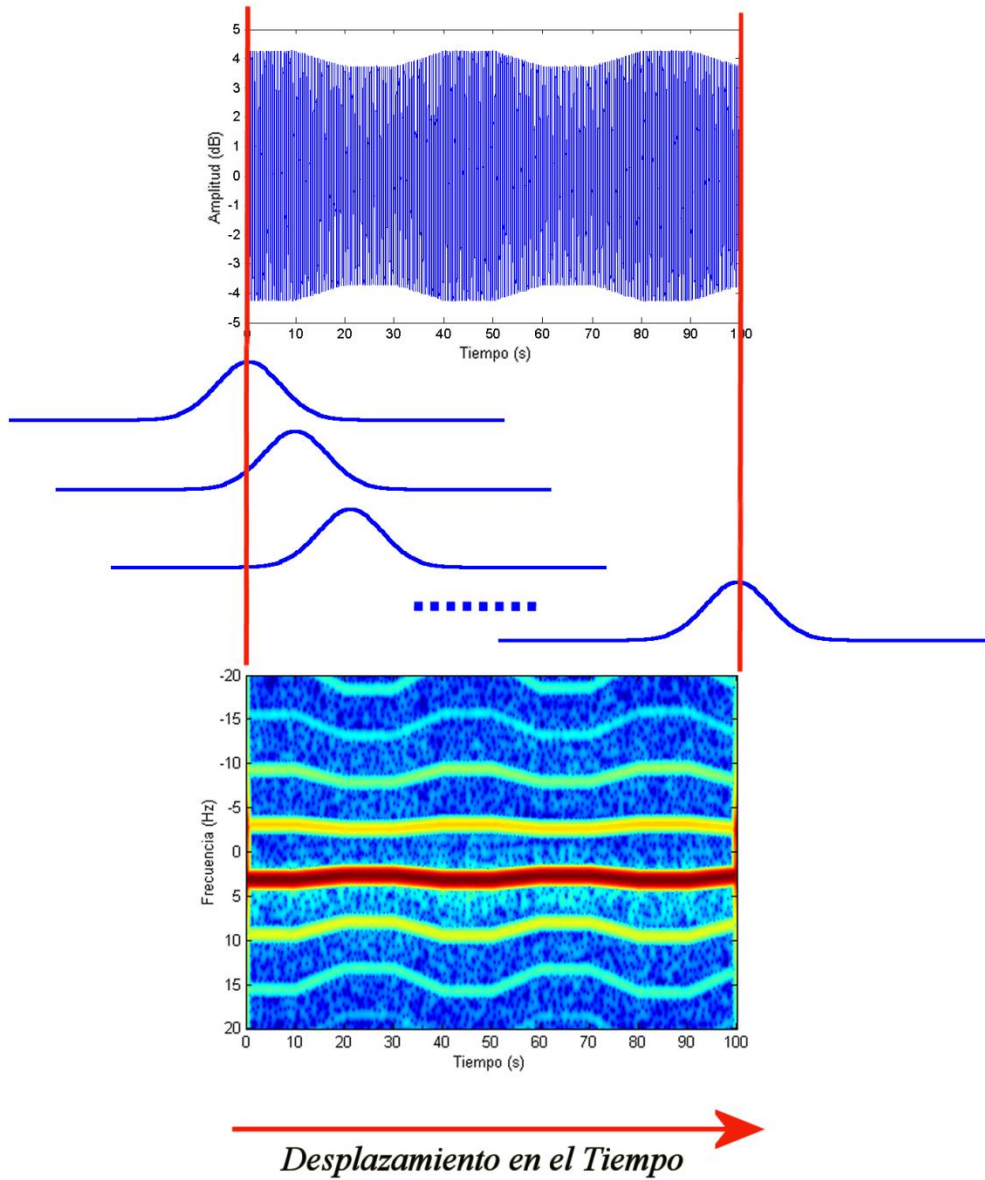


Figura 2.31 Desplazamiento de la función de ventana Gaussiana sobre la señal de corriente original obteniendo así el espacio de tiempo frecuencia STFT.

La siguiente Figura 2.32 muestra un ejemplo del espacio tiempo frecuencia generado por la técnica STFT de una muestra obtenida de una corriente de fase de un motor de inducción operando en régimen transitorio.

En esta imagen se observan claramente marcas amarillentas que van difuminándose. Estas marcas se corresponden con componentes armónicas producidas por un fallo de barra rota.

Es importante ajustar los parámetros de la ventana (la relación de aspecto T/Ω) para obtener la mayor energía del espacio tiempo frecuencia (la mejor representación de fallo). Como ya se ha comentado para cualquier tipo de ventana, igualmente para la ventana Gaussiana la relación de aspecto óptima de la ventana se consigue ajustando los parámetros T y Ω de la ventana para que sea capaz de solapar lo máximo posible la componente frecuencial de fallo. No obstante, la variación de la trayectoria de la componente de fallo cuando se modifica la evolución de la frecuencia de control dificulta la generación de una ventana que mantenga el mayor solapamiento con la trayectoria de estas componentes de fallo (Figura 2.32).

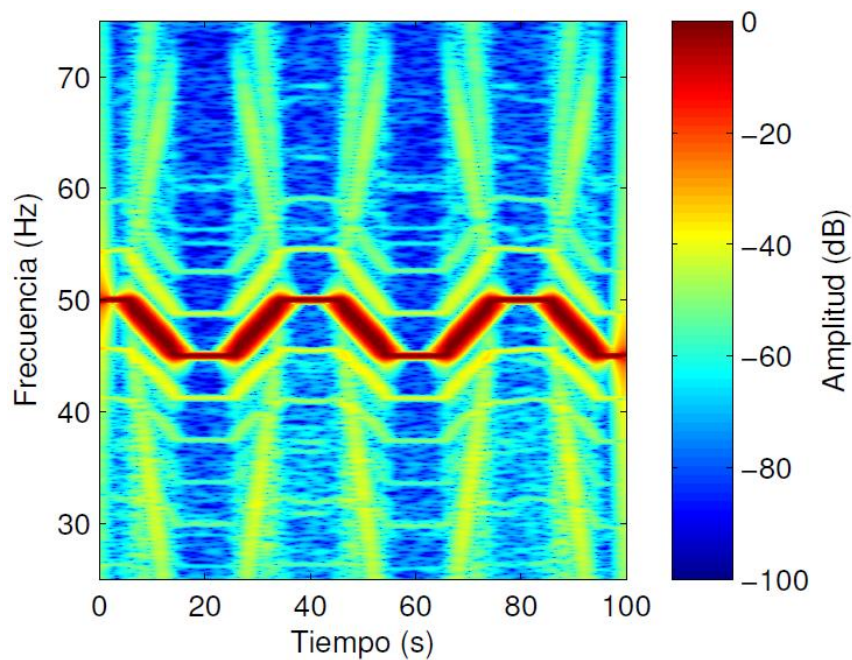


Figura 2.32 Espacio tiempo frecuencia para una corriente de fase de un motor con control escalar con ganancia del deslizamiento operando en régimen transitorio (45-50Hz).

Es decir, de forma directa no es posible obtener la mayor concentración de energía para todos los tramos de las componentes de fallo cuando se opera en un régimen de trabajo

con transitorios. En todo caso, la proporción de la ventana se debe centrar a las posiciones del espacio donde puede realizarse un mayor solapamiento de la componente frecuencial de fallo.

Por tanto ajustar lo mejor posible la proporción entre tiempo y frecuencia de la función de ventana para optimizar la aproximación de la característica representativa de fallo es muy importante pero nada sencilla. Un buen ajuste mostrará una aproximación muy cercana a las amplitudes reales de las características de fallo, siendo esta peor conforme la proporción sea menos idónea.

Se puede modificar la proporción de la ventana para mejorar la resolución de la dimensión de tiempo T o de frecuencia Ω , pero la mejora en una dimensión modifica negativamente la resolución en la otra dimensión.

A continuación se comenta el desarrollo de un algoritmo propuesto recientemente [42] para ajustar lo mejor posible la relación de aspecto entre tiempo y frecuencia de la función de ventana para obtener unas buenas características representativas de fallo.

2.3.2.1.1 Búsqueda del espacio de mínima entropía.

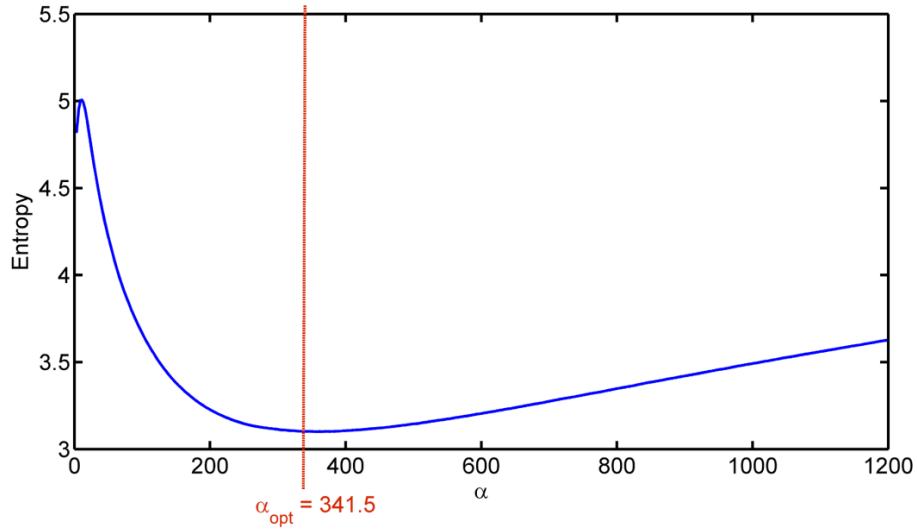
En la investigación de Riera Guasp [42] se propone el análisis de la entropía correspondiente a diversos espacios tiempos frecuencias generados por diversas configuraciones del filtro de ventana para la obtención de un espacio tiempo frecuencia de máxima energía de donde obtener las características representativas de fallo.

La obtención de un espacio con poca entropía se corresponde con un espacio tiempo frecuencia con buena concentración de energía, por tanto se busca la configuración de ventana que desarrolle un espacio tiempo frecuencia con la mínima entropía posible.

En este caso se realizan un proceso de obtención de un vector de mínimas entropías cubriendo un amplio rango de combinaciones distintas de los parámetros de Frecuencia/Tiempo de la función de ventana. En cada variación de la proporcionalidad se calcula la entropía del espacio de tiempo frecuencia generada y se va formando el vector de información de la curva de entropía de espacios respecto de la variación de los parámetros de la función de ventana.

En la Figura 2.33 se muestra un ejemplo de dicha curva de entropía obtenida a partir del espacio tiempo frecuencia de la transformada STFT con diversas configuraciones de un filtro de ventana de tipo Gaussiana.

El cálculo de la entropía del espacio tiempo frecuencia se realiza con la fórmula (2.34), a partir de la matriz de coeficientes de Gabor generada con la fórmula (2.36) sobre la corriente de fase i_r y la ventana de filtro gaussiana g . En este caso m y n representan la posición dentro de la matriz de longitud máxima N .



Fuente: Riera-Guasp (2012) [42]

Figura 2.33 Vector de entropías de STFT respecto del parámetro α de relación proporcional entre la dimensión del espacio tiempo y frecuencia del filtro de ventana Gaussiana.

$$Entropia = -\sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \frac{c_{m,n}^2}{\lambda_{m,n}} \cdot \log \left(\frac{c_{m,n}^2}{\lambda_{m,n}} \right) \quad (2.34)$$

$$\lambda_{m,n} = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} c_{m,n}^2 \quad (2.35)$$

$$c_{m,n} = \sum_{k=0}^{N-1} i_r(k) \cdot g(k-n) \cdot \exp \left(-j \frac{2\pi}{N} \cdot m \cdot k \right) \text{ donde } 0 \leq m, n < N \quad (2.36)$$

Con el vector de entropías generado, la configuración óptima del ajuste del filtro de ventana recae finalmente en el valor de la relación de aspecto T/Ω que genera el espacio tiempo frecuencia con la menor entropía posible (mayor concentración de energía).

En el ejemplo de la Figura 2.33, el valor óptimo de α para la función de filtro de ventana Gaussiana corresponde al punto mínimo de la curva atravesado por el marcador rojo.

2.3.2.2 Transformada Wavelet Discreta.

La transformada Wavelet Discreta (DWT) es una función de descomposición de señales muy utilizada en un amplio abanico de campos relacionados con el estudio de señales de una o más dimensiones. La teoría de ondículas tal como se la conoce actualmente aparece por primera vez en 1909 donde Alfred Haar presenta su tipo de ondícula Haar. La concreción de toda la teoría se desarrolla en los años 80 de ese mismo siglo [49].

A diferencia de las transformaciones comentadas en el apartado 2.3.1 para régimen estacionario donde únicamente se podía analizar el espectro frecuencial resultante, con esta transformada es posible realizar un análisis de la señal tanto en frecuencia como en tiempo, por lo que posibilita el análisis de características transitorias dentro de ese rango de tiempo.

La transformación DWT desarrolla una descomposición de la onda original en un grupo de ondículas aproximadas para distintas escalas frecuenciales. Cada ondícula del grupo muestra las componentes correspondientes a dicha escala de frecuencias dentro de ese rango de tiempo de la señal original.

La descomposición de la señal original se desarrolla por niveles y en teoría no existe un límite máximo de profundidad en la descomposición.

Como se comenta en el libro de Mallat [50], para que la energía de la descomposición sea unitaria se utiliza una familia de ondículas madre normalizada y ortogonal (ψ). Con esta ondícula madre aplicada en la función de escalado (2.37), se desarrolla un filtro espejo de cuadratura [51], que hace la función de filtro paso bajo y filtro paso alto, y con el que se descompone la señal en ondículas. El ajuste de este filtro se realiza mediante los coeficientes wavelet (c_k) de la función descrita en (2.37)

$$\phi(t) = \sum_{k=-1}^{N-2} (-1)^k \cdot c_{k+1} \cdot \psi(2t + k) \quad (2.37)$$

Dependiendo de los rangos de frecuencias a análisis, la descomposición de las ondículas se puede realizar tanto para las ondículas resultantes solo del filtro paso bajo (1D), o como descomposición de la ondícula tanto en su frecuencia baja como en la alta (2D).

Un ejemplo esquemático de descomposición para 3 niveles se muestra en la siguiente Figura 2.34, donde se observa como la ondícula resultante del nivel anterior generado en paso bajo vuelve a ser descompuesta hasta alcanzar el nivel máximo.

El resultado de descomposición de una señal de una dimensión con DWT da lugar a un numero de ondículas de coeficientes de detalle $d_i(t)$ igual al número de niveles más una ondícula de coeficiente de aproximación $a_n(t)$ correspondiente a la descomposición de la parte más baja de las frecuencias resultante en el último nivel de descomposición.

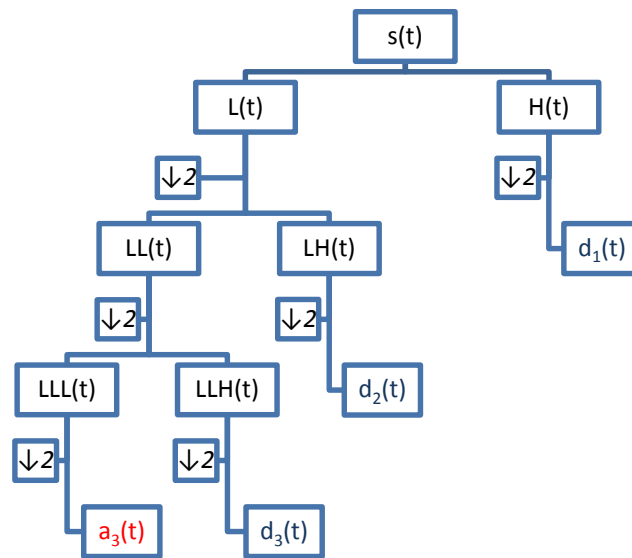


Figura 2.34 Ejemplo de esquema de descomposición en profundidad de tres niveles de una señal mediante wavelets.

La transformación DWT se define matemáticamente en [9] por la función (2.38), donde la variable ψ representa a la ondícula madre y la variable n se corresponde con el número de niveles.

$$DWT_f(i, j) = \frac{1}{\sqrt{a_0^i}} \cdot \sum_{n=-\infty}^{+\infty} f(n) \cdot \psi \left[\frac{n - s_0^i j \tau_0}{s_0^i} \right] \quad (2.38)$$

Como se expone en el mismo trabajo [9], en el caso de las muestras de corrientes obtenidas del estator del motor de inducción, se considera la frecuencia de alimentación f

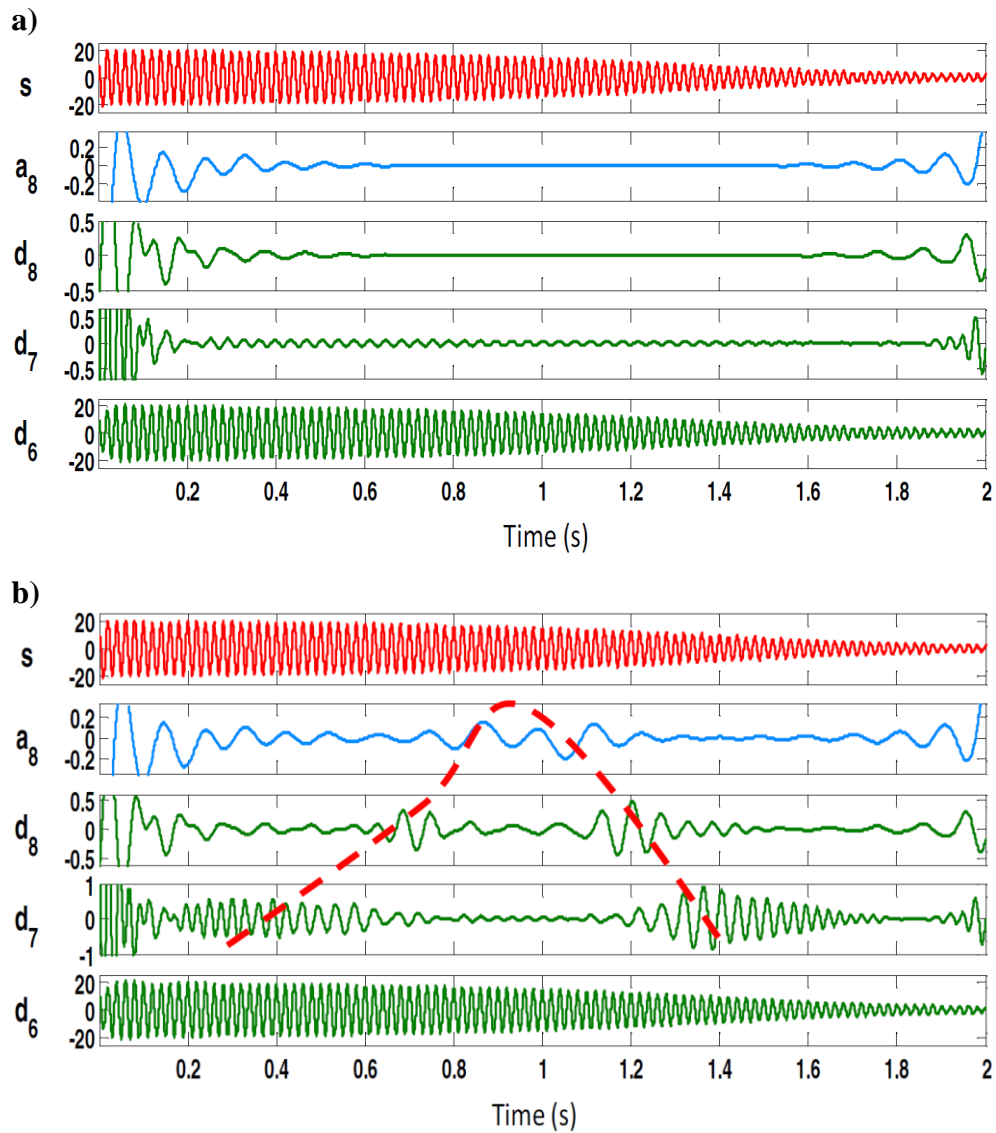
como rango de frecuencia superior en la descomposición DWT de la señal en ondículas. En consecuencia, los rangos de frecuencia desarrollados para los coeficientes de detalle y para el coeficiente de aproximación están dentro de sus respectivos intervalos definidos por (2.39) y (2.40).

$$f(d_n) \in \left[\frac{f}{2^{n+1}}, \frac{f}{2^n} \right] \text{Hz} \quad (2.39)$$

$$f(a_n) \in \left[0, \frac{f}{2^{n+1}} \right] \text{Hz} \quad (2.40)$$

La variable n mostrada en (2.39) y (2.40) se corresponde con el número del nivel de descomposición de la ondícula.

Con esta técnica, la detección de barras rotas en el estator se realiza con la detección del patrón característico que se muestra en las ondículas resultantes al producirse el fallo. Un ejemplo bastante claro de detección de fallo de barra rota se muestra en la Figura 2.35, donde se compara el patrón generado de las ondículas de nivel más bajo (7 y 8) de la señal de corriente de un motor sano y de un motor con dos barras rotas. Este ejemplo ha sido obtenido del artículo de Antonino Daviu [52].



Fuente: Antonino-Daviu et al (2014)[52]

Figura 2.35 Comparación de wavelets del nivel 6, 7 y 8 (8 niveles) para una señal de arranque de un motor de inducción (a) sano y (b) con dos barras rotas en el rotor.

Una de las mayores ventajas de esta técnica sobre los otros vistos, además de poder detectar el fallo en régimen de operación transitoria, radica en que el patrón de fallo que se genera no se ve afectado por las interferencias generadas por las variaciones del par de carga.

Sin embargo, esta técnica de detección también contiene varios inconvenientes importantes. Dada la naturaleza de la transformada DWT donde se divide por filtros paso alto y paso bajo, las frecuencias muy próximas al rango de corte del filtro se ven afectadas por la rampa de corte. Además los límites de rangos de frecuencia entre ondículas vecinas se solapan.

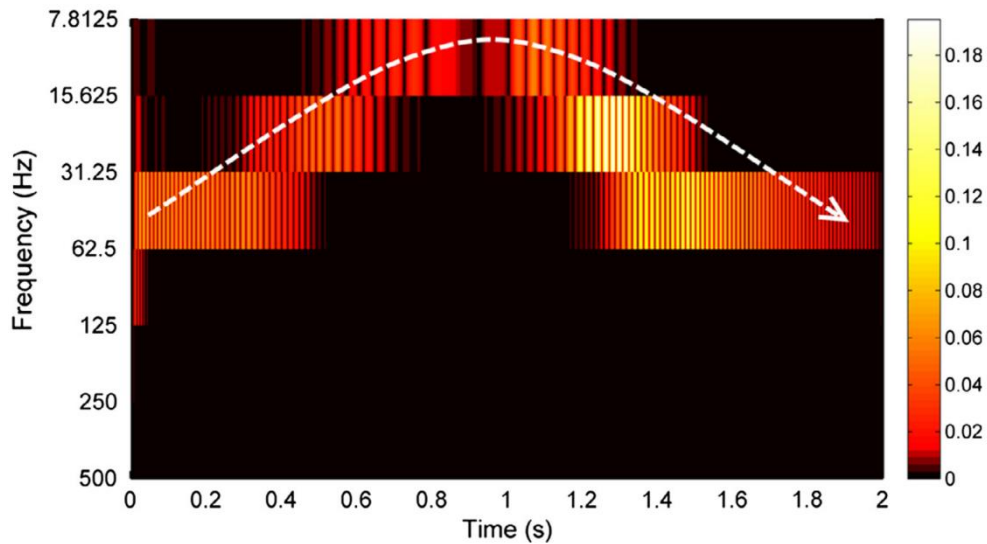
Otro problema usual, comentado en el libro desarrollado por Giri [53] y que se puede observar en la Figura 2.35, es que, aunque la diferencia a simple vista entre motor sano y motor con fallo es bastante evidente, la detección de dicho patrón sobre ondículas separadas es bastante complejo de detectar por un sistema de reconocimiento artificial.

Para paliar este inconveniente en la detección automática, en el mismo libro [53] se expone como solución la construcción de un único espacio bidimensional de tiempo-frecuencia a partir de las ondículas generadas de mayor nivel. En este tipo de representación la amplitud de cada frecuencia en el espacio está representada por las diferentes tonalidades cromáticas. De este modo se simplifica la representación en una única forma de dos dimensiones, mucho más fácil de detectar por un sistema automático que n diversas formas (ondículas) de una dimensión.

En la Figura 2.36 obtenida en [54] se muestra la generación de un espacio de tiempo-frecuencia a partir de las ondículas de una corriente de fase de un motor de inducción con rotura en una de las barras. En este caso, el espacio muestra la asimetría frecuencial que aparece a lo largo del tiempo cuya forma tiene una mayor facilidad para ser detectable por un sistema de visión artificial bidimensional adaptado para ello.

Aunque esta representación en tiempo-frecuencia simplifica la detección automática, muestra varios inconvenientes dados por la transformada DWT. Como se puede observar en las componentes de tiempo-frecuencia de la Figura 2.36, el uso de un ancho de banda fijo condicionado por el carácter diádico de la transformada DWT hace que la resolución de la frecuencia no sea lo bastante precisa para utilizarse como representación de la señal en el espacio de tiempo-frecuencia.

Además, la aparición de variaciones en la amplitud provocadas por la evolución senoidal de la señal en el tiempo ajenas a las amplitudes generadas por las componentes armónicas del fallo hace que sea más difícil de detectar estas últimas.



Fuente: Pineda Sanchez (2013)[54]

Figura 2.36 Representación de un espacio de tiempo-frecuencia en dos dimensiones correspondiente a la evolución de las ondículas de mayor nivel generadas con la transformada DWT a partir de la corriente de arranque de una de las fases de un motor de inducción con una barra rota en el rotor.

2.3.2.3 Transformada Continua Wavelet y Transformada Continua Compleja Wavelet.

Como mejora respecto a la técnica anterior DWT, para paliar los efectos negativos de la representación bidimensional de tiempo-frecuencia, se expone como alternativa en los artículos [55, 56] el uso de la transformada continua wavelet (CWT) y la transformada continua compleja wavelet (CCWT).

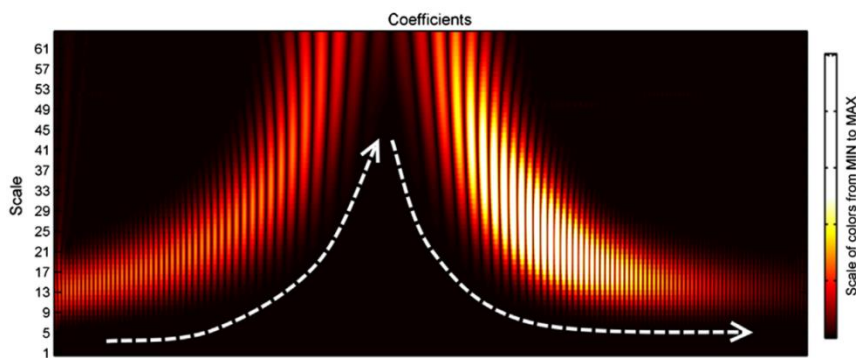
$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(t) \cdot \overline{g(t)} \cdot dt \quad (2.41)$$

$$CWT_f(a, b) = \frac{1}{\sqrt{|a|}} \cdot \int_{-\infty}^{+\infty} f(t) \cdot \overline{\psi\left(\frac{t-b}{a}\right)} \cdot dt \quad \text{donde} \quad \begin{array}{l} a, b \in \mathbb{R} \\ a \neq 0 \end{array} \quad (2.42)$$

Tanto la transformada CWT como la transformada CCWT calculan la densidad de energía existente en cada punto en el espacio tiempo-frecuencia. La fórmula que define

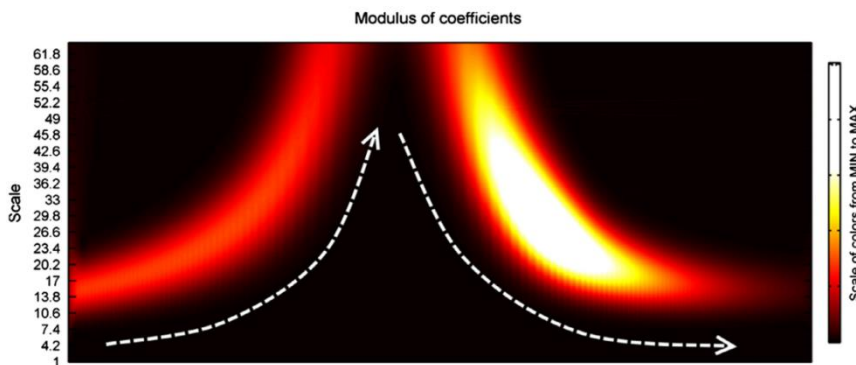
estas transformaciones de energía dentro del espacio de Hilbert (2.41) se muestra en la ecuación (2.42). La única diferencia entre la transformada CWT y la transformada CCWT se encuentra en la función de onda madre $\psi(t)$.

De esta forma, como se puede observar en la Figura 2.37 y Figura 2.38 generada por la transformada CWT y CCWT respectivamente, la representación del espacio tiempo-frecuencia alcanza una mayor resolución respecto a su anterior desarrollo expuesto utilizando transformada DWT (Figura 2.36).



Fuente: Pineda Sanchez (2013)[54]

Figura 2.37 Representación de un espacio de tiempo-frecuencia en dos dimensiones correspondiente a la evolución de las ondículas de mayor nivel generadas con la transformada CWT a partir de la corriente de arranque de una de las fases de un motor de inducción con una barra rota en el rotor.



Fuente: Pineda Sanchez et al (2013) [54]

Figura 2.38 Representación de un espacio de tiempo-frecuencia en dos dimensiones correspondiente a la evolución de las ondículas de mayor nivel generadas con la transformada CCWT a partir de la corriente de arranque de una de las fases de un motor de inducción con una barra rota en el rotor.

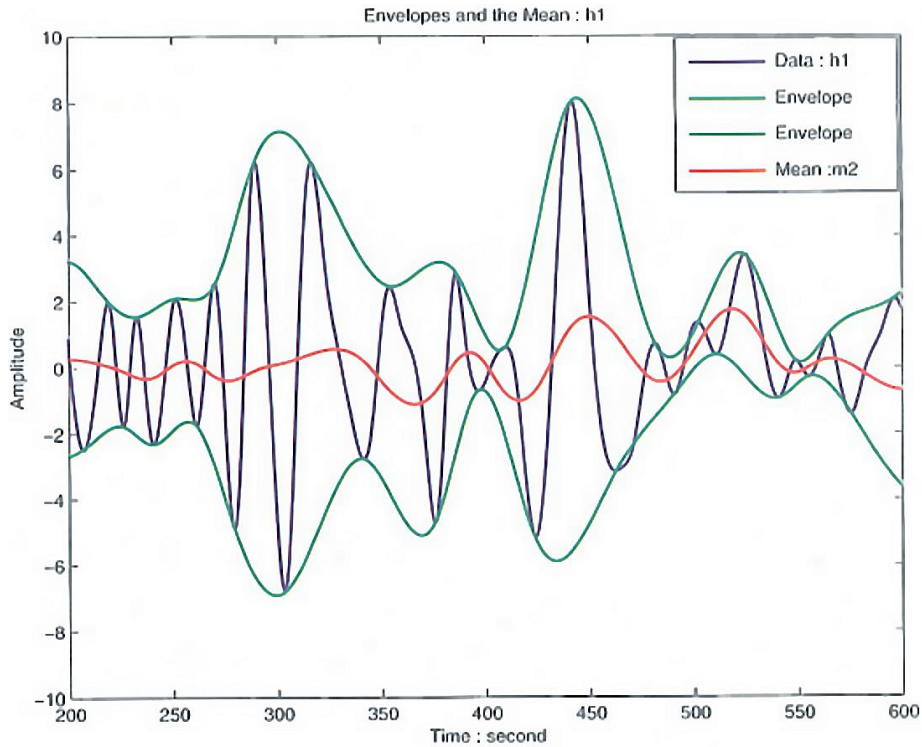
Al utilizar la transformada CWT (Figura 2.37) aparece una discontinuidad secuencial a lo largo del tiempo generado por la alternancia propia de las ondas en si (comprendidas en las ondículas de mayor nivel que representa). Por ello mismo se utiliza como alternativa la transformada CCWT, donde se desarrollan ondículas con valores complejos para poder calcular el módulo de los coeficientes de ajuste de la onda madre, y donde como se observa en la Figura 2.38, se obtienen una mayor resolución de las amplitudes frecuenciales dentro del espacio tiempo-frecuencia.

2.3.2.4 Transformada Hilbert Huang.

Como se comenta en el libro de Huang [55], la transformada de Hilbert Huang (HHT) está pensada para el análisis de las variaciones en señales que no son lineales ni estacionarias. Esta cualidad hace que HHT sea ideal para utilizarse en diversos sistemas de análisis donde las señales a analizar tienen cierta transitoriedad. En el caso de análisis de fallos en motores de inducción se pueden detectar satisfactoriamente daños en las barras del rotor, excentricidades [56] y en cojinetes [57].

Esta transformada realiza una descomposición de una señal en un conjunto de funciones equivalentes aproximadas. Estas funciones se denominan funciones de modo intrínseco (IMF).

Para la detección de cada una de estas funciones IMF se realiza una búsqueda iterativa donde se construye una envolvente superior y otra inferior de la señal original, se calcula la media entre ambas envolventes y se resta esta a la señal hasta obtener la IMF (ejemplo en Figura 2.39). Una vez obtenida la IMF se elimina de la señal original y con la señal residuo resultante se vuelve a iterar el bucle para detectar nuevas IMF. Existen varios criterios para finalizar el algoritmo; bien finaliza cuando la suma de las diferencias pasa un umbral definido, o finaliza cuando los valores que pasan por cero y los valores extremos son prácticamente iguales (se vuelve monótona).

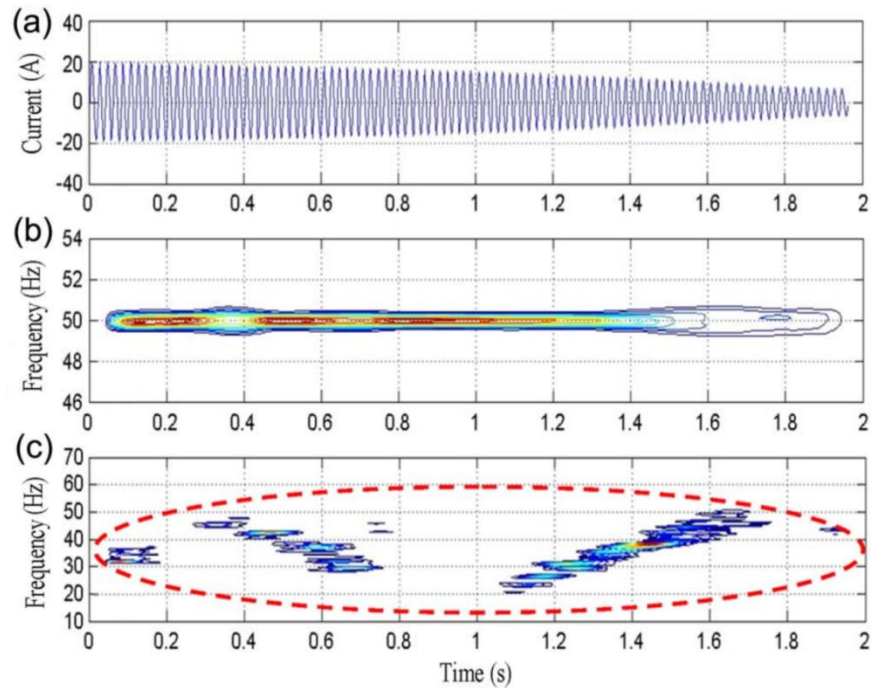


Fuente: Huang y Shen (2005) [55]

Figura 2.39 Señal con una envolvente superior e inferior y el valor medio de las dos envolventes.

Como en el caso de las transformadas en ondículas (apartado 2.3.2.2 y 2.3.2.3), en este caso se tienen en cuenta las funciones IMF que contengan características representativas del fallo a analizar. Para el diagnóstico de dichas características se suele analizar el correspondiente espacio tiempo-frecuencia formado por el espectro de Hilbert-Huang resultante de la IMF en cuestión.

Un ejemplo representativo de descomposición HHT respecto de una muestra de corriente de arranque de un motor con fallo de barra rota se muestra en la Figura 2.40. Como se puede observar, el espectro HH de la IMF2 muestra la característica representativa de fallo por barra rota.



Fuente: Antonino Daviu et al (2009)[86]

Figura 2.40 Representación de los espacios de tiempo-frecuencia del espectro Hilbert-Huang de (b) la IMF1 y (c) la IMF2 obtenidas de la señal original mostrada en (a) con la transformada HHT a partir de la corriente de arranque de una de las fases de un motor de inducción con fallo de barra rota.

2.3.2.5 Distribución de Wigner-Ville.

Como se comenta en Hlawatsch [58], las distribuciones de clase Cohen [59] generan una distribución en tiempo-frecuencia de la densidad de energía de una señal a partir de la susodicha función de Cohen.

El tipo de distribución más simple de las distribuciones de clase Cohen se corresponde con la distribución de Wigner-Ville, donde la función kernel es igual a la unidad. Este hecho se explica con mayor detalle en los documentos de Giri [53] y de Climente Alarcón [60].

En este caso se construye el espacio tiempo-frecuencia calculando de forma local la densidad de energía mediante la fórmula de distribución de Wigner-Ville (2.43) obtenida de [53] y [60].

$$WVD_x(t, \omega) = \frac{1}{2\pi} \cdot \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) \cdot x^*\left(t - \frac{\tau}{2}\right) \cdot e^{-j\tau\omega} d\tau \quad (2.43)$$

En la fórmula que se muestra en (2.43), la variable τ comprende el retraso de la señal y x^* representa al conjugado de x . Además, la integral de esta función se corresponde con la transformada de Fourier respecto de τ en la convolución de la señal, con traslaciones tanto en el tiempo como en frecuencia.

La principal ventaja de esta técnica radica en que su coste temporal es menor en comparación de otros tipos de algoritmos exponenciales con el mismo fin. Además, la resolución del espacio tiempo-frecuencia que muestra es bastante alta.

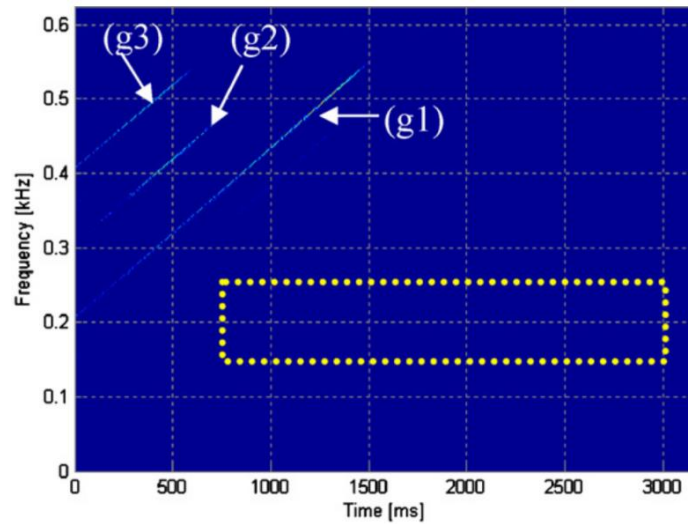
En contraposición, esta técnica se ve afectada por interferencias externas de ruido generando componentes visibles incluso en los momentos en que no debería aparecer ningún flujo de energía. En ciertos casos estas componentes pueden enmascarar parcial o totalmente los patrones característicos de fallo, dificultando la detección de estos.

En las Figura 2.41 y Figura 2.42 que se exponen a continuación, se muestra un ejemplo de generación del espacio tiempo-frecuencia resultante de aplicar la transformación WVD sobre la corriente de fase en el arranque de un motor de inducción sano y sobre la corriente de fase en el arranque de un motor con barras rotas.

En este caso se puede apreciar la notable diferencia cuando se analiza un motor sano (Figura 2.41) y uno con fallo de barra rota (Figura 2.42). El patrón característico de fallo de barra rota se obtiene de la zona delimitada en la Figura 2.42 como $g4$.

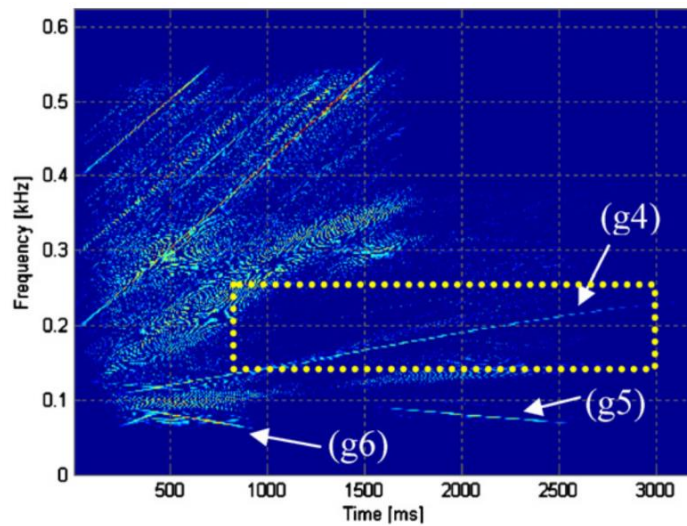
En la Figura 2.41 se muestran tres componentes armónicas $g1$, $g2$ y $g3$ del rotor, normales en un motor sano.

Por otra parte, en la Figura 2.42 para un motor en arranque con fallo de barra rota, además de la componente de fallo $g4$ también se han obtenido dos armónicos descendientes $g5$ y $g6$. Además, en esta misma imagen también se muestran una gran cantidad de términos cruzados (interferencias) sobre los distintos conjuntos de componentes armónicas que podrían llegar a afectar a la detección de la componente armónica de fallo $g4$.



Fuente: Climente Alarcon (2012)[60]

Figura 2.41 Representación de un espacio de tiempo-frecuencia correspondiente a la densidad de energía obtenida mediante la distribución de Wigner-Ville a partir de la corriente de arranque de una de las fases de un motor de inducción sano.



Fuente: Climente Alarcon (2012) [60]

Figura 2.42 Representación de un espacio de tiempo-frecuencia correspondiente a la densidad de energía obtenida mediante la distribución de Wigner-Ville a partir de la corriente de arranque de una de las fases de un motor de inducción con fallo de barra rota.

2.3.2.6 Análisis del Orden Armónico para régimen transitorio (HOTA transitorio).

La siguiente técnica HOTA (Harmonic Order Tracking Analysis) para régimen transitorio es la adaptación de la técnica HOTA visto en el apartado 2.3.1.6 acondicionado para el análisis de fallos en motores de inducción operando en régimen transitorio. Esta adaptación de la técnica se desarrolla principalmente por la sustitución de la transformada FFT (régimen estacionario) por una transformada STFT (régimen transitorio).

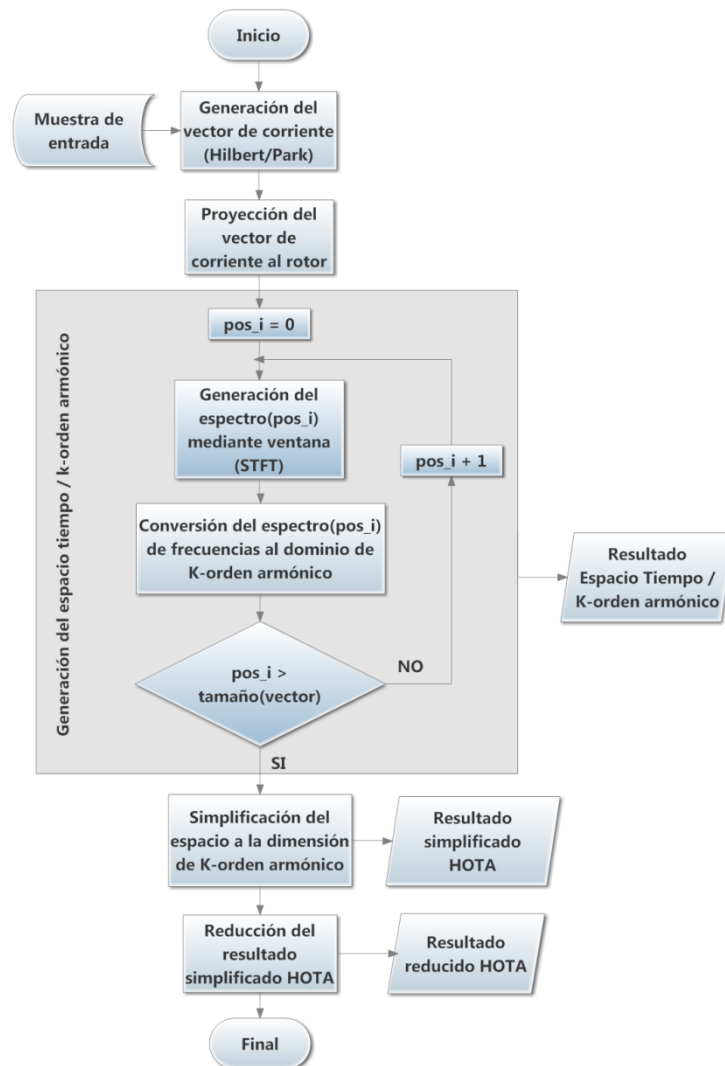


Figura 2.43 Esquema de flujo del funcionamiento de la técnica HOTA para régimen de trabajo transitorio.

Como ya se ha comentado en el apartado 2.3.1.6, tanto el HOTA para régimen estacionario como el HOTA para régimen transitorio es presentado por primera vez por Sapeña (2014)[35].

Al igual que en HOTA para régimen estacionario, como se muestra en el diagrama de flujo del algoritmo referente a la técnica HOTA transitorio en la Figura 2.43, inicialmente se genera el vector de giro de la corriente mediante la transformada Hilbert o Park, a continuación se desarrolla una proyección de las corrientes del estator al rotor (cambio del sistema de referencia desde el estator al rotor).

Con las corrientes proyectadas para cada posición i dentro del vector de corrientes se va desplazando una ventana de filtro (como en la técnica de diagnóstico STFT), generando para esa posición local el espectro resultante. Al igual que ocurría con HOTA estacionario, este espectro de frecuencias se convierte al dominio del k -orden armónico mediante la fórmula (2.32). Como resultado se obtiene un espacio de dos dimensiones formado por el dominio del tiempo y el dominio del k -orden armónico (ejemplo en Figura 2.44 (d)).

Al estar normalizado el espacio tiempo/ k -orden armónico respecto del k -orden armónico, este espacio de dos dimensiones se puede simplificar (promediando los valores en la dimensión del tiempo) a un vector de una dimensión respecto del k -orden armónico, obteniendo un vector resultado de características de fallo compatible con el generado en HOTA estacionario (ejemplo en Figura 2.44 (e)). Al igual que ocurre en HOTA estacionario, este resultado simplificado de HOTA transitorio puede ser reducido para contener solo los datos relevantes (resultado reducido HOTA).

De esta forma, cada valor absoluto del orden k se corresponde con una posible frecuencia característica del fallo generada por dichas componentes armónicas.

La diferencia principal entre HOTA para régimen estacionario y para régimen transitorio radica en que, en el caso del HOTA para régimen estacionario, la obtención del espectro de frecuencias a partir del vector de corrientes y su transformación al dominio de k -orden armónico se realiza una única vez. Sin embargo, en el caso del HOTA para régimen estacionario este proceso se realiza en cada uno de los desplazamientos de la ventana de filtro para cada incremento del tiempo del vector de corrientes. Por tanto, el coste de cálculo de esta técnica para régimen transitorio es mucho mayor que en su equivalente para régimen estacionario.

Aunque, como ya se ha comentado, HOTA transitorio obtiene como resultado un espacio tiempo/ k -orden armónico, este puede promediarse en el dominio del tiempo para obtener el mismo tipo de resultado (un vector de amplitudes ajustadas al k -orden armónico de fallo) que en el caso de HOTA estacionario.

Se utiliza el sistema de conversión a frecuencias de la transformada STFT (apartado 2.3.2.1) por su alta precisión en la representación del espacio tiempo-frecuencia al

utilizarse la función de ventana Gaussiana que garantiza la mayor concentración de energía posible para una señal de duración ilimitada.

Dentro del artículo de Sapena [35] se desarrolla muy claramente un ejemplo representativo de las fases de transformación que realiza este método HOTA transitorio (Figura 2.44).

Como se puede ver en la Figura 2.44 (a), se ha utilizado una muestra obtenidas de una fase de un estator de un motor de inducción trabajando en un rango transitorio de frecuencias entre 40 y 50 Hz.

En el primer paso de HOTA se desarrolla una traslación del sistema de referencia de las corrientes del estator (Figura 2.44 (b)) al sistema de referencia de las corrientes visto desde el rotor (Figura 2.44 (c)).

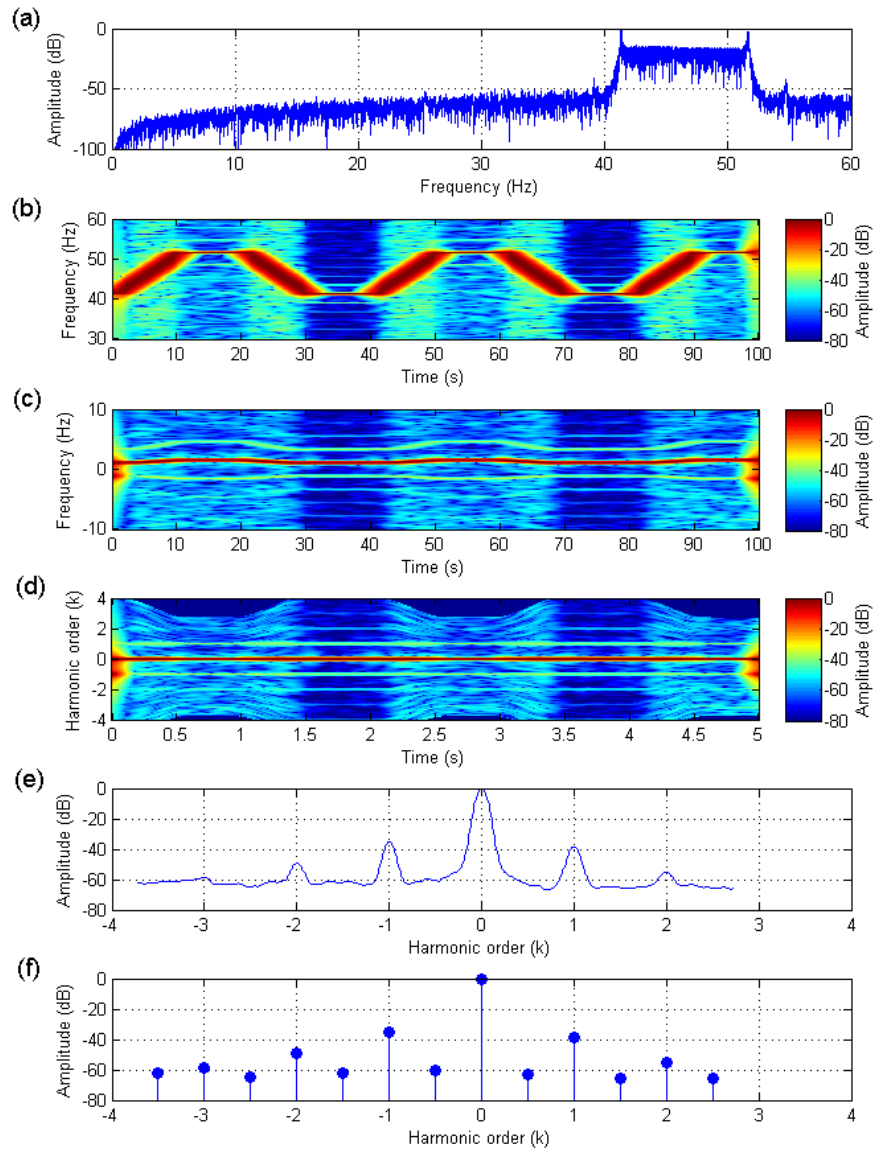
Aunque en la Figura 2.44 (a), (b) y (c) que representa HOTA, se muestra la señal en frecuencias (a) o en el espacio tiempo/frecuencia ((b) y (c)), en realidad el algoritmo no transforma ninguna señal al dominio de frecuencias ni genera ningún espacio de tiempo/frecuencia en (a) y (b) mostradas en la Figura 2.44, las cuales solo se muestran para mejorar de forma visual el entendimiento sobre el proceso de transformación.

Una vez generado el espacio tiempo/frecuencia a partir de las corrientes proyectadas en el rotor (c), se realiza el paso de escalado para cada t tiempo, resultando en un nuevo espacio de características tiempo k -orden (d).

Si bien con este nuevo espacio tiempo k -orden ya es posible valorar la presencia o ausencia de componentes debidas al fallo, como se muestra en la Figura 2.44 (e) se realiza un promedio de las amplitudes a un único vector lo que simplifica todavía más la detección de fallos. Además, como notable ventaja que ya se ha comentado anteriormente, este vector resultante es el mismo que se obtendría con el método HOTA para régimen de trabajo estacionario, por tanto cualquier sistema experto que haga uso de las características de fallo obtenidas mediante HOTA estacionario también es compatible con las características de fallo obtenidas mediante HOTA transitorio.

Finalmente, para minimizar notablemente los requisitos de almacenamiento en memoria y de transmisión de datos, el vector de k -orden armónico se puede reducir hasta contener solo los datos necesarios (normalmente 15 valores) de amplitudes características de fallo y promedios de ruido (Figura 2.44 apartado (f)) pudiendo realizar un diagnóstico de fallos satisfactorio.

Para que este método muestre unos buenos resultados de cara a obtener posibles características de fallo, es esencial para ello desarrollar un espacio tiempo frecuencia STFT con una buena definición respecto a las características de fallo. El ajuste de la técnica STFT para obtener el mejor espacio tiempo/frecuencia representativo de fallos se ha tratado en su respectivo apartado 2.3.2.1.



Fuente: Sapena (2014) [35]

Figura 2.44 Ejemplo de transformación de la corriente de fase a HOTA en régimen transitorio para el diagnóstico de fallos. (a) Frecuencias de la corriente de fase del estator. (b) Espacio tiempo frecuencia STFT (c) Espacio tiempo frecuencia STFT tomando el sistema de coordenadas de referencia del rotor. (d) HOTA transitorio de la corriente de fase. (e) HOTA transitorio simplificado. (f) HOTA transitorio comprimido.

Al igual que en HOTA para régimen estacionario, las componentes características de fallo siempre se muestran en la misma posición independientemente del tiempo. Gracias a esta cualidad de HOTA es mucho más sencillo detectar los patrones de fallo en el espacio tiempo- k orden armónico que los correspondientes patrones de fallo mostrados por otras transformadas como por ejemplo DWT, CWT, CCWT, STFT, Hilbert-Huang y Wigner-Ville.

En resumen, la información resultante (características de fallo) obtenida con la técnica HOTA transitorio aporta varias ventajas como por ejemplo:

- El vector resultante de características comprimido de k -orden (Figura 2.45), es totalmente compatible con el generado por su técnica homónima para régimen estacionario. En consecuencia, cualquier sistema de diagnóstico que utilice el HOTA en régimen estacionario, es compatible con el diagnóstico de fallos en régimen transitorio.
- Este mismo vector resultante de características reducidas (Figura 2.46), concentra toda la información de fallo (componentes armónicas de fallo y promedio del ruido de fondo) en un único vector unidimensional simplificado que mantiene únicamente la información significativa para deducir el posible fallo dentro de un diminuto vector de puntos (Figura 2.46).

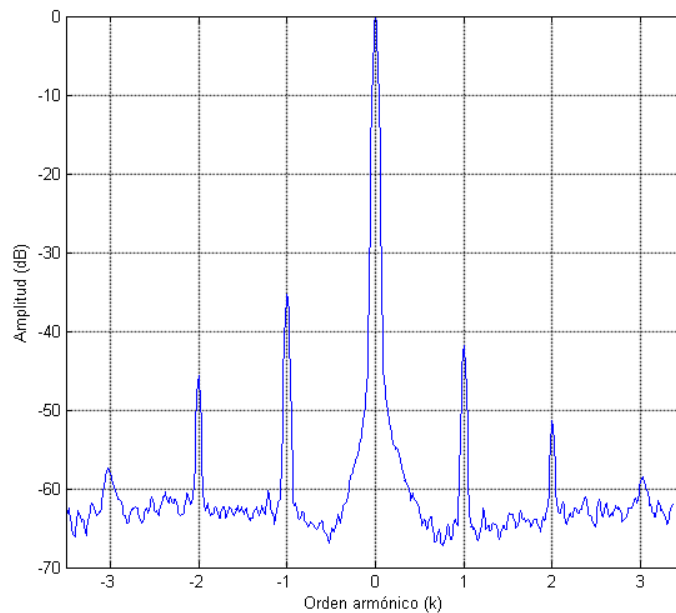


Figura 2.45 Resultado comprimido de HOTA transitorio obtenido a partir de su espacio tiempo-frecuencia de k -orden generado.

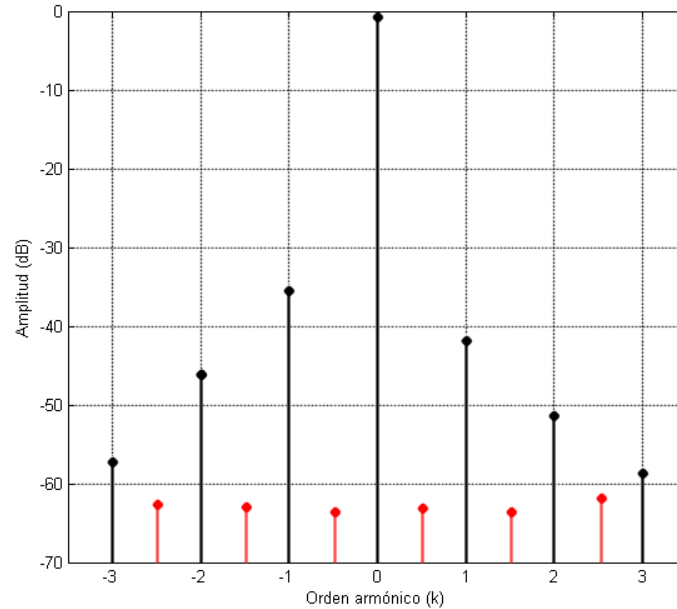


Figura 2.46 Resultado simplificado de HOTA transitorio obtenido a partir del resultado comprimido mostrado en la Figura 2.45. En este caso los valores en negro representan los valores obtenidos para cada componente armónica k mientras que los valores en rojo son valores de media del fondo para conocer el ruido de fondo aproximado de la onda original.

Como se verá posteriormente en el apartado 6.2.1 durante las pruebas experimentales de esta técnica, también tiene sus desventajas, como por ejemplo el coste computacional de calcular el espacio de características tiempo/ k -orden, el gran espacio de memoria requerido durante el desarrollo del espacio tiempo/frecuencia con STFT o la necesidad de ajustar correctamente parámetros de configuración del filtro de ventana de la técnica STFT usada internamente para obtener unas componentes características de fallo muy representativas.

No obstante, estas desventajas de la técnica HOTA para régimen transitorio no son muy diferentes de las desventajas que se dan en las otras técnicas orientadas al régimen transitorio de las analizadas en este apartado 2.3.2.

2.3.2.7 Otras técnicas de detección de fallos en régimen transitorio.

Aunque en el apartado 2.3.2 se ha comentado las técnicas de análisis de fallos más conocidas en el ámbito del diagnóstico de fallos para un régimen de trabajo transitorio, existen otras técnicas que también permiten realizar un análisis satisfactorio sobre diversos fallos en los motores de inducción.

A continuación se exponen algunas de estas otras técnicas basadas en el estudio de la corriente de fase:

- **Instantaneous Frequency (IF) [61].**
- **Polynomial-Phase Transform (PPT)[62].**
- **Diagnosis of faults in the fractional Fourier domain (FrFT) [63].**
- **Teager-Kaiser Energy Operator (TKEO) [54].**
- **Multilayer Park's Vector Approach (MPVA) [29].**

2.3.3 Filtros de la señal.

De forma complementaria a las técnicas de diagnóstico desarrolladas, es posible añadir filtros a estas técnicas de diagnóstico para intentar mejorar la representatividad de las características de fallo obtenidas.

Estos filtros se utilizan normalmente para eliminar efectos no deseados de la onda como por ejemplo la fuga espectral o el ruido de fondo frecuencial.

Existen varios filtros de ventana (Triangular, Hann, Hamming, Bartlett, etc.) para aplicar sobre señales de corriente o para convertir la señal al dominio de frecuencias (FFT, periodograma Welch, periodograma Blackman-Tukey, etc.) con resultados visibles en la detección de características de fallos.

En los siguientes subapartados se desarrollan con más detalle dos filtros destacados en el estado del arte por su uso en la mejora del diagnóstico de fallos [64]. Estos filtros se corresponden con la ventana de Hann y el periodograma de Welch, los cuales son aplicados (en el apartado 5.2) sobre todas las técnicas de diagnóstico para régimen estacionario analizadas en esta tesis.

2.3.3.1 Filtro de ventana de Hann.

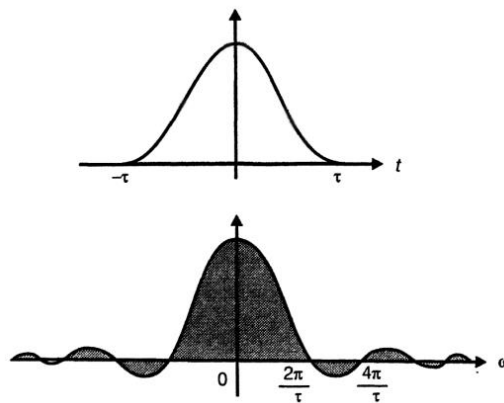
El filtro de ventana de Hann ha mostrado una buena respuesta en trabajos anteriores [24], [65]-[67] de cara a minimizar el efecto del leakage.

Este tipo de ventana genera un filtro en coseno elevado (2.48). En esta ecuación (2.48) la variable n representa el elemento actual filtrado y N se corresponde con el número total de elementos.

Un filtro en coseno elevado es un filtro pulsante de la familia de filtros de Nyquist que reduce al mínimo la distorsión de la señal (ejemplo en Figura 2.47). Este filtro tiene la ventaja de tener una pérdida de resolución muy pequeña y de producir muy poco “alia-

sing” sobre la onda procesada [69]. El “aliasing” se muestra como la superposición periódica de la propia señal al ser muestreada por debajo de la frecuencia de Nyquist.

$$w_{Hann}(n) = \frac{1}{2} \left(1 - \cos \left(2\pi \cdot \frac{n}{N} \right) \right) \text{ siendo } 0 \leq n \leq N \quad (2.48)$$



Fuente: Herrera Perez (2002) [68]

Figura 2.47 Ejemplo gráfico de un filtro de coseno elevado (a) en función del tiempo y (b) en función de la frecuencia.

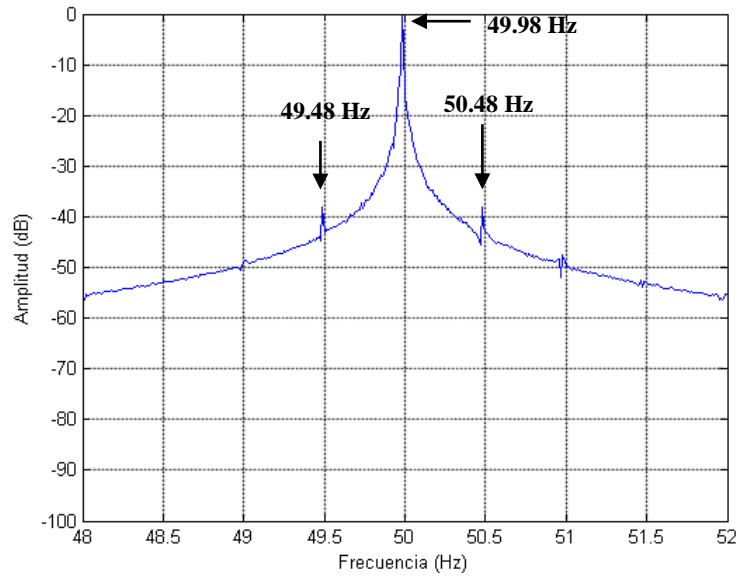


Figura 2.48 FFT de un motor con fallo de barra rota trabajando a una frecuencia de alimentación de 50Hz, donde se aprecia un problema de fuga espectral.

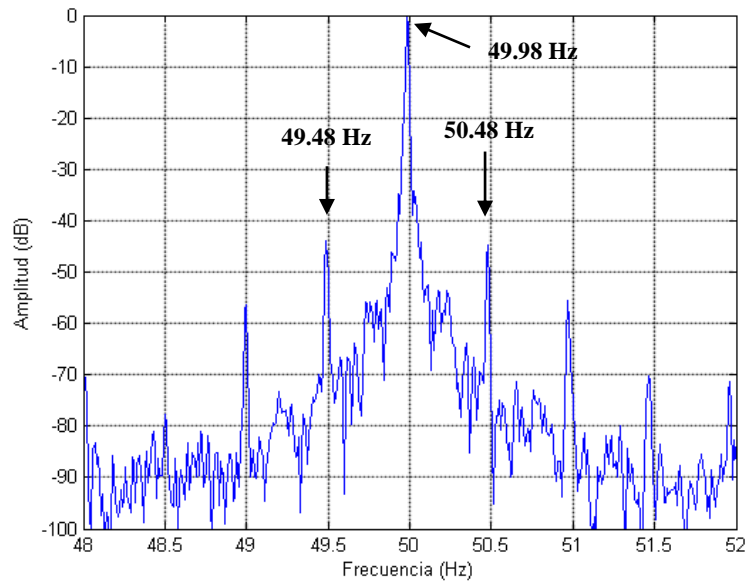


Figura 2.49 FFT de un motor con fallo de barra rota trabajando a una frecuencia de alimentación de 50Hz, donde se ha minimizado la fuga espectral mediante una ventana de Hann.

En la Figura 2.48 y Figura 2.49 se muestra el espectro de frecuencias obtenido con la técnica FFT para una muestra de corriente de un motor con barras rotas. En la Figura 2.49 se ha aplicado una ventana de Hann al espectro de frecuencias, de modo que se puede observar como se ha minimizado el problema de la fuga espectral (leakage).

2.3.3.2 Periodograma de Welch.

Además de la transformada rápida de Fourier (FFT) usada normalmente para la transformación de los vectores de tiempo al dominio de frecuencias, también existen otros métodos de transformación de la señal que obtienen un vector de frecuencias equivalente. Un ejemplo de estas transformaciones alternativas son el estimador del periodograma de Blackman-Tukey [70] y el periodograma de Welch [71].

En el campo del diagnóstico de fallos en motores de inducción, se observa como el periodograma de Welch se ha utilizado como alternativa de forma satisfactoria la transformada FFT en varios estudios [64], [72]-[73].

El periodograma de Welch obtiene una estimación de la densidad frecuencial de la energía de la señal [16]. Como se muestra en la representación de la Figura 2.50, esta transformación se desarrolla dividiendo la señal original en un conjunto de muestras; para cada muestra de este conjunto se genera su espectro de frecuencia y finalmente a todos los espectros de frecuencias de estas subdivisiones se le aplica un factor de solapamiento entre ellos para obtener la estimación media del espectro frecuencial final (periodograma de Welch). De esta forma se disminuye notablemente el ruido de las frecuencias a costa de reducir la resolución de la muestra.

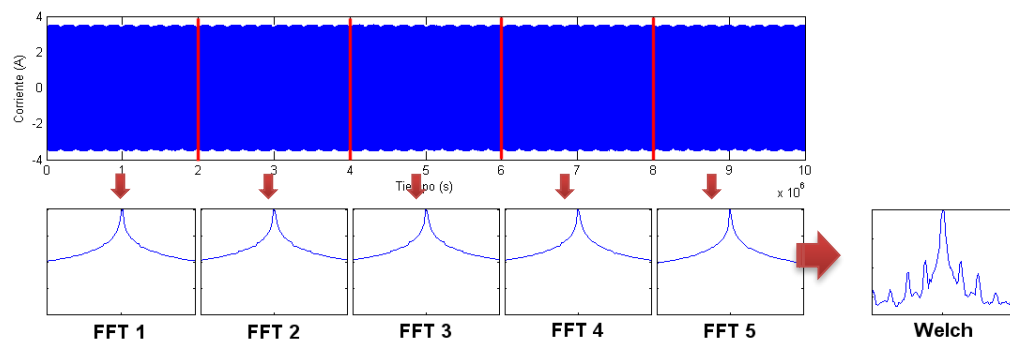


Figura 2.50 Representación del proceso de transformación de una corriente de fase utilizando un periodograma de Welch.

En la Figura 2.51 se muestra el resultado de desarrollar el periodograma de Welch como alternativa a la técnica de diagnóstico FFT. Esta Figura 2.51 se puede comparar con su homónimo mostrado en la anterior Figura 2.48 para la transformada FFT, y en comparación con la anterior Figura 2.49 para la transformada FFT a la que se le ha aplicado una ventana de Hann.

En este caso de la Figura 2.51, el espectro de la señal muestra una disminución notable del ruido de la muestra sacrificando por otra parte la resolución de esta. Además, también se ha minimizado el efecto provocado por la fuga espectral que se mostraba en la Figura 2.48.

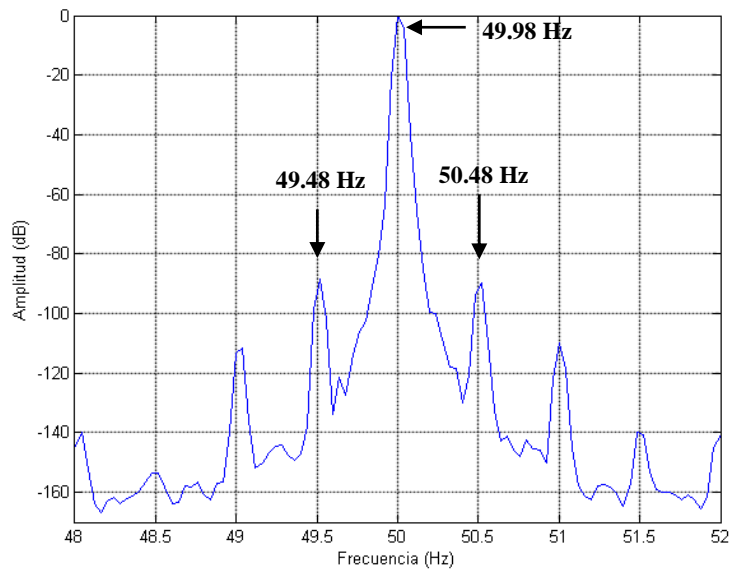


Figura 2.51 Periodograma de Welch de un motor con fallo de barra rota trabajando a una frecuencia de alimentación de 50Hz.

2.4 Sistemas expertos en la detección y clasificación de patrones.

Partiendo de las técnicas de diagnóstico descritas en el apartado 2.3, es posible recopilar un número ingente de tipos de patrones y características identificativas de fallo, obtenidas a partir de las muestras de corrientes de fase de un motor de inducción.

Sin embargo, la identificación del fallo está condicionada a diversos factores que pueden llegar a dificultar ésta identificación de forma correcta, y además es altamente dependiente de los parámetros de los motores, por lo que debe adaptarse cada sistema de detección a la topología del tipo de motor.

Además, en los “momentos iniciales” de la aparición del fallo (fallos incipientes), este puede no ser detectado correctamente al no existir un umbral límite conciso entre la aparición o no de fallo, dándose una pobre precisión en la detección en estos casos. Esos “momentos iniciales”, según el tipo de fallo pueden ser bastante duraderos (desde minutos a incluso meses).

El ruido eléctrico o magnético del entorno también puede dificultar la detección de los fallos, pudiendo ser estos enmascarados o realzados produciéndose en consecuencia una detección deficiente.

Por ello, se busca optimizar las técnicas de diagnóstico para maximizar la precisión de estos. La búsqueda de dicha optimización se realiza, últimamente, con el uso de **sistemas expertos** que mejoran y automatizan las técnicas de diagnóstico.

Estos sistemas expertos contienen técnicas para la resolución de problemas de forma similar a la desarrollada por una inteligencia humana, como por ejemplo el reconocimiento del habla [74] o la medición estructural aproximada del cerebro mediante resonancia magnética [75].

Algunas de las ventajas de estos sistemas expertos son:

- Adaptabilidad del sistema experto a cada problema de forma local, por lo que es posible ajustar el sistema de diagnóstico para diversos tipos de motores y fallos.
- Algunos sistemas expertos usan correlaciones pseudo-estadísticas, de forma que si está diseñado eficientemente aumenta la probabilidad de diagnosticar correctamente la mayoría de casos donde la detección de fallo es dudosa, con lo que se consigue una mayor precisión en la detección de fallos leves o incipientes.
- El coste temporal de la ejecución del sistema experto es únicamente dependiente del coste intrínseco de su algoritmo, estructura y al sistema computacional asociado. Con la tecnología actual, es factible el uso de sistemas expertos en los casos en que sea necesario realizar el diagnóstico en tiempo real;

aunque en la mayoría de sistemas de diagnóstico el diagnóstico en tiempo real no es un requisito indispensable.

- Tanto el aprendizaje como la ejecución del sistema experto puede realizarse de forma automática, evitando la necesidad de interacción con un operador.

Los primeros sistemas expertos aparecen en la década de los 60 del siglo 20, donde ya se reconoce su potencial en proyectos de envergadura como el “General Problem Solver” para la resolución de problemas de criptoaritmética [76] o el proyecto DENDRAL para la identificación automática de estructuras químicas moleculares y desarrollo de hipótesis conforme a los resultados [77].

Con todas sus ventajas, el uso de sistemas expertos se orienta a muchos campos distintos de diversa índole (manufactura, medicina, seguridad, banca, derecho, defensa, etc.), pero sobretodo en el grupo de las ingenierías (aeroespacial, automoción, electrónica, informática, industria, etc.).

En las siguientes secciones se comentan en general los conjuntos formados por los diversos tipos de sistemas expertos destinados al reconocimiento de patrones más conocidos y extendidos en el ámbito de las ingenierías. Posteriormente en el capítulo 3 de esta tesis se exponen con mayor detalle los sistemas expertos más usados para el diagnóstico de fallos en máquinas eléctricas de inducción.

2.4.1 Sistemas expertos basados en deducción lógica.

Esta clase de sistemas expertos es el primero desarrollado y por tanto el más estudiado dentro del campo de la Inteligencia Artificial. Cuando se hace referencia a “sistemas expertos” se puede entender esta denominación como sinónimo de Inteligencia Artificial en general o se puede referir a este conjunto de sistemas expertos clásicos en especial, agrupados en este apartado 2.4.1.

Como se puede observar al comparar los apartados 2.4.1.1 y 2.4.1.2, la estructura y el funcionamiento de este tipo de sistemas expertos varía de uno a otro, pero siempre existe un sistema que resuelve problemas de forma *inteligente* compuesto por:

- Una base de conocimientos lógicos inicial, intrínseca al sistema experto, adquirida por un aprendizaje previo y que en algunos sistemas expertos puede ser modificada posteriormente por el propio sistema para perfeccionar su *conocimiento* sobre el problema,
- Una base de reglas y/o casos lógicos definidos que representan las relaciones existentes entre los conocimientos de la base de conocimientos (datos relacionados).

- Una función de inferencia que a partir de los datos que recibe el sistema experto por sus entradas activa las reglas correspondientes para que, a partir de la base de conocimientos actual, desarrollar la resolución del problema.

A continuación se comentan brevemente algunos de ellos.

2.4.1.1 Lenguajes de programación lógica.

Existen varios lenguajes de programación ideados para el desarrollo de este tipo de sistemas expertos, como por ejemplo el conocido Prolog (Programmation en logique) creado por los investigadores Alain Colmerauer y Philippe Roussel a principio de los años 70 del siglo XX [78] y CLIPS (C Language Integrated Production System) desarrollado por la sección de Inteligencia Artificial de la NASA en 1984.

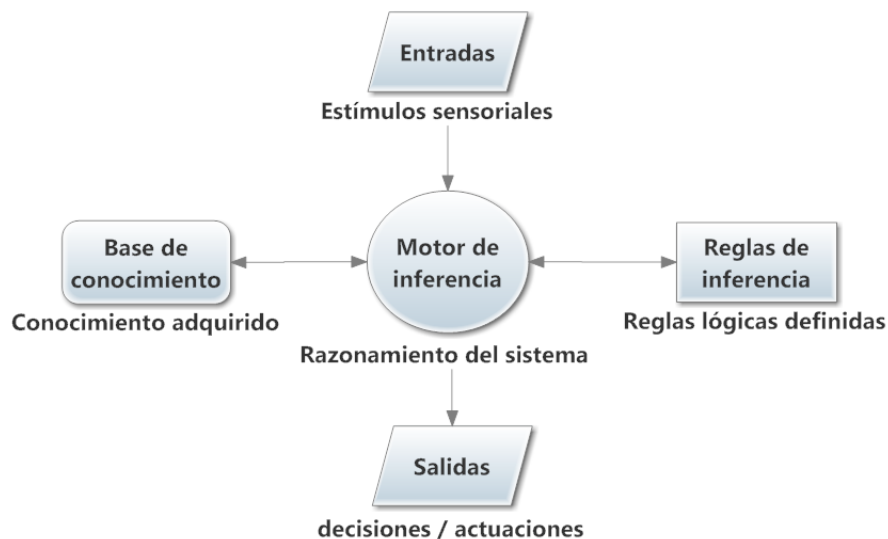


Figura 2.52 Esquema de un sistema experto compuesto por reglas.

Como se representa en la siguiente Figura 2.52, estos lenguajes de programación se estructuran en:

- Una **base del conocimiento** inicial donde se incluye tanto el conocimiento actual como el que va a adquirir en el futuro. En este caso se corresponde con una estructura bien definida donde los conocimientos son catalogados respecto a su tipo y propiedades.

- Un conjunto de **reglas de inferencia** que denota el razonamiento lógico dado por el sistema. En este caso las reglas de inferencia son reglas lógicas de tipo Modus Ponendo Ponens o de tipo Modus Tollendo Tollens [79].
- Las **entradas y salidas** de este sistema se representan a su vez como recolectores sensoriales (datos de entrada) y actuadores (resolución dada por el sistema) del entorno. Los datos obtenidos por la entrada son usados por el motor de inferencia tanto para resolver el problema de forma inteligente como para actualizar su base de conocimientos.

2.4.1.2 Árboles de decisión.

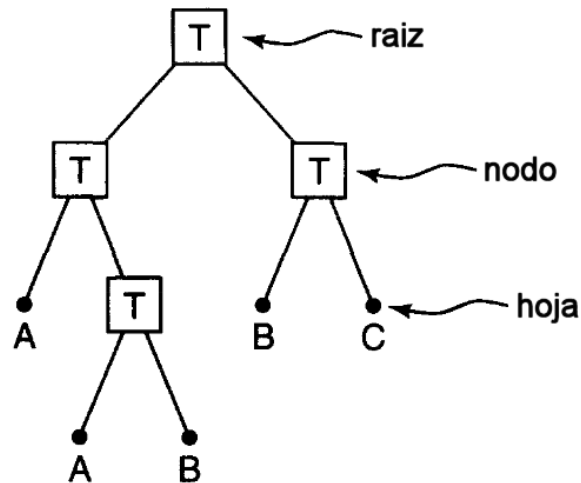
Además de este tipo de representación de sistema experto de aprendizaje lógico visto en la Figura 2.52, existen otros tipos diversos de representaciones, que aunque mantienen la misma esencia lógica, se estructuran de forma más acorde con la naturaleza del problema.

Un ejemplo claro de estas alternativas se observa en los modelos predictivos basados en árboles de decisión. Estas estructuras en árbol están orientadas o bien a la descripción de datos (clasificación) o a la toma de decisiones.

En un árbol de decisión, el conocimiento y las reglas lógicas se encuentran intrínsecos en la estructura del sistema, donde cada división de las ramas representa una nueva concreción hacia la decisión óptima. Dependiendo del conocimiento y los estímulos aportados al sistema, se desarrolla un camino dentro del árbol hasta llegar a la rama final donde se encuentra la salida óptima (decisión) para ese conocimiento y estímulos en concreto.

Por ejemplo, en la Figura 2.53 se muestra un esquema de árbol lógico donde existe un nodo T de un *conocimiento* más genérico que se divide en otros nodos T de conocimientos derivados del anterior pero más específicos. La selección de una u otra rama del conocimiento T depende de los datos de la entrada del sistema experto. El último nodo de la rama (nodo hoja) corresponde finalmente al conocimiento o decisión lógica a la que debe llegar el sistema experto.

Una ventaja de este tipo de sistema experto es que no es necesario definir de forma explícita unas reglas de inferencia, sino que este árbol puede ir generándose a partir de unas muestras representativas del problema. Como contrapartida, la mayor parte del conocimiento sobre el problema (y en algunos casos todo este conocimiento) lo adquiere el sistema experto en el desarrollo inicial del árbol de decisión, siendo posteriormente mucho más complejo de actualizar dichos conocimientos, pudiendo ser necesario incluso la reconstrucción completa del árbol de decisión.



Fuente: Friedl y Brodley (1997)[80]

Figura 2.53 Topología de un árbol de decisión.

2.4.2 Sistemas expertos basados en espacios de características.

Los siguientes sistemas expertos son otra alternativa a la clasificación de patrones de forma inteligente.

En este caso, el sistema requiere de un conjunto formado por los datos de entrada del problema a resolver, de estos datos se tiene que poder obtener alguna información relevante o representativa con la que sea posible resolver el problema, por ello se le da el nombre de características representativas del problema a estos datos de entrada.

Los sistemas expertos basados en espacios de características son sistemas discriminantes donde el resultado (decisión del sistema) se corresponde a una opción dentro de un conjunto discreto de opciones.

Cada una de estas opciones es una clase, especificándose en este caso como a una catalogación de los posibles resultados. Por ejemplo en un sistema de detección de fallos el conjunto de clases puede estar formado por la clase 'motor sano' y la clase 'motor con fallo'.

La diferenciación de características y concreción de clases se efectúa dentro de un espacio virtual definido por cada uno de los tipos de características específicas para la resolución del problema. Cada propiedad de la característica se representa como una dimensión de este espacio. Aunque en teoría no hay un número límite de dimensiones que puede tomar el espacio, en la práctica por límites de espacio y de tiempo de cómputo se seleccionan solo las propiedades más relevantes de las características.

Dependiendo del sistema clasificatorio, se sigue una función de decisión para dilucidar la clase a la que pertenece la muestra a procesar.

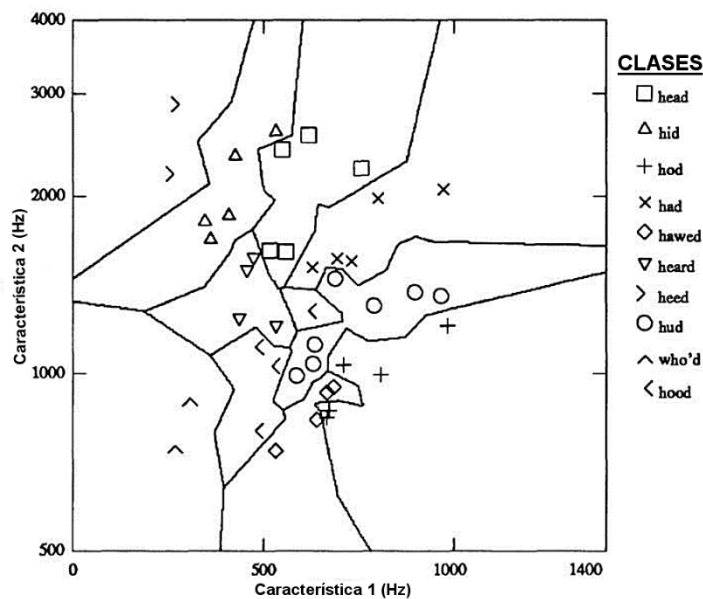
Este grupo de sistemas expertos está compuesto principalmente por los sistemas de clasificación k -NN (k -Nearest Neighbors), SVM (Support Vector Machine), y por todas las variaciones y ajustes desarrolladas para mejorar la clasificación en diversos subgrupos de problemas.

En todo caso, estos sistemas expertos requieren de un aprendizaje supervisado anterior a su uso, partiendo de un conjunto de muestras representativas del problema cuyos resultados son conocidos.

2.4.2.1 K -vecinos más cercanos (k NN).

El sistema experto k -NN fue presentado por Silfman (1951) [81], exponiéndose como un sistema novedoso para el análisis discriminante no paramétrico. Como ya se ha comentado, esta clase de sistemas expertos genera un espacio n -dimensional de características donde el número de dimensiones es igual al número de características.

En la Figura 2.54 se puede ver un ejemplo de representación del espacio bidimensional de un sistema experto k -NN para un problema específico con dos características de entrada.



Fuente: Chang y Lippmann (1997)[82]

Figura 2.54 Espacio de características de un clasificador k -NN donde se muestran los límites de decisión entre clases para un problema de reconocimiento del habla.

En este espacio de la Figura 2.54 se observan elementos de 10 tipos diferentes de clases (head, hid, hod, had, hawed, heard, heed, hud, who'd y hood), donde cada elemento dentro del espacio de características está formado por la característica 1 y la característica 2 (2 Dimensiones). Además, en el espacio de características también se han dibujado las fronteras que separan cada zona correspondiente a una de las clases del resto de zonas de las demás clases.

Inicialmente, dicho espacio es rellenado por el conjunto de muestras de aprendizaje cuya clase es conocida, ya que para la clasificación de las nuevas muestras se tienen en consideración la posición dentro del espacio de las muestras de aprendizaje.

La clasificación de la muestra a resolver depende de la clase a la que pertenecen los ' k ' vecinos que se encuentran más cerca de esta. De este modo, como se observa en la Figura 2.54, se genera una frontera virtual que separa el espacio entre las diversas clases.

En este clasificador se puede adaptar el número de k -vecinos y la función del cálculo de distancia para mejorar el acierto en la clasificación.

Como principal ventaja el número de clases distintas soportadas por este mismo clasificador es prácticamente ilimitado y su entrenamiento es muy simple ya que únicamente es necesario modificar el conjunto de muestras de entrenamiento de dicho clasificador.

Por contrapartida, la misma facilidad para el entrenamiento aporta como desventaja en que, cuanto mayor es el conjunto de muestras de entrenamiento, más pesado y lento se hace el clasificador.

2.4.2.2 Máquina de vectores de soporte (SVM).

Al igual que en el caso del clasificador anterior (apartado 2.4.2.1), este clasificador genera un espacio n -dimensional de características donde el número de dimensiones es igual al número de características. La primera versión conocida del clasificador fue desarrollada por Cortes y Vapnik en los laboratorios de la empresa AT&T en 1995 [83].

Como se observa en la Figura 2.55, dentro de este tipo de clasificadores se calcula en la fase de entrenamiento un hiperplano que, a similitud de una frontera, divide el espacio en zonas. A cada zona se le asigna la clase (resultado) a la que corresponde la mayoría de características representativas que se encuentran dentro de esta zona.

El hiperplano en si se considera como un espacio de una dimensión menos que el espacio que lo contiene, de forma que este hiperplano puede actuar como frontera de separación del espacio que lo divide en las dos zonas (separación de 2 clases). Por ejemplo si el espacio es de 2 dimensiones el hiperplano será una línea (1 dimensión), si el espacio es de 3 dimensiones el hiperplano será una superficie de 2 dimensiones, etc.

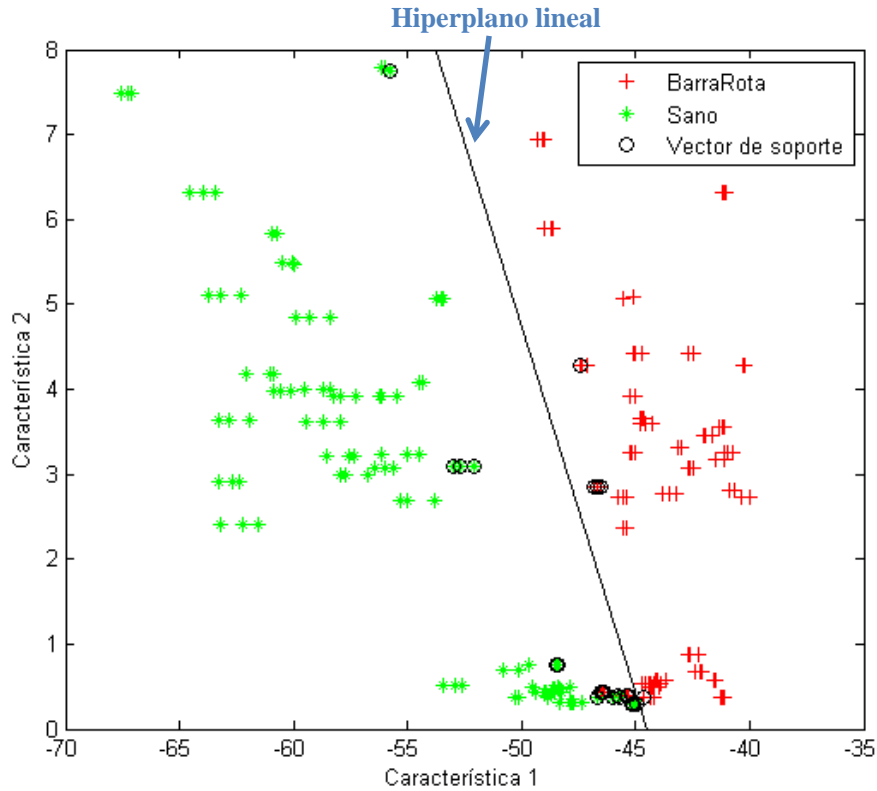


Figura 2.55 Espacio de características de un clasificador SVM donde se observa el hiperplano discriminante separando las muestras de cada clase.

Para la resolución posterior de la clasificación de las muestras desconocidas, se le asigna a esta muestra la clase asignada en la fase de entrenamiento a la zona del espacio donde este el punto correspondiente a las características de la muestra desconocida.

En la siguiente Figura 2.56 se muestra un ejemplo de un espacio bidimensional (2 características) para la resolución de un problema de clasificación. A diferencia de la Figura 2.55, en esta imagen se puede ver claramente como el hiperplano (que en teoría tendría que ser plano) forma una curva divisora de clases que divide el espacio en dos zonas de forma que cada zona contiene el mayor número de muestras correspondientes con la clase de esa zona.

Esta curvatura se calcula en la fase de entrenamiento del hiperplano separador (no lineal) para poder separar mejor las zonas correspondientes a las características de dos clases distintas que no pueden ser separadas mediante un hiperplano lineal.

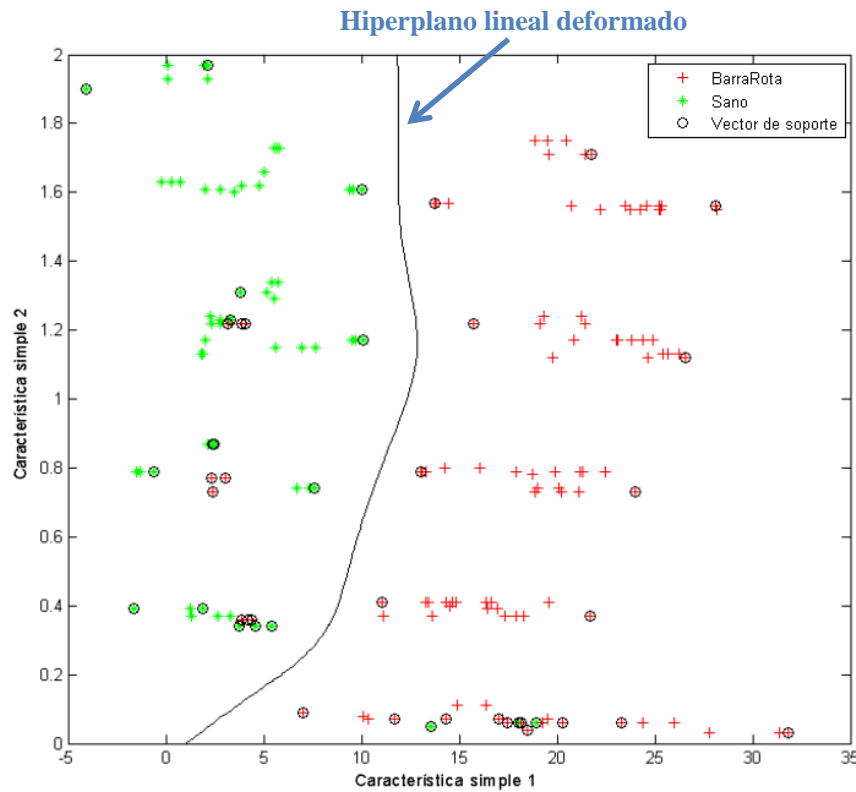


Figura 2.56 Espacio de características de un clasificador SVM donde se observa el hiperplano discriminante y una serie de muestras a las que se les ha asignado una u otra clase dependiendo de la zona en que se sitúan.

En su aprendizaje, se usa un conjunto de muestras de entrenamiento (características representativas) de las que se conoce el resultado correcto del problema (clase). Con este conjunto de muestras representativas se construye el espacio de características.

Como se conoce la clase a la que pertenece cada una de estas características, se ajusta el hiperplano dentro del espacio (entrenamiento) del modo que finalmente queden separadas lo más fielmente las características de una clase respecto de las características de otra clase distinta. Dependiendo de la linealidad del conjunto de datos, se puede personalizar la estructura espacial del hiperplano (su forma) para mejorar la frontera de separación de las clases.

El clasificador con una línea de segregación (no deformado, ver Figura 2.55) representa la versión más sencilla de estos sistemas de clasificación en un hiperplano, donde la línea discriminante divide los datos en dos subzonas asociadas a cada clase.

El clasificador con línea de segregación actúa en todo caso igual que un perceptron simple, definido por McCulloch [84]. El perceptron simple es un sistema experto representado por una red neuronal con solo una capa de neuronas de salida.

Para obtener la curvatura del hiperplano que divide el espacio no se actúa sobre este directamente. El hiperplano inicial es siempre un hiperplano lineal, mientras que el espacio de características es deformado (por ejemplo curvado) mediante unas funciones de transformación llamadas kernels. La deformación del espacio se lleva a cabo para que las diversas clases de características distintas puedan ser separadas lo mejor posible con el hiperplano lineal. Una vez se ha encontrado la mejor configuración de este hiperplano para ese espacio deformado, se aplica la transformación inversa del kernel para convertir el hiperplano lineal del espacio deformado en un hiperplano no lineal dentro del espacio de características original.

Para la resolución posterior de nuevas características (resolución del problema), solo es necesario conocer la configuración del hiperplano dentro del espacio. Por tanto el tamaño de este tipo de sistema experto no es dependiente de las muestras de entrenamiento, y en consecuencia el sistema experto resultante es de tamaño reducido. Por contrapartida, esta misma cualidad dificulta la implementación de detección de más de dos clases; además, en el caso de que el sistema experto necesite ser actualizado con nuevo *conocimiento* del problema se tiene que volver a entrenar todo el sistema con todo el conjunto de muestras para desarrollar el nuevo hiperplano discriminante.

El desarrollo de los sistemas expertos SVM se verá con mayor detalle en el apartado 3.2 y su aplicación al problema abordado en esta tesis se verá en los apartados 5.4 y 6.3.

2.4.3 Sistemas expertos basados en emulación biológica (Redes Neuronales Artificiales).

El sistema experto con un mayor potencial está representado por el cerebro humano, por tanto, es lógico que exista un interés general en intentar emular dicha funcionalidad de forma que se tenga un sistema experto con una excelente capacidad de resolución de problemas no lineales.

Se considera un problema como lineal cuando las restricciones sobre el conjunto de características o la función objetivo a maximizar son lineales. En el caso de que alguna de estas restricciones o la función objetivo no fueran lineales el problema se considera como no lineal.

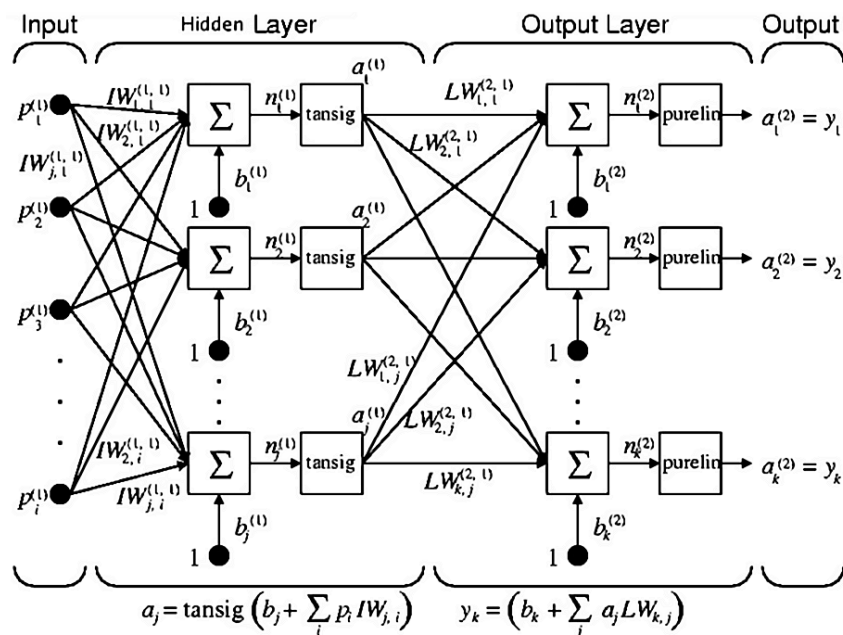
Actualmente, existen dos sistemas de emulación de redes neuronales basados en su funcionamiento físico y en su funcionamiento conceptual. Sin embargo, como sistema experto eficiente se usa únicamente el modelado conceptual, relegando la emulación física de este órgano al estudio médico de la biología humana.

La primera conceptualización del funcionamiento de una red neuronal fue desarrollada por los biofísicos McCulloch y Pitts (1943)[84]. Aunque esta aproximación conceptual ha sido perfeccionada hasta nuestros días, se ha mantenido toda la base de elementos y funcionalidades que aparecen en ese trabajo inicial.

El principal escollo en la evolución de estos sistemas expertos se ha debido a su dificultad para su entrenamiento, ya que su ajuste paramétrico y estructural es el menos intuitivo de todos en comparación con los sistemas expertos citados en el apartado 2.4.1 y 2.4.2.

Ha sido a partir del año 1982 cuando, con el surgimiento de algoritmos de aprendizaje de redes neuronales basados en retropropagación del gradiente, se ha comenzado a mostrar el potencial de este tipo de sistemas expertos; extendiéndose y usándose hasta la actualidad en un número mayor de ámbitos diferentes.

A similitud del sistema biológico, el modelo conceptual representado en la Figura 2.57 se compone de una especie de red compuesta por un número determinado de neuronas interconexas. Además, algunas de estas neuronas tienen configuradas sus dendritas (entradas al sistema experto) o axones (salidas del sistema experto) como interfaz para la comunicación con el exterior.



Fuente: Demuth et al (2008)[85]

Figura 2.57 Esquema de una red neuronal artificial Perceptron multicapa.

En el funcionamiento de esta clase de redes, el estímulo de las entradas provoca una respuesta en la salida del sistema promovido en general por el flujo de transferencia de datos por la conexión entre capas; modelado dicho flujo también dentro de cada capa por los correspondientes pesos, función umbral y función de transferencia específico a cada neurona en donde circula el flujo de datos.

El número mínimo de entradas del sistema dependen de los patrones de información de las muestras (propiedades de las características representativas), pudiendo añadirse nuevas entradas de retroalimentación con información o soluciones anteriores. El número de salidas del sistema es definido por el tipo de respuesta del problema requerido y virtualmente no existe un máximo definido, además estas pueden ser tanto de tipo lógico como paramétrico.

La versatilidad de estas entradas y salidas hace que estos sistemas expertos sean estu-
pendos para la resolución de problemas de diversa índole, como por ejemplo el ajuste y aproximaciones de funciones, reconocimiento y clasificación, y de funciones de agrupación de elementos.

En el caso concreto del problema de la clasificación de características de fallos, estos sistemas expertos permiten el diagnóstico de innumerables fallos distintos pudiéndose identificar también al mismo tiempo el nivel de gravedad de estos. Además, su sistema de resolución de problemas no lineales es mayor a todos los sistemas expertos clasificatorios comentados anteriormente.

Sin embargo, no todo son ventajas, siendo la principal desventaja el algoritmo de aprendizaje, el cual es actualmente factible pero su coste computacional y de espacio de almacenamiento en memoria es elevado.

Aunque algunas versiones de redes neuronales pueden aprender a medida que van evaluando muestras, en el caso de la problemática de la clasificación es más eficiente una red con un entrenamiento previo. En consecuencia, al igual que en los sistemas expertos SVM (apartado 2.4.2.2), la actualización de las muestras de aprendizaje requiere un nuevo entrenamiento completo de todo el sistema experto.

Como se desarrolla posteriormente en el capítulo 3, los sistemas expertos más extendidos de cara al diagnóstico de fallos son actualmente las máquinas de soporte de vectores SVM (apartado 2.4.2.2) y las redes neuronales artificiales (apartado 2.4.3).

El desarrollo de los sistemas expertos ANN se verá con mayor detalle en el apartado 3.3 y su aplicación al problema abordado en esta tesis se verá en los apartados 5.5 y 6.3.

Capítulo 3

Sistemas expertos orientados a la clasificación

3.1 Introducción

La necesidad de discernir sobre posibles resultados en un problema de clasificación/discriminación es una ardua tarea que necesita de un profundo análisis del problema a tratar.

Dentro del marco del diagnóstico de fallos en los motores de inducción, el problema de detección de fallo se trata de un problema de discriminación de aparición/no aparición del fallo. En consecuencia, es necesario utilizar una buena técnica para realizar la detección de forma correcta.

Con un buen sistema de clasificación se puede llegar a detectar el fallo incluso en su fase temprana de aparición, mientras que en caso contrario, con un sistema de clasificación deficiente se puede llegar a errar en la detección en los casos menos claros, aumentando en consecuencia la tasa de aparición de problemas relacionados con dicho fallo. En consecuencia es de vital importancia utilizar un sistema clasificatorio óptimo con respecto al problema.

En sus orígenes, el principal sistema experto utilizado para la detección del fallo se correspondía con la del operario humano entrenado con el conocimiento necesario. Actualmente, este agente se va sustituyendo por complejos sistemas artificiales que denotan cierta inteligencia, ya que, en problemas repetitivos con límites acotados han demostrado su superioridad. Esto es debido a que, al centrarse en un único problema, el

sistema artificial puede ser entrenado hasta llegar a alcanzar un rendimiento casi perfecto respecto al susodicho problema.

Existen varios sistemas artificiales expertos que usan su propia técnica, como por ejemplo los que actúan mediante la implementación de lógica asociativa o por medio de aproximaciones físicas (neuronas). Pero en general todos ellos desarrollan una aproximación del comportamiento de la inteligencia humana.

Observando las últimas líneas de investigación [1], [87]-[89], se puede comprobar una clara orientación hacia las técnicas de diagnóstico cuyo discriminante es un sistema de “Maquina de vectores de soporte” (SVM) y a los sistema expertos de “Redes neuronales artificiales” (ANN). Cada una de estas técnicas tiene sus ventajas y desventajas por lo que dependiendo de los requisitos del sistema se puede seleccionar la más adecuada.

Por tanto, el objetivo de este capítulo se ha centrado en explicar con un mayor detalle la estructura y el funcionamiento de los sistemas expertos que se han utilizado finalmente en esta tesis.

En la primera parte del capítulo se desarrolla el sistema experto basado en Máquinas de Vectores de soporte (SVM). Se comenta su funcionamiento interno, su metodología de clasificación y su método de entrenamiento incluyendo la optimización de coeficientes para maximizar la tasa de aciertos del sistema experto.

En la segunda parte de este capítulo se desarrolla el sistema experto basado en Redes Neuronales Artificiales (ANN). Para esta clase de sistema experto se comenta igualmente su estructura y funcionamiento interno, su método de entrenamiento y posibles ajustes y optimizaciones para maximizar la tasa de aciertos del sistema experto.

3.2 Máquina de vectores de soporte (SVM).

La máquina de vectores de soporte, o “Support Vector Machine” (SVM) en inglés, es un sistema clasificatorio por discriminación por medio de un hiperplano puro (que en los casos de esta tesis equivale a un plano en 2 dimensiones) que divide el *espacio de características n-dimensional* (explicado en el apartado 3.2.1). La primera versión de este clasificador fue publicada por Cortes y Vapnik [83].

Este tipo de clasificador es muy recomendado para su utilización en sistemas que requieren de pocos recursos de cantidad de memoria y de cálculo computacional.

La principal desventaja de este clasificador recae en que únicamente es capaz discriminar entre dos **clases**. En este caso una clase representa una respuesta que puede dar el sistema experto, siendo como mínimo imprescindible para realizar el diagnóstico de fallos dos clases (Motor con fallo, Motor sano).

A continuación se definen algunos de los conceptos que se usan en el sistema experto:

- **Característica representativa del fallo.** Este concepto está asociado a un vector de datos, de dimensión 1 o mayor, con información que puede ser relevante para detectar el fallo. Esta información relevante de fallo se puede usar para el entrenamiento del sistema experto (si se conoce la clasificación final de este) o para la clasificación (diagnóstico de fallo / no fallo).

Cuando este vector de datos **característica representativa del fallo** tiene una longitud mayor que 1 se puede decir que es una **característica representativa compuesta** por unidades de **características simples**. Al menos una de estas características simples dentro del vector tiene que ser representativa del fallo, pudiendo ser el resto también representativo del fallo o no representativas pero que igualmente sean significativas para mejorar la detección del fallo.

Cuando se habla de característica de entrada del sistema experto sin más, se está indicando la característica representativa de fallo compuesta.

En el caso concreto del diagnóstico de fallos en motores, la característica representativa de fallo compuesta se obtiene a partir de las muestras de mediciones obtenidas de los motores en funcionamiento. No hay que confundir que, aunque las características representativas de fallo compuestas se obtienen de las muestras, las muestras no son las características representativas del fallo.

- **Clase.** Una clase se corresponde con un valor discreto que caracteriza a una de las posibles soluciones del sistema experto. Por ejemplo, en el caso de los sistemas de diagnóstico existen como mínimo dos clases distintas (dos posibles soluciones del problema), puede ser de **con fallo** (clase 1) o **sin fallo** (clase 2).
- **Espacio de características de N-dimensiones.** El espacio de características se representa como un espacio donde cada una de sus N dimensiones totales corresponde al tipo de característica simple que contiene el vector de característica representativa de fallo.

Por ejemplo, una característica representativa de fallo compuesta de tres características simples formará un espacio de tres dimensiones, mientras que una característica representativa de fallo compuesta de diez características simples formará un espacio de diez dimensiones.

El espacio de características, puede referirse al espacio de características de entrada formado por las características iniciales obtenidas del problema, también llamado espacio de entrada; o en el caso de que se ha aplicado un kernel de transformación se puede referir a un espacio de Hilbert (que es una generalización del espacio euclideo, es decir, el espacio).

- **Espacio de Hilbert.** Como se explica en el desarrollo de García [90], el espacio de Hilbert es una generalización del **espacio euclideo elevado a un número infinito de dimensiones.**

Por ejemplo, un espacio euclideo se puede definir por su producto escalar (producto interno) mediante la fórmula (3.1) para dos vectores x e y de d dimensiones. A partir de este producto interno se puede desarrollar otros conceptos del espacio euclideo como la ortogonalidad y el ángulo entre vectores (teorema de Pitágoras, la identidad del paralelogramo, la proyección ortogonal, etc.)

$$x \cdot y = x_1 \cdot y_1 + x_2 \cdot y_2 + \dots + x_d \cdot y_d \quad (3.1)$$

Este espacio euclideo puede generalizarse a un espacio de infinitas dimensiones (reales o complejas) definible por su producto interno de infinitas dimensiones. Si el espacio de infinitas dimensiones cumple una norma (L^1 norma, L^2 norma, ∞ norma) se habla de un espacio normado.

El espacio de Hilbert es un espacio particular del espacio normado que cumple la L^2 norma, donde la convergencia de esta norma es de media cuadrática. Su producto interno se define por la fórmula (3.2).

$$\langle x, y \rangle = \sum_{d=1}^{\infty} x_d \cdot \bar{y}_d \quad \text{donde } x, y \in L^2(\text{Norma}) \quad (3.2)$$

- **Kernel de transformación del espacio.** Un kernel de transformación del espacio, también llamado simplemente kernel, se corresponde con una función que transforma el espacio de características de entrada inicial en un espacio de Hilbert donde se cumple la condición de **Mercer** (comentado en el apartado 3.2.3).

La definición del tipo de kernel se realiza mediante el producto interno de ese espacio de Hilbert (ver fórmula (3.2) y (3.5)).

- **Hiperplano discriminante.** Un hiperplano discriminante, también llamado hiperplano, es una superficie contenida dentro del espacio de características.

Este hiperplano tiene la función de dividir el espacio de características en dos zonas dentro de este espacio de características (ver imagen 3.1). A cada una de las dos zonas se le asocia de forma coherente una clase de la solución.

En consecuencia, dentro de un sistema experto de tipo SVM, al poder dividirse el espacio en solo dos zonas únicamente es posible tener una solución de dos clases distintas. En este trabajo estas dos clases están caracterizadas por la clase “motor sano” y la clase “motor con fallo”.

- **Intervalo y subintervalo de coeficientes.** Para el caso de esta tesis, donde se especifica el “intervalo de coeficientes” se refiere expresamente a un dominio de valores acotado por un valor máximo y un valor mínimo al que puede optar como valor el coeficiente de ajuste; Fuera de ese dominio acotado no existen valores para ese coeficiente de ajuste.

Cuando se especifica un “subintervalo de coeficientes” respecto de un “intervalo de coeficientes” anterior se está refiriendo a un subdominio de valores contenido dentro del intervalo principal.

Por ejemplo, el intervalo de valores para el coeficiente C de límite de caja (que se desarrolla posteriormente) asignado es de $[2^{-5} 2^{15}]$. Por tanto un subintervalo válido de este intervalo sería $[2^{-0.5} 2^5]$.

- **Vectores de soporte.** Los vectores de soporte son puntos dentro del espacio de características seleccionados como representativos para la separación de los conjuntos que forman las clases. A partir de los vectores de soporte seleccionados durante el entrenamiento se desarrolla el hiperplano discriminante.
- **Sobreentrenamiento.** Cuando un sistema experto es entrenado por medio de unas características representativas del problema, este puede ser entrenado hasta ajustar lo máximo posible el aprendizaje respecto de las características de entrenamiento.

Sin embargo, si el conjunto de características de entrenamiento no contiene todas las posibles variaciones de características (cosa común en características obtenidas de mediciones de valores continuos), ajustar el entrenamiento del sistema lo máximo posible respecto de las características que se tienen disponibles puede dar lugar a que el sistema experto clasifique de forma errónea nuevas características que no se ajustan totalmente a las características de entrenamiento.

En este caso se dice que el sistema ha sobreaprendido o ha sido sobreentrenado, ya que es capaz de clasificar con una alta tasa de aciertos las características

con las que ha sido entrenado pero obtiene una tasa de aciertos muy inferior cuando son nuevas características del problema desconocidas.

Para minimizar el efecto del sobreentrenamiento se suele separar una parte del conjunto de características totales en un nuevo conjunto (no utilizado para el entrenamiento) destinado a validar el ratio de aciertos del clasificador como si fueran muestras del problema desconocidas.

- **Linealidad y no linealidad entre muestras de distintas clases.** En esta tesis cuando se expone el concepto de “linealidad en la separación entre muestras de la distintas clases”, se refiere a que el conjunto de muestras de una clase y el conjunto de muestras de la otra clase puede ser separado dentro del espacio representativo del problema por una línea recta. En contraposición, cuando se habla de la “no linealidad en la separación entre muestras de la distintas clases” se está especificando dos conjuntos de muestras de dos clases distintas los cuales no es posible separarlos dentro del espacio representativo del problema por una línea recta.

En el caso de los sistemas expertos SVM, se puede decir que la línea de separación se corresponde con el hiperplano discriminante antes de ser deformado.

3.2.1 Método de clasificación del clasificador SVM.

Un clasificador SVM ya entrenado para detectar las distintas clases (ver apartado 3.2.2), contiene un espacio con un número de dimensiones igual al número de características representativas que se pueden obtener de cada muestra. Cada dimensión se corresponde con su respectiva característica, denominándose a este tipo de espacio vectorial **espacio de características**. En la Figura 3.2 se muestra un ejemplo de espacio de características, formado por dos características (2 dimensiones).

Este espacio de características contiene un hiperplano discriminante (función de clasificación) con el que se van a separar una clase de la otra. Este hiperplano es de N dimensiones y su misión es la de seccionar el espacio en dos subzonas representativas de cada clase. Como solo se utiliza un único hiperplano el clasificador se limita a poder discriminar únicamente dos clases. Un hiperplano bien ajustado **tiene que maximizar el margen de separación entre las dos clases**, y en el caso de que los conjuntos de puntos N -dimensionales de estas clases no puedan ser estrictamente separados, minimizar el posible error en la clasificación.

En la Figura 3.1 se muestra un ejemplo de un posible espacio bidimensional (2 características representativas) de un clasificador SVM. Cada dimensión del espacio se corresponde con una de las características simples del vector de característica de fallo compuesto. Este espacio está seccionado por un *hiperplano* que separa las dos zonas representativas de cada una de las dos clases.

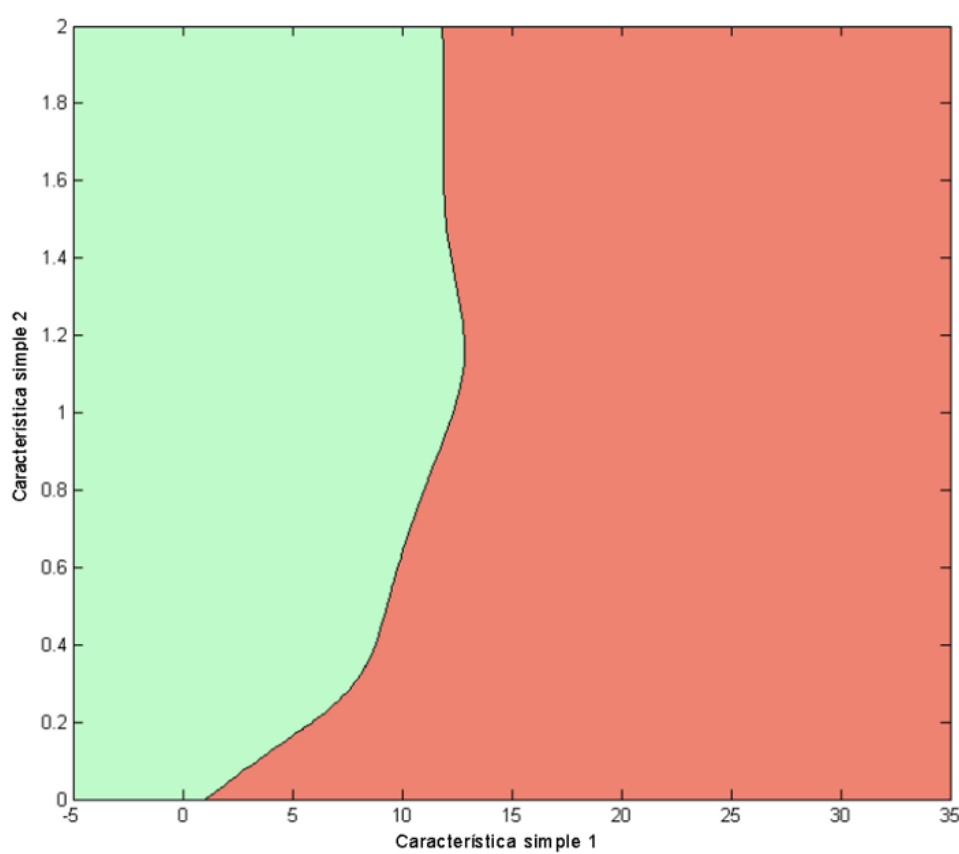


Figura 3.1 Espacio de características de un clasificador SVM donde se observa la frontera del hiperplano discriminante y las dos subzonas definidas para cada una de las clases.

En general, dentro del espacio n-dimensional cada muestra a validar se representa por un punto n-dimensional definido mediante cada una de sus características representativas. Dependiendo de en cuál de las dos zonas del espacio seccionado por el hiperplano se sitúe este punto, obtenido a partir de una muestra, se catalogará este en la clase correspondiente a esa zona. En la Figura 3.2, se muestran puntos correspondientes a una serie de muestras, que han sido clasificados según la zona en que se encuentran respecto del espacio vectorial mostrado en la Figura 3.1.

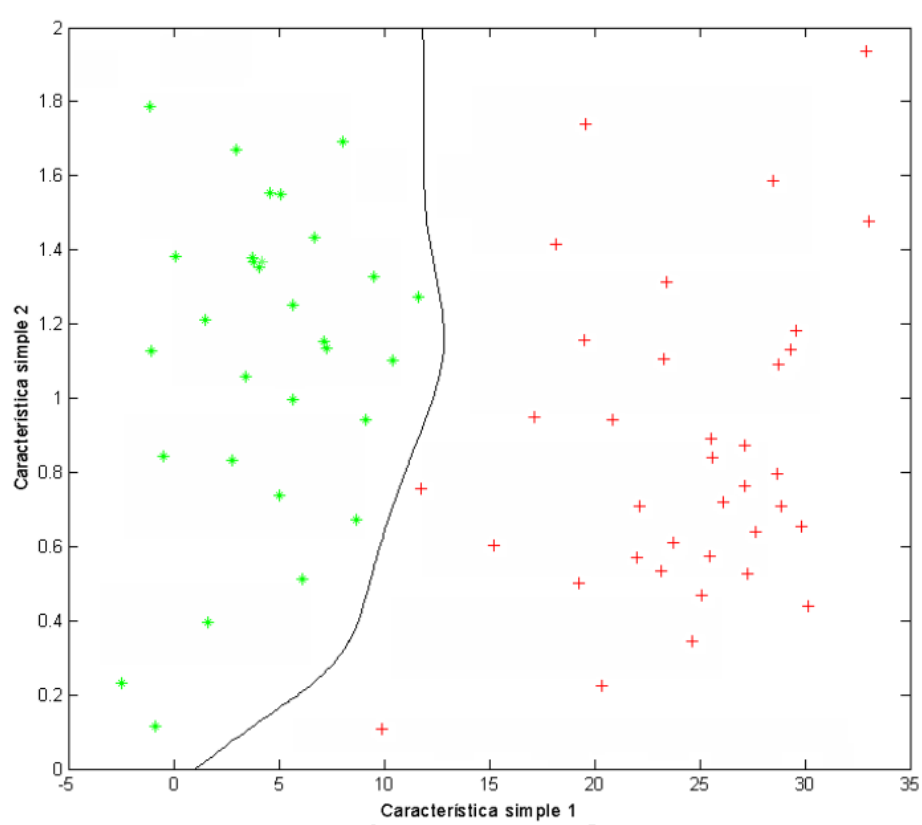


Figura 3.2 Espacio de características de un clasificador SVM donde se observa el hiperplano discriminante y una serie de muestras a las que se les ha asignado una u otra clase dependiendo de la zona en que se sitúan.

3.2.2 Entrenamiento de un clasificador SVM.

Para que este clasificador sea capaz de realizar una clasificación correcta, es necesario que sea entrenado previamente, es decir, requiere de un aprendizaje supervisado.

Los datos para realizar el entrenamiento tienen que ser muestras representativas del tipo de clasificador que se quiere obtener. En cada dato de entrada tiene que estar definido el valor correspondiente a cada una de las características clasificatorias y el valor que indique la clase a la que pertenece (sin fallo / con fallo).

Como ya se ha comentado anteriormente, el sistema experto SVM solo es capaz de realizar la clasificación de dos clases, por lo que solo se puede hacer la detección de un fallo en concreto. Si fuera necesaria la gestión de varios fallos al mismo tiempo existen varias alternativas como por ejemplo:

- El uso de múltiples sistemas expertos SVM entrenados para cada tipo de fallo.
- Usar versiones de SVM adaptadas a múltiples clases (SVM multiclase). Sin embargo, estos sistemas expertos están basados en la implementación de muchos sistemas SVM básicos entrenados para poder realizar una clasificación de una clase contra el resto, o una clasificación de clase contra clase; lo que en ambos casos es poco óptimo.
- Usar otro sistema experto basado en espacio de características capaz de realizar clasificaciones multiclase, como por ejemplo kNN (k-Nearest Neighbors).

Previamente al uso de estos datos para el entrenamiento del clasificador, se tiene que realizar una normalización de los intervalos que abarcan cada conjunto de características.

Esto se realiza ya que en caso contrario, si el intervalo de una característica simple es de órdenes de magnitud mayor que el intervalo de otra característica simple dentro del vector de característica compuesta; el tipo de característica mayor se vuelve más influyente en la clasificación y el tipo de característica menor puede llegar a no ser lo bastante influyente dentro de la clasificación.

Por ejemplo, un caso plausible de detección de fallos mediante dos características básicas distintas donde el intervalo de valores correspondiente a la primera característica se encuentra en el intervalo [0.25 0.5] en Hz y los valores de la segunda característica se encuentran entre [-90 -20] en dB. En este caso el orden de magnitud de la segunda característica es 2 veces mayor al de la primera característica básica de fallo.

En consecuencia, dentro del sistema experto los valores de la segunda característica obtendrían una ponderación mayor en la clasificación respecto a los valores de la primera característica. Para minimizar este efecto se suele utilizar una normalización para todas las características a un intervalo homogéneo entre [-1 1] o [0 1].

Con las serie de muestras representativas disponibles, se intenta ajustar lo mejor posible el hiperplano discriminador para que se obtenga una separación de zonas donde las muestras representativas de cada clase se sitúen dentro de su zona de clase, posicionándose el hiperplano lo más equidistante posible entre las dos clases. Si no se puede realizar una discriminación total de las muestras entre clases, el hiperplano se ajusta para que se maximice el número de muestras representativas clasificadas correctamente.

Para el ajuste del hiperplano se utilizan las muestras más cercanas al *hiperplano* de separación de clases. Estas muestras seleccionadas se denominan **vectores de soporte**.

En la Figura 3.3, se muestra el espacio 2-dimensional de características completado con muestras representativas de dos clases distintas.

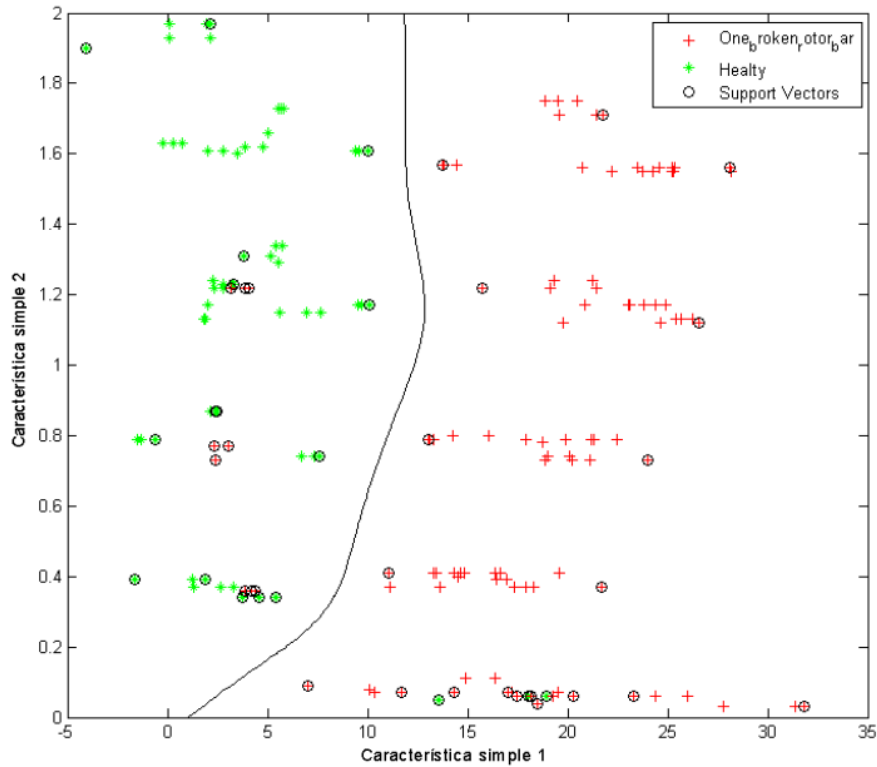


Figura 3.3 Espacio de características de un clasificador SVM donde se muestran las muestras representativas, los vectores de soporte y el hiperplano discriminante que ha sido entrenado.

Las muestras rodeadas con un círculo se corresponden con los vectores de soporte, donde a partir de estos ha sido ajustado el hiperplano representado en negro. La selección de una u otra muestra como vector de soporte depende de los coeficientes (configurados durante el entrenamiento) usados en el sistema experto.

Para mejorar la adaptabilidad del hiperplano de cara a maximizar la discriminación de clases, el clasificador SVM permite varios ajustes de coeficientes, utilizando **kernel** que pueden ser de varios tipos, para la transformación del espacio por lo que lo hace muy versátil para ser utilizado en un amplio rango de datos de entrada con necesidades de separación de conjuntos distintas. Esta personalización de coeficientes se muestra en el siguiente apartado 3.2.3.

Existe una gran variedad de algoritmos destinados a la búsqueda óptima de los mejores parámetros que discriminen lo máximo posible el tipo de características para un problema en concreto. Dentro de este conjunto de algoritmos los más utilizados son los siguientes:

- **Búsqueda de coeficientes en malla [91]-[92].** Como se comenta en Hsu [91], este algoritmo desarrolla una búsqueda acotada dentro del rango máximo posible de los coeficientes. Para maximizar la eficiencia de este algoritmo se realiza una búsqueda en profundidad realizando una subdivisión del intervalo de los coeficientes para explorar la calidad del clasificador dentro de ese subintervalo. De los subintervalos que han mostrado la mejor calidad del clasificador se vuelve a realizar una nueva subdivisión y exploración, y así recursivamente hasta que se cumple alguna de las detecciones de parada (nivel de exploración máximo alcanzado, detectado sobreentrenamiento o no existe mejora apreciable).

La sencillez con respecto a otros métodos de búsqueda de los mejores coeficientes, como por ejemplo la búsqueda en línea, y el buen resultado que ofrece este método de búsqueda hace que sea el más utilizado cuando el número de coeficientes a ajustar no es demasiado elevado. Por el tipo de búsqueda que se realiza se garantiza obtener siempre una solución global.

Cuanto mayor es la profundidad de exploración mayor es el ajuste del hiperplano, lo que puede acabar dando lugar a sobreentrenamiento. No obstante, este sobreentrenamiento se puede detectar al mismo tiempo que se desarrolla la exploración de los intervalos, descartando en tal caso la exploración de estos intervalos donde el clasificador muestra signos de sobreentrenamiento.

Como es lógico, la principal desventaja de este algoritmo deriva en el número de coeficientes a ajustar. A diferencia de otros métodos, como por ejemplo la búsqueda evolutiva, el aumento del número de coeficientes aumenta el coste de cálculo computacional.

En el apartado 5.4.2.1 se comenta con un mayor detalle la implementación de este método de búsqueda.

- **Búsqueda en línea [92].** Este algoritmo desarrollado inicialmente por Chappelle [93], utiliza un algoritmo de gradiente descendiente para minimizar el margen de estimación de error para el conjunto de parámetros definido.

En este caso, dependiendo de las clases, el ajuste del SVM se separa en dos subtipos **que son ajustados por separado** mediante algoritmos de gradiente descendiente donde su coste se calcula mediante una adaptación de la fórmula de Radio/Margen (3.3). En el conjunto denominado L1-SVM se tratan los errores cuya penalización es lineal y en el conjunto L2-SVM se tratan los errores con penalización cuadrática.

$$T = \frac{1}{N} \cdot \frac{R^2}{M^2} \quad (3.3)$$

En esta fórmula de Radio/Margen, T es la cota superior de errores, N se corresponde con el número de muestras, R muestra el radio de la esfera que contiene las características y M representa al margen máximo de separación. Esta fórmula fue expuesta por Vapnik (1998)[94], se asume que todas las características se encuentran dentro de una esfera de radio R y se busca el margen de separación máximo entre clases para el ajuste del tipo “Leave One Out” (validación cruzada de uno a muchos). Se entiende que cuanto mayor sea el margen de separación, la diferenciación entre clases será mejor y por tanto la cota máxima de errores será menor.

La principal ventaja de este método radica en su bajo coste, ya que solo necesita desarrollar unos pocos clasificadores SVM para obtener la solución final. Esto lo hace idóneo para casos de gran envergadura donde el número de parámetros a ajustar es alto.

Como desventajas se encuentran su mayor complejidad de desarrollo respecto a otros métodos (por ejemplo la búsqueda en malla) y es posible que las soluciones obtenidas no sean las mejores ya que esta función radio/margen (que obtiene la relación entre el margen y el radio de los vectores de soporte seleccionados)[92] en ciertos casos puede llegar a devolver una solución local.

- **Búsqueda evolutiva [92].** Este algoritmo es un algoritmo de búsqueda estocástica donde se desarrolla un ajuste de parámetros mediante algoritmos genéticos.

Se desarrollan diversos ciclos donde existen sistemas expertos con un “código genético” que se corresponde con sus parámetros de ajuste personalizados.

Con el avance de los ciclos se desarrollan “mutaciones” en clasificadores, re-combinado de clasificadores, nuevos clasificadores y eliminación de clasificadores menos aptos. Conforme avanzan los ciclos se mantienen los clasificadores más aptos (con ratios de fallo bajos), hasta alcanzar un ajuste de parámetros final aceptable de mínimo ratio de fallos.

Cuanto mejor sea el conjunto inicial de parámetros menor será el tiempo y coste de obtención de la solución objetivo [95]. No obstante, el mejor conjunto inicial de parámetros para comenzar la búsqueda cambia para cada problema y

no existe una técnica explícita para el cálculo de los mejores valores de estos parámetros iniciales.

Si el número de coeficientes a ajustar es pequeño lo recomendable es el uso del algoritmo de búsqueda en malla, aportando la mejor solución en un tiempo aceptable. Si el número de coeficientes es muy grande, se puede utilizar la estrategia de búsqueda en línea o la búsqueda evolutiva para alcanzar una posible solución en un menor tiempo, aunque hay que evitar en tal caso posibles soluciones locales.

En nuestro caso, como se verá posteriormente para cualquier sistema experto SVM que se desarrolla en esta tesis siempre es necesario ajustar solamente 2 coeficientes, el coeficiente de límite de caja C y el coeficiente correspondiente al kernel utilizado (d para kernel polinomial o σ para kernel Gaussiano). En consecuencia, dado el bajo número de coeficientes a ajustar se ha utilizado el algoritmo de búsqueda en malla que garantiza una solución próxima a la solución global.

3.2.3 *Kernels del clasificador SVM.*

Un hiperplano puro es un plano (geométrico) con el que se pueden separar dos clases distintas que sean linealmente equidistantes a similitud de un perceptron simple. Un perceptron simple es una red neuronal de una sola capa de salida (se comenta en el apartado 3.3).

En el caso de los sistemas expertos SVM, el significado de plano es menos estricto que el utilizado en geometría. En este caso un hiperplano (plano de X dimensiones) se corresponde con una superficie que sirve de frontera entre clases pero que no tiene por qué ser en especial plana, sino que puede acabar siendo un hiperplano curvo.

No en todos los casos las características correspondientes a clases diferentes pueden ser separadas linealmente por medio de un hiperplano. En el caso del clasificador SVM, la solución que se utiliza se corresponde con aplicar un kernel de transformación del espacio de características para mejorar la distancia de separación entre conjuntos de distintas clases.

Un kernel de transformación del espacio está formado por una función (kernel) que deforma el espacio n -dimensional para intentar mejorar la distancia discriminativa entre clases. Dependiendo del tipo de datos se utilizará el tipo de kernel más adecuado.

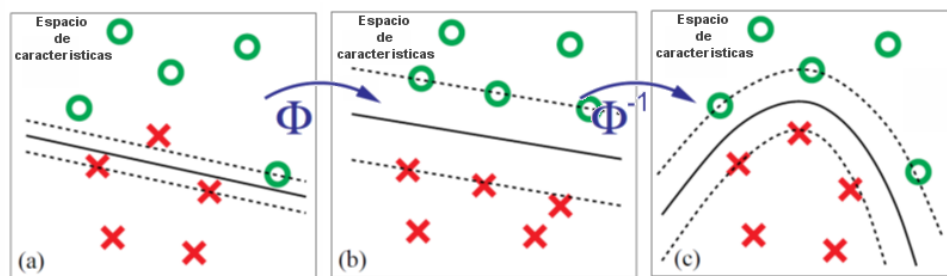
En la Figura 3.4 se puede observar con claridad un ejemplo visual de este tipo de transformación entre espacios y optimización del hiperplano de separación.

En el caso (a) de la Figura 3.4, se muestra el espacio de características original que se forma con las características relevantes de cada muestra. Como se puede observar, las características correspondientes a las dos clases de características no pueden ser sepa-

radas completamente según la clase usando un hiperplano puro. En este caso un plano en 2 dimensiones.

En este caso **(b)** mediante un kernel de transformación se ha transformado el espacio. Dentro de este nuevo espacio donde la separación entre características ha variado, se ha creado y ajustado un hiperplano lineal con el que se discriminan las muestras. Una vez obtenido el hiperplano, se utiliza el mismo kernel de transformación para realizar la transformación inversa del espacio, volviendo al espacio normal de características mostrado en el caso **(c)**. Al realizar la transformación inversa entre planos, también se transforma el hiperplano desarrollado, deformándose en este caso para adaptarse a la forma discriminatoria dentro del espacio de características original, obteniéndose finalmente el hiperplano no lineal con la mejor discriminación.

En el caso **(c)** se observa el mismo espacio original de características, pero en este caso el hiperplano ha sido curvado de forma que discrimina a la perfección las dos clases de elementos. Como se ha comentado en el párrafo anterior, para obtener este hiperplano primero se ha aplicado un kernel de transformación que ha modificado el espacio, obteniéndose el espacio de características del caso **(b)** en el que las muestras sí pueden ser separadas linealmente.



Fuente: Zien et al (2000)[96]

Figura 3.4 Representación de un espacio 2D con su correspondiente hiperplano: (a) Espacio original con un hiperplano lineal. (b) Espacio de características transformado con un kernel donde se muestra un hiperplano lineal. (c) Espacio original con un hiperplano deformado.

Así pues, para desarrollar un hiperplano utilizando un kernel de transformación, primero se construye el hiperplano (hiperplano lineal) dentro del espacio transformado por el kernel, donde la distancia entre las dos clases ha sido mejorada respecto del espacio original.

Una vez que se ha ajustado la posición del hiperplano con los vectores de soporte para obtener la mejor separación, se aplica la transformación inversa del kernel entre espa-

cios, de forma que el anterior hiperplano lineal del espacio deformado por el kernel se convierte en un hiperplano deformado dentro del espacio de características original.

Como se comenta en el trabajo de Cortes [83], de acuerdo a la teoría de Hilbert Schmidt [97] se pueden considerar la forma genérica del producto escalar de un espacio de Hilbert (kernel) como se expresa en (3.4), y por tanto cualquier función simétrica $K(x_i, x_j)$ que esté contenida dentro de la L2-norma se puede expresar con la forma de (3.5). En este caso Φ está representado por una autofunción y λ por un autovalor.

Para que un Kernel pueda ser considerado válido tiene que asegurar que todos los autovalores λ que genera el producto escalar del espacio de características son positivos. Para garantizar que estos coeficientes son positivos tienen que cumplir con la condición de **Mercer**. Es decir, para un tipo de autofunción Φ y una expansión en series del kernel (3.5) solo se cumple esta condición si para cualquier $g(x)$ cuya integral (3.6) es finita, se cumple la restricción (3.7) de que todos sean mayores o igual a 0.

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (3.4)$$

$$K(x_i, x_j) = \sum \lambda \cdot \Phi(x_i) \cdot \Phi(x_j) \quad (3.5)$$

$$M = \int g(x)^2 dx \quad \text{siendo } M \neq \pm\infty \quad (3.6)$$

$$\int K(xy) \cdot g(x) \cdot g(y) \cdot dx \cdot dy \geq 0 \quad (3.7)$$

Hay que tener en cuenta que no todos los tipos de datos pueden ser separados linealmente aunque se utilice un kernel de transformación. Además, en muchos casos esta imposibilidad de discriminación es debido a características seleccionadas para formar el espacio que no son representativas del problema o cuando las muestras que se utilizan para entrenar no son de buena calidad.

Dependiendo del tipo de transformación del kernel, este puede contener coeficientes ajustables para mejorar la transformación del espacio, de forma que se puede mejorar la discriminación de muestras modificando estos coeficientes.

Aunque se puede desarrollar cualquier tipo de kernel personalizado, existe un conjunto de kernels ya predefinidos por defecto con los que prácticamente se puede discriminar de forma muy eficiente cualquier conjunto de datos.

A continuación se muestran los tipos de kernel más populares, en estos kernels (3.8), (3.9), (3.10), (3.11) y (3.12) los coeficientes x_i y x_j se corresponden con los vectores del espacio de entrada:

- **Kernel lineal (3.8).**

$$K(x_i, x_j) = x_i^T \cdot x_j \quad (3.8)$$

Fuente: Hsu et al (2000)[91]

Este tipo de kernel es el más simple, considerado como un caso especial de kernel que no se puede utilizar en todas las ocasiones. Como se observa por el tipo de producto interno que desarrolla este kernel (3.8) como un tipo de espacio de Hilbert, este se corresponde con el espacio Euclideo.

Este kernel solo se utiliza en los tipos de problemas de clasificación donde la separación entre características de clases distintas se puede hacer con un hiperplano puro. No se recomienda este kernel si el tipo de clases a discriminar no es lineal o es desconocida.

No obstante, este tipo de kernel tiene un consumo de recursos muy bajo en comparación con el resto de kernels que aquí se exponen, por lo que es recomendable su uso si las clases a discriminar cumplen con las restricciones de discriminación lineal entre conjuntos de clases.

- **Kernel cuadrático (3.9).**

$$K(x_i, x_j) = (1 + x_i^T x_j)^2 \quad (3.9)$$

Fuente: Manning et al (2008)[98]

Este kernel es el segundo más rápido después del kernel lineal. Como se puede observar, este kernel es un caso especial del kernel polinomial, que se ve a continuación, donde el valor de la variable d es 2.

- **Kernel polinomial (3.10).**

$$K(x_i, x_j) = (1 + x_i^T x_j)^d \quad (3.10)$$

Fuente: Hsu et al (2000) [91]

Este kernel es uno de los más populares y es suficiente para la mayoría de casos. Su coste de ejecución es mayor respecto del kernel cuadrático y del kernel lineal, pero menor que el kernel de función de base radial Gaussiana.

Las variables x_i y x_j representan la posición i y j respectivamente dentro del espacio de características, el parámetro d define el orden de la curva polinomial. Cuando el valor del coeficiente d es 1 este kernel se comporta como un kernel lineal, y si el valor de d es 2 el kernel es un kernel cuadrático. Cuanto mayor es el orden polinómico (d) mayor es la posible curvatura del hiperplano que se obtiene.

- **Kernel de función base radial (RBF) (3.11).**

$$K(x_i, x_j) = \exp \left[- \left(\frac{\|x_i - x_j\|}{2\sigma} \right)^2 \right] \quad (3.11)$$

Fuente: Hsu et al (2000) [91]

Este kernel desarrollado por Boser et al (1992) [99], también conocido como kernel Gaussiano por la función de base radial gaussiana que utiliza, es el kernel más recomendado como kernel **por defecto** para el entrenamiento del SVM. Con este kernel se transforma el espacio en un espacio definido por el peso probabilístico de la distancia Euclidea entre dos puntos del espacio. El parámetro σ se corresponde con el factor de escala de la función RBF.

Aunque el coste computacional de este es superior al kernel polinomial, también se consigue que la flexibilidad del hiperplano discriminante sea mayor, pudiendo llegar en algunos casos a discriminar mejor las características relevantes. En contrapartida, al permitirse una mayor flexibilidad también se puede llegar a dar lugar un cierto **sobre-aprendizaje**.

La variable σ funciona como factor de escala de la función de base radial. El rango de su valor puede ser cualquier número real positivo.

En los casos particulares donde el número de características es muy grande se recomienda utilizar un kernel con mayor linealidad en vez de este.

- **Kernel sigmoideo (3.12).**

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (3.12)$$

Fuente: Hsu et al (2000) [91]

El coeficiente γ se corresponde con el margen superior del escalado de características en el perceptron multicapa y su valor tiene que ser superior a 0. El coeficiente r se corresponde con el margen inferior del escalado de características en el perceptron multicapa y su valor tiene que ser inferior a 0.

Este kernel es también conocido como kernel de tangente **hiperbólica** o kernel **perceptron multicapa** por su funcionamiento como perceptron. Este kernel es el equivalente a una red neuronal de dos capas (ver apartado 3.3).

Su desarrollo a partir de las redes neuronales lo hizo popular en el momento de su aparición en 1995 [100], pero estudios posteriores como por ejemplo el de Lin (2003) [101] han concluido que, en general, no existe mayor ventaja respecto a un kernel de tipo RBF.

3.2.4 Coeficientes de ajuste del clasificador SVM.

Además de un buen entrenamiento, el ajuste correcto de coeficientes es un paso importante de cara a obtener un clasificador SVM de buena calidad. Dependiendo del tipo de entrenamiento, existen diversos algoritmos de optimización de coeficientes capaces de obtener la mejor solución o una aproximación de esta.

- **Limitación de la ventana de ajuste (Box constraints (C)).**

En el proceso de ajuste del hiperplano, es posible modificar el número aceptable de elementos mal clasificados del hiperplano por medio de la variable C que limita la ventana de ajuste del hiperplano. Poder definir un margen de error es importante, ya que con datos del mundo real no siempre se va a poder realizar una discriminación completa entre clases.

Como se expone en el artículo de Martínez Ruedas [102], el coeficiente C mantiene los valores de los **multiplicadores de Lagrange** α_i (para minimizar el problema L1 o L2 norma) dentro del límite marcados en (3.13) (entre 0 y C). De este modo el sistema experto SVM expresado por la fórmula (3.14) queda restringido.

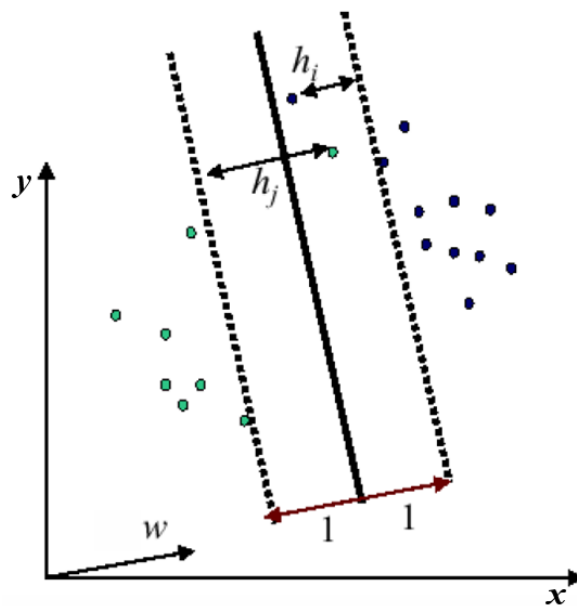
$$0 \leq \alpha_i \leq C \quad \forall i$$

$$\sum_{i=1}^N \alpha_i \cdot z_i = 0 \quad (3.13)$$

$$f(\mathbf{x}) = \text{sign}\left(\sum_{\substack{\text{vectores} \\ \text{soporte}}} \alpha_i \cdot z \cdot K(x_i, x_j) + b\right) \quad (3.14)$$

Fuente: Martínez Ruedas (2006)[102]

Con valores grandes de C el ajuste del hiperplano se intenta realizar con una separación más estricta entre las clases. Con valores próximos a 0 en el ajuste del hiperplano se acepta un mayor número de clasificaciones erróneas en la búsqueda del mejor hiperplano. En el ejemplo de la Figura 3.5 se observa cómo el ajuste del coeficiente C ha permitido que la configuración del hiperplano en ese momento permita que la muestra con la distancia h_j se encuentre clasificada de forma errónea dentro de la zona de la otra clase.



Fuente: Martínez Ruedas (2006) [102]

Figura 3.5 Margen de posicionamiento del hiperplano. Cuanto mayor es el valor de C la separación entre clases es más estricta, cuanto menor es C se permite un mayor número de clasificaciones erróneas.

Si las muestras de ambas clases son totalmente discriminantes, la mejor sugerencia de ajuste del hiperplano es la separación estricta donde se maximiza la distancia entre estas clases. Si no se puede definir un hiperplano que pueda separar estrictamente todas las muestras de las clases, es preferible ceder cierto margen de error con tal de obtener la clasificación correcta de la mayoría de las muestras.

Este parámetro C no depende del kernel sino que corresponde a la función de ajuste del hiperplano durante el entrenamiento del sistema experto SVM, por lo que no se ve afectado por el tipo de kernel utilizado.

- **Selección del Kernel y de sus coeficientes.**

En la siguiente Tabla 3.1 se muestra el conjunto de kernels más populares, junto a sus coeficientes personalizables (en el caso de que tenga alguno).

Tabla 3.1. Definición de los coeficientes disponibles en cada kernel.

Tipo de Kernel	coeficientes
Lineal (3.8)	
Cuadrático (3.9)	
Polinomial (3.10)	<p>d: Define el orden de la curva polinomial. Su valor tiene que ser entero > 0. Por defecto su valor es 3.</p> <p>Un valor de 1 hace que el kernel se comporte como un kernel lineal. Un valor de 2 hace que se comporte como un kernel cuadrático.</p> <p>La cota superior de d es $+\infty$, pero no se suelen asignar valores > 6 ya que, aunque mejora la curvatura del hiperplano, los límites de discriminación prácticamente no varían.</p>
Función de base radial Gaussiana (3.11)	<p>σ: Factor de escala de la función de base radial. Su valor tiene que ser un número real positivo. Por defecto su valor es 1.</p> <p>Todo valor obtenido por el kernel de base radial gaussiana se encuentra acotado entre $0 < K_{ij} \leq 1$ [91].</p>
Sigmoideo (3.12)	<p>γ: Coeficiente del margen superior del escalado de características en el perceptron multicapa. Su valor tiene que ser > 0. Por defecto su valor es 1.</p> <p>r: Coeficiente del margen inferior del escalado de características en el perceptron multicapa. Su valor tiene que ser < 0. Por defecto su valor es -1.</p> <p>Como se comenta en Vapnik [83], no son válidos todos los posibles valores acotados por estos dos coeficientes.</p>

3.2.5 Ventajas y desventajas del clasificador SVM.

En definitiva, un clasificador SVM tiene sus ventajas y limitaciones. Estas cualidades positivas y negativas se deben tener en cuenta para poder generar un sistema experto clasificatorio que cumpla con los requisitos.

Las ventajas de este tipo de clasificador son:

- Su sencillez para desarrollar y entrenar el clasificador. El sistema de entrenamiento del clasificador SVM es un problema de programación cuadrática (QP) que puede ser resuelto de forma eficiente con herramientas informáticas [102].
- Si las muestras características son lo bastante significativas y no tienen ruido que pueda entorpecer la detección del posible fallo, entonces es improbable que se de sobreentrenamiento.
- Es un sistema clasificatorio válido para discriminación lineal y no lineal de las muestras.
- Su requisito de memoria es bajo. El clasificador entrenado solo necesita almacenar la configuración del hiperplano.
- Dependiendo del sistema de ajuste de coeficientes, se puede tener un ajuste aceptable con un coste computacional relativamente pequeño.
- Aunque el conjunto de muestras sea pequeño, si estas son representativas del problema, el clasificador obtenido suele adecuarse bastante bien al problema.

Las desventajas de este clasificador son:

- La versión original sin adaptaciones únicamente es capaz de realizar una discriminación booleana entre dos clases distintas. Actualmente existen versiones de SVM para más de dos clases pero no tienen un coste óptimo respecto de otros clasificadores del mismo tipo [102], como por ejemplo el clasificador k -NN.
- Como en cualquier clasificador que utiliza un espacio de características, el coste computacional y de espacio crecen conforme crece el número de características a discriminar.
- A diferencia de otros sistemas expertos basados en espacios de características, como por ejemplo el *k Nearest Neighbors* (kNN) [81], cualquier variación de las muestras de entrenamiento, ya sea por mejorar el ratio de aciertos del sistema experto o añadir nuevas muestras con nuevo conocimiento inexistente anteriormente, obliga siempre a reentrenar todo el clasificador SVM para adaptarse a las nuevas muestras.

3.3 Redes neuronales artificiales.

Una red neuronal artificial (ANN), como se puede deducir por su propio nombre, está inspirada en el funcionamiento de las redes neuronales biológicas de los animales.

Dentro de esta denominación se pueden encontrar dos conjuntos distintos de redes neuronales:

- **Simulación física del modelo de red neuronal biológico.**

En este conjunto se encuentran los diseños de redes neuronales que simulan físicamente la estructura y funcionamiento de una red neuronal biológica.

El desarrollo de estas redes neuronales se realiza principalmente para su estudio dentro del ámbito biosanitario, de forma similar al desarrollo y simulación de otras clases de órganos como por ejemplo el corazón. En consecuencia, su función principal no está ligada a su uso como sistema de deducción inteligente, como por ejemplo la clasificación de elementos.

Dado que el modelo en este caso simula de forma física la función biológica, al igual que cualquier otro simulador de tejido vivo tanto el tamaño que ocupa como el tiempo de ejecución de su simulación son bastante altos.

- **Simulación conceptual del modelo de red neuronal biológico.**

A diferencia del conjunto anterior de ANN, en este caso la construcción de la red neuronal se centra en el concepto y funcionalidad abstracta de la red. Es decir, aunque entre una ANN de este tipo y una red biológica no hay ninguna similitud física, el funcionamiento operacional y los resultados en base a los mismos datos de entrada serán similares.

Un ejemplo claro de similitud operacional equivalente entre las redes neuronales artificiales y las biológicas se puede apreciar en el trabajo de Machado [103], donde se duplica funcionalmente la red neuronal biológica de un nematodo *Caenorhabditis elegans* en una red neuronal artificial equivalente.

Al no tenerse en este caso en cuenta la simulación físico biológica, el tamaño de la red es mucho más reducida y el tiempo de ejecución mucho menor.

La creación de este tipo de redes se centra principalmente en el desarrollo de sistemas expertos para el tratamiento complejo de datos. Dentro de este grupo se encuentran las redes neuronales de tipo perceptron multicapa, redes retroalimentadas de Hopfield y redes no supervisadas basadas en mapas autoorganizados (SOM).

Dado que esta tesis se centra en el uso de sistemas expertos para el perfeccionamiento en el análisis de fallos, el tipo de simulación neuronal por medio de modelo conceptual se corresponde con el más idóneo para este caso y en consecuencia es el único tipo que se va a desarrollar en profundidad en esta tesis. En este caso la implementación del sistema experto desarrollada es la del **perceptron multicapa**.

Como se comenta en el libro de Rich [104] sobre la inteligencia artificial, la evolución de los sistemas expertos de tipo ANN ha sufrido de varios altibajos desde su primera aparición en 1943 de la mano de los biofísicos McCulloch y Pitts [84].

Aunque ya en los años 40 del siglo XX se realizaron las primeras investigaciones, es a partir del año 1958 cuando se empieza a vislumbrar su gran potencial gracias en buena parte a las investigaciones sobre el perceptron de Rosenblatt [105] y sobre los Madalines de Widrow y Angell (1962)[106].

Sin embargo la limitada funcionalidad del perceptron comentada en la revista del MIT por Minsky y Papert (1969) [107], el cual es capaz de simular funciones OR pero no XOR, supuso un duro revés a esta línea de investigación.

Finalmente en torno al año 1982, con el surgimiento de los nuevos conceptos de perceptrón multicapa y el uso de algoritmos de aprendizaje por retropropagación del error para su entrenamiento [108], el campo de las ANN vuelve a coger fuerza hasta ser un referente en sistemas de aprendizaje artificial en la actualidad.

Una de las mayores ventajas de las ANN como sistemas de clasificación respecto a otros sistemas expertos, como por ejemplo el SVM visto en el apartado 3.2, se centra en su capacidad de simular funciones complejas de discriminación multiclase (mayor a dos) respecto de datos tanto lineales como no lineales, pudiendo clasificar características difícilmente discriminables mediante fronteras de conjuntos.

A continuación se definen algunos de los conceptos que se usan en el sistema experto:

- **Característica representativa del fallo.** Al igual que en el caso del sistema experto SVM la característica de fallo se asocia a un vector de datos de dimensión 1 o mayor con información que puede ser relevante para detectar el fallo. Esta información relevante de fallo se puede usar para el entrenamiento del sistema experto (si se conoce la clasificación final de este) o para la clasificación (diagnóstico de fallo / no fallo).

Cuando este vector de datos **característica representativa del fallo** tiene una longitud mayor que 1 se dice que es una **característica representativa compuesta** por unidades de **características simples**.

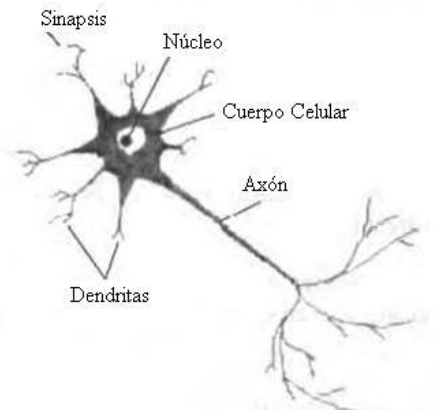
- **Clase.** En este sistema experto también se puede gestionar las soluciones del problema como clases correspondientes cada una a un valor discreto que caracteriza a una de las posibles soluciones del sistema experto. Pero además y aunque no se utiliza en esta tesis, a diferencia de un sistema experto SVM, un

sistema experto ANN también puede tener una respuesta de salida continua para resolver problemas donde las soluciones no pueden ser discretas.

- **Red Neuronal artificial.** Una red neuronal artificial es una estructura compleja compuesta de neuronas (nodos) agrupados en capas ordenadas en entradas, cero o más capas ocultas y una capa de salida. (Figura 3.9). La entrada al sistema experto en si no tiene una capa de neuronas asociada, sino que la entrada se corresponde con la primera capa que tiene la red (si no tiene capas ocultas la capa de salida también sería de entrada de datos).
- **Capa de neuronas.** Una capa de neuronas es una secuencia en paralelo de neuronas, que en el caso que se trata en esta tesis del perceptron multicapa, estas neuronas no están conectadas a otras neuronas de la misma capa pero si están conectadas a las neuronas de la capa anterior (o de la entrada si es la primera capa) y con la capa posterior (o con las salidas si esta es la capa de salida).
- **Neurona.** Es la unidad básica de la red neuronal. Este elemento es una unidad que procesa los datos que le llegan y envía el resultado generado en base a su configuración. Este elemento se desarrolla con más detalle en el apartado 3.3.1 de este capítulo.

3.3.1 Unidad básica de una red neuronal.

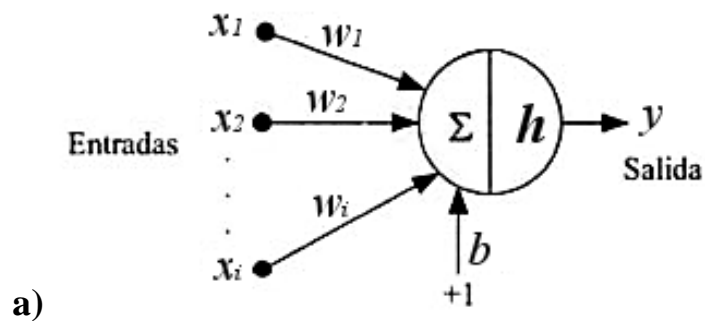
La unidad básica de una red neuronal se corresponde con la **neurona**, tanto en su forma biológica (Figura 3.6) como en su modelo artificial equivalente (Figura 3.7).



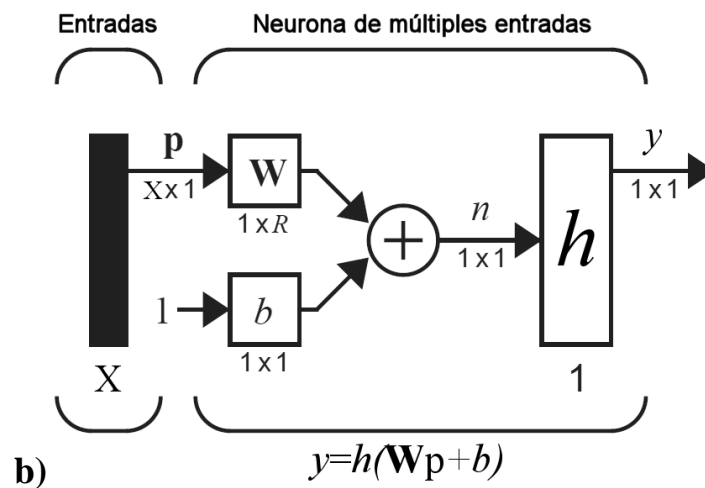
Fuente: Isasi y Galván (2004)[109]

Figura 3.6 Representación ilustrativa de una neurona de tipo biológico.

Aunque en el ámbito biológico existen diversas clases de neuronas dependiendo de ciertas características morfológicas como la longitud del axón o el número de dendritas, en general la función más genérica que desempeñan todas las neuronas es muy similar.



Fuente: Martín y Sanz (2001)[110]



Fuente: Hagan et al (1996)[111]

Figura 3.7 Modelo de una neurona artificial a) McCulloch-Pitts b) Adalina.

En el caso de la representación de las neuronas (a) y (b) de la Figura 3.7, estas son totalmente equivalentes, a pesar de que provienen de dos estudios diferentes. Dado que las diferencias físicas no son aplicables, el modelo de **neurona artificial** representado tanto en (a) como (b) de la Figura 3.7 estaría compuesto por:

- **Entradas ($x_i(t)$).** Receptor de datos de la neurona, equivalente a las dendritas de una neurona biológica.
- **Pesos asociados a las entradas (w_i).** Estos pesos están asociados a las entradas, también denominados pesos simpáticos. Este elemento desarrolla una ponderación de la entrada, de forma que desarrollando la función de propagación con estos pesos se define si se permite o no la activación de la neurona. El valor de cada peso es el elemento clave que define la especialización de la red neuronal y se define en su fase de entrenamiento.
- **Regla de propagación ($p(t)$).** Esta función define la propagación de la ponderación de las entradas y los pesos de la neurona. Para el cálculo del parámetro de propagación correspondiente se suele usar habitualmente, pero no necesariamente, una función sumatorio de tipo lineal $x_i(t)$ con ponderación mediante dichos pesos w_i (3.15) o una función radial de segundo orden no lineal (3.16).

$$h(t) = \sum w_i \cdot x_i(t) \quad (3.15)$$

$$h(t) = \sqrt{\sum (x_i(t) - w_i)^2} \quad (3.16)$$










Fuente: Hagan et al (1996)[111]

- **Umbral (b).** Esta variable se corresponde con el valor de activación de la neurona. En este caso la neurona únicamente se activa si el valor resultante de la regla de propagación es mayor al valor del umbral. Esta variable es optativa.
- **Función de transferencia ($h(t)$).** Llegado el caso en que la neurona es activada, se utiliza el valor resultante de la regla de propagación dentro de la función de transferencia para generar finalmente el valor de salida en la neurona artificial. Aunque se puede utilizar cualquier tipo de función de transferencia específica, existen predefinidas una serie de funciones genéricas preestablecidas ampliamente aceptadas con las que se puede resolver la mayoría de problemas con una buena tasa de aciertos.

En la siguiente Tabla 3.2 se muestran algunas de estas funciones de transferencia ($h(t)$) más comunes. La variable n representa el valor de la entrada, mientras que la variable a representa el valor de salida.

En la función de transferencia Gaussiana también se observa el coeficiente de ajuste σ que ajusta el factor de escala de la función y el valor medio μ de las entradas.

Tabla 3.2. Funciones de transferencia ($y(t)$) más utilizadas.

Nombre	Relación entre entrada y salida	Función
Escalón	$\begin{cases} a = 0 & n < 0 \\ a = 1 & n \geq 0 \end{cases}$	
Escalón simétrico	$\begin{cases} a = -1 & n < 0 \\ a = +1 & n \geq 0 \end{cases}$	
Lineal	$a = n$	
Lineal positiva	$\begin{cases} a = 0 & n < 0 \\ a = n & n \geq 0 \end{cases}$	
Lineal saturada	$\begin{cases} a = 0 & n < 0 \\ a = n & 0 \leq n \leq 1 \\ a = 1 & n > 1 \end{cases}$	
Lineal saturada simétrica	$\begin{cases} a = -1 & n < -1 \\ a = n & -1 \leq n \leq 1 \\ a = 1 & n > 1 \end{cases}$	
Sigmoidal	$a = \frac{1}{1 + e^{-n}}$	
Tangente hiperbólica sigmoidal	$a = \frac{e^n - e^{-n}}{e^n + e^{-n}}$	
Gaussiana	$a = \exp\left(-\frac{(n - \mu)^2}{2\sigma^2}\right)$	
Competitiva	$\begin{cases} a = 1 & \text{neurona con } n \text{ máximo} \\ a = 0 & \text{el resto de neuronas} \end{cases}$	

Fuente: Hagan et al (1996)[111]

3.3.2 Topología de red neuronal supervisada.

La estructura de una red neuronal artificial supervisada está formada principalmente por grupos de neuronas entrelazadas (ver Figura 3.9), de modo que esta red en la mayoría de sus formas va a tener una o más entradas para la introducción de datos y una o más salidas por las que se obtendrá el resultado.

Teóricamente, mientras se respete el número de entradas y salidas de la red, se puede realizar cualquier red de conexión entre las neuronas internas. No obstante, en la práctica la topología estándar que se desarrolla por su facilidad de implementación, facilidad de entrenamiento y por su capacidad para resolver una gran cantidad de problemas; está compuesta por una red de capas de neuronas con entrelazamiento entre capas y propagación hacia la salida.

De esta topología estándar se distinguen los siguientes subtipos de redes dependiendo de la naturaleza de los datos:

3.3.2.1 Datos sin relación secuencial.

Los datos sin relación de secuencialidad son aquellos datos que se obtienen de un mismo proceso o función pero donde no existe ninguna relación dentro del conjunto de datos respecto de cada uno con su anterior o posterior inmediato. Es decir, no existe ninguna significancia por la que sea necesario mantener la secuencia de datos ordenada según su adquisición dentro del conjunto.

Un ejemplo de este tipo de datos son aquellas características generadas para la implementación de una red neuronal con función clasificatoria. Aunque dichas características sean reordenadas esto no afectará al funcionamiento ni a la solución generada por la red.

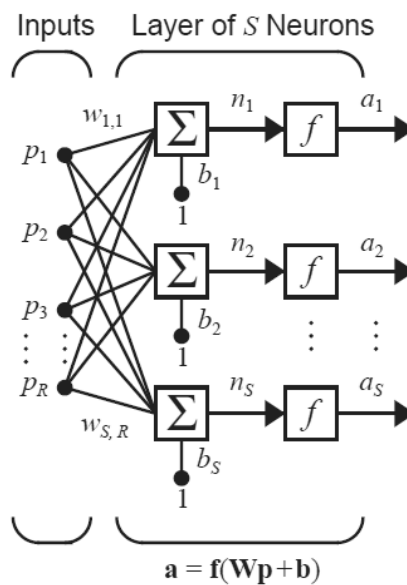
Las redes neuronales supervisadas entrenadas con estos tipos de datos suelen desarrollarse para realizar:

- **Funciones de ajuste o aproximación.** Se entrena la red para generar una serie de salidas determinadas por los datos de entrada. Un ejemplo de este tipo es una red entrenada para funcionar como filtro de señal.
- **Funciones de reconocimiento de patrones y clasificación.** A partir de unos patrones de entrada (características relevantes), se entrena a la red neuronal para que sea capaz de deducir a que elemento corresponde dicho patrón. Un ejemplo bastante famoso de este tipo se corresponde con una red entrenada para clasificar botellas de vino.
- **Funciones de agrupación.** En este caso se entrena la red para agrupar datos con similares características. Este tipo de red se usa mucho en el tratamiento de grandes conjuntos de datos para agrupar datos como por ejemplo los existentes en una base de datos de encuestas o perfiles.

Como se muestra en la Figura 3.9 una red neuronal de este grupo se compone de un vector de entradas de datos, un número variable de capas ocultas (pudiendo ser ninguna) y una capa de salida con un número de neuronas igual a las salidas configuradas en la red.

En cualquier caso, de forma estricta, las salidas de una capa de la red únicamente pueden conectar con las entradas de la siguiente capa, con única excepción de la capa de salida donde sus salidas se utilizan siempre como salidas de la red.

Una red neuronal que no contiene ninguna capa oculta se denomina Perceptron (ver Figura 3.8) y está indicada para aquellos problemas que pueden ser discriminados linealmente, pues la única capa de neuronas (la capa de salida) solo es capaz de discriminar de forma lógica como lo haría una función OR.



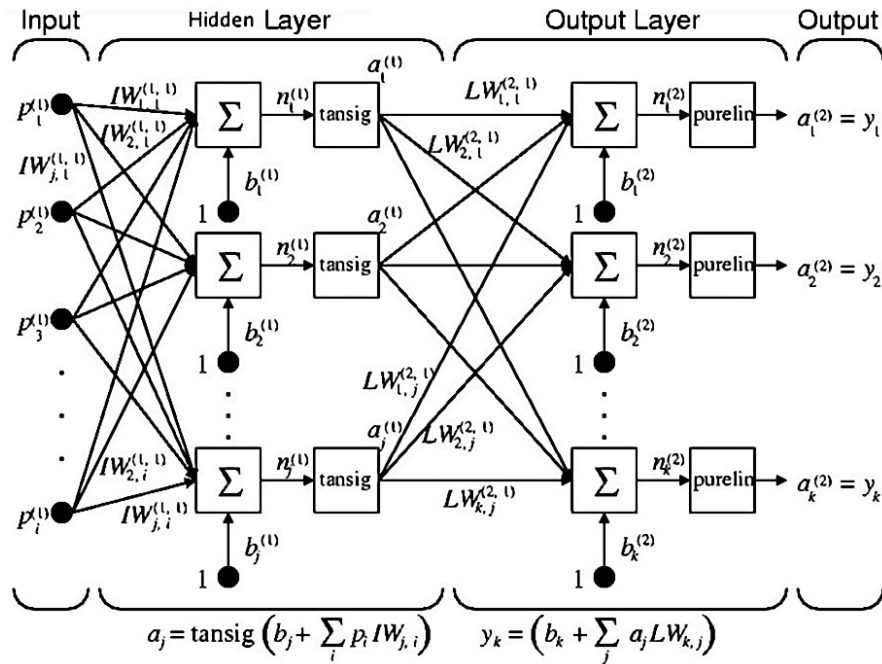
Fuente: Hagan et al (1996)

Figura 3.8 Esquema de una red neuronal artificial Perceptron.

Cuando la red neuronal incluye una o más capas ocultas esta se denomina Perceptron multicapa MLP (Figura 3.9). Esta red puede configurarse para realizar funciones más complejas que el simple Perceptron, pues cada capa oculta funciona como una XOR.

Sin embargo, hay que tener en cuenta que conforme crece el número de capas ocultas aumenta la complejidad de la red y en consecuencia aumenta la dificultad para entre-

narla correctamente de cara al problema propuesto. Por ello, siempre se suele buscar la funcionalidad óptima de la red con el menor número de capas posibles.



Fuente: Demuth et al (2008) [85]

Figura 3.9 Esquema de una red neuronal artificial Perceptrón multicapa.

En la mayoría de problemas una única capa oculta es suficiente para resolverlos. Solo en algunos casos en que la no linealidad en la separación de los conjuntos de características de entrada es muy alta es necesario un número mayor de capas.

En el caso de las muestras obtenidas de motores para el análisis de fallos no existe una relación secuencial entre estas. Es decir, da lo mismo el momento en que se obtienen las muestras y el orden en que se procesan estas, tanto para el entrenamiento del sistema experto como para la clasificación posterior de nuevas muestras.

En consecuencia, el tipo de redes neuronales que se desarrollan como sistemas expertos ANN en esta tesis se centra exclusivamente en la estructura de red desarrollada en este apartado 3.3.2.1 para sistemas expertos ANN donde los datos de entrada no están relacionados secuencialmente.

3.3.2.2 Datos relacionados secuencialmente.

Los datos relacionados de forma secuencial mantienen una significancia dependiente del dato anterior o posterior a este. La omisión o cambio de posición de algún dato genera una disfunción en los resultados de la red neuronal. La relación más común entre los datos de este tipo es su secuencia temporal.

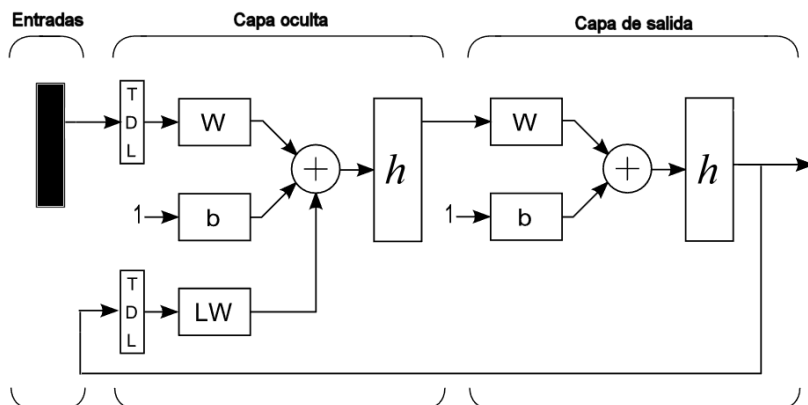
Las redes neuronales supervisadas entrenadas con estos tipos de datos suelen desarrollarse para realizar labores de predicción o mejoras en la precisión del resultado partiendo de los resultados obtenidos anteriormente. Ejemplos de uso de estas redes se dan en la predicción de cambio de valores en sistemas financieros, en predicción de modelos dinámicos de sistemas físicos y en sistemas de control.

La tipología de estas redes es similar a la tipología comentada en el apartado 3.3.2.1, con la salvedad en que, dependiendo del tipo, la entrada de la red puede ser realimentada con su salida.

En este caso se añaden elementos adicionales al sistema de red neuronal visto hasta ahora, por ejemplo en la Figura 3.10 se muestran elementos TDL (Time delay lines) para retrasar un ciclo los resultados, y nuevos pesos a la capa de entrada de tipo LW (layer weights) que se corresponde con la salida anterior del sistema experto.

Las redes de este tipo se pueden categorizar en tres grupos:

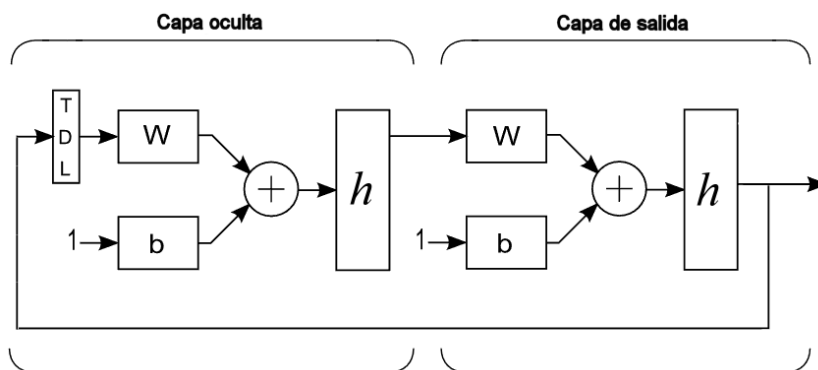
- **Red autorregresiva con entradas externas (NARX).** Como se muestra en la Figura 3.10. Este tipo de red funciona como un Perceptron multicapa con retroalimentación de las salidas generadas anteriormente por la secuencia. De esta forma la red obtiene información adicional con dichas salidas anteriores además de las entradas actuales, realizando una nueva predicción en consecuencia.



Fuente: Demuth et al (2008)[85]

Figura 3.10 Esquema de una red neuronal artificial NARX.

- **Red autorregresiva (NAR).** Este tipo de red neuronal se corresponde con otra variante de red retroalimentada donde, a diferencia de la red NARX, no existen neuronas de entrada y el reajuste de la predicción de la salida se realiza únicamente a partir de los resultados de la salida anteriores. Esta sutil diferencia se puede observar en el esquema de la Figura 3.11 para una red neuronal NAR respecto al esquema de la Figura 3.10 de la red neuronal NARX.



Fuente: Demuth et al (2008) [85]

Figura 3.11 Esquema de una red neuronal artificial NAR.

- **Redes de Entrada - Salida.** Se usa solo en caso de que no se puedan mantener las salidas antiguas para realizar la predicción, optimizándose estas únicamente a partir de las entradas. Evidentemente, si no se pueden mantener estos valores tampoco pueden ser realimentados en la red. En este caso este tipo de red sigue el mismo esquema sin realimentación de entradas del Perceptron multicapa mostrado en la Figura 3.8.

3.3.3 Entrenamiento de la red.

Como ya se ha visto en otros sistemas expertos como por ejemplo el SVM en el apartado 3.2.2, para que la red neuronal sea capaz de discernir sobre el problema propuesto es necesario que aprenda como hacerlo.

En el caso en que se centra esta tesis de redes neuronales supervisadas, se realiza un entrenamiento previo de la red una vez que su estructura se encuentra definida. Para su entrenamiento se hace uso de ejemplos resueltos que son representativos del problema.

Los puntos relevantes para un correcto entrenamiento se centran en su capacidad de generalización (que se comenta en el apartado 3.3.4), en la selección del algoritmo de

entrenamiento (ajuste de pesos y umbrales desarrollado en el apartado 3.3.3.1) y en la función de coste (desarrollada en el apartado 3.3.3.2).

3.3.3.1 Algoritmos de entrenamiento.

Los elementos principales de una red neuronal que definen la supuesta inteligencia de esta sobre el problema específico se centra en los pesos y los umbrales asignados a cada neurona de cada una de las capas.

Como ya se ha explicado en el apartado 3.3.1, los pesos se encargan de controlar el reajuste de los datos y su propagación entre la entrada y la salida; y los umbrales actúan como sesgo de activación en cada neurona. Un ajuste ideal de pesos y umbrales generará la solución correcta (o próxima) a partir de los datos iniciales del problema asignados a las entradas.

El algoritmo de ajuste de pesos, también denominado **algoritmo de entrenamiento**, es el encargado de ajustar correctamente dichos pesos (entrenar la red neuronal) a partir de los casos resueltos representativos del problema aportados al sistema.

Para medir el rendimiento de los sucesivos ajustes que desarrolla esta clase de algoritmos para minimizar el error se usa en cada iteración (epoch) una función de estimación del coste para calcular el error. Estas funciones se comentan en el apartado 3.3.3.2.

Como se comenta en el artículo desarrollado por Sharma [112], los diversos algoritmos de entrenamiento más usados para una **red neuronal perceptron multicapa** (conocida en inglés por *multilayer perceptron* (MLP)) se pueden agrupar en 3 grupos dependiendo de su similitud en su implementación para realizar dichos ajustes. La similitud de muchos de ellos recae en que algunos son versiones con optimizaciones y mejoras aplicadas sobre los originales.

No obstante existen otros tipos de algoritmos de entrenamiento menos óptimos para redes MLP pero aplicables a estructuras de redes neuronales más exóticas donde estos algoritmos generalizados no son operativos. Un ejemplo de este tipo se muestra en el estudio de Bertona [113] respecto del uso de sistemas genéticos para desarrollar el entrenamiento de redes neuronales.

3.3.3.1.1 Algoritmos de retropropagación del gradiente descendiente.

Este grupo de algoritmos es el más estudiado dado que el algoritmo de retropropagación del gradiente descendiente perteneciente a este grupo, definido por Werbos [108], es uno de los primeros algoritmos usados para el entrenamiento de las redes neuronales MLP.

Estos algoritmos reajustan los pesos y umbrales a partir del gradiente negativo obtenido con la función de rendimiento. Los tipos de algoritmo pertenecientes a este grupo son:

- **Retropropagación del gradiente descendiente.** Este algoritmo calcula el gradiente descendiente del error de la red neuronal y reajusta los pesos en sentido inverso desde la salida a la entrada de la red, hasta minimizar por completo el error y maximizar el ratio de aciertos de la red.
- **Retropropagación del gradiente descendiente con momento.** Este algoritmo adapta de forma heurística el ratio de aprendizaje para mejorar el tiempo de aprendizaje, la convergencia y el riesgo de las soluciones locales respecto del algoritmo clásico de retropropagación [114].
- **Retropropagación flexible del gradiente descendiente.** El siguiente algoritmo es una optimización de los anteriores para minimizar el efecto que se da en redes donde existen capas de neuronas definidas con funciones de transferencia sigmoideal (Tabla 3.2).

En este tipo de capas, el ajuste del gradiente puede llegar a tomar valores muy pequeños por lo que deja de influir de forma efectiva sobre los pesos y umbrales a reajustar [115]. Para ello, solo se tiene en cuenta el signo de las derivadas (no su magnitud) para determinar la actualización en los ajustes de los pesos y umbrales.

3.3.3.1.2 Algoritmos de retropropagación del gradiente conjugado.

Como ya se ha comentado en el apartado 3.3.3.1.1, los algoritmos de retropropagación del gradiente realizan la convergencia a partir de las direcciones descendientes del gradiente. Como diferencia a estos, los algoritmos del gradiente conjugado buscan la convergencia en las direcciones de los conjugados del gradiente, realizando una búsqueda lineal en cada iteración en busca del gradiente mínimo.

Aunque estos tipos de algoritmos requieren mayor memoria obtienen unos tiempos de convergencia mejores. Los tipos de algoritmo pertenecientes a este grupo son:

- **Retropropagación del gradiente conjugado escalado.** En vez de realizar una búsqueda lineal en cada ciclo de ajuste del aprendizaje, este algoritmo usa un mecanismo de escalado para realizar la búsqueda de ajustes lo que lo hace más rápido que otros algoritmos que usan el gradiente conjugado. Como inconveniente, el número de iteraciones para llegar a la convergencia es mayor [116].
- **Retropropagación del gradiente conjugado con actualización de Fletcher-Reeves.** Este algoritmo utiliza para la búsqueda lineal el ratio Z existente entre el mínimo cuadrado del gradiente actual respecto del gradiente de la iteración anterior (ecuación 3.17) para hallar la dirección actual de búsqueda del gradiente $d(x)$ a partir de este ratio Z , el gradiente actual $g(x)$ y la dirección de búsqueda anterior $d(x-1)$ (ecuación 3.18) [117].

El espacio usado por este algoritmo es similar al anterior de la retropropagación del gradiente conjugado escalado. La función $L2_norm$ se corresponde con la norma L_2 (3.19) también conocida como la $L2$ norma, norma euclídea o norma de mínimos cuadrados, la cual minimiza la suma de los cuadrados de los valores de x .

$$Z = \frac{L2_norm(g(x))}{L2_norm(g(x-1))} \quad (3.17)$$

$$d(x) = -g(x) + d(x-1) \cdot Z \quad (3.18)$$

$$L2_norm(x) = \sqrt{\sum_{i=1}^n |x_i|^2} \quad (3.19)$$

Fuente: Sandhu y Chhabra (2011)[118]

- **Retropropagación del gradiente conjugado con actualización de Polak-Ribière.** En la búsqueda lineal para hallar la dirección de búsqueda, este algoritmo calcula el ratio Z del producto interno de la diferencia entre el gradiente actual $g(x)$ y el gradiente anterior $g(x-1)$ respecto del mínimo cuadrado de dicho gradiente anterior (ecuación 3.20) [85]. Este ratio Z se aplica a la misma fórmula mostrada en la ecuación 3.19 para calcular la nueva dirección de búsqueda.

$$Z = \frac{(g(x) - g(x-1)) \cdot g(x)}{L2_norm(g(x))} \quad (3.20)$$

Fuente: Sandhu y Chhabra (2011) [118]

3.3.3.1.3 Algoritmos casi Newtonianos.

Los algoritmos Newtonianos desarrollan en cada iteración la inversa de la matriz Hessiana respecto al índice de rendimiento correspondiente a los pesos y umbrales. La convergencia es más rápida en términos de iteraciones, también lo es el tiempo necesario para calcular dicha matriz para el caso de una red neuronal de tipo perceptrón multicapa (MLP). La matriz Hessiana del algoritmo Newtoniano es una matriz cuadrada

que contiene las segundas derivadas parciales respecto de una función de entrada, en este caso una función de coste (apartado 3.3.3.2).

En la práctica se utiliza el método denominado ‘casi Newtoniano’. Este método calcula una aproximación de la matriz Hessiana en cada iteración a partir de dos gradientes del error desarrollados en sucesivas iteraciones. En este caso el resultado es menos preciso pero requiere de un tiempo de cómputo mucho menor.

Estos algoritmos, diferentes a los algoritmos anteriores que utilizan gradientes, suelen dar mejores optimizaciones en la mayoría de casos. Sin embargo, como se comenta en Sandhu [118], si la función diferenciable del punto estacionario no tiene una expansión de la serie de Taylor cuadrática cercana a la óptima no se garantiza la convergencia hacia el resultado.

Los tipos de algoritmo pertenecientes a este grupo son:

- **BFGS (Broyden-Fletcher-Goldfarb-Shanno).** Este algoritmo está basado en una búsqueda local donde, a partir de una solución aleatoria inicial, esta va variando hasta encontrar una solución mejor. En todo caso, para conseguir el resultado óptimo el gradiente final tiene que llegar a ser cero [114].

Este algoritmo converge rápidamente y es muy eficiente en redes pequeñas, como contrapartida requiere mucha capacidad tanto de cómputo como de almacenamiento a causa del cálculo del gradiente y de la aproximación a la matriz Hessiana [119].

- **Retropropagación Levenberg-Marquardt.** En cada una de las iteraciones de este algoritmo se minimizan aquellas funciones multivariantes que se correspondan con la suma de los cuadrados de otras funciones de valores reales no lineales. De este modo la función de rendimiento va disminuyendo en cada iteración hasta llegar al margen especificado.

Al igual que en el caso del algoritmo BFGS, este algoritmo tiene una convergencia muy rápida, sobretodo en redes neuronales donde el rendimiento de la red es dependiente del error cuadrático medio. En este caso se considera una buena opción para redes de tamaño medio, pero igualmente requiere mucha capacidad tanto de cómputo como de almacenamiento [119].

3.3.3.2 Funciones de coste.

Los algoritmos de entrenamiento comentados en el punto anterior requieren de una función de coste para calcular la tasa de error de aprendizaje en cada iteración. Valorando el error de aprendizaje en la iteración actual los algoritmos de entrenamiento pueden realizar los ajustes correspondientes para disminuir dicho error en la siguiente iteración.

Las **funciones de coste** dependen del problema a resolver, por lo que el tipo de función a utilizar no está limitado y puede personalizarse cuanto sea necesario. Sin embargo, se suele usar una serie de funciones de coste generalizadas que son suficientes para la mayoría de los problemas planteados. A continuación se muestran las más conocidas:

- **Suma del error cuadrático (SSE).** En este caso, como se observa en la fórmula (3.21) el rendimiento de la red se calcula con la suma de la diferencia de los errores entre la salida deseada y la salida obtenida al cuadrado. En este caso \hat{y}_i corresponde a las salidas deseadas e y_i corresponde a las salidas actuales. La variable n se corresponde con el número de datos de entrenamiento disponibles.

$$SSE = \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3.21)$$

Fuente: Sharifi y Mohebbi (2012) [120]

- **Error cuadrático medio (MSE).** Esta función (3.22) obtiene la media de la suma del error cuadrático obtenido para todas las características del conjunto del que se está calculando el error.

$$MSE = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n} \quad (3.22)$$

Fuente: Sharifi y Mohebbi (2012) [120]

- **Suma del error absoluto (SAE).** A diferencia de la función de coste SSE (3.18), en esta función SAE (3.23) el error de la red se calcula con la suma del valor absoluto de la diferencia entre los errores de la salida deseada y de la salida real obtenida.

$$SAE = \sum_{i=1}^n |\hat{y}_i - y_i| \quad (3.23)$$

Fuente: Chacon et al (2010)[121]

- **Error absoluto medio (MAE).** Esta función (3.24) obtiene el promedio de la función de rendimiento SAE (3.23) respecto al número de datos de entrenamiento.

$$MAE = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n} \quad (3.24)$$

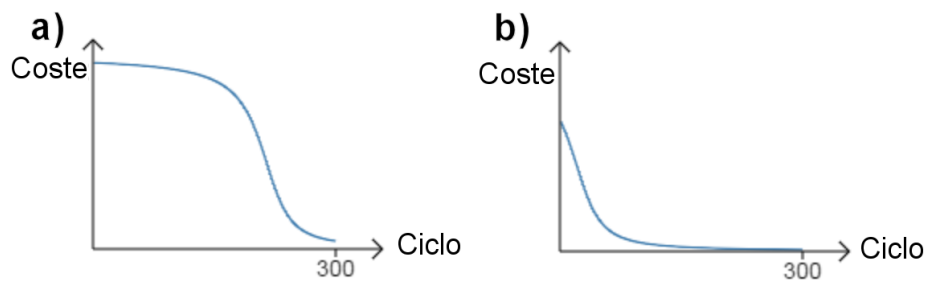
Fuente: Chacon et al (2010)[121]

- **Entropía cruzada del error.** Cuando el error obtenido es grande y se usa alguna de las funciones de coste comentadas anteriormente, entonces el aprendizaje de la red se vuelve lento (gráfica (a) de la Figura 3.12). Esto es debido a que las derivadas parciales de dicha función de coste respecto del peso y del umbral son muy pequeñas para las capas de neurona con función sigmoideal.

Este efecto se puede evitar usando la función de coste del error por entropía cruzada (3.25), donde cuanto mayor es el error también es mayor la velocidad de aprendizaje (gráfica (b) de la Figura 3.12).

$$CE = - \frac{\sum_{i=1}^n [\hat{y}_i \cdot \ln(y_i) + (1 - \hat{y}_i) \cdot \ln(1 - y_i)]}{n} \quad (3.25)$$

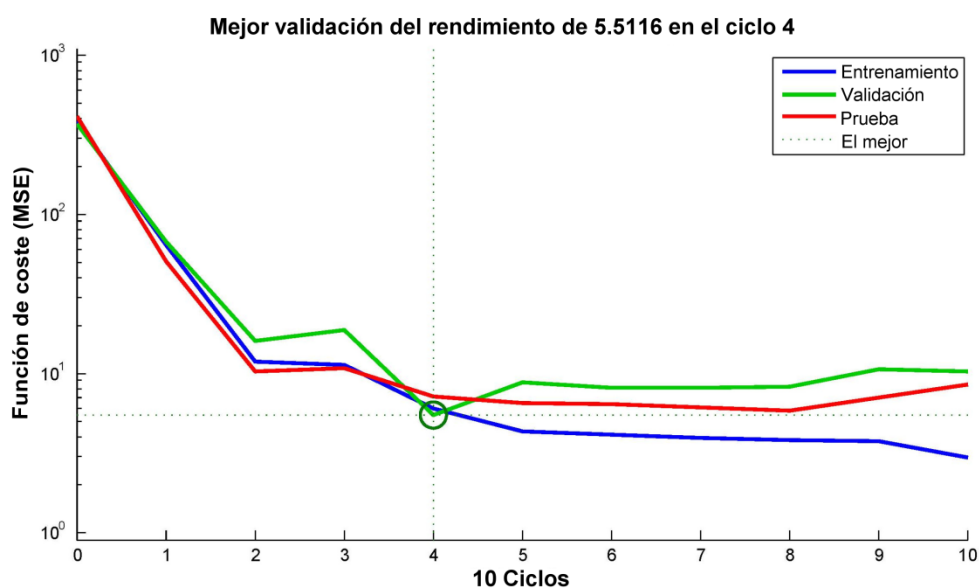
Fuente: Nielse (2015) [122]



Fuente: Nielsen (2015) [122]

Figura 3.12 Grafica que refleja la evolución del error de aprendizaje en cada ciclo calculado por la función de coste en cada iteración del entrenamiento cuando el error es grande para una a) función de coste MSE, b) función de coste por entropía cruzada.

Un ejemplo de esta disminución del error de aprendizaje en cada iteración del algoritmo de entrenamiento, obtenido del estudio de Hassan.Esfahani [123], se muestra en la curva de entrenamiento de la gráfica de la Figura 3.13. Este tipo de entrenamiento con tres conjuntos de entrenamiento/validación/prueba se corresponde con el algoritmo de entrenamiento con detención temprana que se desarrolla en el apartado 3.3.4.2.



Fuente: Hassan.Esfahani et al (2015)[123]

Figura 3.13 Gráfica que refleja la evolución del error de aprendizaje en cada iteración del entrenamiento para los grupos de entrenamiento, validación y prueba.

En esta imagen se muestra:

- **Numero de ciclo de entrenamiento (en el eje de abscisas).** El algoritmo de entrenamiento funciona por ciclos (epoch), donde en cada ciclo intenta disminuir el error obtenido mediante la función de coste empleada.
- **Error obtenido por la función de coste (en el eje de ordenadas).** En cada ciclo se intenta reducir el error calculado con la función de coste (que se comentan a continuación), que en el caso de ejemplo de la Figura 3.13 la función de coste corresponde al error cuadrático medio (Mean Squared Error (MSE)).
- **Curva de entrenamiento (training).** Durante el entrenamiento la red neuronal se entrena utilizando un conjunto de características representativas selecciona-

das para el entrenamiento. Durante el entrenamiento se intenta reducir en cada ciclo el error cometido al ajustar las neuronas de la red para que las salidas de esta se correspondan con sus características de entrada. Como ya se ha comentado, el error de ajuste de las neuronas de la red en cada ciclo se calcula con la función de coste.

- **Curva de validación (validation).** Durante el entrenamiento mientras se intenta minimizar el error con las características de entrenamiento, al mismo tiempo se usa otro conjunto de características como validación.

De estas características de validación también se obtiene su error para cada ciclo, pero no se usa para minimizar el error en el siguiente ciclo, pero sí para evitar el sobre aprendizaje. Este conjunto de validación se utiliza como características para evitar de forma local el sobreentrenamiento.

- **Curva de prueba (test).** Además del conjunto de validación se usa un conjunto de pruebas. La curva de error de este conjunto se obtiene a parte del conjunto de entrenamiento y de validación.

El conjunto de prueba se utiliza para evitar el sobreentrenamiento de forma global y no afecta en ningún momento al entrenamiento ni validación en cada ciclo.

El conjunto de prueba, junto a la curva de validación y la curva de entrenamiento, se utiliza para obtener el ciclo donde la configuración de la red neuronal tiene un error mínimo pero sin llegar a sobreentrenar el sistema.

Una separación notable entre el error calculado de la curva de entrenamiento respecto de la curva de validación y de test, donde el conjunto de entrenamiento obtiene un error mucho menor, indica un claro sobreentrenamiento del sistema.

Por tanto, se toma como **entrenamiento óptimo** de la red neuronal al ciclo de entrenamiento donde el error es mínimo y al mismo tiempo las tres curvas de entrenamiento, validación y test confluyen manteniéndose próximas. Es decir, se obtiene como red neuronal entrenada final a la red neuronal cuya configuración se ha generado en el ciclo de entrenamiento que cumpla con esas premisas (error mínimo sin sobreentrenamiento).

Como ya se ha comentado, en cada ciclo (epoch) de la gráfica de la Figura 3.13 se ha minimizado el error de la curva de entrenamiento hasta alcanzar el mínimo sin sobreentrenamiento, que en este ejemplo se obtiene en el ciclo 4 del entrenamiento. Este ciclo óptimo de ajuste del entrenamiento se muestra marcado con un círculo donde se fusionan las tres curvas.

3.3.4 Mejora en la generalización del aprendizaje.

Al igual que se ha comentado en otros sistemas expertos, como por ejemplo en el apartado 3.2.2 del clasificador SVM, es necesario evitar el sobreentrenamiento del sistema experto. Es decir, además de realizar el aprendizaje con entrenamiento para maximizar su rendimiento, hay que cerciorarse que el sistema neuronal es capaz de resolver correctamente nuevos patrones de datos del problema no estudiados con anterioridad; lo que se suele llamar generalización.

Para mejorar dicha generalización existen varias recomendaciones en redes neuronales [114] que afectan tanto a la estructura de la red neuronal como a su metodología de entrenamiento.

3.3.4.1 Estructura de la red ajustada al problema.

En el caso de que el conjunto de datos para entrenamiento abarque todos o casi todos los posibles patrones correspondientes al problema se puede decir que el sobreentrenamiento que pueda aparecer es muy pequeño o nulo. En cambio, si el conjunto de datos para entrenamiento no es lo bastante denso, y si el dimensionado de la red no es óptimo puede llegar a afectar al rendimiento de esta red.

Si el número de capas y/o neuronas es más pequeño del mínimo necesario para desarrollar la función correspondiente al problema la tasa de aciertos alcanzable será menor que si tuviera el número de capas y de neuronas suficiente. Si por el contrario la red neuronal es más grande de lo necesario puede desarrollar funciones más complejas, pero como se comenta en Hagan [111], también puede dar lugar a un sobreentrenamiento además de un mayor coste de entrenamiento y de espacio.

Por tanto, la mejor estructura posible en todo caso del estudio de es aquella que obtiene el mejor coste con el menor número de capas y neuronas. Por desgracia, actualmente no existe un algoritmo directo con el cual hallar la mejor estructura para cualquier tipo de red.

En la mayoría de casos se realiza un ajuste de la estructura por medio de búsqueda iterativa de la mejor combinación de capas/neuronas partiendo de reglas heurísticas.

Como se comenta en Bishop [124], se considera como mínimo una capa oculta y como máximo dos capas ocultas por los siguientes motivos:

- Teóricamente con dos capas ocultas se puede resolver cualquier tipo de **problema no lineal** de ajuste, clasificación o predicción que converja hacia la solución.
- La mayoría de los problemas pueden ser resueltos con una única capa oculta.
- El número de resultados locales aumenta drásticamente con el aumento del tamaño de la red.

- Cada nueva capa oculta propaga el error de la anterior haciendo que el gradiente varíe de forma más inestable.
- Cuantas más capas tenga la red el entrenamiento de esta se hace más costoso.

De forma genérica, el algoritmo desarrolla los siguientes pasos para buscar la mejor combinación estructural.

- 1) Inicialmente se estudia el número de neuronas óptimo para resolver el problema con redes neuronales de una sola capa.
- 2) Para intentar agilizar la detección del número de neuronas óptimo se utiliza una regla heurística para definir el número de neuronas aconsejado desde el cual iniciar la búsqueda. Sin embargo, actualmente no hay un consenso sobre un tipo de regla específico.

Las reglas más comunes para definir el número inicial de neuronas ($n^{\circ}_{neuronas}$) son:

- El valor medio del número de entradas E y salidas S (3.26).

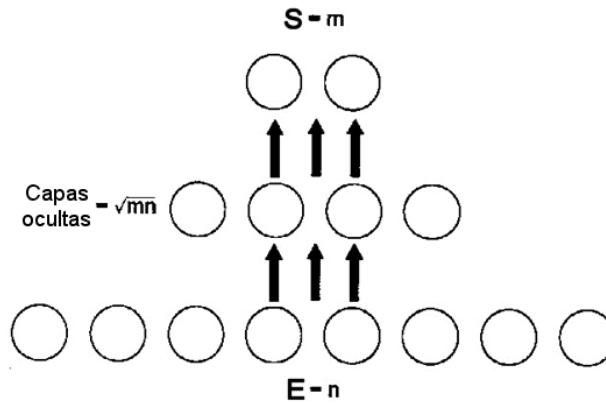
$$n^{\circ}_{neuronas} = \frac{E + S}{2} \quad (3.26)$$

Fuente: Blum y Rivest (1992)[125]

- La pirámide geométrica [125], donde el número de cada capa se define formando una estructura triangular (piramidal) entre las entradas y las salidas (ejemplo en Figura 3.14). Para una sola capa, la fórmula de la pirámide geométrica para calcular el número inicial de búsqueda ($n^{\circ}_{neuronas}$) viene dado por (3.27) donde E es el número de las neuronas de entrada y S el número de las neuronas de salida.

$$n^{\circ}_{neuronas} = \sqrt{E \cdot S} \quad (3.27)$$

Fuente: Blum y Rivest (1992)[125]



Fuente: Blum y Rivest (1992) [125]

Figura 3.14 Ejemplo gráfico de la definición inicial del número de neuronas para una capa oculta mediante la regla de la pirámide.

- Regla basada únicamente en el número de características por patrón (número de entradas) (3.28).

$$n^{\circ}_{neuronas} = \frac{2}{3} \cdot E \quad (3.28)$$

Fuente: Diez et al. (2001)[126]

Aun así, estas reglas heurísticas solo sirven para agilizar la búsqueda en ciertos casos, pero independientemente de la regla utilizada el resultado final debe ser el mismo para cualquier caso.

- 3) Partiendo de dicho número de neuronas inicial se busca el número óptimo de neuronas en la capa oculta con incrementos/decrementos de una neurona. El objetivo es obtener el número mínimo de neuronas con el cual la red neuronal desarrolla el mejor rendimiento (mínimo error).
- 4) En el caso de que la red neuronal no de unos buenos resultados a pesar de tener un gran número de neuronas en la capa oculta, se incrementa la búsqueda a dos capas ocultas.
- 5) Para este nuevo caso de dos capas ocultas; para definir el número inicial de neuronas a explorar en ambas capas se puede utilizar, o bien la misma regla de

promedio (3.26) para cada capa oculta, la misma regla (3.28), o la regla de la pirámide geométrica para la primera capa (3.29) y la segunda capa (3.30).

$$n^{\circ}_{\text{Capa_oculta_1}} = S \cdot \left(\sqrt[3]{\frac{E}{S}} \right)^2 \quad (3.29)$$

$$n^{\circ}_{\text{Capa_oculta_2}} = S \cdot \sqrt[3]{\frac{E}{S}} \quad (3.30)$$

Fuente: Blum y Rivest (1992)[125]

- 6) Como ya se ha comentado anteriormente, no existe ninguna demostración teórica por la que se necesite más de dos capas para resolver el problema. Si el resultado no es óptimo, se asume que el problema no es resoluble de forma eficiente.

3.3.4.2 Entrenamiento con detención temprana.

En cualquier entrenamiento de una red neuronal es necesario conocer el punto en que la curva de minimización del error pasa del aprendizaje al sobreentrenamiento. Para ello, el método de parada temprana ([88], [111], [114]) es uno de los más generalizados para obtener ese número de ciclos de iteración (epoch) óptimo donde la red neuronal minimiza la tasa de error sin sobreentrenamiento.

En este caso el conjunto de datos de entrada se divide en tres grupos:

- **Un conjunto para entrenamiento (T).** Este grupo es utilizado para realizar el entrenamiento de la red. Normalmente es el conjunto más grande de datos.
- **Un conjunto para validación (V).** Este grupo es utilizado para detectar el sobreentrenamiento durante el entrenamiento.

En cada ciclo de entrenamiento de la red, se comprueba el rendimiento de esta con el conjunto de validación. Al no haber sido la red entrenada con este grupo de datos se está analizando su comportamiento ante el nuevo conjunto de datos.

Como se muestra en la anterior Figura 3.13 que representa la evolución de las curvas de entrenamiento, prueba y validación; durante el aprendizaje de la red el rendimiento de esta respecto del conjunto de entrenamiento y el de valida-

ción es similar. Conforme la red deja de generalizar y sobreentrena el conjunto de entrenamiento, disminuye la tasa de error en este grupo pero aumenta la tasa de error en el grupo de validación.

En consecuencia, se busca el ciclo de aprendizaje con menor tasa de error tanto para el grupo de entrenamiento como para el grupo de validación. Como este método es aproximado, aunque se detecte el mejor ciclo normalmente se suelen analizar unos cuantos ciclos más por si se encuentra una mejor solución.

Es de vital importancia tener en cuenta que los datos de validación deben ser lo más heterogéneos y representativos del conjunto de posibles patrones de características del problema.

- **Un conjunto para pruebas (P).** Aunque el conjunto de validación se utiliza para calcular la tasa de error de la red neuronal sobre muestras desconocidas, esto se realiza de manera local para cada red entrenada.

Dado que el conjunto de entrenamiento y el de validación puede (y es recomendable) cambiar con cada nueva red entrenada, se utiliza el conjunto de pruebas para analizar el error mínimo obtenido por este conjunto en cada ciclo de entrenamiento. A diferencia del conjunto de entrenamiento y de validación, este conjunto de prueba solo influye en el entrenamiento cuando se ha de seleccionar la configuración final de la red neuronal.

Si el error calculado por la función de coste entre el conjunto de entrenamiento, conjunto de validación y conjunto de prueba no diverge demasiado se considera que la red se encuentra bien entrenada para ese ciclo de entrenamiento.

En cualquier caso, si se utiliza este algoritmo de ajuste de parámetros de la red neuronal, se recomienda que el algoritmo de entrenamiento no minimice el error demasiado rápido. De esta forma se puede detectar de forma eficiente el mejor ciclo de entrenamiento con la menor tasa de error de entrenamiento y validación.

3.3.4.3 Reentrenamiento de la red neuronal artificial.

Como se ha comentado en el apartado anterior, si se utiliza para la detección de sobreentrenamiento la parada temprana, los datos a utilizar se dividen en tres grupos, uno de entrenamiento, otro de validación y otro de prueba.

En tal caso, es recomendable realizar el entrenamiento de distintas versiones de la misma estructura de red neuronal variando las muestras contenidas en el conjunto de entrenamiento (T), en el conjunto de validación (V) y en el conjunto de pruebas (P). Es decir, variando el reparto del conjunto total de muestras disponibles entre los tres conjuntos (T, V, P).

De este modo disminuye el riesgo de obtener un mal entrenamiento por una mala elección en el reparto de los datos en los tres conjuntos. En este caso la tasa de error será el promedio obtenido de todas ellas.

Por otro lado, independientemente de la estructura de la red neuronal y del reparto de datos en grupos, la función de entrenamiento utilizada en cada caso (ver apartado 3.3.1) inicia el entrenamiento de la red con unos pesos y umbrales iniciales seleccionados de forma aleatoria, limitados dentro de un rango.

Por ejemplo, como se comenta en el capítulo 3 del libro de Nielsen [122], para inicializar los pesos y umbrales generalmente se suele utilizar una función de selección de variables gaussianas aleatorias normalizadas entre el intervalo 0 y 1, intentando que la media de la distribución gaussiana para los pesos y umbrales sea próxima a cero y la desviación estándar próxima 1.

En este caso para una misma red, dependiendo del valor inicial del peso y umbral, su entrenamiento puede ser más rápido o más lento, y obtener una mayor o menor tasa de fallos.

Como no se garantiza que la misma red con cualquier combinación de pesos y umbrales iniciales den los mejores resultados, se recomienda también reentrenar esta red (con misma estructura y división de datos) con distintos pesos para obtener el promedio de error en este caso.

En resumen, para mejorar la generalización se recomienda tanto reentrenar la misma red con diversos repartos del conjunto de datos, y además se recomienda reentrenar ésta misma red con diversas configuraciones de pesos y umbrales.

3.3.5 Ventajas y desventajas de las redes neuronales artificiales.

Como cualquier sistema experto, las redes neuronales tienen sus ventajas y desventajas a tener en cuenta a la hora de seleccionar el mejor sistema experto para clasificación que cumpla con los requisitos.

Las ventajas de las redes neuronales son:

- a) Su gran adaptabilidad a la resolución de problemas tanto lineales como no lineales así como su robustez ante datos de entrada inusuales.
- b) Si el conjunto de muestras características es grande y de calidad, no se da sobreentrenamiento; y en el caso de que este conjunto no sea muy grande existen métodos para generalizar (apartado 3.3.4) minimizando el efecto del sobreentrenamiento.
- c) En general, el requisito de memoria es relativamente bajo para un clasificador de una red neuronal de tipo MLP con máximo de dos capas ocultas como los que se proponen en esta tesis.

- d) Estas redes neuronales pueden dar soluciones dentro del ámbito continuo y discreto. En el ámbito de resultados discretos, como por ejemplo clasificaciones de datos, no hay ningún límite en el número distinto de clases que puede procesar.
- e) Aunque el conjunto de muestras sea pequeño, si estas son representativas del problema, el clasificador obtenido suele adecuarse bastante bien al problema.

Las desventajas de las redes neuronales son:

- a) El coste computacional necesario para desarrollar un entrenamiento correcto es alto teniendo en cuenta que hay que realizar una diversidad de entrenamientos sobre diversas variaciones de redes neuronales para descartar el sobreentrenamiento y soluciones locales.
- b) El entrenamiento se realiza en base a estimaciones proporcionadas por métodos aproximados de validación cruzada de n iteraciones fijas, por lo que se obtiene una aproximación de la calidad del clasificador menos exacta que por ejemplo la que se obtiene con el entrenamiento por validación cruzada de uno a muchos en el caso de clasificadores SVM.
- c) En el caso de las redes neuronales supervisadas, que son las que se tienen en cuenta en esta tesis, cualquier variación de las muestras de entrenamiento originales obliga a realizar un reentrenamiento completo para adaptarse a dicha variación.

3.4 Conclusiones.

Finalmente en esta tesis, como se verá en el capítulo 0, se ha optado por implementar en el sistema automático el ajuste de sistemas expertos basados en SVM y ANN. En ambos casos se ha optado por la implementación de los algoritmos más robustos utilizados para su entrenamiento.

En el caso de entrenamiento de los sistemas expertos basados en SVM se ha utilizado para el ajuste de coeficientes del SVM y del kernel una **búsqueda de coeficientes en malla**. Los kernels usados en estos casos son el **kernel polinomial** y el **kernel de función de base radial gaussiana**.

En el caso de entrenamiento de los sistemas expertos basados en ANN, se ha optado por utilizar una estructura de red neuronal de tipo perceptron multicapa donde no hace falta retroalimentación de los datos. Para el entrenamiento de la red se ha optado por el algoritmo de retropropagación del gradiente conjugado escalado y se ha utilizado una función de coste del tipo **entropía cruzada del error**.

La configuración de estos sistemas expertos autogenerados se desarrolla con más detalle en el apartado 5.4 y 5.5.

En el **Anexo 1** de esta tesis se ha insertado información adicional sobre cómo crear un clúster de redes neuronales, donde en teoría se obtiene una pequeña mejoría en la tasa de aciertos respecto al uso de una sola red neuronal. Este clúster de redes neuronales puede desarrollarse como un trabajo futuro para mejorar un poco los ratios de aciertos obtenidos con este tipo de sistemas expertos.

Capítulo 4

Banco de ensayos de los motores

4.1 Introducción

En el transcurso de este capítulo se establece la descripción del banco de ensayos que ha sido utilizado con el fin de ejecutar las pruebas de validación experimental correspondientes a las técnicas de diagnóstico descritas en el presente trabajo.

El banco de ensayos proporciona la posibilidad de emular (funcionar de la misma manera) funcionamientos del motor tanto en régimen estacionario como en régimen no estacionario.

Entre otras funcionalidades, el banco de ensayo permite que el motor pueda conectarse directamente a la red, a tensión por debajo de la nominal, a distinto régimen de frecuencia a través de sendos convertidores de frecuencia controlados mediante configuraciones variadas, e incluso operando en modo generador, aunque en esta tesis no se empleará dicho funcionamiento.

El objetivo de este capítulo es diseñar un banco de ensayo que permita reproducir condiciones industriales de los motores de inducción

4.2 El banco de ensayos.

El método científico establece que toda hipótesis o marco validado teóricamente debe ser sometido a evaluación empírica. Esta validación debe ser repetitiva de modo que

sirva para la confirmación de los resultados previstos en dicha hipótesis. La validación experimental permite la posibilidad de que la hipótesis sea no certera, es decir, buscar al menos un contraejemplo que invalide el marco teórico. Si con cada ensayo y repetición se verifican los resultados teóricos, la ausencia de contraejemplo conlleva una mayor robustez de la hipótesis planteada que presentara mayor grado de certidumbre. Cuando un marco teórico ha sido ensayado lo suficiente como para alcanzar dicho grado de certidumbre la hipótesis pasa a considerarse teoría.

Con el fin de dotar de una base robusta a los fundamentos teóricos de las técnicas de diagnóstico desarrolladas, una vez expuestos tales fundamentos, se ha creado una plataforma de pruebas empíricas que permita dar soporte experimental al marco teórico de dichas técnicas. Esta plataforma de pruebas experimentales se establece como el “Banco de Ensayos”, que será descrito a lo largo del presente capítulo.

Un banco de ensayos, plataforma de pruebas, de experimentación, etc., consiste en una instalación que dispone el equipamiento e instrumentación que permiten medir, cuantificar, y acotar, los parámetros y características de una máquina simulando su régimen de funcionamiento en condiciones reales. Dicha plataforma posibilita la comprobación exhaustiva, rigurosa, precisa, exacta, repetible y definida del nuevo marco o método formulado teóricamente.

Por lo tanto, en base a la definición de “banco de ensayos” y a que las técnicas de diagnóstico que han sido desarrolladas cubren cualquier condición de trabajo que se pueda presentar, se infiere que dicho banco debe ser capaz de repetir cualquier estado de funcionamiento que la máquina pueda presentar en el mundo real. Cuanta más versatilidad proporcionemos al banco, mejor podrán simularse los varios regímenes de funcionamiento que puedan presentarse en un marco real.

Como se ha mencionado, la validez de las técnicas desarrolladas comprende cualquier régimen de funcionamiento en circunstancias reales, es decir, ya sea régimen estacionario como transitorio, trabajando con carga y/o velocidad constante o bien carga y/o velocidad variable, funcionando bajo diversos niveles de tensión (bien en situaciones de topología estrella/triángulo o bien por problemas inherentes a la red), estableciendo la conexión mediante convertidores de frecuencia controlados mediante diversos métodos, con conexión directa a la red, e incluso operando en modo generador.

La reproducción de las condiciones de funcionamiento real de la máquina debe ser una de las principales características de la cuales esté provisto el banco de ensayos. Por consiguiente, el banco ha sido provisto de los elementos necesarios para reproducir todas y cada una de tales condiciones. En el apartado 4.3 se detalla los aspectos de dichos elementos.

Por consiguiente, se ha instalado un banco de ensayos en el laboratorio de control de máquinas eléctricas del Departamento de Ingeniería Eléctrica de la Universitat Politècnica de València, con el objeto comprobar de manera exhaustiva, rigurosa, precisa,

exacta, repetible y definida cada una de las técnicas de diagnóstico desarrolladas en el marco teórico de este trabajo.

Para poder lograr los objetivos propuestos, dicho banco se ha dotado de automatismos que permiten realizar los mismos ensayos bajo las condiciones definidas rigurosamente, tantas ocasiones como se requiera.

De esa manera es posible comprobar y comparar los resultados en un mismo ensayo llevado a cabo con diversos motores que presenten el mismo o distinto tipo de fallo. Lo cual permite contrastar con las respuestas obtenidas en distintas circunstancias y, principalmente, con la parametrización obtenida de la respuesta a ensayo de una máquina sana, probando la validez del marco teórico de las técnicas de diagnóstico en cualquier circunstancia o condición de funcionamiento posible de la máquina en la vida real. Una ilustración del banco de ensayos puede verse en la Figura 4.1.

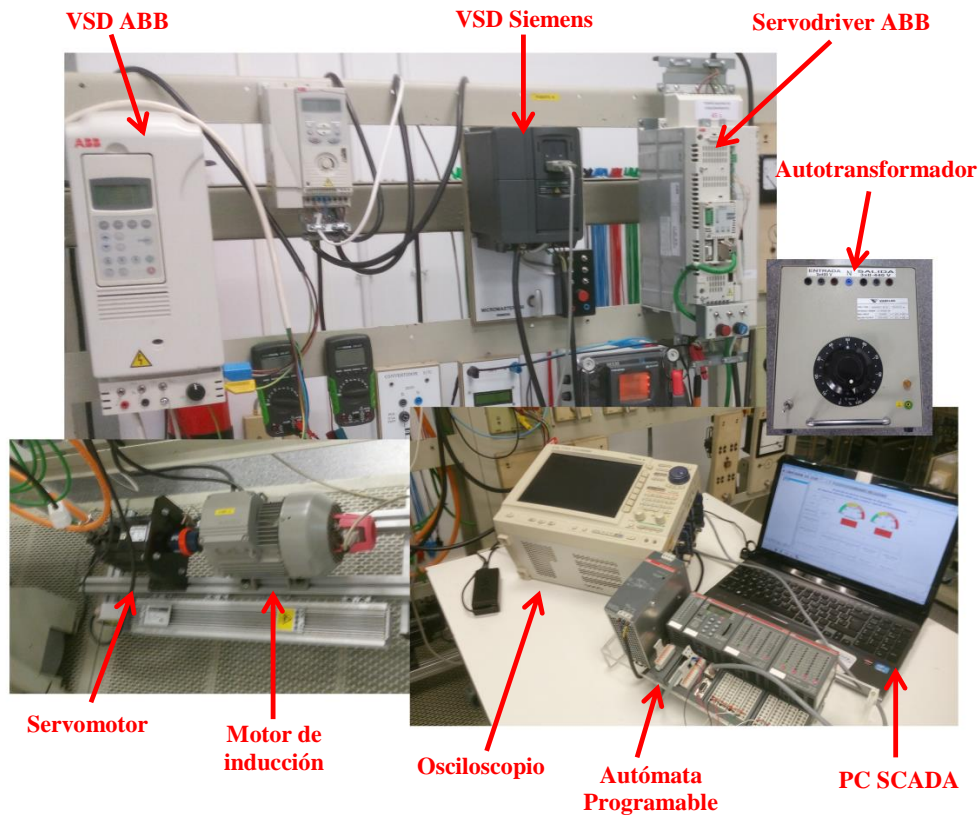


Figura 4.1 Ilustración del banco de ensayos.

4.3 Elementos que forman el banco de ensayos.

A continuación, se elabora la descripción de los elementos que conforman el banco de ensayos desarrollado:

- Un motor de imanes permanentes del tipo síncrono (servomotor).
- Convertidor de frecuencia (variador AC) marca ABB modelo ACSM1, para controlar el régimen de trabajo del motor síncrono.
- Varios motores asíncronos de inducción que presentan las siguientes fallas de funcionamiento: barra rota, excentricidad, y fallo en el cojinete. Además se dispone de otro conjunto de motores en perfecto estado. Los motores son los elementos de ensayo disponibles todos ellos de las mismas características nominales e idéntica serie de fabricación.
- Dos convertidores de frecuencia que controlan el motor sometido a ensayo:
 - e) ABB modelo ACS 800 que implementa dos tipos de control: escalar y control directo del par motor (DTC o “Direct Torque Control”).
 - f) Siemens modelo M440 que implementa tres tipos de control: Escalar, escalar con compensación de deslizamiento y vectorial sin sensor de velocidad.
- Autotransformador que permite someter al motor a diferentes tensiones de ensayo.
- Un autómatas programable.
- Un osciloscopio.
- Un ordenador.

El banco de ensayos se compone de 3 subsistemas funcionales distintos:

- Sistema de potencia (4.3.1).
- Sistema de control automatizado (4.3.2).
- Sistema de adquisición de datos (4.3.3).

En adelante se mostrarán las características más importantes de los elementos que componen tales sistemas y su función específica dentro del banco de pruebas.

4.3.1 Sistema de potencia.

Se procede a mostrar los elementos que integran el sistema de potencia en el banco de ensayos (Figura 4.2).

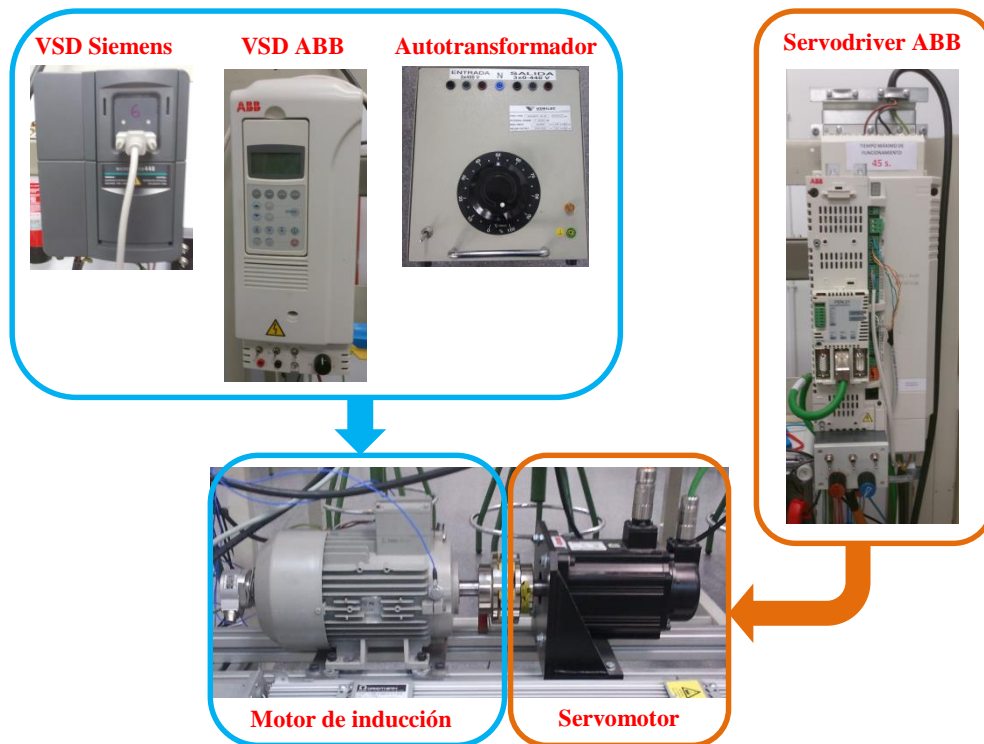


Figura 4.2 Sistema de potencia.

4.3.1.1 Motores de inducción.

El elemento clave en un banco de ensayos es el equipo que se va a someter a prueba. En este caso los objetos de ensayo son motores de inducción del tipo “jaula de ardilla”. Sin embargo, el banco de ensayos se encuentra plenamente capacitado para poder probar cualquier clase de maquina eléctrica rotativa siempre que esté dentro de un rango de potencia tolerado por el motor de imanes permanentes que funciona como carga.

Los motores de inducción objetos de ensayo, tipo jaula de ardilla, son de marca Siemens, modelo 1LA7090-2AA10 [127], los cuales son idénticos al motor mostrado en la Figura 4.3.

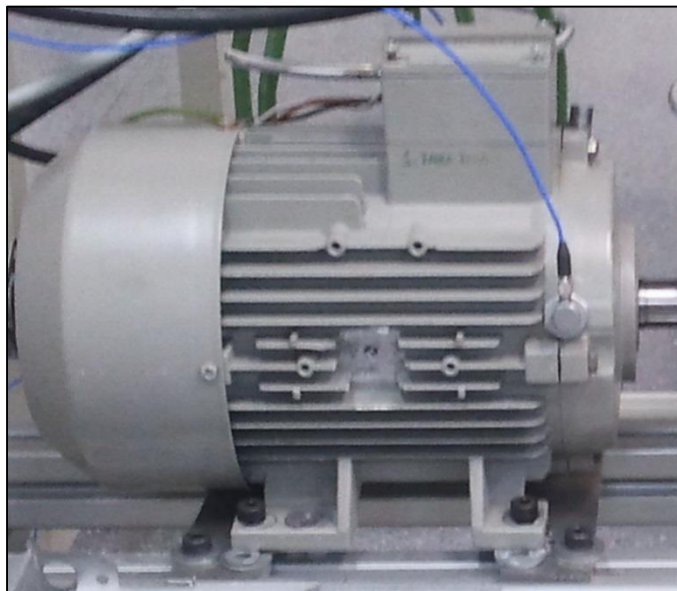


Figura 4.3 Motor Siemens modelo 1LA7090-2AA10.

Las características eléctricas más importantes se muestran en la Tabla 4.1 y han sido obtenidas a partir de la placa de características tal y como se muestra en la Figura 4.4.

Tabla 4.1. Características eléctricas del motor Siemens modelo 1LA7090-2AA10.

Tensión nominal	230V Δ / 400V Y	Frecuencia	50 Hz
Potencia nominal	1.500 W	Pares de polos	1
Corriente nominal	5,7A Δ / 3,3A Y	Par nominal	5 Nm
Velocidad nominal	2.860 r.p.m	Factor de potencia	0,85
Deslizamiento nominal	4,67 %	Rendimiento	77,2 %



Figura 4.4 Placa de características del motor Siemens modelo 1LA7090-2AA10.

Los motores con los cuales se han realizado los ensayos se corresponden a dos grupos, motores sin fallo y motores con fallo. Para el desarrollo de los sistemas clasificadores de esta tesis se han empleado concretamente dos tipos de motores, sano y con barra rota, tal y como se muestran en la Figura 4.5.



Figura 4.5 Fotografía que muestra el rotor sano y el rotor con rotura en una de sus barras.

Motor Sano (“sin fallo”). Se ha utilizado como patrón un motor prácticamente en la misma condición en la que se encontraba al fabricarse, no se ha realizado ninguna merma artificial de sus características y los únicos defectos que pudieran surgir serían inherentes a su proceso de fabricación y a las imperfecciones en la uniformidad y homogeneidad (o pureza) de las materias primas empleadas. Además, estos motores han sido validados de forma práctica como libre de fallos y que opera en perfectas condiciones.

Motor con barra rota. Partiendo de un motor sano se le ha practicado la rotura de una barra rota. De esta forma, se ha generado de forma artificial el fallo de barra rota. En la Figura 4.5 se puede apreciar el rotor sano del motor y el rotor con fallo donde se ha practicado el mecanizado sobre una de las barras del rotor.

4.3.1.2 Auto-transformador.

La máquina debe de poder ser probada en el banco de ensayos estando directamente conectada al suministro de red.



Figura 4.6 Autotransformador de la marca Verilec modelo 3ARC 5-2.



Figura 4.7 Placa de características del autotransformador Verilec 3ARC 5-2.

Gracias al auto-transformador empleado de la marca Verilec tipo 3ARC 5-2 [128], es posible recrear un significativo número de condiciones de funcionamiento con tensiones reales con las que la máquina rotativa puede verse implicadas operando en un entorno industrial. En la Tabla 4.2 se muestran las características técnicas de este auto-transformador obtenidas a partir de la placa de características mostrada en la Figura 4.7.

Tabla 4.2. Características técnicas del auto-transformador Verilec 3ARC 5-2.

Tensión de entrada	3 x 400V	Tensión de salida	3 x (0 - V_{Entrada})
Corriente de entrada	10 A	Corriente de salida	10 A
Frecuencia de entrada	50 Hz	Frecuencia de salida	50 Hz
Potencia	7.500 VA	Tipo	3ARC 5-2

4.3.1.3 Variadores de frecuencia.

Los variadores de frecuencia (también conocidos como unidades de variación de la velocidad VSD, o variadores AC) se emplean en innumerables aplicaciones industriales por su versatilidad para controlar un amplio abanico de motores de diversa índole.

Tal versatilidad es debida a los muchos métodos de control específicos que implementan en mayor o menor medida. Debido al significativo papel que estos convertidores desempeñan en el entorno industrial, el banco de ensayos se ha dotado de dos convertidores de frecuencia de distinto modelo.

Los variadores utilizados provienen de dos fabricantes de referencia en entornos industriales, los cuales son Siemens y ABB. Seguidamente se muestran las características más destacables de cada convertidor, y los métodos de control utilizados en los ensayos.

Variador de frecuencia de la marca Siemens. El variador de frecuencia de Siemens empleado en el banco corresponde al modelo Micromaster 440 de potencia nominal 2,2kW [129] y que se ilustra en la Figura 4.8.

Sus características más destacables se han obtenido de su placa de características (que se puede observar en la Figura 4.9) y son las que a continuación se exponen en la Tabla 4.3.



Figura 4.8 Variador de frecuencia marca Siemens modelo Micromaster 400.

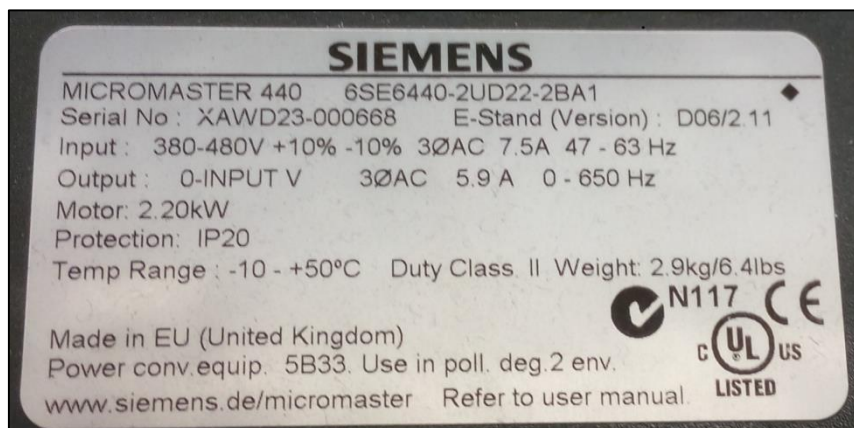


Figura 4.9 Placa de características del variador de frecuencia marca Siemens modelo Micromaster 400.

Tabla 4.3. Características técnicas del variador Siemens modelo Micromaster 400.

Tensión de entrada	380 - 480V ±10%	Tensión de salida	0 - V _{Entrada}
Corriente de entrada	7,5 A	Corriente de salida	5,9 A
Frecuencia de entrada	47 - 63 Hz	Frecuencia de salida	0 - 650 Hz
Potencia	2.200 W	Versión	D06/2.11

En los ensayos que se han realizado han sido utilizados tres métodos de control distintos con este variador:

- **Control escalar:**

Este tipo de control pretende que el flujo se mantenga constante. En consecuencia, las variables frecuencia y tensión se encuentran interrelacionadas, esto es, una variación en frecuencia implica una variación de tensión con el fin de mantener constante dicho flujo. La dependencia de la tensión inducida con respecto a la frecuencia y el flujo corresponde a la ecuación (4.1)

$$V = 4.44 \cdot f \cdot N \cdot \phi_m \quad (4.1)$$

$$\phi_m = k \cdot \frac{V}{f} \quad (4.2)$$

En esta ecuación (4.1) se puede despejar la variable ϕ_m que representa el flujo de magnetización (4.2). La variable V se corresponde con la tensión de alimentación, f se corresponde con la frecuencia, N representa al número de espiras en serie y k es un parámetro dependiente de la construcción de la máquina y su diseño.

Este método de control (conocido también como control tensión/frecuencia o V/f) pretende mantener el flujo magnético invariable en el entrehierro de la máquina y así conseguir alcanzar el **par máximo** en cualquier configuración (es decir, funcionando a la frecuencia deseada). La principal ventaja de este método es su operación en lazo abierto por lo que no precisa de ninguna medida de la velocidad.

Sin embargo el inconveniente más importante se manifiesta en el funcionamiento en la zona de frecuencias bajas, donde la tensión puede no ser lo suficientemente elevada como para que el motor supere la resistencia del par o incluso el momento inercial de giro.

La Figura 4.10 muestra la gráfica de la relación entre Tensión y frecuencia de un control de tipo escalar. No existe compensación alguna dependiendo del nivel de la carga. Por ello el deslizamiento y en consecuencia la velocidad mecánica de giro del motor será en principio desconocida y dependerá de la carga que se oponga al giro del motor. Así pues no resulta adecuado el empleo de este método de control cuando la aplicación exija como requisito insalvable el control preciso de la velocidad de la máquina.

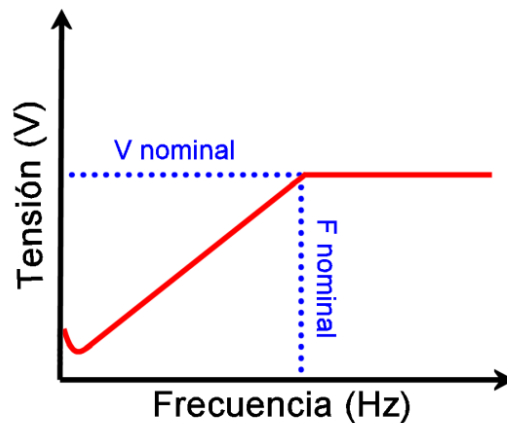


Figura 4.10 Gráfica de relación V/f de un control escalar.

- **Control escalar con compensación de deslizamiento (Scalar + Slip):**

La Figura 4.11 muestra la relación V/f del control escalar con compensación de deslizamiento. Este control es similar al control escalar, no obstante en este tipo de control cuando se detecta un incremento de la corriente solicitando un par mayor, se pasa del punto de trabajo 1 al punto 2 para compensar el aumento de la carga y así evitar la pérdida de velocidad.

El control del variador supervisa la corriente del motor ya que dicha corriente la suministra el variador mismo. El incremento de tensión y frecuencia de trabajo se proporciona con el fin de mantener la velocidad del eje de la máquina

invariable de manera independiente de la oposición al giro que la carga esté manifestando.

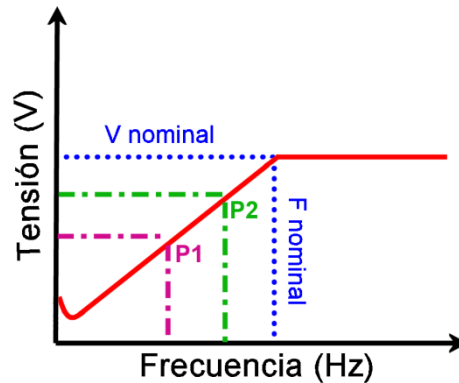


Figura 4.11 Gráfica de relación V/f de un control escalar con compensación del deslizamiento.

- **Control vectorial (sin sensor de velocidad):**

Para este tipo de control el convertidor necesita un modelo de aproximación funcional de la máquina [130]. De manera habitual se parametrizan las características de la máquina a partir de una función interna del convertidor. Este tipo de control utiliza un sistema cartesiano de coordenadas aplicando la transformada de Clark-Park (por lo general los ejes en cuadratura d-q).

A raíz de este modelo se deriva una máquina bifásica en la que las variables senoidales trifásicas, trabajando en régimen estacionario, son transformadas en valores constantes siempre que el sistema trifásico se encuentre equilibrado y sea simétrico.

El control vectorial tiene como ventaja principal el desacople de las variables eléctricas del motor de inducción. Por consiguiente, se logra un control del par y velocidad de manera independiente, semejante al tipo de control utilizado para una máquina de continua.

Como ventaja añadida, la precisión y la respuesta dinámica se ven mejoradas en comparación con el control del tipo escalar. Concretamente, la referencia de la máquina medida con el campo giratorio en el entrehierro es correspondida en un sistema de referencia que gira con el vector de flujo en el rotor, Ψ_R provocando que la componente de corriente directa i_d controle el flujo magnético y la componente en cuadratura i_q , controle el par.

Sin embargo los inconvenientes más destacables del tipo de control vectorial se pueden enumerar de la siguiente manera:

- a) Se necesitan conocer los valores de algunos parámetros de construcción y de diseño del motor para poder obtener la aproximación matemática.
- b) Este sistema de control necesita de una capacidad de proceso de datos y operaciones aritméticas en tiempo real elevada. No obstante no suelen emplearse frecuencias de muestreo elevadas para este tipo de control (a lo sumo unos cuantos kHz) y el flujo de entrada/salida de datos es bajo relativamente en comparación con las capacidades de proceso de los dispositivos de tecnología actual, lo cual indica que este inconveniente no representa una limitación seria en la implementación adecuada de este tipo de control.

Variador de frecuencia de la marca ABB. El variador de frecuencia de la marca ABB empleado en el banco corresponde al modelo ACS800-01-0005-3+E200+L502 el cual puede observarse en Figura 4.12. Su potencia nominal es de 3kW.



Figura 4.12 Variador de frecuencia marca ABB modelo ACS800-01-0005-3+E200+L502.

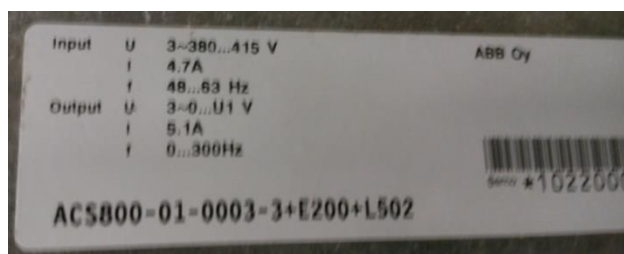


Figura 4.13 Placa de características del variador de frecuencia marca ABB modelo ACS800-01-0005-3+E200+L502.

Tabla 4.4. Características técnicas del variador ABB modelo ACS800-01-0005-3+E200+L503.

Tensión de entrada	380 - 415V \pm 3%	Tensión de salida	0 - V _{Entrada}
Corriente de entrada	7,9 A	Corriente de salida	8,5 A
Frecuencia de entrada	48 - 63 Hz	Frecuencia de salida	0 - 300 Hz

Las características eléctricas más destacables se pueden observar en la placa de características de la Figura 4.13 y en la Tabla 4.4.

Los métodos de control que este modelo en concreto implementa y que han sido utilizados en los ensayos son dos:

- **Control escalar:**

Este método es el mismo tipo de control que el correspondiente en el convertidor de frecuencia Siemens y que ya ha sido mostrado con anterioridad.

- **Control directo del par (Direct Torque Control o DTC):**

Este control es similar en complejidad al control vectorial antes descrito.

Éste permite el control del motor en lazo abierto lo que implica que no es necesario muestrear la velocidad de la máquina para conocerla en cualquier momento. Las variables controlables son el par del motor y el flujo magnético.

Este método requiere un sistema de procesamiento de señales de alta velocidad y gran capacidad de operaciones matemáticas. Dentro de este sistema se implementa un modelo adaptativo matemático de la máquina lo cual posibilita ade-

cuar la respuesta de la máquina dependiendo de la configuración del control deseada.

Las ventajas más destacables del control DTC son:

- a) Rapidez de respuesta del par. Este tipo de control consigue la respuesta más rápida de todos los métodos de control.
- b) Reproducción precisa del par bajo la misma configuración. Bajo la misma referencia el par de salida puede reproducirse con la precisión suficiente.
- c) Velocidad precisa. El margen de error entre la velocidad de referencia y la velocidad real es el menor margen de todos los métodos de control actuales.
- d) En lugar de implementar un modulador, que introduce retrasos en la respuesta de control, se implementa una tabla de estados que permite alcanzar más rápidamente el control deseado obteniendo un margen dinámico de respuesta mejorado en comparación con el resto de métodos.

4.3.1.4 Variador de frecuencia marca ABB modelo ACSM1 (Driver del servomotor).

El control del motor síncrono de imanes permanentes, que hace la función de carga del motor ensayado, ha sido implementado usando un variador de frecuencia modelo ACSM1-04AS-024A-4+L516 de la marca ABB [131]. Este se muestra en la Figura 4.14 y sus características en la Tabla 4.5.



Figura 4.14 Variador de frecuencia marca ABB modelo ACSM1-04AS-024A-4+L516.

Tabla 4.5. Características técnicas del variador ABB modelo ACSM1-04AS-024A-4+L516.

Tensión de entrada	3 x 380 - 480V	Tensión de salida	3 x 0 - $V_{Entrada}$
Corriente de entrada	20,2 A	Corriente de salida	24 A
Frecuencia de entrada	48 - 63 Hz	Frecuencia de salida	0 - 500 Hz

En la Figura 4.15 se muestra el módulo FEN21 con el que viene equipado este variador. Este módulo sirve para su control. Dispone de una salida que permite emular el tren de pulsos de un encoder a partir de los datos obtenidos del resolver.

Esta característica es de gran utilidad ya que en las técnicas desarrolladas en el presente trabajo se necesita conocer la posición angular de la máquina y/o la velocidad instantánea. La resolución de pulsos por revolución de dicha salida es programable, lo que permite la adaptación de la señal a varios sistemas de adquisición y procesamiento de señales con diferentes frecuencias de muestreo.

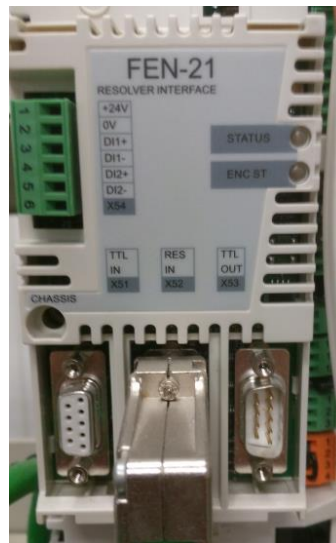


Figura 4.15. Módulo FEN21 del variador de frecuencia marca ABB modelo ACSM1-04AS-024A-4+L516.

4.3.1.5 Motor de imanes permanentes de tipo síncrono.

La carga mecánica es un elemento crucial en todo sistema en el que interviene una máquina eléctrica.

El motor síncrono de imanes permanentes de ABB modelo MS4839N4008E43C10 [132] de la Figura 4.16 se ha seleccionado para representar la carga a la que se somete el motor de ensayo por la versatilidad y las ventajas que presenta en comparación con los otros métodos para simular una carga (servofrenos eléctricos, sistemas mecánicos de arrastre, generadores, etc.).

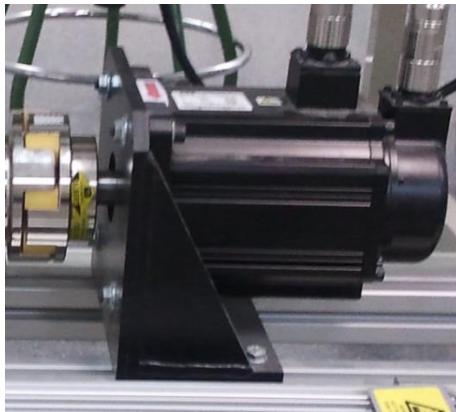


Figura 4.16. Motor síncrono de imanes permanentes.



Figura 4.17. Placa de características del motor de imanes permanentes.

Tabla 4.6. Características técnicas del motor de imanes permanentes.

Potencia nominal	4.900 W	Frecuencia	200 Hz
Corriente nominal	14,4 A	Velocidad nominal	3.000 r.p.m
Corriente nominal	5,7A Δ / 3,3A Y	Par nominal	15,5 Nm
Tensión a velocidad nominal (3.000 r.p.m.)	203,6V		

La Tabla 4.6, que ha sido obtenida a partir de la placa de características del motor de imanes permanentes de la Figura 4.17, muestra los parámetros más relevantes de dicho motor.

Se puede apreciar que ha sido elegido un servomotor de potencia mayor que los motores de inducción objetos de ensayo. Esta elección ha sido basada considerando dos criterios.

Por un lado se asegura que en los ensayos el par resistente ofrecida por este motor sea suficiente para someterlos a las diversas condiciones de trabajo

Por otra parte el sobredimensionado se justifica al tratar de dotar la suficiente versatilidad al banco de pruebas en vistas de que en el futuro sea posible el ensayo de máquinas de mayor potencia que las ensayadas en este trabajo. Debido a esa potencia superior, y como medida de protección, se ha procedido a limitar el par resistente que ofrece esta máquina en los ensayos al par nominal de los motores objeto de estudio de este trabajo

En los ensayos se ha utilizado el control de par para simular cargas de tipo constante, pulsantes y rampas de carga.

4.3.2 Sistema de control automatizado.

El banco de ensayos se ha automatizado para llevar a cabo varias funciones:

- Una de las funciones se basa en dotar a los ensayos de repetitividad en los experimentos. Al someter las diversas repeticiones del ensayo a las mismas condiciones de funcionamiento y con el mismo motor se pueden comparar entre sí los resultados del mismo ensayo o bien, con distintos motores, comparar los resultados de un mismo experimento al que se someten motores que presentan varias fallas.
- Otra función es obtener un sistema autónomo no precisando intervención humana más allá de las tareas de sustitución del motor a ensayar. Por lo tanto, en cuanto éste se ha instalado en el banco, el sistema realiza la tarea de supervisar y completar los ensayos, y almacenando los resultados de forma automática.

La Figura 4.18 muestra el diagrama del sistema de control de automatizado. Éste se compone de un autómata programable (PLC) que controla mediante contactores los variadores para ejecutar el ensayo y el osciloscopio para comenzar la adquisición de los datos. También se encuentra disponible un ordenador personal que da soporte al PLC por medio del sistema de Supervisión del Control y Adquisición de Datos (SCADA), de forma que mediante dicho SCADA se pueden programar y monitorizar los diversos ensayos del banco de pruebas. El ordenador personal y el autómata programable se describen dentro de este capítulo.



Figura 4.18. Diagrama del sistema de control automatizado.

4.3.2.1 El autómata programable.

Para lograr la automatización de las tareas del banco de ensayos se ha elegido la implementación mediante un autómata programable. Este tipo de control es el más ampliamente utilizado en entornos industriales gracias a su fiabilidad y resistencia a las condiciones de operación en dichos entornos.

Una de las características de un autómata programable es que se ha fabricado para funcionar de manera ininterrumpida, previendo las excepciones en el funcionamiento

normal del automatismo. Para el banco de ensayos se ha elegido el autómata programable modular de ABB (Figura 4.19) cuyo modelo de referencia es PM583ETH [133]

Ya que es modular es posible actualizar su hardware en caso necesario añadiendo tarjetas. Entre otros interfaces, las tarjetas comprenden módulos de E/S ya sea digitales o analógicos, módulos de comunicaciones o incluso tarjetas con funcionalidades específicas como la modulación por ancho de pulsos (en inglés Pulse Width Modulation, PWM), entradas de pulsos en frecuencia como pueden ser los encoders, etc.



Figura 4.19. Autómata programable de la marca ABB modelo PM583ETH.

Al sistema utilizado se le ha dotado de los siguientes módulos adicionales:

- DA554-CM. Tarjeta de 8 canales digitales bidireccionales configurables como entradas o como salidas.

DA501. Una tarjeta de expansión de E/S. Los canales de esta tarjeta son configurables, y dentro de este módulo se encuentran disponibles entradas y salidas tanto analógicas como digitales.

La pieza clave del sistema de control la representa el autómata programable. Es el que tiene la labor de controlar todos los sensores y actuadores periféricos, es decir, los variadores de frecuencia, el auto-transformador, el servodriver, los contactores de potencia y el osciloscopio. Cada elemento es controlado por el autómata de la siguiente manera:

- Variadores de frecuencia: Establece los periodos de marcha y paro, la frecuencia de referencia y el sentido en el que girará el motor.
- Autotransformador: Establece su conexión, desconexión y la variación de la tensión de entrada; por lo que permite controlar la alimentación directa a la red del motor mediante este componente.
- Servodriver: Señala el modo de control en que debe operar. Controla la puesta en marcha y el par resistente.

- Contactores de potencia: La actuación tanto en la conexión como en la desconexión de éstos es controlada por el autómata programable de manera que se decide cuál de los convertidores de frecuencia controla el motor objeto de ensayo. Como medida añadida, impiden que el motor pueda estar conectado a dos convertidores a la vez para mayor seguridad.
- Osciloscopio: El autómata controla el inicio y la finalización de la adquisición y almacenamiento de las señales de monitorización.

4.3.2.2 Ordenador personal.

El equipo informático es utilizado para dar soporte de actualización del programa interno del autómata programable y para la aplicación SCADA que se ha desarrollado en él. Así pues, se utiliza un equipo portátil marca ASUS (Figura 4.20) que cumple los requisitos mínimos que el fabricante de dicho software establece.

Las características de equipo informático son:

- Procesador i7 modelo i7-3632QM. Velocidad de proceso de 2200MHz. Cuatro núcleos independientes de dos hilos por núcleo.
- Memoria RAM de 8GB DDR3.
- Gráfica marca Intel modelo HD Graphics 4000.
- Disco duro mecánico de 750GB a 4500 rpm de giro. Bus SATA3 de 600MB/s.
- Sistema operativo Windows 7 versión profesional de 64 bits con *service pack* SP1.



Figura 4.20. Ordenador portátil de la marca ASUS con SCADA de CodeSys.

Para la programación del autómatas programable se utiliza el entorno integrado de desarrollo “Control Builder Plus versión 2.1”, basado en CodeSys Control Builder Plus.

Este entorno de desarrollo ha sido diseñado para la programación eficiente de autómatas y soporta todos los lenguajes de programación especificados en el estándar IEC 61131-3 además del lenguaje de programación gráfico CFC (Continuous Functions Chart).

Dentro de este entorno IDE se pueden añadir visualizaciones SCADA y de esa manera se ha implementado el panel de configuración y control del banco de pruebas.

4.3.3 Subsistema de medida y adquisición de datos.

Con toda probabilidad se pueden considerar a los equipos de adquisición de datos como elementos completamente imprescindibles del banco ya que permiten atestiguar que el ensayo se efectúa tal y como ha sido previsto. La Figura 4.21 muestra el diagrama del sistema de adquisición de datos donde el osciloscopio recopila los datos de tensiones y corrientes a través de las sondas acopladas.

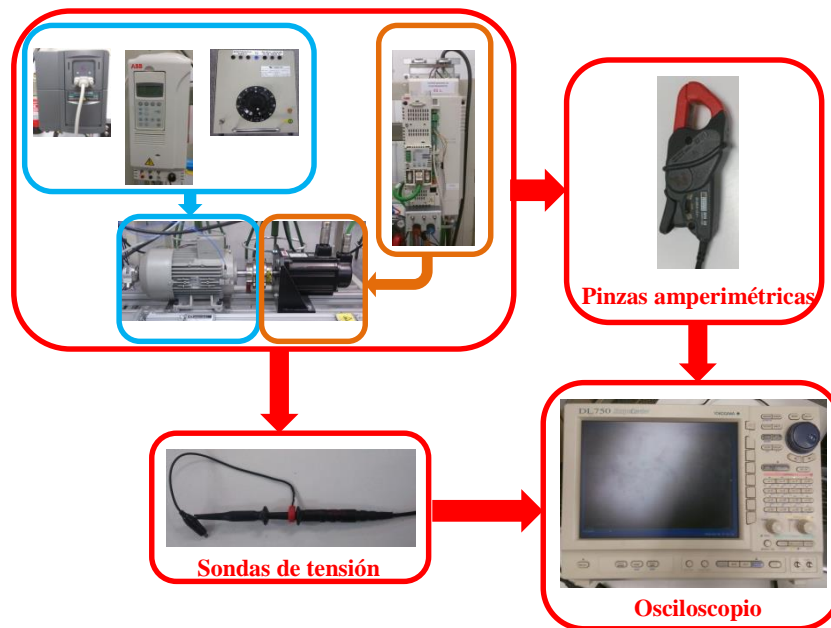


Figura 4.21. Sistema de adquisición de datos.

Las técnicas de diagnóstico se basan en el análisis de la respuesta de una corriente del motor ensayado. No obstante, y teniendo en consideración que los resultados de los

ensayos puedan utilizarse como referencia en futuras investigaciones en los cuales las muestras de datos puedan ser de otra índole, se han adquirido un conjunto de señales adicionales para este fin.

La adquisición de varias señales permite disponer de un registro más completo de los parámetros del experimento, para usar como base de datos en vistas de investigaciones futuras, que únicamente adquiriendo las señales de interés de este trabajo.

A continuación se exponen las señales que el sistema registra en dicha base de datos:

- Valores de las tres tensiones en los electrodos del motor.
- Valores de las tres corrientes que circulan por esos puntos.
- Valor de una corriente de línea del servomotor. Ya que dicha corriente es proporcional al par desarrollado por él, sirve para tener constancia del par resistente ofrecido por dicho servomotor.
- Captura de los pulsos generados por la tarjeta FEN-21 y que como anteriormente se ha detallado, corresponden a la emulación de un encoder.

Con el fin de adquirir todos estos valores se han usado los instrumentos y dispositivos que se describen en los siguientes apartados.

4.3.3.1 El Osciloscopio.

El modelo de osciloscopio empleado para la captura de las señales corresponde con la referencia DL750 de la marca Yokogawa [134] y que la Figura 4.22 muestra a continuación.



Figura 4.22. Osciloscopio Yokogawa modelo DL750.

Se trata de un osciloscopio portátil de gama alta y de altas prestaciones. Se pueden destacar las siguientes características:

- Puede capturar hasta 16 canales, que están compuestos por tarjetas modulares aisladas y comprenden los siguientes modelos:
 - a) Modelo 701250: Velocidad de muestreo de 10MS/s. Resolución de 12 bits. Independiente del resto de tarjetas.
 - b) Modelo 701251: Velocidad de muestreo de 1MS/s. Resolución de 16 bits.
 - c) Modelo 701255: Velocidad de muestreo de 10MS/s. Resolución de 12 bits.
 - d) Modelo 701260: Módulo capaz de soportar altas tensiones y velocidad de adquisición de 100kS/s y 16 bits de resolución.
 - e) Modelo 701275: Especializado en la medida de acelerómetros y de voltaje. Tiene un ancho de banda de 100kHz y una resolución de 16 bits.
 - f) Modelo 701280: Módulo contador/frecuencímetro que permite conocer las tasas de pulsos por segundo procedentes de un encoder.

En la imagen Figura 4.23 se muestra una fotografía de dichos módulos.

- Dispone de una gran cantidad de puertos para conectar con diversos interfaces como pueden ser impresoras, teclados, almacenamiento de masa externos, etc. En la Figura 4.24 se muestran las conexiones de los puertos que soporta.
- Puertos de comunicaciones con soporte para varios protocolos: UART, Ethernet, USB, etc.
- Posee la funcionalidad de servidor FTP, sistema de compartición de ficheros desde conexión remota que permite poner a disposición de las señales adquiridas desde cualquier parte del mundo por medio de Internet.



Figura 4.23. Tarjetas modulares del osciloscopio para muestreo.



Figura 4.24. Conexiones de los puertos del osciloscopio.

4.3.3.2 Sondas de Tensión.

La Figura 4.25 muestra una fotografía de una de las sondas utilizadas para la adquisición de los valores de las señales.



Figura 4.25. Sonda para la adquisición de valores.

Ya que en los ensayos se medirán señales de tensión elevada, hasta 400Vrms (560Vp), se utilizan sondas capaces de soportar esas tensiones (Categoría III) a la vez que reducen las señales en un factor $\times 100$ a modo de protección por sobretensión de las entradas de los módulos del osciloscopio.

Las sondas utilizadas de la marca HAMEG modelo H53 usan un cable apantallado coaxial de impedancia 50Ω que les confiere cierta inmunidad al ruido electromagnético parásito. Como se puede ver en la imagen, además poseen en la punta de un anclaje retráctil que permite el agarre firme de la sonda a la conexión que se debe supervisar.

4.3.3.3 Pinzas Amperimétricas.

En la Figura 4.26 se puede apreciar una fotografía de las pinzas amperimétricas utilizadas para la medida de las corrientes. Estas pinzas amperimétricas corresponden al modelo MN39 de la marca Chauvin Arnoux [135].

No es necesario el seccionado del cable en el punto de medida ya que para tomar los valores de corriente sólo se necesita que el cable atravesase el interior de la pinza. Ésta facilita el acceso del cable a su interior mediante un resorte móvil que su apertura. Estas pinzas permiten seleccionar dos rangos de medida: De 0 a 20A ($\times 10\text{mV/A}$) y de 0 a 200A ($\times 100\text{mV/A}$).



Figura 4.26. Pinza amperimétrica utilizada en el banco de ensayos.

A continuación se exponen sus principales características en la Tabla 4.7:

Tabla 4.7. Características técnicas de la pinza amperimétrica.

Modelo	MN 60
Ancho de banda	40 Hz → 40 kHz
Precisión de medida	$\pm 1.5\% + 5 \text{ mV}$
Tamaño máximo del conductor	20mm
Categoría de seguridad	CAT I 300 V, CAT III 600 V

4.4 Base de datos de muestras de ensayos.

Con ayuda del banco de ensayos comentado en el apartado 4.2, se ha generado una completa base de muestras obtenidas de los motores sanos y con fallo de barra rota explicados en este apartado 4.4.

Esta base de datos de muestras está orientada a su uso por parte del sistema generador de sistemas de diagnóstico óptimo para obtener el mejor sistema de diagnóstico en relación a las características de los motores estudiados. Además, parte de las muestras de esta base de datos también han sido desarrolladas para validar la robustez de los sistemas de diagnóstico óptimos resultantes generados por el sistema principal.

Como igualmente se ha expuesto en el apartado anterior 4.3.2, para garantizar que las muestras obtenidas tanto para motor sano como para motor de barra rota se efectúan fielmente bajo la misma configuración y duración se ha automatizado el proceso de generación de las muestras del banco de ensayos mediante un sistema de control por autómatas programables.

Como en el momento de la creación de esta base de datos todavía no se conocía con exactitud que técnicas de obtención de características iban a ser implementadas dentro del sistema, se optó por sacar aquellas magnitudes que pudieran ser usadas por las técnicas de diagnóstico para la obtención de características relevantes de fallo. A continuación se comentan las magnitudes medidas para cada muestra obtenidas mediante el osciloscopio (apartado 4.3.3.1) del banco de ensayos:

- Medidas de las tensiones de las tres fases del motor de inducción obtenidas con las sondas de tensión (apartado 4.3.3.2)
- Medidas de las corrientes de las tres fases del motor de inducción obtenidas con las pinzas amperimétricas (apartado 4.3.3.3)
- Medidas de la corriente de una de las fases del motor síncrono (apartado 4.3.1.5) obtenidas con pinza amperimétrica. Con esta medida se obtiene el porcentaje relativo de carga que desarrolla el motor síncrono.
- Medidas de los pulsos por vuelta generados por el variador de frecuencia del motor síncrono (apartado 4.3.1.4). Este variador de frecuencia que controla el motor síncrono además genera también una trama de pulsos por vuelta equivalentes al que se obtendría usando un encoder. La resolución de este encoder simulado ha sido configurado en 720 pulsos por vuelta.

En total han sido generadas **245 muestras** de motores variando el sistema de control de operación y las cargas de trabajo. La mitad de estas muestras pertenecen al grupo de **motores sanos (126 muestras)** y la otra mitad (**119 muestras**) al grupo de motores con fallo de barra rota. La frecuencia máxima de adquisición de cada muestra es de 100 KHz (tiempo de muestreo de 100 segundos).

Como se explica en el apartado 2.3, a excepción de la técnica de obtención de características representativas de fallos en régimen estacionario ‘Aproximación al Vector de Park Extendido’, todas las demás técnicas requieren de una única fase de corriente para generar la característica de fallo. Por tanto, considerando cada fase de corriente de cada muestra por separado se tiene en conjunto un total de **735 muestras de fase**.

En las siguientes subsecciones se explica la distribución del conjunto de muestras de la base de datos para cada caso en especial (detección en régimen estacionario, detección en régimen transitorio y medidas para la comprobación de los sistemas de diagnóstico obtenidos).

4.4.1 Ensayos para la generación de sistemas de diagnóstico en régimen estacionario.

Del total de muestras en régimen estacionario se han desarrollado **36 muestras de motor sano** y **36 muestras de motor con fallo de barra rota**. Esto hace un total de **72 muestras de motor (216 muestras de fase)** operando en régimen estacionario.

El conjunto de estas muestras para régimen estacionario se han usado tanto para el desarrollo de sistemas de diagnóstico óptimos para régimen estacionario explicado en el capítulo 0, como, en unión de las muestras en régimen transitorio (apartado 4.4.2) el análisis comparativo de los sistemas de diagnóstico óptimos para régimen de trabajo transitorio desarrollado en el capítulo 6 de esta tesis respecto de los sistemas de diagnóstico estacionarios.

En las Tabla 4.8 para conexión por auto transformador, Tabla 4.9 para conexión mediante variador Siemens y Tabla 4.10 para conexión mediante variador ABB; se representan con mayor detalle las variaciones de parámetros de las muestras generadas sobre el tipo de control/tensión utilizada, frecuencia de trabajo constante y par de carga.

Tabla 4.8. Variaciones distintas de parámetros generadas con conexión por autotransformador.

TENSIÓN DE ALIMENTACIÓN (V)	PAR DE CARGA (%)
300 V	0 %
	50 %
	70 %
400 V	0 %
	50 %
	100 %

Tabla 4.9 Variaciones distintas de parámetros generadas con conexión mediante variador de la marca Siemens.

SISTEMA DE CONTROL	FRECUENCIA DE TRABAJO (Hz)	PAR DE CARGA (%)
VECTORIAL	25 Hz	0 %
		35 %
		70 %
	50 Hz	0 %
		50 %
		100 %
ESCALAR	25 Hz	0 %
		35 %
		70 %
	50 Hz	0 %
		50 %
		100 %
ESCALAR CON COMPENSACIÓN DEL DESLIZAMIENTO	25 Hz	0 %
		35 %
		70 %
	50 Hz	0 %
		50 %
		100 %

Partiendo de los diversos sistemas de conexión: por alimentación directa mediante auto transformador (apartado 4.3.1.2), por variador de frecuencia Siemens (apartado 4.3.1.3), y por variador de frecuencia ABB (apartado 4.3.1.3); se han generado distintas muestras variando los parámetros del ensayo correspondientes al sistema de control, la frecuencia de trabajo y el par de carga aplicado. De este modo se ha obtenido un amplio abanico de posibles casos de funcionamiento en régimen de trabajo estacionario tanto para motores sanos como motores con fallo de barra rota.

Tabla 4.10. Variaciones distintas de parámetros generadas con conexión mediante variador de la marca ABB.

SISTEMA DE CONTROL	FRECUENCIA DE TRABAJO (Hz)	PAR DE CARGA (%)
ESCALAR	25 Hz	0 %
		35 %
		70 %
	50 Hz	0 %
		50 %
		100 %
CONTROL DIRECTO DEL PAR (DTC)	25 Hz	0 %
		35 %
		70 %
	50 Hz	0 %
		50 %
		100 %

4.4.2 Ensayos para la generación de sistemas de diagnóstico en régimen transitorio.

En el caso de las muestras para el diagnóstico en régimen transitorio, se han generado **90 muestras de motor sano y 83 muestras de motor con fallo** de barra rota. En **total 173 muestras de motor (519 muestras de fase)** operando en régimen transitorio.

Mediante los sistemas de control por variador, se ha variado la frecuencia de trabajo en rampas de 10 segundos de duración, creando así transitorios en las corrientes de fase por medio de la variación de la frecuencia. En la Figura 4.27 se visualiza una corriente de una fase correspondiente a una muestra con transitorios entre los 40 y 50 hertzios.

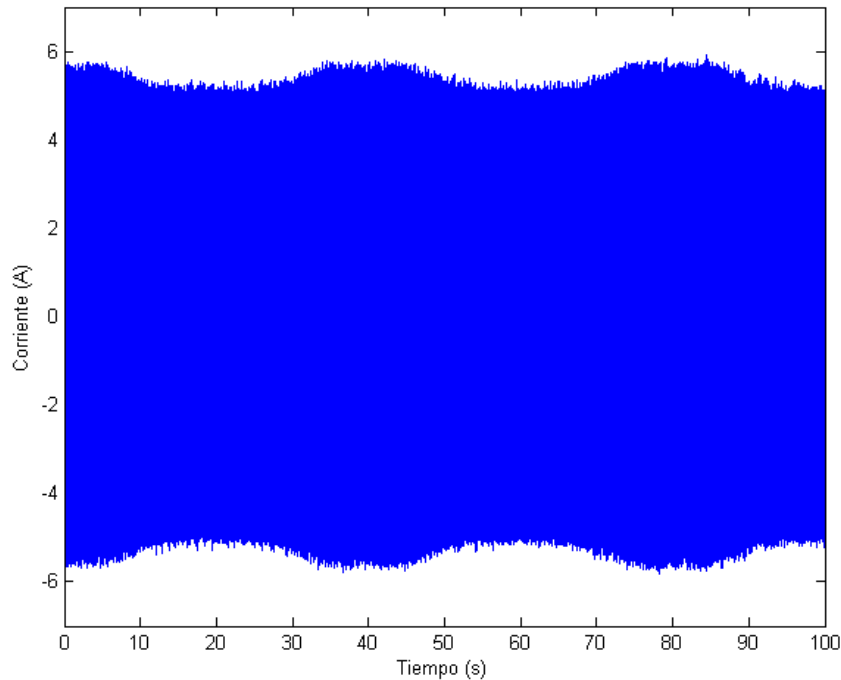


Figura 4.27. Muestra de corriente de fase con control DTC, carga al 100% y con transitorios de 10 segundos de duración.

También se ha generado transitorio por medio de la variación en rampa del par de carga, independientemente del régimen de trabajo estacionario o transitorio aplicado. En la siguiente Figura 4.28 se visualiza una corriente de una fase correspondiente a una muestra con una frecuencia de alimentación constante a 50Hz y un par de carga transitorio entre el 25% y el 100%.

Al igual que en el proceso de obtención de muestras en régimen estacionario, para este conjunto de muestras en régimen transitorio, se han ido variando los parámetros del ensayo correspondientes al sistema de control, el rango de frecuencias de trabajo, el rango de los transitorios sobre la frecuencia de trabajo y el par de carga aplicado para cada caso.

Las muestras obtenidas de motores sanos y motores con fallo para régimen transitorio se han generado exclusivamente para el desarrollo de sistemas de diagnóstico óptimos para régimen transitorio (capítulo 6), ya que lógicamente no son de aplicación en el desarrollo de sistemas de diagnóstico óptimos para régimen estacionario (capítulo 0).

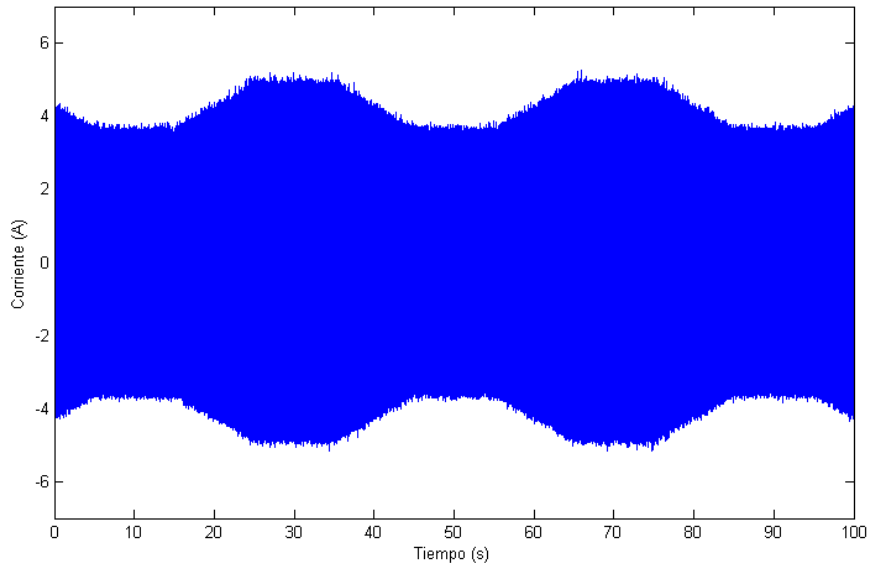


Figura 4.28 . Muestra de corriente de fase de un motor con fallo de barra rota, con control DTC, frecuencia de alimentación constante a 50Hz y carga en rampa de 10 segundos de duración entre el 25 y el 100% del par de carga.

En la Tabla 4.11 (para conexión mediante variador Siemens), y Tabla 4.12 (para conexión mediante variador ABB); se esquematiza el conjunto de variaciones de parámetros que han sido configurados (control/tensión utilizada, rango de frecuencias de operación y par de carga) para el desarrollo de las muestras transitorias por medio de la variación de la frecuencia de alimentación.

En el caso de las muestras transitorias, sin variación en la frecuencia de alimentación, pero con rampas en el par de carga, se ha aplicado la siguiente configuración de parámetros para conexión por auto transformador (Tabla 4.13), para conexión mediante variador Siemens (Tabla 4.14) y para conexión mediante variador ABB (Tabla 4.15).

Tabla 4.11. Variaciones distintas de parámetros generados con conexión mediante variador de la marca Siemens introduciendo transitorios de 10 segundos de duración.

SISTEMA DE CONTROL	RANGO DE FRECUENCIA (Hz)	PAR DE CARGA (%)
VECTORIAL	Mínima: 20 Hz Máxima: 25 Hz	0 %
		35 %
		70 %
	Mínima: 40 Hz Máxima: 50 Hz	0 %
		50 %
		100 %
	Mínima: 45 Hz Máxima: 50 Hz	0 %
		50 %
		100 %
ESCALAR	Mínima: 20 Hz Máxima: 25 Hz	0 %
		35 %
		70 %
	Mínima: 40 Hz Máxima: 50 Hz	0 %
		50 %
		100 %
	Mínima: 45 Hz Máxima: 50 Hz	0 %
		50 %
		100 %
ESCALAR CON COMPENSACIÓN DEL DESLIZAMIENTO	Mínima: 20 Hz Máxima: 25 Hz	0 %
		35 %
		70 %
	Mínima: 40 Hz Máxima: 50 Hz	0 %
		50 %
		100 %

	Mínima: 45 Hz Máxima: 50 Hz	0 %
		50 %
		100 %

Tabla 4.12. Variaciones distintas de parámetros generados con conexión mediante variador de la marca ABB introduciendo transitorios de 10 segundos de duración.

SISTEMA DE CONTROL	RANGO DE FRECUENCIA (Hz)	PAR DE CARGA (%)
ESCALAR	Mínima: 20 Hz Máxima: 25 Hz	0 %
		35 %
		70 %
	Mínima: 40 Hz Máxima: 50 Hz	0 %
		50 %
		100 %
	Mínima: 45 Hz Máxima: 50 Hz	0 %
		50 %
		100 %
CONTROL DIRECTO DEL PAR (DTC)	Mínima: 20 Hz Máxima: 25 Hz	0 %
		35 %
		70 %
	Mínima: 40 Hz Máxima: 50 Hz	0 %
		50 %
		100 %
	Mínima: 45 Hz Máxima: 50 Hz	0 %
		50 %
		100 %

Tabla 4.13. Variaciones distintas de parámetros generadas con conexión por autotransformador, con variaciones en rampa de la carga de 10 segundos de duración por rampa.

TENSIÓN DE ALIMENTACIÓN (V)	RANGO DE LA RAMPA DE CARGA (%)
300 V	Mínima: 25 % Máxima: 35 %
	Mínima: 25 % Máxima: 50 %
	Mínima: 25 % Máxima: 70 %
400 V	Mínima: 25 % Máxima: 50 %
	Mínima: 25 % Máxima: 75 %
	Mínima: 25 % Máxima: 100 %

Tabla 4.14. Variaciones distintas de parámetros generadas con conexión mediante variador de la marca Siemens, con variaciones en rampa de la carga de 10 segundos de duración por rampa.

SISTEMA DE CONTROL	FRECUENCIA DE TRABAJO (Hz)	RANGO DE LA RAMPA DE CARGA (%)
VECTORIAL	25 Hz	Mínima: 25 % Máxima: 35 %
		Mínima: 25 % Máxima: 50 %
		Mínima: 25 % Máxima: 70 %
	50 Hz	Mínima: 25 % Máxima: 50 %

		Mínima: 25 % Máxima: 75 %
		Mínima: 25 % Máxima: 100 %
ESCALAR	25 Hz	Mínima: 25 % Máxima: 35 %
		Mínima: 25 % Máxima: 50 %
		Mínima: 25 % Máxima: 70 %
	50 Hz	Mínima: 25 % Máxima: 50 %
		Mínima: 25 % Máxima: 75 %
		Mínima: 25 % Máxima: 100 %
ESCALAR CON COMPENSACIÓN DEL DESLIZA- MIENTO	25 Hz	Mínima: 25 % Máxima: 35 %
		Mínima: 25 % Máxima: 50 %
		Mínima: 25 % Máxima: 70 %
	50 Hz	Mínima: 25 % Máxima: 70 %
		Mínima: 25 % Máxima: 70 %
		Mínima: 25 % Máxima: 70 %

Tabla 4.15. Variaciones distintas de parámetros generadas con conexión mediante variador de la marca ABB, con variaciones en rampa de la carga de 10 segundos de duración por rampa.

SISTEMA DE CONTROL	FRECUENCIA DE TRABAJO (Hz)	RANGO DE LA RAMPA DE CARGA (%)
ESCALAR	25 Hz	Mínima: 25 % Máxima: 35 %
		Mínima: 25 % Máxima: 50 %
		Mínima: 25 % Máxima: 70 %
	50 Hz	Mínima: 25 % Máxima: 50 %
		Mínima: 25 % Máxima: 75 %
		Mínima: 25 % Máxima: 100 %
CONTROL DIRECTO DEL PAR (DTC)	25 Hz	Mínima: 25 % Máxima: 35 %
		Mínima: 25 % Máxima: 50 %
		Mínima: 25 % Máxima: 70 %
	50 Hz	Mínima: 25 % Máxima: 50 %
		Mínima: 25 % Máxima: 75 %
		Mínima: 25 % Máxima: 100 %

Finalmente, en el último caso de las muestras transitorias, se han generado muestras con variación de la frecuencia de alimentación y rampas en el par de carga. Las diversas configuraciones de los parámetros aplicadas en este caso se han realizado mediante variador Siemens (Tabla 4.16) y mediante variador ABB (Tabla 4.17).

Tabla 4.16. Variaciones distintas de parámetros generados con conexión mediante variador de la marca Siemens introduciendo transitorios de 10 segundos de duración y con variaciones en rampa de la carga de 10 segundos de duración por rampa.

SISTEMA DE CONTROL	RANGO DE FRECUENCIA (Hz)	RANGO DE LA RAMPA DE CARGA (%)
VECTORIAL	Mínima: 40 Hz Máxima: 50 Hz	Mínima: 25 % Máxima: 75 %
		Mínima: 50 % Máxima: 100 %
ESCALAR	Mínima: 40 Hz Máxima: 50 Hz	Mínima: 25 % Máxima: 75 %
		Mínima: 50 % Máxima: 100 %
ESCALAR CON COMPENSACIÓN DEL DESLIZAMIENTO	Mínima: 40 Hz Máxima: 50 Hz	Mínima: 25 % Máxima: 75 %
		Mínima: 50 % Máxima: 100 %

Tabla 4.17. Variaciones distintas de parámetros generados con conexión mediante variador de la marca ABB introduciendo transitorios de 10 segundos de duración y con variaciones en rampa de la carga de 10 segundos de duración por rampa.

SISTEMA DE CONTROL	RANGO DE FRECUENCIA (Hz)	RANGO DE LA RAMPA DE CARGA (%)
ESCALAR	Mínima: 40 Hz Máxima: 50 Hz	Mínima: 25 % Máxima: 75 %
		Mínima: 50 % Máxima: 100 %
CONTROL DIRECTO DEL PAR (DTC)	Mínima: 40 Hz Máxima: 50 Hz	Mínima: 25 % Máxima: 75 %
		Mínima: 50 % Máxima: 100 %

4.5 Conclusiones.

Como se ha expuesto en esta tesis, se ha desarrollado un número ingente de muestras provenientes de las corrientes de fases de varios motores de inducción con las mismas características pero distinto estado de salud (sano y con barras rotas).

Se ha optimizado y automatizado el sistema de generación de características para que las muestras obtenidas tengan la rigurosidad requerida en este tipo de experimentos. Es decir, se ha asegurado que el proceso de generación de las muestras se ha mantenido dentro de unos parámetros establecidos (tiempo de la muestra, control del motor, control del par resistente, etc.) monitorizados y controlados por un sistema PLC (apartado 4.3.2.1), eliminando la necesidad del desarrollo de los ensayos mediante un operador humano (con menos precisión en el control y la monitorización en comparación con el PLC).

Además, una vez generado la base de datos de muestras, han sido analizadas todas ellas personalmente para garantizar que no ha habido ningún fallo durante la generación de estas que invalide su uso.

En las muestras que se ha detectado alguna anomalía, se ha vuelto a generar una nueva muestra con la misma configuración del ensayo y con la misma precisión en las condiciones del ensayo que permite el PLC).

Finalmente se ha obtenido una base de muestras de ensayos idóneos para validar de forma experimental el sistema de generación de sistemas de diagnóstico de fallos, para el caso concreto del fallo de rotura de barras en los motores de inducción.

Capítulo 5

Generador de detectores de fallo en máquinas de inducción en régimen estacionario

5.1 Introducción

Dentro de un buen sistema de diagnóstico de fallo es importante tanto la obtención de buenas características representativas de los fallos a diagnosticar como la mejor interpretación de estas de cara a obtener la mayor información posible y más precisa.

Como se ha comentado dentro del estado del arte en el apartado 2.1, actualmente existe una gran variedad de técnicas con las que obtener una gran diversidad de características representativas. Aunque estas características ya de por sí aportan bastante conocimiento sobre el fallo, un estudio más profundo puede ayudar a mejorar la precisión en la detección de los casos de fallo más dudosos o aportar más datos al respecto.

Un sistema de diagnóstico de fallos para motores de inducción mediante sistemas expertos está formado por una técnica de extracción de las posibles características de fallos (apartado 2.3) y un sistema inteligente entrenado para clasificar y discernir sobre esas características (apartado 2.4).

Como ya se ha expresado en los artículos de Bellini [10] y Samira[136], la configuración óptima del sistema de diagnóstico de tipo experto no es algo trivial y es muy dependiente de las características del motor a diagnosticar y de su operativa de trabajo. Un sistema de diagnóstico bien configurado requiere su tiempo, por lo que estos se desarrollan principalmente para los casos de máquinas (motores) muy importantes por

su coste tanto de fabricación (gran potencia) como por su importancia dentro del proceso productivo.

En consecuencia, es de gran interés tener un sistema automático capaz de generar de forma óptima el mejor sistema de diagnóstico de fallos para cualquier tipo de motor de inducción. Las ventajas de este sistema son la minimización del coste económico y tiempo necesarios para el mantenimiento, la eliminación de la subjetividad humana en el proceso, la obtención de la mejor precisión (o cercana a ella) sin sobredimensionado del coste computacional y, como consecuencia de las anteriores, estudiar la posibilidad de la implementación de sistemas de diagnóstico para motores más allá del campo de los sistemas críticos.

A lo largo de este capítulo se van a exponer las características de mayor relevancia para poder detectar posibles fallos en régimen estacionario, así como el desarrollo del generador de sistemas expertos basados en Máquinas de Vectores de Soporte (SVM) y Redes Neuronales (ANN), con el que se pretende desarrollar un sistema óptimo.

En este capítulo solamente se utilizan técnicas válidas para régimen estacionario (apartado 2.3.1). El caso correspondiente al diagnóstico de fallos en régimen transitorio se desarrolla en el capítulo 6.

5.2 Características relevantes en régimen estacionario.

Partiendo de las investigaciones estudiadas en el estado del arte para la obtención de indicadores para la detección de fallos por barra rota en régimen estacionario (apartado 2.3.1), donde se han descrito todas las características correspondientes a estas técnicas usando como señales de muestras todas aquellas compatibles con el régimen estacionario que se encuentran disponibles en la base de datos de muestras generada por el banco de ensayos comentado en el capítulo 4.

En este apartado 5.2, se menciona las técnicas de diagnóstico utilizadas, enfocadas a su aplicación en sistemas expertos, así como las características representativas generadas y seleccionadas para utilizarse en el entrenamiento y la validación de los sistemas expertos de diagnóstico.

Además de las características representativas directas, en ciertos casos también se han escogido características correspondientes a los parámetros de funcionamiento, las cuales aunque no dan información directa del fallo afectan de alguna forma a las características representativas, pudiendo mejorar la detección de estas últimas.

Con el uso de sistemas expertos aplicados a todas estas técnicas se espera potenciar los siguientes puntos:

- Automatizar la detección de fallos de forma eficiente.
- Obtener el mejor tipo de diagnóstico de fallos acorde a los requisitos.
- Mejora en la precisión de la detección en los casos dudosos. Un ejemplo se muestra en la Figura 5.1, en este caso de ejemplo el sistema experto SVM bien entrenado puede mejorar la precisión cuando el valor del deslizamiento es mínimo llegándose casi a confundir las características de fallo con las características sanas.
- Mejora en la precisión de la detección cuando existe una carga pulsante con una frecuencia muy próxima a la frecuencia del fallo.

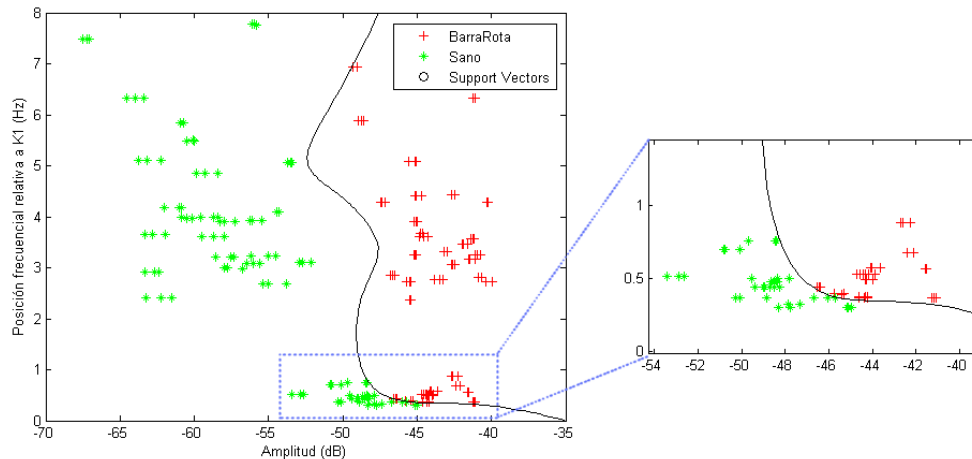


Figura 5.1. Espacio de características de un sistema experto SVM donde se ha ampliado la zona de conflicto (características de clases distintas muy próximas) originada por unas muestras de corriente con un deslizamiento muy bajo y una amplitud muy cercana.

Se han incluido aquellos casos (3 últimos puntos anteriores) en los que la experiencia obtenida en trabajos anteriores en el grupo de investigación ha mostrado que había dificultades en la detección del fallo.

5.2.1 Técnica de Análisis de la corriente estatórica mediante FFT.

Esta técnica bastante extendida en el ámbito de estudio del diagnóstico de fallos, se basa en el análisis del espectro frecuencial generado mediante la transformada rápida de Fourier. Se obtienen los indicadores representativos a partir de las frecuencias gene-

radas por la asimetría que se genera por la descompensación de los campos magnéticos al producirse la rotura de una o más barras rotas.

Estas frecuencias se calculan con la ecuación (5.1), donde s representa el deslizamiento y k se corresponde a un valor constante igual o mayor a 1.

$$f_{BrokenBars} = [1 \pm 2(k \cdot s)] \cdot f \quad k = 1, 2, 3, 4, 5, \dots \quad (5.1)$$

Como ya se comentó en el apartado 2.3.1.1, las frecuencias de fallo más significativas se dan para $k=1$, disminuyendo su significancia conforme aumenta k .

En la Figura 5.2 y Figura 5.3 se muestra un ejemplo generado a partir de un par de muestras de la base de datos. Las gráficas de frecuencia de estas señales muestran las diferencias que se dan en dos señales de un mismo tipo de motor con accionamiento electrónico para el caso en que el motor está sano y el caso en que este muestra un posible fallo de barra rota.

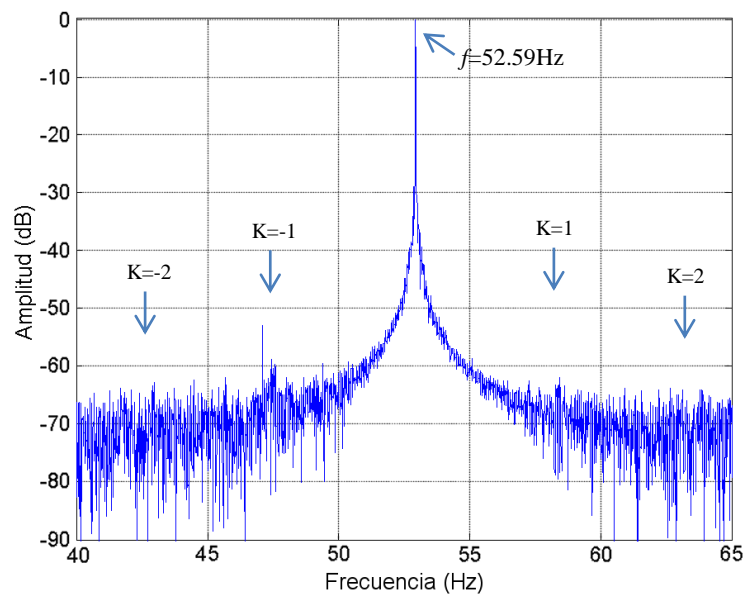


Figura 5.2. Resultado de la FFT aplicado a la corriente de un motor sano alimentado con un convertidor de frecuencia con control DTC y referencia de 50Hz.

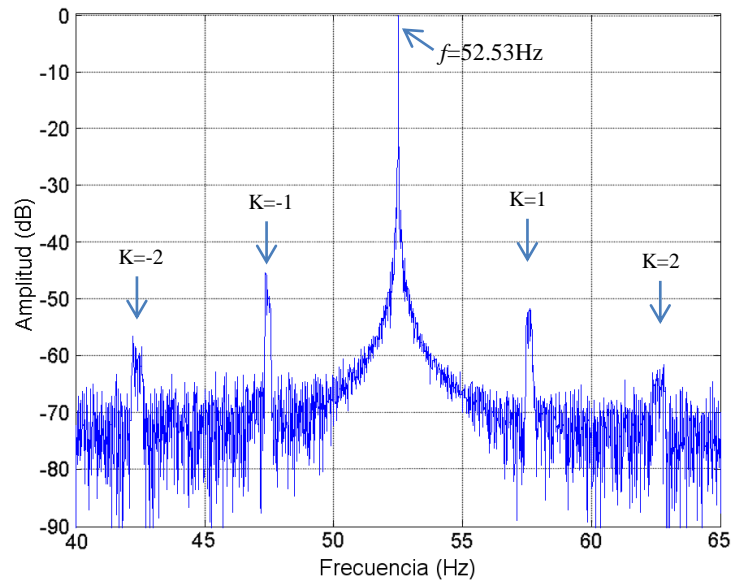


Figura 5.3. Resultado de la FFT aplicado a la corriente de un motor con fallo de barra rota alimentado con un convertidor de frecuencia con control DTC y referencia de 50Hz.

En la Figura 5.3 se aprecia claramente un aumento de la amplitud dentro de la posición frecuencial de fallo para los valores -2, -1, 1 y 2 de k en comparación del análisis de la corriente para un motor sano (Figura 5.2) donde no se aprecia aumento significativo de la amplitud para ningún valor de k .

También se observa en la Figura 5.3 como en los valores de k en -2 y 2 la amplitud característica del fallo es mucho menor respecto de los valores de k en -1 y 1.

Las **características significativas** de la técnica FFT se corresponden con:

- **Características significativas representativas del fallo de barra rota:**
 - **Las amplitudes de las dos frecuencias de fallo obtenidas para $k=\pm 1$.** Las amplitudes de las frecuencias características para valores de $k > 1$ se han descartado ya que estas últimas son menos significativas.
 - **El valor medio de las amplitudes a las frecuencias de fallo para $k=\pm 1$.** Aunque las frecuencias obtenidas en $k=1$ suelen tener una amplitud parecida, suelen variar dentro de un pequeño rango. El valor medio de las amplitudes de las frecuencias para $k=1$ y $k=-1$ obtiene una característica ponderada entre las dos anteriores

▪ **Características significativas NO representativas de fallo:**

- **El deslizamiento (s).** Aunque el deslizamiento no es una característica representativa del fallo, este permite separar mejor las características representativas según el régimen de funcionamiento estacionario; sobre todo teniendo en consideración que esta técnica de detección de fallos tiene problemas para discernir correctamente el fallo cuando el deslizamiento es muy bajo.
- **La distancia entre las componentes frecuenciales de fallo.** Analizando la fórmula (5.1) se observa que las posiciones de las frecuencias características de fallo son dependientes del deslizamiento (s), el cual a su vez se obtiene a partir de los parámetros de funcionamiento correspondientes a la frecuencia de alimentación y la velocidad mecánica. Por tanto la distancia en la que se separan las componentes frecuenciales de fallo pueden ser usadas como características significativas ya que derivan del deslizamiento.

En consecuencia para esta tesis se ha considerado relevante como característica a tener en cuenta para el sistema experto el valor medio de las amplitudes para las frecuencias de fallo en $k=\pm 1$, y el valor obtenido del deslizamiento.

5.2.2 Componente alterna del módulo de la señal analítica obtenida mediante Hilbert.

Esta técnica evita en mayor medida los problemas de detección que tiene la técnica FFT cuando el deslizamiento es muy bajo y sobre todo elimina el efecto de la fuga espectral.

Para ello se genera una señal analítica (AS) a partir de una corriente de fase estatórica a la que se le aplica la transformada de Hilbert (Figura 5.4). A esta señal analítica se elimina su componente continua caracterizada por el vector de giro calculado mediante la transformada de Hilbert sobre la fase de corriente estatórica. Esta componente continua se representa en la Figura 5.5 como el vector de giro principal indicado por la flecha azul, de forma que solo quedan las componentes alternas entre ellas las generadas por el fallo (vector amarillo en la Figura 5.5).

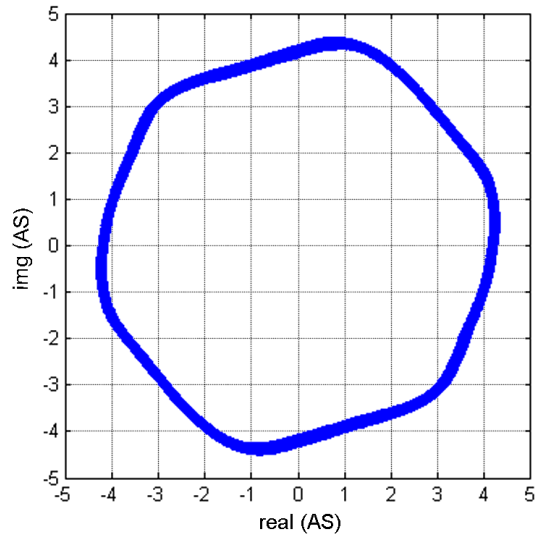


Figura 5.4. Señal analítica obtenida mediante la transformada de Hilbert.

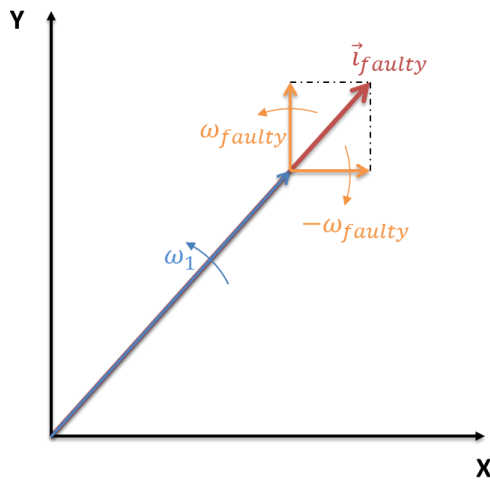


Figura 5.5 Vector de giro de la componente continua (en azul) y de la componente alterna (en amarillo).

En (5.2) se muestra la fórmula para generar la señal analítica respecto de una corriente de fase con barra rota.

Como se comentó en el apartado 2.3.1.5, en este caso I_M representa la amplitud máxima, n_b representa el número de barras rotas, N representa el número de barras del rotor, s representa el deslizamiento, k tiene un valor constante y f es la frecuencia de alimentación del motor.

Finalmente en (5.3) se elimina, restando el valor medio de la señal, la componente continua de la señal analítica.

$$\vec{i}_{BarraRota}(t) = I_M e^{j\omega t} \cdot \left(1 + \frac{n_{br}}{N} \cos(2\pi(2ksf)t) \right) \quad (5.2)$$

$$AC_{SA} = \vec{i}_{BarraRota}(t) - \left| \vec{i}_{BarraRota}(t) \right| \quad (5.3)$$

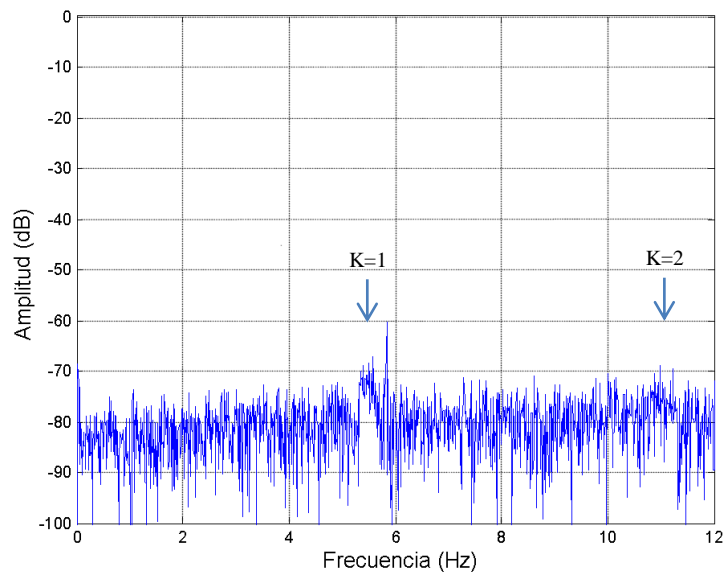


Figura 5.6. Resultado de la componente alterna del módulo de la señal analítica correspondiente a una muestra de corriente de fase obtenida de un motor sano con control electrónico configurado a una frecuencia de trabajo de 50Hz.

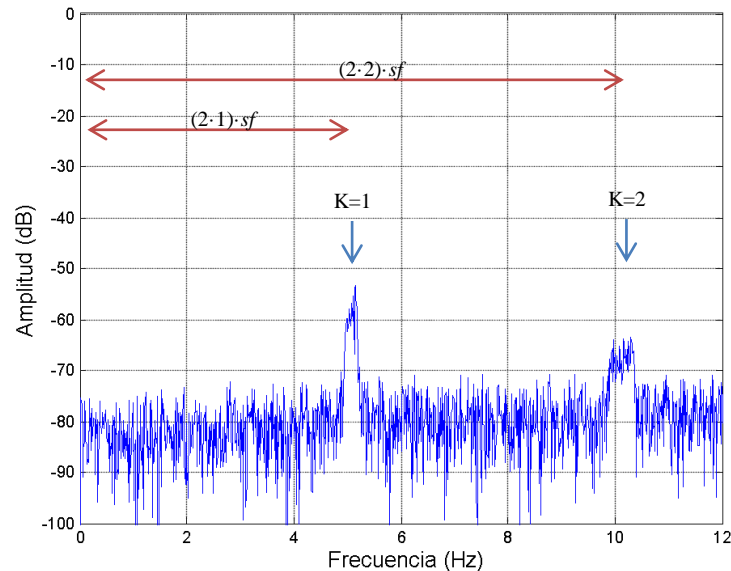


Figura 5.7. Resultado de la componente alterna del módulo de la señal analítica correspondiente a una muestra de corriente de fase obtenida de un motor con fallo y con control electrónico configurado a una frecuencia de trabajo de 50Hz.

En los espectros de frecuencia mostrados por la Figura 5.6 y Figura 5.7 se observa un ejemplo de motor sano y uno con fallo generados ambos con muestras la base de datos. En este caso se puede ver claramente la diferencia cuando aparece el fallo en motores del mismo tipo, uno sano y otro con fallo, en las mismas condiciones de trabajo.

Las **características significativas** en este caso se corresponden con:

- **Características significativas representativas del fallo de barra rota:**
 - **La amplitud obtenida en la posición de la frecuencia de barra rota reflejada en la componente alterna del módulo de la señal analítica para $k=1$.** Como se muestra en la Figura 5.7, esta frecuencia se encuentra a una distancia en frecuencia desde el origen de $2ksf$ siendo s el deslizamiento y f la frecuencia de control.
- **Características significativas NO representativas de fallo:**
 - **El deslizamiento (s).** En este caso el deslizamiento también puede ser considerado para separar mejor las características representativas, el cual aunque no contiene información del fallo, sí que puede ser utilizada para mejorar la detección de la característica significativa de fallo en el sistema experto.

- **La distancia entre las componentes frecuenciales de fallo.** Como se ha visto anteriormente en la ecuación 5.1 para obtener la frecuencia de fallo de barras rotas, el cálculo de esta frecuencia también depende del deslizamiento y por tanto también puede ser utilizada como característica significativa.

Finalmente, se ha formado una característica representativa de fallo compuesta de la característica simple de amplitud frecuencial del fallo y de la distancia entre las componentes frecuenciales de fallo.

5.2.3 *Análisis del Orden Armónico, versión para régimen estacionario. (HOTA estacionario).*

Esta técnica, desarrollada con mayor detalle en el apartado 2.3.1.6, genera el vector de corrientes y traslada los puntos de referencias de coordenadas de este vector desde el estator al rotor, y finalmente reescala las frecuencias respecto del valor k (k -orden) correspondiente a la fórmula de fallo en barra rota (5.1).

Con la fórmula representada en (5.4) se calculan las posiciones de frecuencias de fallo desde el sistema de referencia del rotor. Con la fórmula (5.5) se desarrolla el cambio de escala a la distribución en k -orden.

$$\omega_{\text{BarraRota}}^r = \omega^r 2k + \omega^r \quad \text{siendo} \quad \begin{matrix} k = \pm(1, 2, 3, 4\dots) \\ \omega^r = s \cdot f \end{matrix} \quad (5.4)$$

$$k = \frac{\omega^{r'} - \omega^r}{2\omega^r} \quad (5.5)$$

En la ecuación (5.4) la variable k representa el número de armónico de fallo, que como se indica es un número entero positivo o negativo. La variable ω^r muestra la frecuencia de rotación de la alimentación multiplicado por el deslizamiento (vista desde el sistema de referencia del rotor) y la variable $\omega_{\text{BarraRota}}^r$ contiene las frecuencias de fallo características vistas también desde el eje de coordenadas del rotor.

En la ecuación (5.5) se muestra la variable $\omega^{r'}$ correspondiente al espectro de la señal analítica respecto del rotor, y la variable k resultante del k -orden armónico del fallo especificado en (5.4).

En la Figura 5.8 y Figura 5.9 se muestra la gráfica del espectro desarrollado con HOTA para motor sano y motor con fallo. Como se puede observar, el resultado es parecido al obtenido con la FFT pero sustituyendo la escala de frecuencias por una escala k-orden. Esta sutil diferencia aporta enormes ventajas para HOTA tanto en régimen estacionario como transitorio, posibilitando el análisis de señales independientemente de los parámetros de funcionamiento de cada motor.

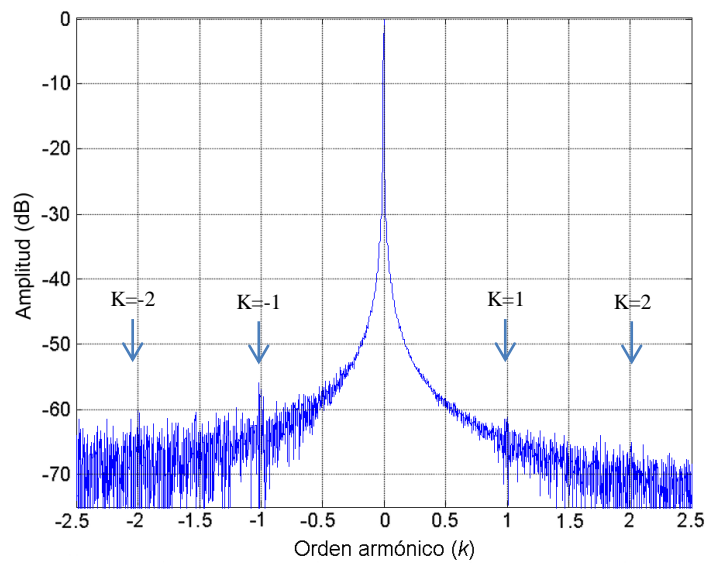


Figura 5.8. Resultado de HOTA correspondiente a una muestra de corriente de fase obtenida de un motor sano con control electrónico configurado a una frecuencia de trabajo de 50Hz.

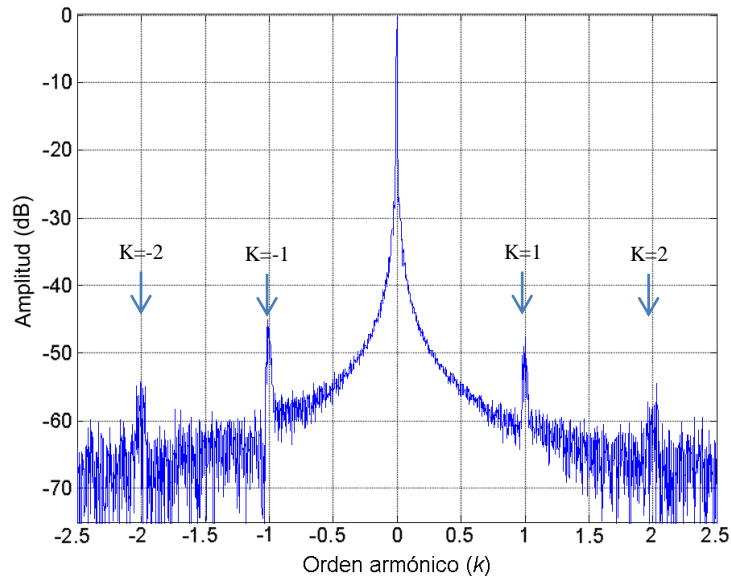


Figura 5.9 Resultado de HOTA correspondiente a una muestra de corriente de fase obtenida de un motor con fallo de barra y con control electrónico configurado a una frecuencia de trabajo de 50Hz.

Las **características significativas** para la técnica HOTA se corresponden con:

- **Características significativas representativas del fallo de barra rota:**
 - **Las amplitudes de las dos frecuencias de fallo obtenidas para $k=\pm 1$.** Las amplitudes de las frecuencias características para valores de $k > 1$ se han descartado por ser menos significativas.
 - **El valor medio de las amplitudes a las frecuencias de fallo para $k=\pm 1$.** Se obtiene la ponderación de las amplitudes de fallo de mayor relevancia.

Las **características significativas** relevantes seleccionadas para detectar la rotura de barras se corresponden con las dos amplitudes de las frecuencias de fallo por separado obtenidas para $k=\pm 1$. Las frecuencias características para valores de $k > 1$ y $k < -1$ se han descartado por ser de menor significativas.

En este caso, de forma optativa se podría llegar a calcular de forma independientemente a HOTA el valor de deslizamiento para usarse como característica secundaria no representativa del fallo. No obstante, esto eliminaría una de las principales bazas de

HOTA como técnica de obtención de características independiente respecto de los parámetros y ajustes del motor (en este caso del deslizamiento).

5.2.4 Cepstrum.

Según la técnica de Cepstrum que se desarrolla con mayor detalle en el apartado 2.3.1.4, y como se puede ver en la fórmula (5.1), la variable k genera una secuencia periódica de aparición de fallo con una misma distancia expuesta en la ecuación (5.6) donde el deslizamiento (s) es importante para cada componente armónica definida por la variable k .

$$f(k) = k \cdot 2 \cdot f \cdot s \quad (5.6)$$

En este caso, al desarrollar la transformada de Fourier sobre el espectro de frecuencias (transformada inversa de Fourier), si existe un fallo de barra rota este se mostrará como una componente frecuencial cuya frecuencia está condicionada por el deslizamiento.

Como se muestra en (5.7) la doble transformación de Fourier no es pura. La primera transformación FFT de la corriente de fase se absolutiza, se normaliza y se convierte a escala logarítmica para resaltar mejor esa característica repetitiva de fallo antes de aplicar la transformada inversa FFT^{-1} .

$$I_{Cepstrum} = \left| FFT^{-1} \left(\log_{10} \left(\left[\frac{|FFT(i_{fase}(t))|}{\max(|FFT(i_{fase}(t))|)} \right]^2 \right) \right) \right)^2 \quad (5.7)$$

A continuación, se muestran Figura 5.10 y Figura 5.11 de motor sano y motor con fallos obtenidos de dos muestras de motores de la base de datos local ajustados a 50Hz con un control DTC y una carga del 100%.

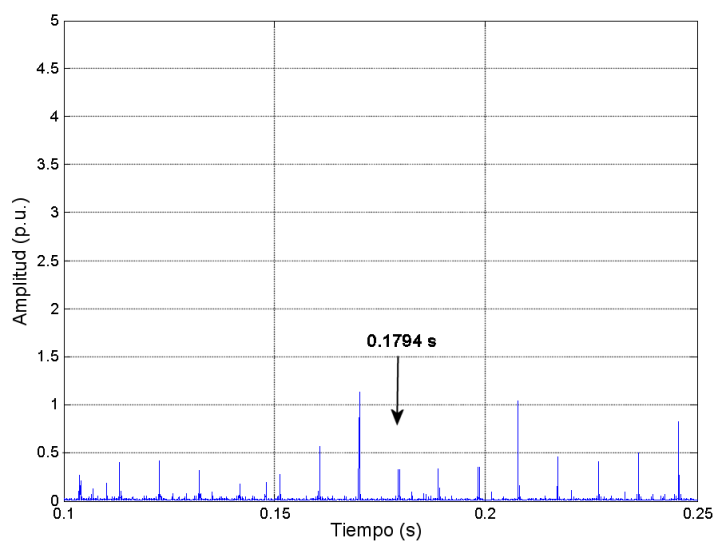


Figura 5.10 . Resultado de Cepstrum correspondiente a una muestra de corriente de fase obtenida de un motor sano con control electrónico configurado a una frecuencia de trabajo de 50Hz.

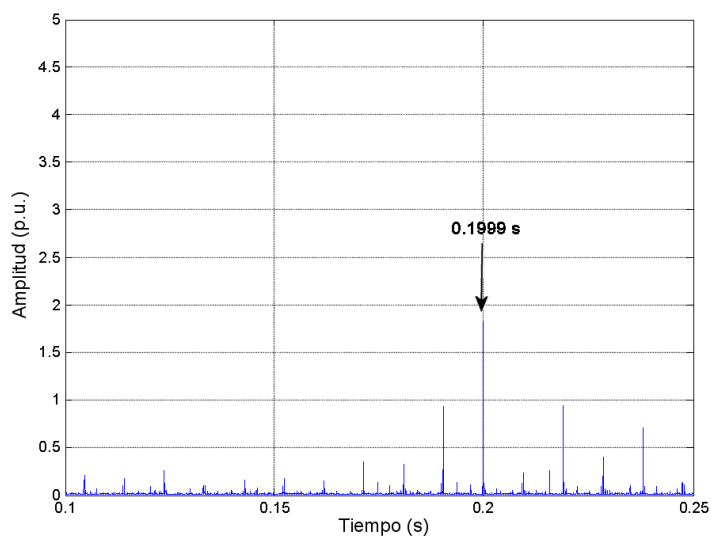


Figura 5.11 Resultado de Cepstrum correspondiente a una muestra de corriente de fase obtenida de un motor con fallo de barra rota y con control electrónico configurado a una frecuencia de trabajo de 50Hz.

Las **características significativas** en la técnica Cepstrum se corresponden con:

- **Características significativas representativas del fallo de barra rota:**
 - **La amplitud obtenida en el instante de tiempo que se corresponde con la periodicidad producida por el propio fallo.** La única característica que representa al fallo en este caso es únicamente esta amplitud generada por la frecuencia de repetición entre las componentes armónicas de orden k . A diferencia de las anteriores técnicas, en el resultado final no existe una familia de componentes armónicas de fallo.
- **Características significativas NO representativas de fallo:**
 - **El deslizamiento (s).** Como en casos anteriores, el deslizamiento puede ser utilizado para mejorar la detección de la característica significativa de fallo en el sistema experto.
 - **El instante de tiempo en el que se muestra la periodicidad producida por el fallo (R).** Este instante de tiempo se encuentra a una distancia de $1/(2sf)$ siendo s el deslizamiento y f la frecuencia de control.

Finalmente, como características significativas para este trabajo se ha utilizado el instante de tiempo correspondiente a la periodicidad de las componentes de fallo y su amplitud correspondiente.

5.2.5 Vector de Park Extendido (EPVA).

El vector de Park reconstruye el vector de rotación a partir de las tres fases de corriente del motor. Con las fórmulas del vector de Park (5.8) y (5.9) se obtiene el vector de puntos de las dos dimensiones (real e imaginario) para su representación. Esta técnica se explica más detalladamente en el apartado 2.3.1.2 y 2.3.1.3.

$$i_D = \sqrt{\frac{2}{3}}i_U - \frac{1}{\sqrt{6}}i_V - \frac{1}{\sqrt{6}}i_W \quad (5.8)$$

$$i_Q = \frac{1}{\sqrt{2}}i_V - \frac{1}{\sqrt{2}}i_W \quad (5.9)$$

Eliminando el valor medio de dicho vector se elimina la frecuencia fundamental, manteniéndose las frecuencias de fallo. Para detectar dicha existencia de fallo se construye el espectro del vector y se comprueba la amplitud que aparece en la posición correspondiente a la frecuencia de fallo, la cual se corresponde a su vez con el resultado dado por la anterior ecuación (5.6) para un valor de $k=1$ desde el origen.

A continuación, se muestran la Figura 5.12 y Figura 5.13 generadas con EPVA, correspondientes a un motor sano y a un motor con fallo, obtenidas de dos muestras de la base de datos local.

Las **características significativas** de la técnica EPVA se corresponden con:

- **Características significativas representativas del fallo de barra rota:**
 - **La amplitud obtenida en la posición de la frecuencia de barra rota reflejada en la componente alterna del módulo de la señal analítica para $k=1$.** Esta técnica obtiene un resultado similar al que se obtiene con la técnica de la “componente alterna del módulo de la señal analítica mediante transformada Hilbert”. Sin embargo, este resultado obtenido se desarrolla a partir de las tres fases de corriente, por lo que lo hace más preciso en el caso de que las características de fallo en alguna de las tres fases sea menos representativas que en las otras fases. Como desventaja el coste de cálculo es mayor.
- **Características significativas NO representativas de fallo:**
 - **El deslizamiento (s).** Se utilizada para mejorar la detección de la característica significativa de fallo en el sistema experto.
 - **La distancia entre las componentes frecuenciales de fallo.** Esta distancia en frecuencia también depende del deslizamiento y también puede ser utilizada como característica significativa.

En este caso se ha utilizado la distancia entre las componentes frecuenciales de fallo y la amplitud obtenida en la posición de la componente de fallo $k=1$.

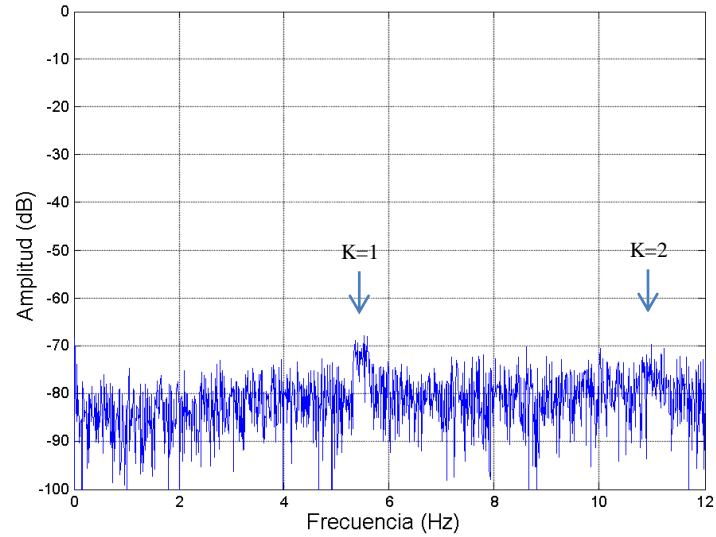


Figura 5.12. Resultado de EPVA correspondiente a las tres fases de corriente de fase obtenida de un motor sano trabajando a 50Hz.

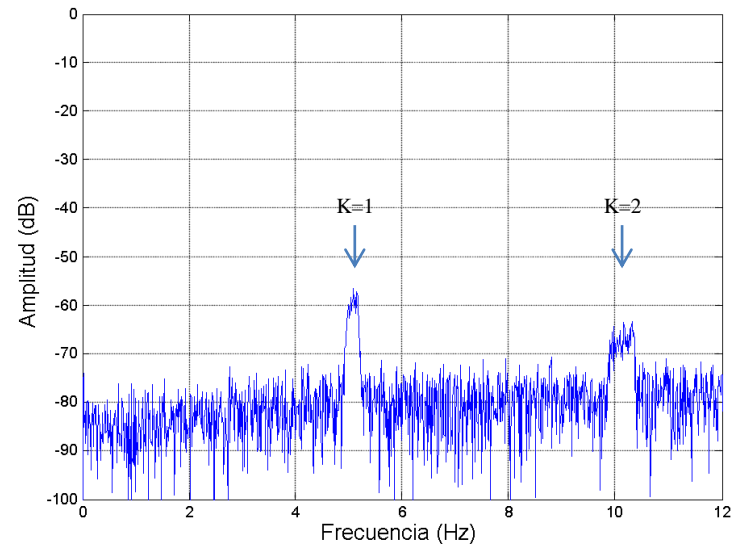


Figura 5.13 Resultado de EPVA correspondiente a las tres fases de corriente de fase obtenida de un motor con fallo de barra rota trabajando a 50Hz.

5.3 Modificaciones de las técnicas de diagnóstico de fallos.

Además de las características obtenidas mediante las técnicas de diagnóstico de fallos, como se ha comentado en el apartado 2.3.3, existen otros filtros y transformaciones entre dominios que en ciertos casos pueden mejorar la detección de las características de fallo.

En esta tesis como **novedad** se han aplicado estas otras variantes de estas técnicas modificadas con los filtros y transformaciones antes mencionados de forma que se ha realizado también una comparativa en la que se puede discernir sobre si realmente estas modificaciones mejoran la significancia del fallo dentro de las características representativas de fallos.

Una de estas modificaciones ha consistido en añadir una ventana de filtro de Hann para mejorar la resolución de las frecuencias con baja amplitud (apartado 2.3.3.1).

Otra modificación de estas técnicas se ha basado en sustituir la transformada de Fourier FFT para obtener el espectro en frecuencias de la corriente de fase por un periodograma de Welch que obtiene la estimación de la energía de la señal también como espectro en frecuencias (apartado 2.3.3.2).

Estas modificaciones se han aplicado a todas las técnicas analizadas dentro de este capítulo 5 para régimen estacionario.

Por tanto se ha realizado el análisis de las siguientes técnicas comentadas:

- Análisis de la corriente.
- Análisis de la componente alterna (AC) del módulo de la señal analítica (AS).
- ‘Cepstrum’.
- ‘Extended Park’s Vector Approach’ (EPVA).
- ‘Harmonic Order Tracking Analysis’ (HOTA) para régimen estacionario.

Donde para cada una de estas técnicas se ha analizado 4 modificaciones distintas:

- Técnica con transformación FFT.
- Técnica con transformación FFT y ventana de filtro de Hann.
- Técnica con periodograma de Welch.
- Técnica con periodograma de Welch y ventana de filtro de Hann.

5.4 Gestión de fallos mediante máquina de vectores de soporte.

El clasificador SVM es un buen candidato como posible sistema experto para detección de la existencia de posibles fallos. Este clasificador es muy utilizado en sistemas donde únicamente es necesario realizar una discriminación booleana, como por ejemplo en cualquier caso básico de detección de fallo.

Principalmente se valora su uso por su versatilidad en la separación de patrones de clases distintas donde la separación entre los patrones de estas clases no puede realizarse de forma lineal. Además, también se tiene en cuenta su buena velocidad de decisión y su bajo requisito de espacio de almacenamiento.

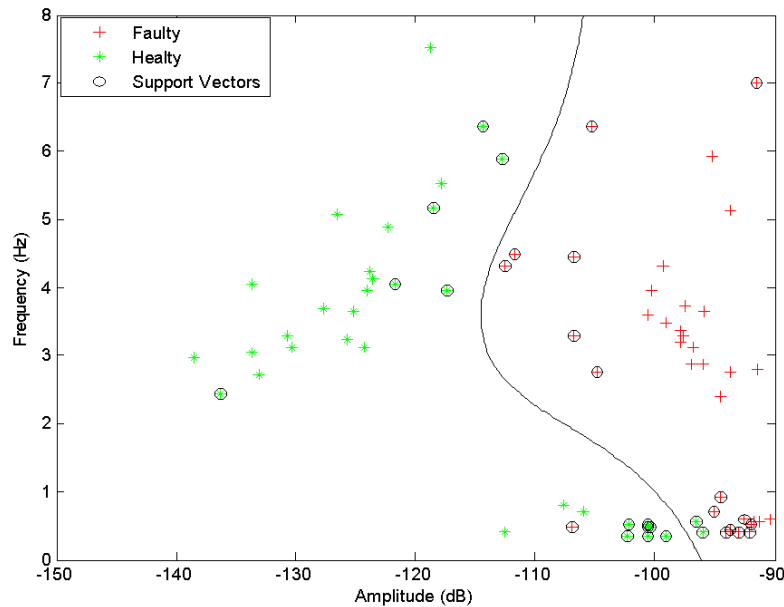


Figura 5.14 Espacio de características de un clasificador SVM con su hiperplano discriminatorio. Las características representativas de este espacio han sido obtenidas, a partir de muestras de motores con distintos niveles de fallo de barra rota, con la técnica EPVA usando un periodograma de Welch y una ventana de Hann.

Como ya se ha comentado en el apartado 3.2, el sistema experto SVM se incluye dentro del conjunto de clasificadores expertos basados en espacios de características. Como se puede ver en la Figura 5.14, se desarrolla un hiperplano con el que se intenta separar todo lo posible las características de una clase respecto de las características de la otra clase al dividirse el espacio en dos partes.

En las siguientes subsecciones se expone como ha sido desarrollado en esta tesis, este tipo de sistema experto clasificatorio, así como las validaciones que se han realizado y los resultados obtenidos. Para el desarrollo de dichas pruebas prácticas se han utilizado las muestras de motores operando en régimen estacionario del conjunto de bases de datos comentados en el apartado 4.4.1.

5.4.1 Estudio de los coeficientes del sistema.

Para poder ajustar correctamente el sistema de diagnóstico es necesario conocer de antemano los posibles coeficientes de ajuste con los que perfeccionar la detección del fallo.

Como se muestra en el apartado 5.2, las diversas técnicas de diagnóstico de las que se obtienen las características representativas de fallos no requieren del ajuste de parámetros o coeficientes para obtener de forma correcta la característica representativa de fallo. sin embargo, en el caso del sistema experto por Máquina de Vectores de Soporte (SVM) del apartado 3.2, la configuración de los coeficientes de ajuste es algo esencial si se quiere obtener el diagnóstico más preciso posible.

El tipo de ajuste de coeficientes que se implementa en este caso, debido a su versatilidad, es un algoritmo de **búsqueda de coeficientes en malla**, el cual ha sido comentado en el apartado 3.2.2.

El estudio desarrollado por Hsu [91] se usa como referente para realizar un buen ajuste de coeficientes en clasificadores SVM mediante la búsqueda en malla. En él se analiza el intervalo en que se encuentran los valores de cada coeficiente y los métodos de ajuste de estos coeficientes sin llegar a sobreentrenar al clasificador.

Como se ha comentado el apartado 3.2.3 sobre los diversos tipos de kernel, el **kernel polinomial** (5.10) puede representar todos los casos de transformación del espacio de características que son capaces de generar tanto el **kernel lineal** como el **kernel cuadrático**.

Tanto el **kernel de función de base radial Gaussiana** (5.11) como el **kernel sigmoideo** son más versátiles que el kernel polinomial y pueden discriminar correctamente problemas con una mayor no linealidad.

$$K(x_i, x_j) = (1 + x_i^T x_j)^d \quad (5.10)$$

$$K(x_i, x_j) = \exp \left[- \left(\frac{\|x_i - x_j\|}{2\sigma} \right)^2 \right] \quad (5.11)$$

Fuente: Hsu et al (2000)[91]

Según se resuelve en el estudio de Lin [101], un kernel de función de base radial es más apropiado que un kernel sigmoideo, ya que este último en algunos casos no se cumple las restricciones impuestas por el espacio de Hilbert comentado en el apartado 3.2.3. Además con el ajuste correcto de coeficientes un kernel de función de base radial Gaussiana puede ser tan versátil como un kernel sigmoideo, por lo que dentro de esta tesis se ha descartado el uso de este kernel sigmoideo.

En tal caso, **se ha optado por usar la implementación de SVM con un kernel polinomial (5.10) y con un kernel de función de base radial Gaussiana (5.11)**, ya que pueden generar todos los hiperplanos posibles, por lo que de cara a obtener el clasificador óptimo solo se requiere del estudio de estos kernels. El uso de otros kernels puede realizarse en trabajos futuros de cara a minimizar el coste temporal y espacial

Para el coeficiente global C que limita la ventana de ajuste (Box constraints) permitiendo un número mayor o menor de muestras clasificadas erróneamente (comentado con más detalle en el apartado 3.2.4) para obtener un buen ajuste del coeficiente C se sugiere según el estudio de Lin [101] de un intervalo de búsqueda del coeficiente entre un mínimo de 2^{-5} y un máximo de 2^{15} .

En caso de utilizar un **kernel de función de base radial (RBF)** el intervalo de búsqueda que se establece para este coeficiente σ en el mismo estudio de Lin se establece entre 2^{-15} y 2^3 .

Para un SVM con kernel polinomial, según el trabajo de Cristianini [137] que desarrolla diversos kernels para el clasificador SVM, el rango del coeficiente del orden polinomial d para un kernel polinomial se encuentra entre 1 (lineal) y 6. Con valores mayores a 6, la curvatura del hiperplano se vuelve despreciable.

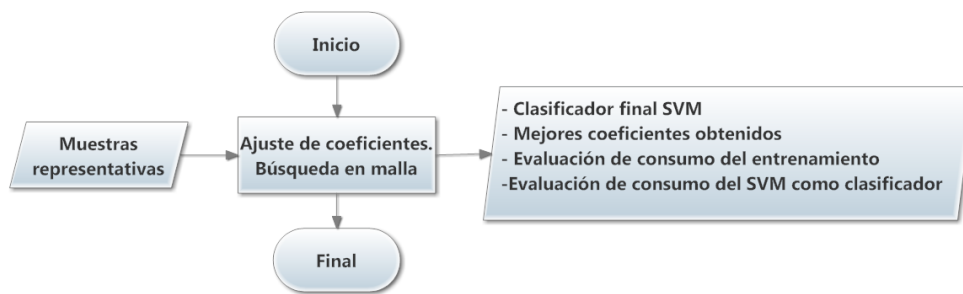
5.4.2 Algoritmo de generación del sistema de diagnóstico.

Para el desarrollo del sistema experto de diagnóstico se utiliza un aprendizaje supervisado, de forma que se requiere tener disponible un conjunto de muestras representativas del problema. En este caso ese conjunto de muestras se corresponde con las diversas bases de datos comentadas en el apartado 4.4.1.

Por consiguiente, como se expone en el esquema mostrado en la Figura 5.15, este algoritmo para generar el sistema de diagnóstico SVM óptimo requiere de un vector de entradas para las muestras de motores correspondientes a los dos tipos de motores (con fallo y sin fallo).

A partir de las muestras representativas se realiza un ajuste de los coeficientes por búsqueda en malla, de forma que se crea un nuevo sistema experto por cada modificación del coeficiente de búsqueda y se realiza su entrenamiento y validación.

Una vez finalizado el algoritmo este devuelve el sistema experto de diagnóstico SVM ya entrenado y ajustado con los coeficientes que han conseguido el mayor porcentaje de acierto. Junto con éste SVM se devuelve la evaluación del tiempo que tarda el clasificador en clasificar y el espacio que ocupa, así como también el tiempo de cálculo necesario y el espacio de memoria que ha sido requerido por el algoritmo.



Algoritmo de generación del sistema experto SVM de diagnóstico de fallos

Figura 5.15. Algoritmo principal de generación del mejor sistema experto SVM entrenado para la detección de fallos respecto de las muestras aportadas.

5.4.2.1 Algoritmo de ajuste de coeficientes por búsqueda en malla.

Siguiendo con el estudio de Hsu [91], se puede realizar un buen ajuste de coeficientes eficaz para un clasificador SVM mediante una búsqueda iterativa como si se tratara de explorar los nodos de una malla explorándola en profundidad, llegando finalmente al ajuste con mejor precisión (ejemplo de búsqueda en la Figura 5.16).

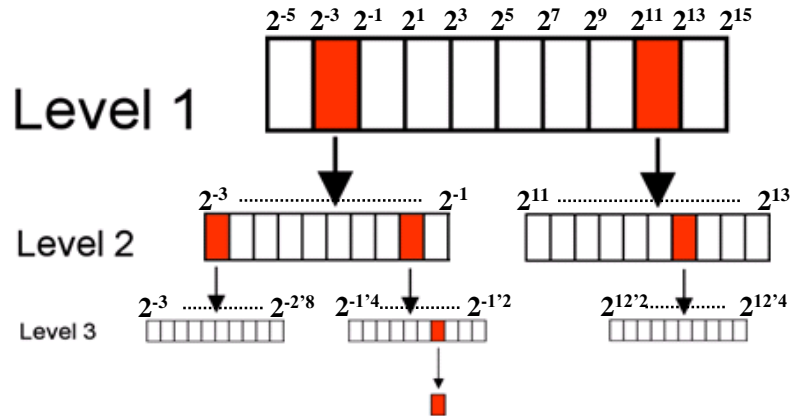
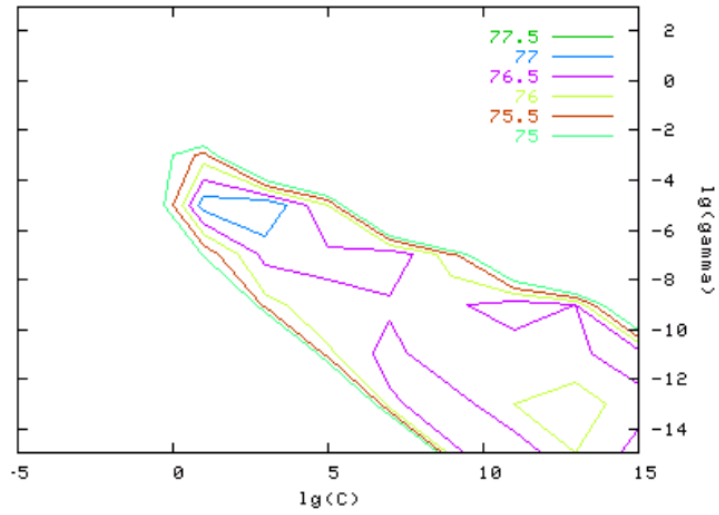


Figura 5.16. Ejemplo de exploración en profundidad para el coeficiente C .

De este modo se descarta la comprobación innecesaria de rangos de parámetros menos óptimos, ahorrándose un valioso tiempo respecto a una búsqueda completa del mejor ajuste del clasificador para todos los posibles valores de los parámetros.

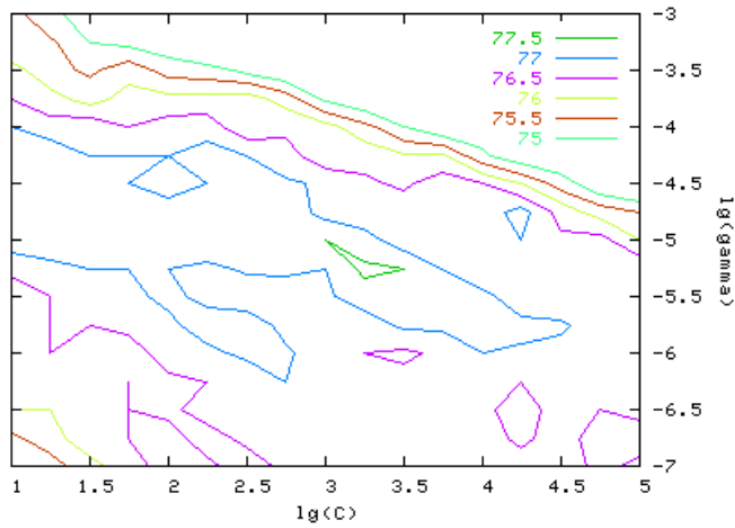
A grosso modo, como se muestra en la Figura 5.16, la búsqueda en malla es similar a dividir el intervalo definido por los valores máximos y mínimos que puede contener el coeficiente, generando para cada una de esas divisiones un sistema experto optimizado de forma que se obtiene la tasa de aciertos para el coeficiente medio de esa división del intervalo. El intervalo de búsqueda depende del coeficiente a ajustar: coeficiente C $[2^{-5} 2^{15}]$, coeficiente σ de kernel de función de base radial $[2^{-15} 2^3]$, coeficiente d de kernel polinomial $[1 6]$.

Los intervalos que han obtenido la mejor tasa de aciertos en el nivel actual pueden volver a ser explorados de forma iterativa en un nuevo nivel, de forma que el valor del coeficiente a optimizar se va concretando con cada nueva iteración.



Fuente: Hsu et al (2000) [91]

Figura 5.17. Búsqueda en malla para $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ y $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$.



Fuente: Hsu et al (2000) [91]

Figura 5.18 Búsqueda fina en malla para $C = 2^1, 2^{1.25}, \dots, 2^5$ y $\gamma = 2^{-7}, 2^{-6.75}, \dots, 2^{-3}$.

Esta forma de exploración de coeficientes también se desarrolla bastante bien en el ejemplo gráfico mostrado de la Figura 5.17 y Figura 5.18 que se comenta en el artículo [91]. En el ejemplo se observan los conjuntos que forman las diversas configuraciones de coeficientes para una búsqueda inicial (Figura 5.17) y la mejora de la precisión del clasificador cuando se exploran los conjuntos seleccionados en la siguiente iteración (Figura 5.18). En este ejemplo se realiza igualmente el mismo proceso pero para dos dimensiones (2 coeficientes), siendo en este caso de ejemplo el mejor intervalo que se explora aquel que obtiene el mayor porcentaje (77.5%).

En la iteración inicial del ejemplo de la Figura 5.16 se parte de los rangos descritos para el coeficiente global C , el cual se fija en el intervalo entre 2^{-5} y 2^{15} y se divide el intervalo que abarca el exponente (-5, 15) entre 10, obteniendo 11 valores del parámetro a explorar ($2^{-5}, 2^{-3}, 2^{-1}, 2^1, 2^3, 2^5, 2^7, 2^9, 2^{11}, 2^{13}, 2^{15}$).

A continuación se desarrolla como se realiza este ajuste del coeficiente C junto al coeficiente σ a ajustar del kernel seleccionado:

- **Ajuste del coeficiente global C con el coeficiente σ del kernel de función de base radial Gaussiana.**

Si el kernel es de función de base radial gaussiana (5.11), el intervalo se fija entre 2-15 y 23 y el intervalo que abarca el exponente (-15,3) se divide en 11 valores del parámetro a explorar (2-15, 2-13.2, 2-11.4, 2-9.6, 2-7.8, 2-6, 2-4.2, 2-2.4, 2-0.6, 21.2, 23).

Por cada combinación posible obtenida de emparejar cada valor del coeficiente C con los valores posibles del coeficiente σ (121 combinaciones) en un kernel de función de base radial, se entrena un clasificador SVM y se mide su capacidad de acierto para esta primera iteración.

De entre todos los clasificadores entrenados con cada pareja de coeficientes, se consideran para la siguiente iteración solo aquellos parámetros cuyos clasificadores han obtenido el mejor porcentaje de acierto en las pruebas de entrenamiento (apartado 5.4.2.2). Se vuelve a iterar con los nuevos intervalos de coeficientes seleccionados.

Para los parámetros C y σ (kernel RBF), en cada nueva iteración estos se dividen en los correspondientes subvalores según las parejas de parámetros seleccionadas como las mejores en el nivel anterior; considerándose como su margen inferior y superior el valor inmediatamente inferior y superior respectivamente.

Por ejemplo, en un caso de desarrollo con kernel de función de base radial, si se ha obtenido el mejor clasificador con kernel RBF para el conjunto de parámetros

$C=2^1$ y $\sigma=2^{-6}$ después de explorar los rangos iniciales, los nuevos valores a explorar en esta nueva iteración serían:

- Para el coeficiente C : 2^{-1} , $2^{-0.6}$, $2^{-0.2}$, $2^{0.2}$, $2^{0.6}$, $2^{1.21.4}$, $2^{1.8}$, $2^{2.2}$, $2^{2.6}$ y 2^3 .
- Para el coeficiente σ : $2^{-7.8}$, $2^{-7.44}$, $2^{-7.08}$, $2^{-6.72}$, $2^{-6.36}$, 2^{-6} , $2^{-5.64}$, $2^{-5.28}$, $2^{-4.92}$, $2^{-4.56}$ y $2^{-4.2}$.

La exploración en profundidad tiene que continuar solo mientras el porcentaje de aciertos aumenta. Cuando el porcentaje de aciertos alcanza su máximo sin cambios se debe parar de profundizar ya que en caso contrario solo se consigue sobreentrenar al clasificador.

Como excepción, puede darse el caso en que el clasificador se sobreentrena antes de alcanzar el porcentaje de aciertos máximo, por lo que esa rama de exploración se tiene que abortar. Este tipo de sobreentrenamiento se explica en el siguiente apartado 5.4.2.3.

Los coeficientes finales deseados son aquellos que obtienen finalmente el clasificador con la máxima precisión sin llegar a sobreentrenarlo.

▪ **Ajuste del coeficiente global C con el coeficiente d del kernel polinomial.**

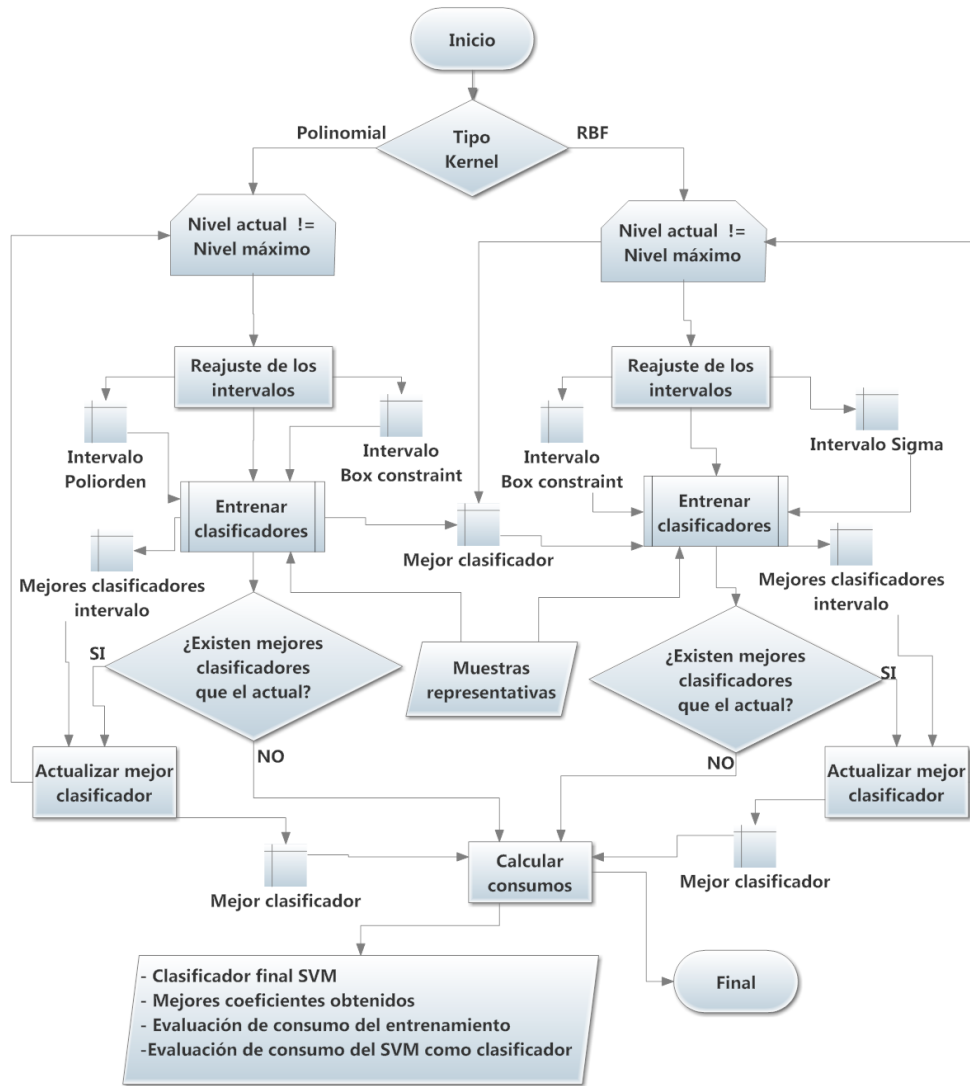
Si el kernel es polinomial (5.10), su rango de búsqueda se fija entre valores del 1 al 6 de tipo entero con incrementos unitarios, resultando en 6 valores a explorar (1, 2, 3, 4, 5,6)

Por cada combinación obtenida de emparejar cada valor del coeficiente C con los valores del coeficiente d (55 combinaciones) para un kernel polinomial, se entrena el clasificador SVM y se mide su capacidad de acierto.

De entre todos los clasificadores entrenados se consideran para la siguiente iteración los coeficientes de C cuyos clasificadores han obtenido el mejor porcentaje de acierto y se vuelve a iterar con los intervalos del coeficiente seleccionado. En este caso no tiene sentido subdividir el coeficiente del orden polinomial d del kernel polinomial. En este caso se desarrolla todo el intervalo (de 1 a 6) ya que el número de valores posibles es muy reducido y son valores enteros acotados.

La exploración en profundidad continúa hasta que el porcentaje de aciertos no se vea incrementado o se dé el caso en que el clasificador es sobreentrenado (ver apartado 5.4.2.3).

Los coeficientes finales son aquellos que obtienen el clasificador con mayor tasa de aciertos pero sin mostrar sobreentrenamiento.



Ajuste de coeficientes mediante la búsqueda en malla.

Figura 5.19. Esquema del algoritmo de ajuste de coeficientes por búsqueda en malla para el kernel polinomial y para el kernel RBF.

En la Figura 5.19 se muestra un esquema de flujo del algoritmo desarrollado para el ajuste de parámetros por búsqueda en malla tanto para el kernel polinomial como para el kernel de función de base radial.

En el apartado 3.2.2 también se comentan otros algoritmos de ajuste de coeficientes distintos con sus ventajas y desventajas respecto a este tipo de ajuste de coeficientes. Sin embargo, como ya se ha explicado anteriormente se ha seleccionado este algoritmo de búsqueda en malla porque es el más idóneo cuando el número de coeficientes a ajustar no es elevado.

5.4.2.2 Entrenamiento y validación del clasificador SVM.

Dentro de cada una de las pruebas de ajuste realizadas con cada valor de la serie de coeficientes seleccionada en el apartado anterior 5.4.2.1 tanto si se utiliza un kernel polinomial como un kernel RBF, es necesario medir el porcentaje de acierto del clasificador para saber su precisión con respecto a los coeficientes utilizados.

Para realizar esta medición no es posible entrenar directamente un clasificador con todas las muestras, ya que la validación posterior con las mismas muestras daría un porcentaje de aciertos muy alto pero erróneo, dado que el clasificador estaría en tal caso ajustado exactamente para este grupo de muestras pero no necesariamente para otras muestras incógnita.

Para estimar con buena aproximación la precisión del clasificador se usa comúnmente la técnica de validación cruzada. Esta técnica de validación es un sistema de valoración de la calidad de un clasificador utilizando un conjunto de muestras de las que se conoce a priori la clase a la que pertenece cada muestra.

Con este sistema de valoración del clasificador, si el conjunto de muestras no es discreto y abarca todo el conjunto posible de características se obtiene un valor aproximado de la calidad del acierto del clasificador. Dependiendo del método de validación cruzada utilizado la aproximación hacia la precisión del clasificador entrenado será de mayor o menor precisión.

Una de las ventajas relevantes del uso de la validación cruzada recae en que, con este tipo de validación, se evita el sobreentrenamiento por sobreajuste de los coeficientes a máxima precisión

Los dos métodos más comúnmente utilizados de validación cruzada son:

- **Validación cruzada de uno a muchos.**

Con este método cada una de las muestras originales se comprueba con un clasificador entrenado con el resto de muestras, por tanto el número de clasificadores a entrenar es igual al número de muestras del sistema.

Para cada muestra, se entrena el clasificador con el resto de muestras y se comprueba si dicho clasificador clasifica la muestra correctamente (representación del proceso en Figura 5.20). Finalmente se tiene en cuenta como número de aciertos las clasificaciones acertadas después de procesar todo el conjunto de muestras.

La ventaja de este algoritmo respecto a otros es que obtiene la mejor aproximación sobre el porcentaje de aciertos del clasificador. Como desventaja, el coste de ejecución depende del número de muestras, es decir, este coste es de orden lineal $\Theta(n)$. Esto lo hace impracticable cuando el número de muestras es muy elevado y, en consecuencia, también lo es el coste computacional.

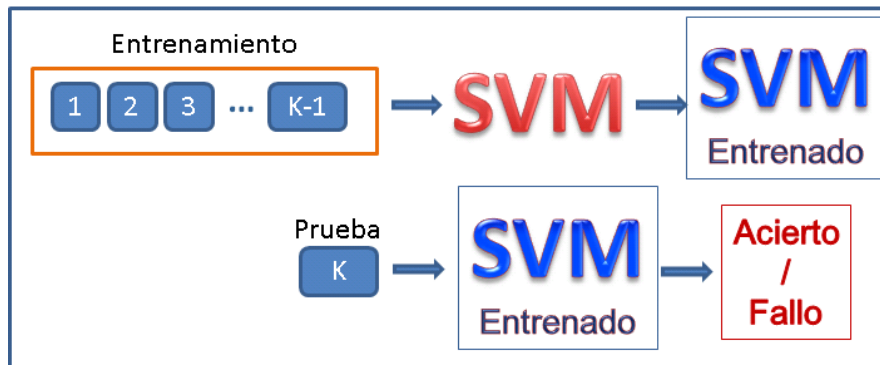


Figura 5.20 Representación del entrenamiento y clasificación para un ciclo del algoritmo de validación cruzada de uno a muchos.

- **Validación cruzada de entrenamiento/prueba (training/test).**

En este método el conjunto de muestras se divide en 2 subconjuntos. Un conjunto de muestras de entrenamiento para entrenar al clasificador (entre el 60 y 80 por ciento de las muestras) y un conjunto de muestras de prueba (las muestras restantes). Con las muestras de entrenamiento se entrena el clasificador para posteriormente procesar con estas las muestras de prueba como de tipo desconocido, obteniendo así el porcentaje de acierto aproximado del clasificador (representación del proceso en Figura 5.21). La elección de muestras para cada conjunto normalmente se realiza al azar.

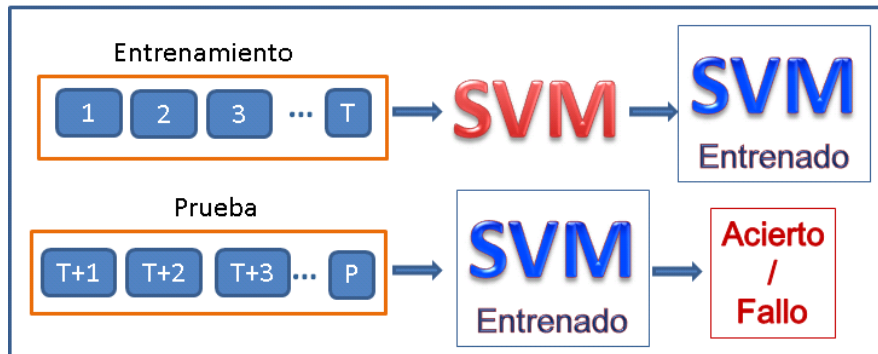


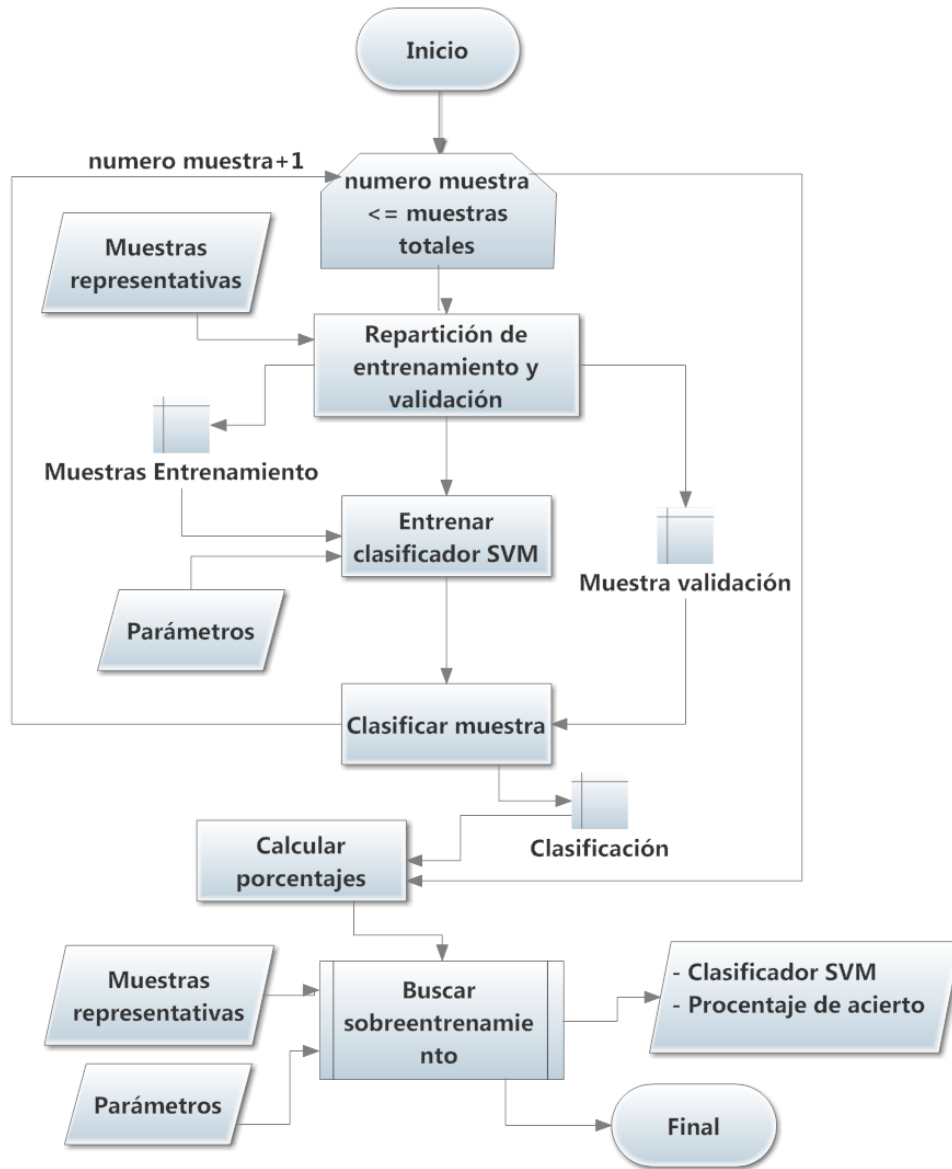
Figura 5.21 Representación del entrenamiento y clasificación para un ciclo del algoritmo de validación cruzada de entrenamiento/prueba.

Este proceso se suele realizar varias veces variando las muestras de los conjuntos para obtener la media de aciertos respecto de los resultados de todas las repeticiones.

La principal ventaja de este método es su bajo coste, ya que este es de coste constante de orden $\Theta(k)$, siendo k el número constante de repeticiones. Por el contrario, el resultado obtenido puede llegar a ser menos aproximado que utilizando la validación cruzada de uno a muchos.

En consecuencia, si el número de muestras no es muy alto, es recomendable utilizar la validación cruzada de uno a muchos antes que este método, para así obtener la mejor precisión posible.

En el esquema de la Figura 5.22 se muestra el algoritmo de entrenamiento y validación del SVM por validación cruzada de uno a muchos, obteniéndose el porcentaje de aciertos del clasificador para unos coeficientes y muestras de entrada concretos.



Entrenamiento y validación del sistema experto.

Figura 5.22 Algoritmo de entrenamiento y validación usando la metodología de validación cruzada de uno a muchos.

5.4.2.3 Sobreentrenamiento por falta de significancia de fallo en las muestras.

Si las muestras a estudio reflejan unas características poco significativas del fallo (por ejemplo por un exceso de ruido en la onda) se dificulta notablemente la separación de los conjuntos.

En tal caso el entrenamiento del clasificador se ve forzado a desarrollar hiperplanos discriminante mucho más complejos para obtener una mejor separación entre clases.

Este caso sucede cuando las características representativas de fallos no son muy discriminantes entre clases o cuando las muestras que se aportan no representan de forma significativa las características necesarias para resolver el problema de clasificación. En la Figura 5.23 se muestra un ejemplo de clasificador con sobreentrenamiento respecto a un conjunto de características poco discriminantes entre clases.

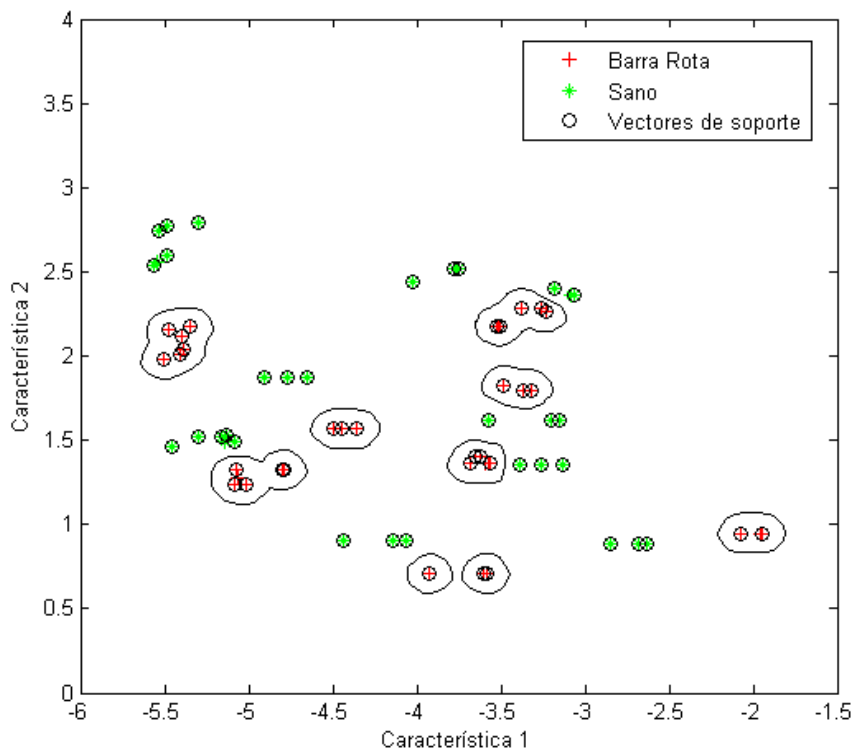


Figura 5.23. Ejemplo de espacio de características del clasificador SVM con hiperplano sobreentrenado.

Se puede detectar dicho fenómeno cuando existen dentro del espacio de características varios conjuntos para separar una misma clase, conteniendo en su mayoría un grupo reducido de muestras de esa clase.

En el caso del algoritmo de ajuste de coeficientes explicado en el apartado 5.4.2.1, si la dificultad es alta, es posible que se de sobreentrenamiento incluso antes de alcanzar el ratio de aciertos máximo posible por la búsqueda en profundidad de la malla. En tal caso hay que detectar para cada SVM generado si existe sobreentrenamiento y en caso afirmativo se ha de finalizar la exploración para esa rama de ajuste de coeficientes.

5.5 Gestión de fallos mediante redes neuronales artificiales.

El uso de sistemas expertos por redes neuronales es un campo muy amplio y avanzado, sobretodo en el tema de la clasificación de elementos. Su uso abarca la mayoría de disciplinas tecnológicas y se adentra incluso en disciplinas de otras índoles, como medicina [138], lingüística [139], psicología [140], etc.

Este tipo de clasificador es bastante robusto, fiable y preciso para la correcta detección y clasificación de los posibles fallos existentes, incluso en patrones de clasificación donde no es posible realizar una separación lineal entre clases.

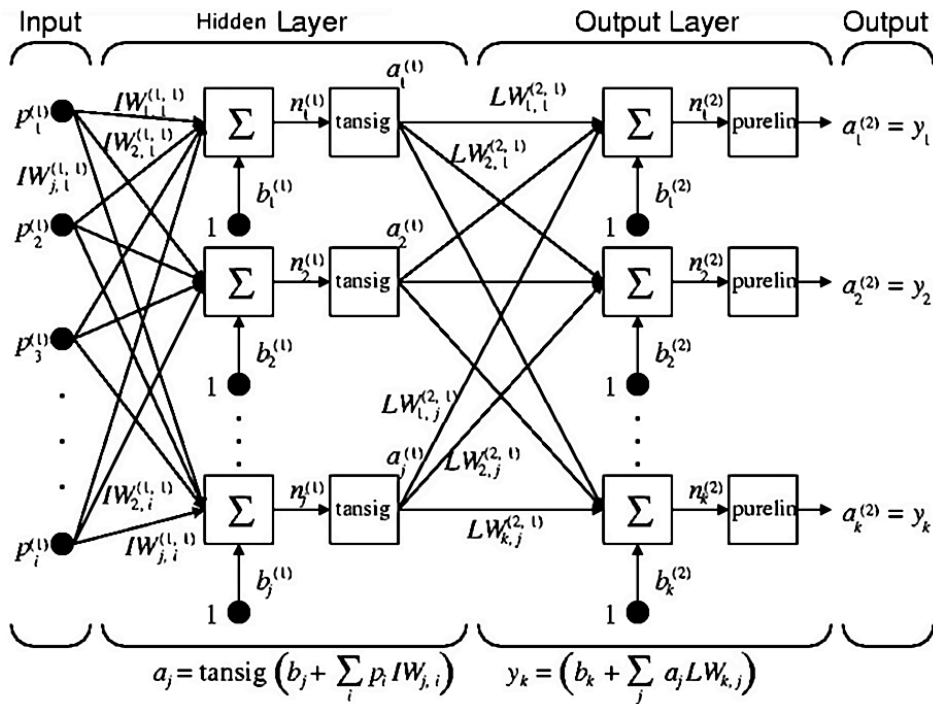
A diferencia de otros tipos de clasificadores, como por ejemplo el SVM, las redes neuronales artificiales permiten realizar una discriminación estricta multiclase, pudiendo ser entrenadas para detectar múltiples estados más allá del simple resultado booleano de afirmación/negación del fallo.

Como ya se ha expuesto dentro del apartado 3.3 de forma más detallada, los sistemas expertos de tipo ANN más comunes son los denominados Perceptron Multicapa.

Como se muestra en la Figura 5.24 este tipo de redes neuronales está compuesto por una serie de capas (ninguna, una o más de una capa oculta y una capa final de salida) conectadas secuencialmente cada capa oculta con su siguiente capa inmediata. Hay que tener en cuenta que las entradas a la red neuronal no poseen una capa propia, sino que se asocian a la primera capa oculta de la red o, en el caso de que no existan capas ocultas las entradas de la red neuronal se asignan a las entradas de la capa de salida.

Cada capa está compuesta de un número de neuronas donde se evalúa en cada una el valor de entrada y, dependiendo de este y de la configuración del peso y umbral de la neurona, puede llegarse a generar un nuevo valor por la salida.

De este modo a partir de los datos de entrada se acaba obteniendo un resultado a la salida de la red como respuesta. El tipo de respuesta ante una entrada concreta depende de la topología y configuración interna de las neuronas.



Fuente: Demuth et al. (2008)[85]

Figura 5.24 . Esquema de una red neuronal artificial Perceptron multicapa.

En las siguientes subsecciones se realiza el ajuste de los diversos coeficientes ajustables, la metodología de entrenamiento utilizada, las validaciones que se han realizado y los resultados obtenidos con la implementación de este sistema experto. Las muestras usadas para el desarrollo del sistema de diagnóstico son aquellas generadas a partir de motores sanos y con fallo de barra rota operando en régimen estacionario, las cuales se encuentran disponibles en las bases de datos comentada en el apartado 4.4.1.

5.5.1 Estudio de los parámetros del sistema.

En el caso de las redes neuronales artificiales, existen parámetros en todos los niveles de desarrollo de la red neuronal que nos van a permitir mejorar notablemente la precisión de esta de cara al diagnóstico del fallo.

Al igual que en el caso de otros sistemas expertos, el sistema experto ANN se combina con las técnicas de diagnóstico (apartado 5.2) de las que se obtienen las características de fallo para generar el mejor sistema de diagnóstico. En consecuencia la configuración

de los parámetros que definen tanto su estructura como su entrenamiento es algo esencial a la hora de generar un buen sistema de diagnóstico.

5.5.1.1 Parámetros topológicos de la red.

Como se ha comentado en el apartado 3.3.2, la topología de la red es altamente influyente en la implantación correcta de la resolución del problema. Por ejemplo, un número escaso de neuronas en la capa oculta limitará la capacidad de aprendizaje de la red, mientras que un número mayor al necesario puede dar lugar a un mayor sobreentrenamiento de la red.

- **Número de capas ocultas.** Aunque para desarrollar una clasificación estricta solo es necesaria la capa de salida, se suele implementar al menos una capa oculta para que la red pueda detectar relaciones complementarias de tipo XOR (relación de clasificación entre datos no linealmente separables) sobre los datos de entrada.

Según Hagan [111], con una única capa oculta se pueden resolver la mayoría de problemas de clasificación. Con un mayor número de capas aumenta el coste computacional de la red tanto en su aprendizaje como en su uso, el gradiente del error durante el entrenamiento es más inestable, pueden aparecer más soluciones locales y la red tiende a un mayor sobreentrenamiento.

Según Bishop [124] no existe ninguna razón práctica por la que cualquier problema de clasificación resoluble no pueda implementarse en una red neuronal con un máximo de dos capas, siendo únicamente necesaria la segunda capa solo en los casos en que la dificultad en separar los datos en sus respectivas clases es muy alta.

Por tanto, para cubrir todos los problemas de detección de fallos (clasificación) se ha considerado en esta tesis como mínimo una capa oculta para detectar posibles relaciones entre datos de entrada y un máximo de dos capas ocultas.

- **Número de neuronas por capa oculta.** Actualmente, no existe una fórmula que define exactamente el número de neuronas idóneo en las capas ocultas para el problema propuesto.

Como se comenta en Beale [114], se puede utilizar el algoritmo de entrenamiento de retropropagación por regularización Bayesiana para que la red se ajuste al problema sin sobreentrenamiento, incluso aunque esta red esté sobredimensionada con un mayor número de neuronas, por lo que no es necesario optimizar a la baja el número de neuronas en las capas ocultas.

El algoritmo de entrenamiento de retropropagación por regularización Bayesiana es un algoritmo que calcula de forma automática una aproximación de

los mejores valores de los coeficientes (no se realiza exploración de entrenamiento, validación y prueba). El ajuste automático se realiza mediante técnicas estadísticas de métodos Bayesianos desarrollados en la tesis doctoral de MacKay [141] para modelos adaptativos.

No obstante, para realizar dicho entrenamiento es necesario usar una **función de coste** (apartado 3.3.3.2) derivada del cálculo del error cuadrático (SE), como por ejemplo el MSE (5.12). Según el artículo de Golik [142], la **función de coste por entropía cruzada** (CE) (5.13) suele mostrar una mayor optimización local para los problemas de clasificación de elementos que la función de coste basada en “SE”. Únicamente en el caso de que la inicialización de pesos sea buena para “SE” se puede obtener una leve mejoría del aprendizaje respecto a “CE”.

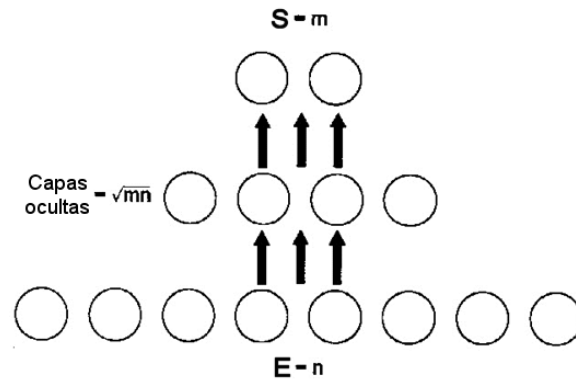
$$MSE = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n} \quad (5.12)$$

Fuente: Sharifi y Mohebbi (2012)[120]

$$CE = - \frac{\sum_{i=1}^n [\hat{y}_i \cdot \ln(\hat{y}_i) + (1 - \hat{y}_i) \cdot \ln(1 - \hat{y}_i)]}{n} \quad (5.13)$$

Fuente: Nielsen (2015) [122]

Como se comenta más adelante en el apartado 5.5.1.2.1, para mejorar el aprendizaje en este caso de la clasificación de elementos es mucho más idóneo el cálculo del coste mediante entropía cruzada, ya que no es posible obtener resultados óptimos con el algoritmo para entrenamiento de retropropagación por regularización Bayesiana. Por tanto, para obtener el número de neuronas óptimo en cada capa oculta se ha optado en esta tesis por una búsqueda estocástica con optimización heurística usando la regla de la pirámide (ejemplo en Figura 5.25); la cual ha sido comentada anteriormente en el apartado 3.3.4.1.



Fuente: Díez et al (2001) [126]


Figura 5.25. Ejemplo gráfico de la definición inicial de búsqueda del número óptimo de neuronas para una capa oculta mediante la regla de la pirámide.

- **Tipo de función de transferencia de cada capa.** Como ya se ha comentado en el apartado 3.3.1, existen el conjunto por defecto de funciones de transferencia (Tabla 5.1) con las que es posible resolver prácticamente cualquier problema en los que se pueden usar redes neuronales artificiales. La variable n de esta tabla representa el valor de la entrada, mientras que la variable a representa el valor de salida.

Dependiendo del tipo de capa (oculta, entrada, salida) es conveniente configurar una función de transferencia adecuada para mejorar la capacidad de resolución de la red. Para que el entrenamiento sea correcto, es necesario que las funciones de transferencia de las capas ocultas y de la salida sean continuas y derivables.

Acogiéndome nuevamente al estado del arte aportado por otros investigadores; según Beale [114], la mejor configuración de funciones de transferencia para la resolución de problemas de clasificación recae en usar (Tabla 5.1) la función tangente hiperbólica sigmoidea en todas las capas ocultas y una función "softmax" para la capa de salida.

Tabla 5.1 Funciones de transferencia recomendadas para problemas de clasificación de patrones.

Nombre	Relación entre entrada y salida	Función
Tangente hiperbólica sigmoidea	$a = \frac{e^n - e^{-n}}{e^n + e^{-n}}$	
Competitiva 'softmax'	$\begin{cases} a = 1 & \text{neurona con } n \text{ máximo} \\ a = 0 & \text{el resto de neuronas} \end{cases}$	

Fuente: Hagan et al (1996)[111]

Esta función de transferencia “softmax” básicamente es una función de transferencia competitiva donde la salida máxima se configura a 1, mientras que el resto de salidas se configuran a 0.

Este tipo de configuración de funciones de transferencia es usado por defecto en varios entornos matemáticos profesionales (como por ejemplo Matlab) para redes neuronales orientadas a problemas de clasificación.

- **Preprocesado de las entradas y salidas de la red.** Aplicando un preajuste de los datos en la entradas y/o salidas de la red neuronal también se puede mejorar la **tasa de aciertos** de la red.

Si dentro del vector de una característica de fallo compuesta, el intervalo de una característica simple es de órdenes de magnitud mayor que el intervalo de otra característica simple; el tipo de característica mayor se vuelve más influyente en la clasificación y el tipo de característica menor puede llegar a no ser lo bastante influyente dentro de la clasificación. No obstante, los sistemas expertos ANN son bastante robustos con datos de escalas muy dispares.

El preprocesador de datos más común en este tipo de redes tanto en sus entradas como salidas se corresponde con una función de ponderación de los datos de cada característica ajustándose entre un intervalo [-1,1] o [0,1], y una función de filtro para la mejora del rendimiento donde se anulan aquellas entradas/salidas donde los datos se mantienen constantes.

5.5.1.2 Función de entrenamiento.

Al igual que diseñar una buena estructura de red, también es importante realizar un buen entrenamiento para llegar a obtener un sistema inteligente de calidad. En tal caso, hay que analizar concienzudamente que algoritmo entrenador a utilizar y su configuración más idónea para realizar el entrenamiento de cada red.

Como ya se comentó en el apartado 3.3.3.1, existen varios tipos de algoritmos de entrenamiento, cada uno con sus cualidades y sus defectos. En el caso de las redes neuronales de tipo perceptron multicapa, es común usar algoritmos de retropropagación del gradiente.

Dentro del conjunto de algoritmos de retropropagación del gradiente existe a su vez varias adaptaciones del algoritmo original expuesto por Werbos [108].

Se han realizado varias investigaciones relacionadas con estos algoritmos de entrenamiento para discernir cuales de estos son más idóneos para aplicarlos en los problemas de reconocimiento de patrones.

Por ejemplo, el trabajo realizado por Mustafidah [143] realiza una comparativa de reconocimiento de patrones sobre clasificación con diversos algoritmos de retropropagación. Como resultado del estudio se expone el subgrupo de algoritmos que obtiene los mejores resultados. Este grupo está formado en orden del mejor al peor del grupo por el algoritmo de:

1. Retropropagación Levenberg-Marquardt.
2. Retropropagación del gradiente conjugado Powell-Beale.
3. Retropropagación del gradiente conjugado escalado.
4. Retropropagación del gradiente conjugado con actualización de Fletcher-Reeves.

Otro estudio de comparación de algoritmos de entrenamiento para la clasificación mediante redes neuronales de hematomas cerebrales, desarrollado por Sharma y Venugopalan [112], presenta como buenos candidatos para el entrenamiento de cara a la detección de patrones a los algoritmos de **retropropagación Levenberg-Marquardt**, la retropropagación del gradiente conjugado escalado y la retropropagación resiliente. En las conclusiones de este estudio se comenta además la notable velocidad de convergencia de los algoritmos de retropropagación Levenberg-Marquardt y de la retropropagación del gradiente conjugado escalado respecto del resto de algoritmos de entrenamiento.

A partir de las conclusiones de los estudios analizados parece ser que la mejor opción es el algoritmo de retropropagación Levenberg-Marquardt. Pero según el estudio de Beale [114], conforme aumenta la complejidad de la red este algoritmo pierde eficiencia, requiere unos recursos de espacio mayores a otros algoritmos y su muy rápida convergencia del error puede dar resultados inconsistentes (óptimos locales) en algunos casos de clasificación de patrones. Además, el algoritmo solo funciona con funciones de coste por cálculo del error cuadrático (apartado 3.3.3.2), expresiones (3.15) y (3.16), y funciones derivadas de este.

En el caso del algoritmo de retropropagación resiliente, este algoritmo es uno de los que menor requisito de memoria requiere. Sin embargo, como se muestra en la Figura 5.26 (a), conforme disminuye el error al avanzar los ciclos de entrenamiento también disminuye el rendimiento de este algoritmo para mejorar dicho entrenamiento.

Analizando el algoritmo de aprendizaje por retropropagación del gradiente conjugado escalado, según el mismo estudio de Beale [114], este algoritmo ofrece una buena tasa de aciertos en la mayoría de problemas (incluido la clasificación de patrones) usando unos requisitos de memoria modestos.

El algoritmo de **entrenamiento por** retropropagación del gradiente conjugado escalado converge casi tan rápido como el algoritmo de retropropagación Levenberg-Marquardt, pero no lo bastante como para llegar a saltarse posibles caminos de minimización del error cuando la convergencia está próxima a ser la óptima. Además, en contraposición del algoritmo de retropropagación resiliente, en este **caso la degradación del rendimiento es mucho menor** conforme **se minimiza el error** cuando se usa una función de coste por entropía cruzada del error (ver Figura 5.26 (b)).

Aunque la velocidad de convergencia del algoritmo de entrenamiento por retropropagación del gradiente conjugado escalado es algo menor que el algoritmo de retropropagación Levenberg-Marquardt y sus requisitos de memoria son mayores al algoritmo de retropropagación resiliente; el algoritmo de entrenamiento por retropropagación del gradiente conjugado escalado es una buena opción.

Según los datos obtenidos de Møller [116], su velocidad de entrenamiento es buena (de complejidad computacional de orden cuadrático $\Theta(3N^2)$) en comparación con el resto de algoritmos de entrenamiento, sus requisitos de memoria son bajos (de complejidad computacional de orden lineal $\Theta(N)$) y el entrenamiento desarrolla un aprendizaje excelente de la red neuronal, todo ello sin los inconvenientes comentados en los otros algoritmos de aprendizaje.

Por tanto se ha seleccionado el algoritmo de retropropagación del gradiente conjugado escalado por sus ventajas comentadas anteriormente respecto a otros tipos de algoritmos de entrenamiento por sus ventajas en la resolución de problemas de clasificación.

5.5.1.2.1 Configuración del algoritmo de entrenamiento por retropropagación del gradiente conjugado escalado.

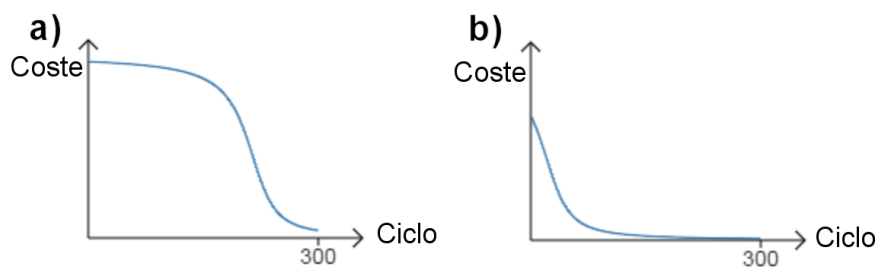
El algoritmo de entrenamiento por retropropagación del gradiente conjugado escalado requiere de una serie de parámetros con los que se puede optimizar tanto el tiempo como la calidad de su entrenamiento.

En este caso concreto de resolución de problemas de tipo clasificatorio se han seleccionado los siguientes parámetros relevantes:

- **Función de coste.** En general la función de coste más utilizada es la que calcula el error cuadrático medio (MSE); para el caso concreto de la clasificación de elementos. No obstante, como ya se expuso anteriormente en el apartado 3.3.3.2 y según el estudio de Nielsen [122], el uso de una función de coste por entropía cruzada del error (5.14) obtiene unos resultados similares a la función MSE. No obstante, a diferencia de la función MSE, con la función por entropía cruzada cuanto mayor es el error mayor es la velocidad de aprendizaje, y cuanto más cerca se encuentra el gradiente de la convergencia más lento se vuelve el aprendizaje (Figura 5.26) por lo que finalmente el rendimiento es mayor y la probabilidad de acabar en mínimos locales es menor.

$$CE = - \frac{\sum_{i=1}^n [\hat{y}_i \cdot \ln(y_i) + (1 - \hat{y}_i) \cdot \ln(1 - y_i)]}{n} \quad (5.14)$$

Fuente: Nielsen. (2015) [122]



Fuente: Nielsen. (2015) [122]

Figura 5.26. Grafica que refleja la evolución del error de aprendizaje en cada ciclo calculado por la función de coste en cada iteración del entrenamiento cuando el error es grande para una a) función de coste MSE, b) función de coste por entropía cruzada.

- **Coefficientes específicos del algoritmo.** El algoritmo de entrenamiento por retropropagación del gradiente conjugado escalado es un algoritmo de optimización heurística. En concreto para este algoritmo iterativo se puede configurar el valor inicial de σ (σ_1) que determina el cambio de los pesos para la aproximación de la segunda derivada, y el valor inicial de λ (λ_1) que regula la indefinición del Hesiano [116]. En esta fórmula (5.15) E' es un vector de gradientes locales, k representa el ciclo actual, w se corresponde con el vector de pesos, p se corresponde con un vector de conjugados que se determinan de forma recursiva mientras se ejecuta el algoritmo.

$$s_k = -\frac{E'(w_k + \sigma_k \cdot p_k) - E'(w_k)}{\sigma_k} + \lambda_k \cdot p_k \quad (5.15)$$

Fuente: Møller (1993) [116]

El algoritmo completo de entrenamiento por retropropagación del gradiente conjugado escalado se puede consultar en el anexo 2.

El intervalo idóneo para un valor de σ_1 se encuentra entre 0 y 10^{-4} , un valor mayor fuera de este intervalo ralentiza la convergencia del algoritmo.

El intervalo idóneo para un valor de λ_1 se encuentra entre 0 y 10^{-6} . Cuanto mayor es el valor de λ inicial se desarrolla un reajuste del error por ciclo más pequeño.

Según Beale [114], los valores óptimos para estos parámetros intrínsecos al algoritmo son de $5e^{-5}$ para el parámetro σ y de $5e^{-7}$ para el parámetro λ . En esta tesis se usan los valores óptimos recomendados por este estudio.

- **Rendimiento mínimo del gradiente (parada temprana).** Cuando el gradiente se hace demasiado pequeño, el ajuste obtenido se hace despreciable y puede llegar incluso a desarrollarse un sobreentrenamiento por un ajuste demasiado preciso del gradiente. Por tanto, se define un valor mínimo del rendimiento del gradiente donde se finaliza el entrenamiento cuando es alcanzado este valor.

El rango definido para el mínimo gradiente se especifica entre el $1e^{-5}$ y el $1e^{-10}$, y en particular para el algoritmo de retropropagación del gradiente conjugado escalado el valor recomendado es de $1e^{-6}$ [114],

En esta tesis se ha usado como parámetro de parada el valor recomendado de $1e^{-6}$ de rendimiento mínimo del gradiente.

- **Número máximo de fallos de validación.** Aunque en general el algoritmo de entrenamiento intenta minimizar la función de error, esta convergencia no es estricta, sino que pueden darse algún aumento del error en el proceso.

Permitiéndose un número limitado de fallos de clasificación (aumento del coste de error) en el proceso de entrenamiento se permite superar pequeños aumentos de la función de error, obteniéndose un mejor entrenamiento si en general la convergencia del error se mantiene.

El valor asignado suele ser de 5 o 6 fallos. Un valor menor puede dar lugar a encontrar una solución local, mientras que un valor mayor puede acabar generando un sobreentrenamiento del problema.

En esta tesis se ha asignado un valor de 6 como el número máximo de fallos de validación.

- **Parámetros de limitación del entrenamiento.** No en todos los casos se asegura que el entrenamiento no haga oscilar indefinidamente la función de error. Para aquellos casos en que no hay convergencia es necesario definir unos criterios de parada para que el entrenamiento no se haga eterno. En este caso se define un número máximo de ciclos de entrenamiento (por defecto 1000 ciclos) y un tiempo máximo de entrenamiento (por defecto 10 minutos).

Estos parámetros se configuran con un buen margen ya que en el peor de los casos tardará un poco más en finalizar el entrenamiento no convergente, mientras que si el valor de parada fuera ajustado podría darse el problema de finalizar el entrenamiento antes de alcanzar un posible resultado óptimo.

En esta tesis se ha mantenido los valores por defecto que limitan de forma global el entrenamiento (1000 ciclos de entrenamiento o 10 minutos de entrenamiento como máximo).

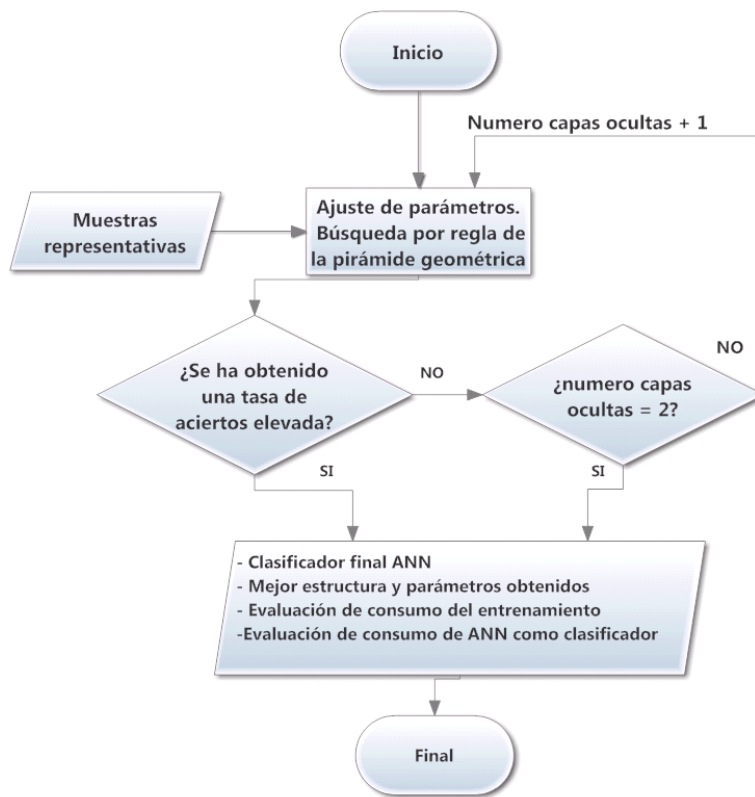
5.5.2 Algoritmo de generación del sistema de diagnóstico.

Como el problema a resolver es de clasificación, se da por hecho que no existe una relación de secuencialidad entre los datos de entrada y por tanto, al igual que el sistema experto de Maquinas de vectores de soporte del apartado 5.4, se ha realizado un aprendizaje supervisado ya que es lo óptimo para la resolución de este tipo de problemas.

La generación de un sistema de diagnóstico óptimo mediante redes neuronales es otra línea de búsqueda del mejor sistema de diagnóstico en paralelo al generador del sistema de diagnóstico óptimo mediante máquinas de vectores de soporte (apartado 5.4).

En este caso se utiliza el mismo conjunto de muestras representativas del problema obtenido a partir de todos los ensayos comentados en el apartado 4.4.1.

Como se presenta a continuación en el esquema mostrado en la Figura 5.27, se requiere de un vector de entrada del sistema, en donde se procesan los datos de entrada para obtener el clasificador óptimo local, los coeficientes, los parámetros y las evaluaciones de la tasa de aciertos.



Algoritmo de generación del sistema experto ANN de diagnóstico de fallos

Figura 5.27. Algoritmo principal de generación del mejor sistema experto ANN entrenado para la detección de los fallos correspondientes a las muestras representativas del problema.

Se inicia la búsqueda del mejor clasificador óptimo para una red compuesta por una única capa, si el resultado no es satisfactorio se incrementa el número de capas ocultas hasta un máximo de dos.

Cuando ha encontrado una red óptima o ha alcanzado el número mayor de capas ocultas el algoritmo devuelve el sistema experto ANN que ha resuelto el problema con la mayor tasa de aciertos, los parámetros con los que se ha generado el sistema experto óptimo, el tiempo de entrenamiento, el tiempo de clasificación y el coste de memoria.

5.5.2.1 Algoritmo de búsqueda de la red neuronal óptima.

Actualmente no existe ninguna fórmula o técnica con la que se pueda calcular directamente la estructura (número de capas ocultas, número de neuronas en cada capa oculta) óptima de una red neuronal respecto de un problema en concreto a resolver. Para el caso de la capa de entrada el número de neuronas se corresponde con el número de características distintas, y en el caso de la capa de salida el número de neuronas se corresponde (en este problema de clasificación) con el número de clases distintas que tiene el problema.

En el caso de esta tesis, el número de neuronas de entrada siempre es de dos porque el número de características simples que tiene una característica compleja obtenida a partir de las muestras es de dos características. La estructura óptima de cada sistema experto ANN respecto de cada técnica de diagnóstico de fallos empleada se ajusta de forma automática usando el algoritmo propuesto en este apartado 5.4.2.1.

Si el número de capas ocultas o el número de neuronas por capa oculta es bajo, se puede estar limitando la capacidad de aprendizaje por debajo de la tasa de aciertos óptima para este tipo de sistema experto. Si por el contrario se sobredimensiona el número de capas ocultas o de neuronas en las capas, puede darse el caso contrario y obtener una red sobrentrenada, donde la tasa de aciertos real será inferior a la obtenida.

En consecuencia, se desarrollan varias redes neuronales entrenadas con distintas estructuras de red (distinto número de capas o distinto número de neuronas por capa). Cada estructura distinta configurada se entrena y ajusta para obtener la mayor tasa de aciertos para esa misma estructura de red. Del conjunto de redes neuronales obtenidas con diferentes estructuras se asume que la configuración estructural idónea de la red para resolver el problema de diagnóstico se corresponde con la estructura de la red que muestra una mayor tasa de aciertos.

Como demuestra el estudio de Lippman [144], la propiedad de aproximación universal de las funciones para una red Perceptron Multicapa identifica una red de dos capas con un número de neuronas de $3 * E$ ($E =$ número de neuronas de entrada) como suficiente para la resolución de cualquier problema con convergencia.

Además, también se corrobora, mediante una variante del teorema de Kolmogorov [145], que incluso con una única capa con un número de neuronas de $2 * E + 1$ donde las funciones de transferencia sean no lineales, monótonas crecientes y continuas resul-

ta suficiente para la resolución de la mayoría de problemas con convergencia hacia una solución.

Por tanto, teniendo en cuenta dicho estudio de Lippman para la implementación del sistema de búsqueda de la mejor red neuronal, se realiza una búsqueda inicial con un número mínimo de 1 capa, y en el caso de que la tasa de fallos de la mejor red obtenida sea inferior al 90% se realiza una nueva búsqueda teniendo en consideración 2 capas ocultas (esquema en la imagen 5.24).

Como ya se ha comentado anteriormente, en este caso la configuración inicial de los parámetros de búsqueda se ha realizado mediante la regla de la pirámide geométrica [125]. Esta regla no es un axioma de cómo se tiene que empezar siempre la búsqueda, sino que al igual que las otras configuraciones de inicio usuales (apartado 3.3.4.1), solo es una opción en la que se asume que la mayoría de problemas requieren de una red neuronal donde la capa de entrada tiene el mayor número de neuronas y dicho número se va reduciendo en cada capa oculta hasta llegar a la capa de salida con el menor número de neuronas.

5.5.2.1.1 Búsqueda de la red neuronal óptima de una capa oculta.

El número inicial de neuronas para comenzar la búsqueda del número óptimo para esa capa oculta se realiza mediante la regla de la pirámide geométrica para redes con una única capa oculta (5.16). En esta fórmula $n^{\circ}_{inicial_oculta}$ es el número de neuronas de la capa oculta, E es el número de las neuronas de entrada y S el número de las neuronas de salida.

$$n^{\circ}_{inicial_oculta} = \sqrt{E \cdot S} \quad (5.16)$$

Fuente: Blum y Rivest (1992) [125]

El número mínimo a explorar es de 1 neurona en la capa oculta y el número máximo es el doble del especificado por Lippman [144] (5.17), ya que se cumple los requisitos de no linealidad, monótona creciente y continua al usar las funciones de transferencia del tipo tangente hiperbólica sigmoidea en las capas ocultas.

$$n^{\circ}_{maximo_oculta} = 4 \cdot E + 1 \quad (5.17)$$

Fuente: Lippman (1987) [144]

Como se expone en el siguiente esquema de la Figura 5.28, partiendo del número inicial de neuronas se exploran con incrementos y decrementos unitarios el aumento de la tasa de aciertos respecto a los ya explorados.

El algoritmo de búsqueda finaliza, bien cuando se ha alcanzado el límite superior e inferior, o en el caso de que el incremento o decremento del número de neuronas no mejore la tasa de aciertos con respecto a lo anteriormente explorado, aplicando un margen de exploración de 6 unidades en incremento o decremento de neuronas.

Es decir, en la rama de búsqueda con incrementos donde las tasas de errores de las redes neuronales con un rango superior hasta 6 unidades de neuronas no muestran mejoría con respecto a valores menores, se finaliza la rama de búsqueda con incrementos. Y de forma análoga, si en la rama de búsqueda con decrementos las tasas de errores de las redes neuronales con un rango inferior de hasta 6 unidades no muestran mejoría con respecto a valores superiores, se finaliza dicha rama de búsqueda decremental.

Para calcular la mejor tasa de acierto para un número definido de neuronas en la capa oculta, se desarrolla el submétodo de búsqueda y entrenamiento expuesto en el siguiente esquema de la Figura 5.28, diseñado para encontrar la mejor configuración de parámetros y conexiones para ese número de neuronas ocultas en concreto.

En el caso de que el rendimiento obtenido con una sola capa no sea lo bastante elevado (> 90% de aciertos), se vuelve a realizar la búsqueda pero construyendo redes neuronales de dos capas ocultas (apartado 5.5.2.1.2).

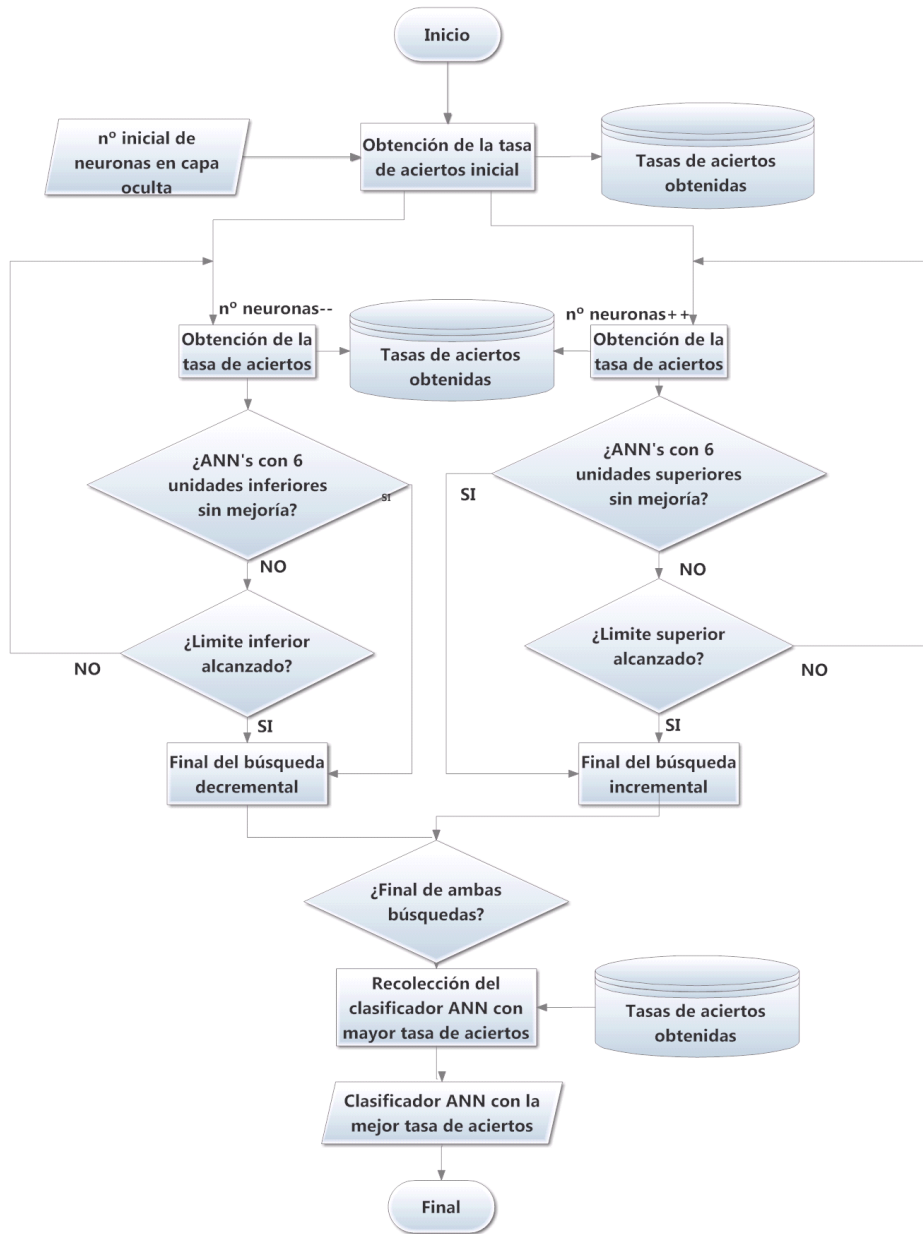


Figura 5.28. Algoritmo de búsqueda del mejor clasificador ANN para una estructura de red neuronal de una sola capa oculta.

5.5.2.1.2 Búsqueda de la red neuronal óptima con dos capas ocultas.

En este caso de búsqueda, la regla de la pirámide geométrica define la fórmula (5.18) para calcular el número de neuronas inicial de la primera capa oculta y la fórmula (5.19) para calcular el número de neuronas inicial de la segunda capa oculta.

$$n^{\circ}_{\text{Capa_oculta_1}} = S \cdot \left(\sqrt[3]{\frac{E}{S}} \right)^2 \quad (5.18)$$

$$n^{\circ}_{\text{Capa_oculta_2}} = S \cdot \sqrt[3]{\frac{E}{S}} \quad (5.19)$$

Fuente: Blum y Rivest (1992) [125]

El número mínimo de neuronas por capa a explorar sigue siendo de 1 neurona en cada capa oculta y el número máximo a explorar para las dos capas ocultas es el especificado por Lippman [144] (5.20) como el doble del número de entradas.

$$n^{\circ}_{\text{maximo_oculta}} = 2 \cdot E \quad (5.20)$$

Fuente: Lippman (1987) [144]

En la búsqueda del mejor sistema experto ANN:

- Para el ajuste del número de **neuronas de la primera capa oculta**, la ejecución del algoritmo de búsqueda es parecido al esquema de la Figura 5.28, salvo que se sustituye el cálculo de la tasa de aciertos por el subalgoritmo (Figura 5.28) para ajustar también el mejor número de neuronas de la segunda capa oculta (Figura 5.29).
- Para obtener el mejor número de neuronas en la segunda capa para un número en concreto de neuronas en la primera capa, se repite el mismo tipo de búsqueda que se ha comentado para redes de una única capa (Figura 5.28), dejando siempre el número de neuronas de la primera capa invariante.

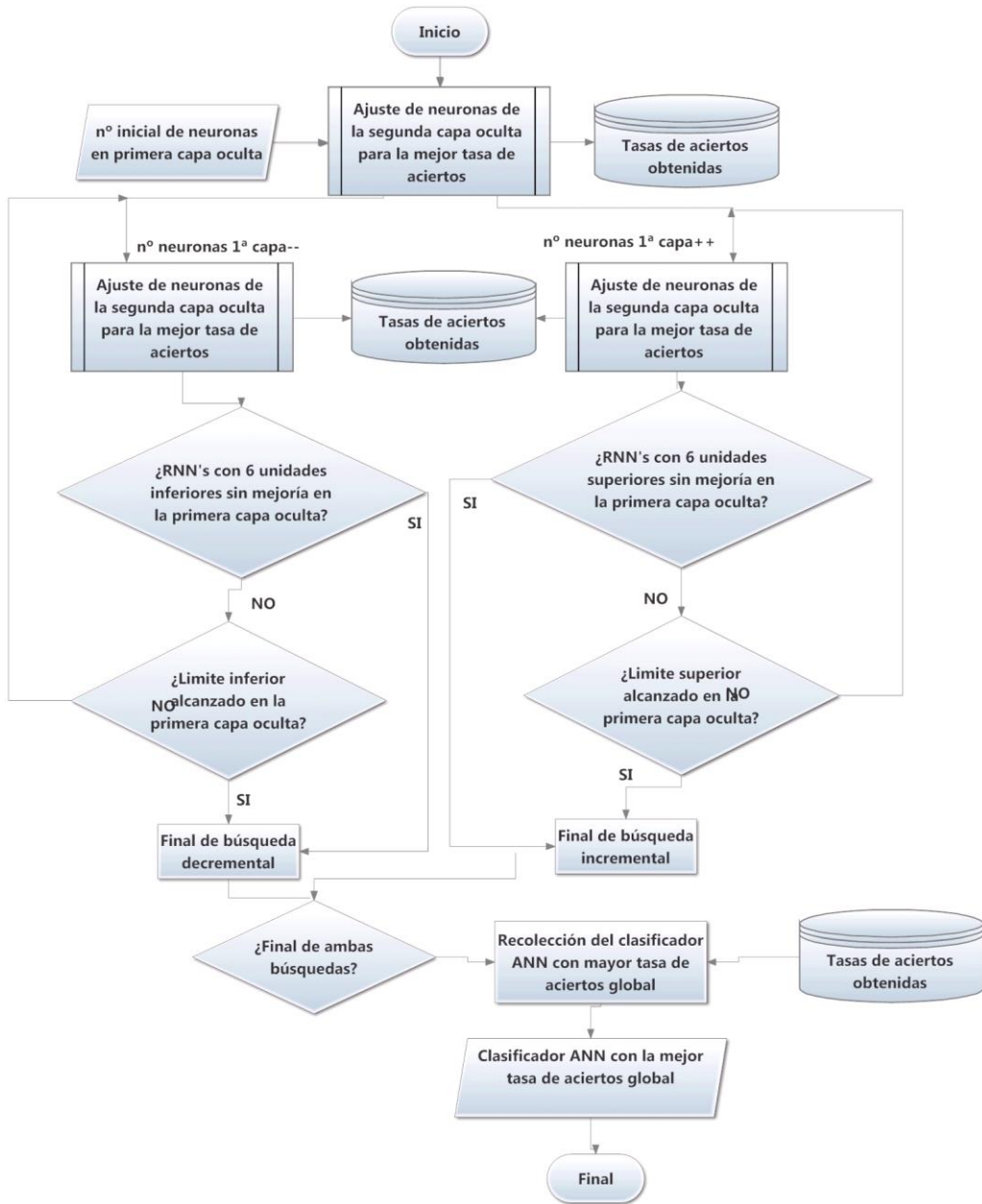


Figura 5.29 Algoritmo de búsqueda del mejor clasificador ANN para una estructura de red neuronal de dos capas ocultas.

El algoritmo de búsqueda finaliza cuando se ha alcanzado el límite superior e inferior para la primera capa oculta o cuando el incremento o decremento del número de neuronas no mejora la tasa de aciertos con respecto a los anteriores explorados respetando el margen de búsqueda.

En caso de que el rendimiento obtenido con dos capas ocultas en la red siga sin ser lo bastante alto ya no se continúa con tres capas ocultas. Dado que, como se expone en el estudio de Lippman [144] y el de Bishop [124], no existe justificación teórica por el que un problema resoluble con redes neuronales necesite un número mayor de dos capas ocultas.

Por tanto, se elige como solución final la red neuronal artificial que ha obtenido la mejor tasa de aciertos en conjunto tanto para los resultados con una capa oculta como en el caso, si se diera, de las redes neuronales de dos capas ocultas.

5.5.2.2 Entrenamiento y validación de la red neuronal óptima para una red estructural concreta.

Para cada forma estructural que se explora en el algoritmo de búsqueda es necesario calcular el ratio de acierto que puede llegar a alcanzar dicha estructura con un buen entrenamiento.

Al igual que ocurría en el caso de los sistemas expertos por SVM, no es posible entrenar simplemente al sistema experto ANN con todas las muestras que se tienen, ya que, en tal caso, al darse un más que probable sobreentrenamiento de la red, el porcentaje de fallo al resolver posteriormente los problemas se incrementaría.

El método más común para obtener una buena aproximación sobre la tasa de aciertos máxima que se puede conseguir es el de validación cruzada, comentada anteriormente en el apartado 5.4.2.2. De esta forma, se realizan múltiples entrenamientos sobre la misma tipología de red, pero variando el número de muestras usado tanto en su entrenamiento como en su posterior prueba de acierto.

Aunque la validación cruzada de uno a muchos es la que suele mostrar una aproximación mayor sobre el resultado más próximo, como indica Beale [114], dado el sistema de configuración inicial y métodos de aprendizaje esto no se cumple para el caso de las redes neuronales artificiales. En consecuencia, para obtener un sistema experto próximo al óptimo sin sobreentrenamiento en el caso de las redes neuronales se usa la **validación cruzada de entrenamiento/validación/prueba**.

5.5.2.2.1 Validación cruzada de entrenamiento / validación / prueba.

El siguiente método de validación cruzada es similar ha comentado en el apartado 5.4.2.2 para el caso de validación cruzada de entrenamiento / prueba, con la diferencia que se añade un nuevo conjunto de validación.

Con el nuevo conjunto de validación se ajusta el mejor entrenamiento del sistema experto ANN que obtiene la mayor tasa de aciertos durante el entrenamiento, mientras que el conjunto de prueba es usado para calcular el ratio de aciertos final después de que se ha entrenado completamente el sistema experto ANN.

El conjunto de muestras inicial se divide en 3 subconjuntos de muestras. Un conjunto de muestras de entrenamiento para entrenar al clasificador (entre el 60 y 80 por ciento de las muestras), un conjunto de muestras de prueba (entre un 10 y un 20 por ciento de las muestras) y un conjunto de muestras de validación (las muestras restantes). En el apartado 5.5.2.2.2 se explica cómo se realiza el particionado de los tres subconjuntos.

En cada iteración de la validación cruzada, se usa el conjunto de muestras de entrenamiento para entrenar el clasificador, de forma que este aprende del problema a resolver mediante dichas muestras.

Durante el entrenamiento, las muestras del conjunto de validación se usan para minimizar el sobreentrenamiento, pero sin ser usadas a su vez para entrenar la red neuronal. Se puede decir que se usan como un conjunto de datos virtualmente desconocido que se va usando para validar en cada iteración del ajuste del gradiente, si el gradiente converge haciendo que la tasa de error del aprendizaje sea menor pero la tasa de error en el conjunto de validación para esa iteración se mantiene o se incrementa, se entiende pues que se está dando un sobreentrenamiento respecto de las muestras de entrenamiento y por tanto en este caso se finaliza el entrenamiento.

Una vez que el clasificador es entrenado se somete a clasificar las muestras del conjunto de prueba para obtener la tasa de aciertos final de la red (representación en Figura 5.30). Si la diferencia en la tasa de aciertos entre el conjunto de validación y de test es superior a un 10% no se da por bueno el aprendizaje, ya que en este caso la red solo habría limitado el rango de muestras desconocidas al disponible en el conjunto de validación, y por tanto se ha de discriminar esa versión entrenada de la red neuronal.

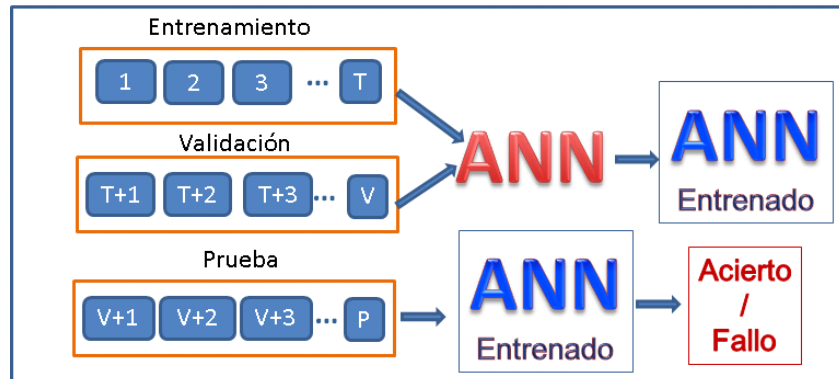


Figura 5.30 Representación del entrenamiento y clasificación para un ciclo del algoritmo de validación cruzada de entrenamiento/prueba.

En el siguiente esquema de la Figura 5.31 se muestra el algoritmo de validación cruzada por entrenamiento / validación y prueba del sistema experto ANN.

Como ya se ha comentado anteriormente, a diferencia de la validación cruzada de uno a muchos, este tipo de validación desarrolla un número de iteraciones acordado e independiente del número de muestras. Esto es una ventaja ya que independientemente del número de muestras, el orden computacional se mantiene constante $\Theta(I*k)$, siendo k el número de repeticiones.

Variando las muestras entre los diversos conjuntos se obtiene una aproximación a la tasa de aciertos global.

Cuanto mayor es el número de repeticiones mayor es la precisión en la aproximación al mejor resultado, pero también es mayor el coste de computo. En este trabajo se ha definido un total de 100 iteraciones por cada estructura de red neuronal.

Hay que tener en consideración que la inicialización de los pesos para entrenamiento en cada neurona de la red se realiza de forma aleatoria dentro de un rango de valores mínimos próximos a cero. En consecuencia, dos redes neuronales artificiales con la misma estructura, mismo número y disposición neuronal, mismo conjunto de entrenamiento / prueba / validación y mismos ajustes de entrenamiento pueden mostrar leves variaciones en el resultado de aciertos final.

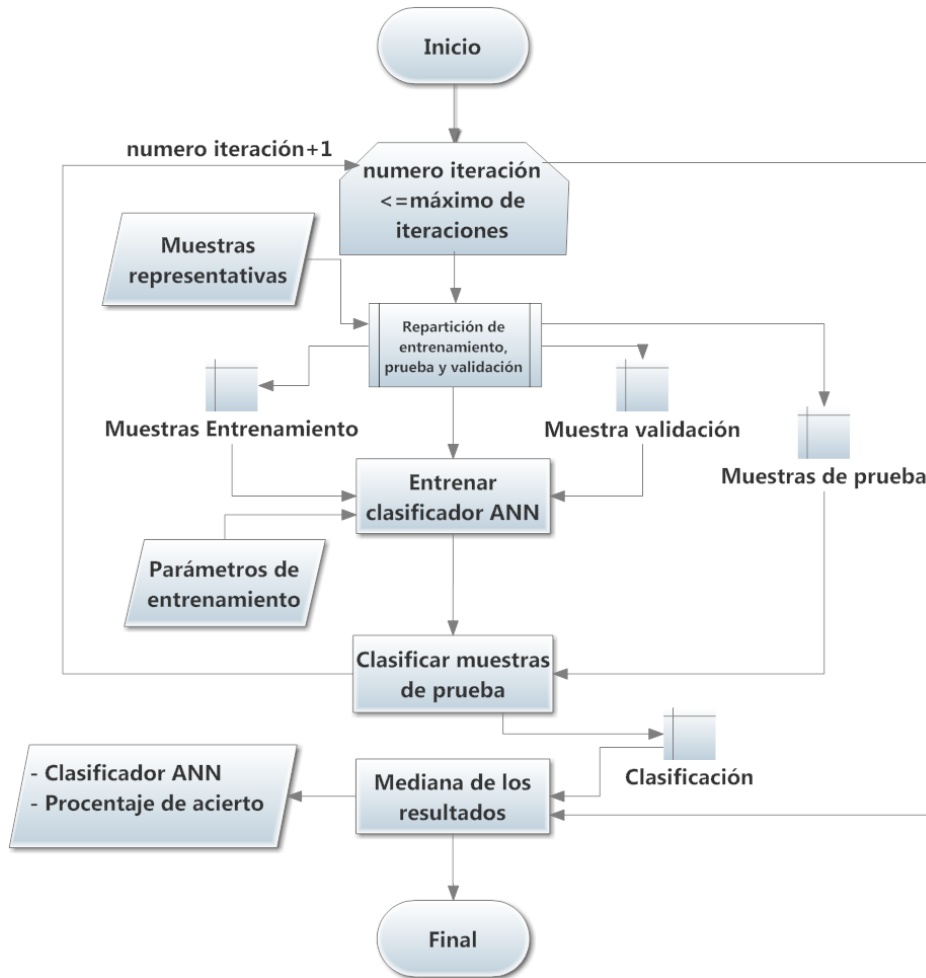


Figura 5.31 Algoritmo de aprendizaje por validación cruzada de entrenamiento, prueba y validación.

Para minimizar el efecto de la inicialización aleatoria de valores mínimos en los pesos, en cada iteración, se realiza 10 veces el mismo entrenamiento de la propia red manteniendo los mismos conjuntos de entrenamiento, de prueba y de validación. De los resultados obtenidos se utiliza el sistema experto ANN con la mejor tasa de aciertos.

Por tanto, teniendo en cuenta las 100 iteraciones y las 10 repeticiones por cada iteración, en conjunto para una sola configuración estructural de la red se realizan 1000

entrenamientos para hallar su mejor ajuste para obtener la mayor tasa de aciertos sin llegar al sobreentrenamiento.

5.5.2.2.2 Algoritmo de particionado de los conjuntos de entrenamiento, prueba y validación.

Para el entrenamiento de las redes neuronales artificiales no es posible distribuir de forma aleatoria las muestras entre los conjuntos para realizar la validación cruzada mediante conjuntos de entrenamiento, prueba y validación.

Esto es debido a que, para que el entrenamiento llegue a buen término, al menos el conjunto destinado al entrenamiento tiene que ser representativo de todo el rango que puede ser abarcado por el conjunto de características de dichas muestras correspondientes al problema a resolver.

En consecuencia, el algoritmo de particionado implementado en este caso para la distribución de muestras se encarga de distribuir las muestras entre los diversos conjuntos, de forma que el rango de las características representativas sea lo más extendido respecto del rango que abarca, manteniendo un factor de aleatoriedad local.

Este algoritmo separa inicialmente las muestras en grupos correspondientes a las clases disponibles. Cada uno de dichos grupos es procesado por separado obteniendo un conjunto de entrenamiento, prueba y validación para cada clase; y uniéndose finalmente los grupos de cada clase a sus respectivos globales de entrenamiento, prueba y validación.

Para cada clase se sigue la rutina marcada en el esquema de la Figura 5.32. Primero se asignan los elementos para el grupo de entrenamiento que es el más importante, relegándose los otros dos al conjunto de muestras restantes.

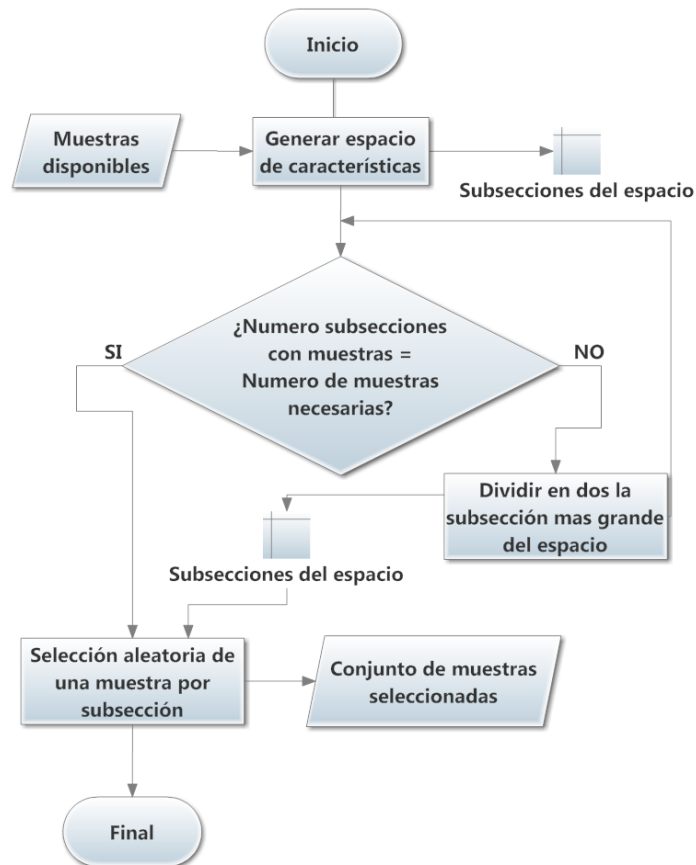


Figura 5.32. Algoritmo de selección de muestras para formar un conjunto donde la representatividad de las muestras abarcan todo el rango del problema.

Como se observa en el algoritmo de la Figura 5.32, se parte de un espacio de características inicial donde se insertan las muestras (ejemplo en Figura 5.33). Este espacio se va dividiendo en cada iteración en subsecciones hasta alcanzar el número de subsecciones (en el que al menos existe una muestra) igual al número de muestras que debe contener el conjunto de entrenamiento.

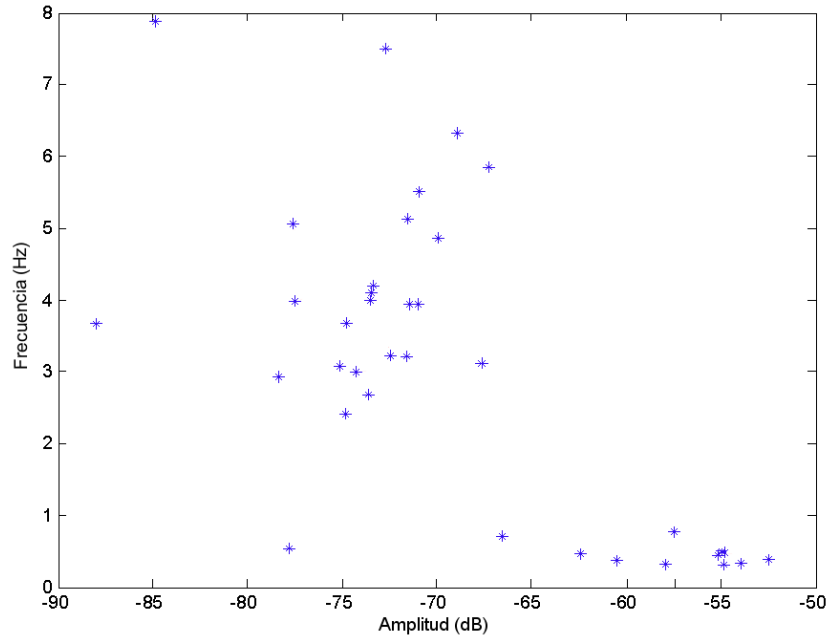


Figura 5.33. Representación de un espacio de características para un motor sano obtenidas por la técnica EPVA.

La división del espacio dentro de una iteración se realiza sobre la subsección de mayor tamaño, seleccionando una dimensión al azar y dividiendo ésta para formar las dos nuevas subsecciones del espacio. Las subsecciones que no tienen muestras no son consideradas para la cuenta de muestras ni para subdividirse y se descartan.

Una vez que el subespacio es dividido en el número adecuado de subsecciones (igual al número de muestras que debe contener el conjunto de entrenamiento) sin contar las subsecciones sin muestras, se elige al azar una muestra de cada subsección, obteniéndose finalmente un conjunto de muestras de entrenamiento bastante representativas del espacio de características que abarca el conjunto de muestras. Ejemplo de división del espacio de características en la Figura 5.34.

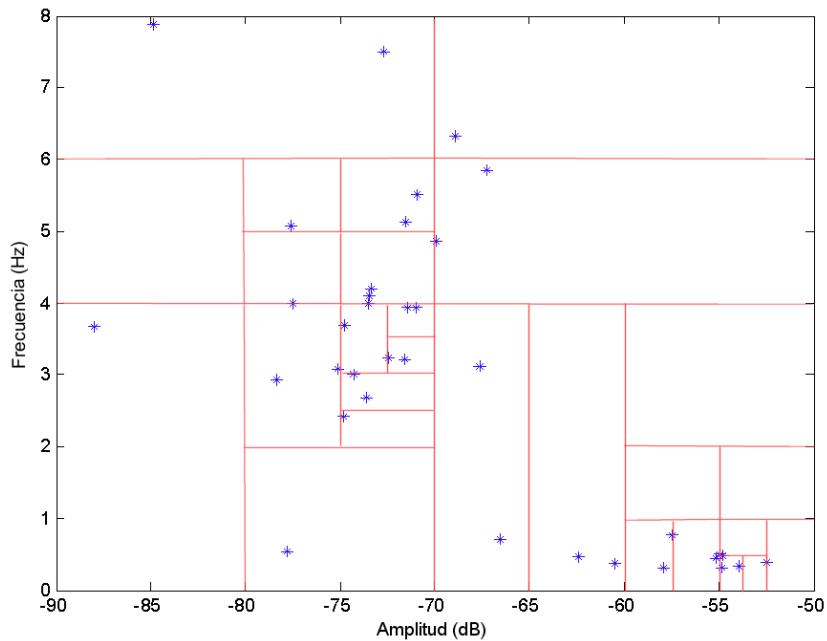


Figura 5.34 Representación de un ejemplo de división del espacio de características para un motor sano obtenidas por la técnica EPVA dividido en 23 subconjuntos de 2 elementos (los subconjuntos vacíos no se tienen en cuenta).

Para el caso de las muestras de prueba y de validación se realiza el mismo proceso de división y selección del espacio.

Como restricción laxa, se intenta inicialmente que cada subsección se limite como mínimo a un número de 3 muestras, una para entrenamiento, una para prueba y una para validación. En el caso de que esta limitación no pueda ser cumplida, se va decreciendo el número de muestras hasta alcanzar su valor mínimo de uno por subsección.

5.6 Entorno de desarrollo empleado del sistema generador y optimizador de sistemas de diagnóstico de fallos.

La implementación del sistema completo de desarrollo de sistemas de diagnóstico mediante el uso de sistema experto de tipo SVM y ANN se ha llevado a cabo en un equipo informático tipo de ordenador personal con las siguientes características:

- Procesador i7, marca Intel, modelo i7-930. Velocidad de proceso de 2966MHz (135MHz de bus multiplicado por 22). Cuatro núcleos independientes de dos hilos por núcleo. Cache L1 de 32KB, L2 de 256KB y L3 de 8192KB
- Placa Base Gigabyte, modelo GA-X58A-UD3R v2. Chipset: Intel Tylersburg X58.
- Memoria RAM de 12GB. Marca Kingston, Modelo 99U5471-001.A00LF, Bus DDR3 a 667MHz
- Tarjeta gráfica NVIDIA, modelo GTX 750 Ti. Memoria RAM dedicada de 2GB. Bus PCI-E 16x. Versión del procesador Maxwell GM107 de 640 núcleos gráficos operando a 1GHz.
- Disco duro mecánico de 1863GB. Marca Seagate. Modelo ST2000DL004 HD204UI. Velocidad de giro 5400 rpm. Bus SATA2 de 300MB/s.
- Disco duro mecánico de 1863GB. Marca Seagate. Modelo ST32000542AS. Velocidad de giro 5900 rpm. Bus SATA2 de 300MB/s.
- Sistema operativo Windows 7 versión profesional de 64 bits con *Service Pack SP1*.

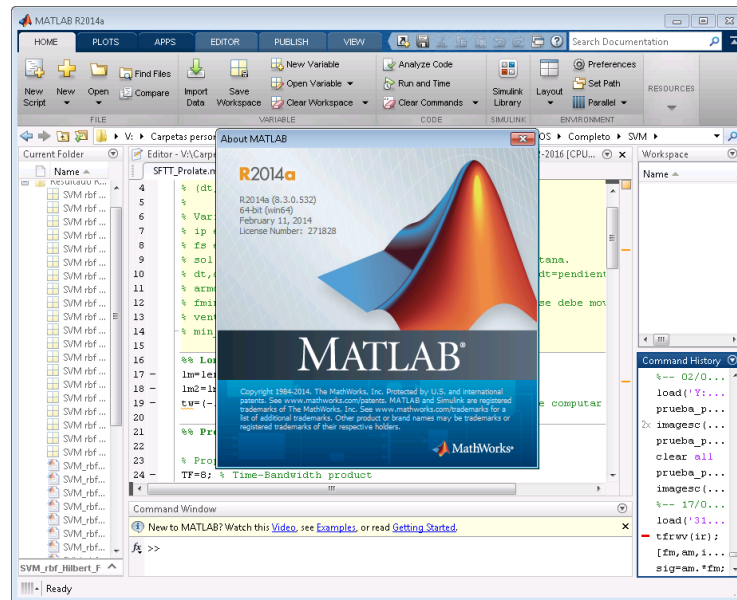


Figura 5.35. Entorno de desarrollo matemático MATLAB R2014a.

El entorno de desarrollo usado para implementar todo el sistema ha sido la herramienta de software matemático *MATLAB* (Figura 5.35), versión R2014a (v8.3.0.532) de 64bits.

El lenguaje utilizado para la programación del sistema es el lenguaje interpretado propio de *MATLAB*, haciendo uso de su potencial de programación orientada a objetos (*POO*).

La selección de dicha plataforma para la construcción del sistema ha sido motivada por su versatilidad para el desarrollo matemático, por sus amplios paquetes *Toolbox* de librerías específicas y en general por su gran uso dentro del ámbito científico y tecnológico.

Las *Toolbox* utilizadas en este caso son:

- *Signal Processing Toolbox*. Esta *Toolbox* se ha usado por sus múltiples librerías orientadas hacia el análisis y procesamiento de señales.
- *Statistics and Machine Learning Toolbox*. Esta *Toolbox* se ha usado por sus librerías para el desarrollo, entrenamiento y aplicación de los sistemas expertos del tipo SVM.
- *Neural Network Toolbox*. Esta *Toolbox* se ha usado por sus librerías para el desarrollo, entrenamiento y aplicación de los sistemas expertos del tipo ANN.

En el anexo 3 de esta tesis se ha añadido una reseña al intento de acelerar el entrenamiento de los sistemas expertos mediante la paralelización computacional en lenguaje CUDA. Sin embargo, como se explica en el anexo 3, no se han obtenido mejoras en el tiempo de cálculo necesario respecto a la implementación original de la aplicación.

Este problema se ha dado con el hardware gráfico utilizado con soporte CUDA (NVIDIA GeForce GTX 750Ti), el cual tiene mucha potencia de cálculo en coma flotante de precisión simple pero muy poca potencia de cálculo cuando la precisión en coma flotante es doble (la '*Statistics and Machine Learning Toolbox*' utiliza solo el tipo de coma flotante de doble precisión para realizar los cálculos).

5.7 Resultados obtenidos del sistema autónomo para régimen estacionario.

En este apartado **se exponen y comentan todos los datos significativos de los que se pueden deducir el mejor sistema de diagnóstico global** entre todas las soluciones de sistemas de diagnóstico incluyendo tanto la técnica de diagnóstico como el sistema

experto, y de forma local **para cada técnica con su respectivo sistema experto también se ha conocido su mejor solución**. Además, **también se han obtenido otros datos relevantes**, como los **parámetros óptimos de ajuste o el tiempo requerido para el diagnóstico**.

En resumen, para obtener estos resultados se han utilizado las siguientes características de fallo según la técnica de diagnóstico empleada. En todos los casos se han utilizado dos características simples para formar un espacio bidimensional, obteniéndose cada conjunto de características de las técnicas comentadas en el apartado 5.2. Es decir, del análisis de la corriente (apartado 5.2.1), de la componente alterna del módulo de la señal analítica (apartado 5.2.2), del análisis del orden armónico (apartado 5.2.3), de la técnica Cepstrum (apartado 5.2.4) y de la técnica EPVA (apartado 5.2.5).

- Para el **análisis de la corriente estatórica** se ha utilizado como primera característica simple el valor medio obtenido de las amplitudes de las componentes armónicas más representativas ($k = 1$ y $k = -1$). Como segunda característica simple se ha utilizado la distancia en frecuencia existente entre la frecuencia de alimentación y la frecuencia de fallo para $k = 1$. Características utilizadas:
 - **Amplitud media de $k=1$ y $k=-1$.**
 - **La distancia entre las componentes de fallo.**

- Para la técnica de diagnóstico basada en la **componente alterna del módulo de la señal analítica obtenida mediante Hilbert** se ha utilizado como primera característica simple la amplitud de la componentes armónica de fallo para $k=1$. Como segunda característica se ha utilizado la distancia existente desde el origen hasta la frecuencia de fallo para $k = 1$. Características utilizadas:
 - **Amplitud de la componente frecuencial $k=1$.**
 - **La distancia entre las componentes de fallo.**

- Para la técnica de **análisis del orden armónico (HOTA)** se ha utilizado como primera y segunda característica simple la amplitud de la componente armónica de fallo para $k = 1$ y para $k = -1$ respectivamente. Características utilizadas:
 - **Amplitud de la componente frecuencial $k=1$.**
 - **Amplitud de la componente frecuencial $k=-1$.**

- Para la técnica de diagnóstico **Cepstrum** se ha usado como primera característica simple la amplitud obtenida en la posición de tiempo que ha sido generada por la repetición de las componentes armónicas de fallo. Como segunda característica se ha utilizado la distancia existente entre el origen y esta posición en tiempo. Características utilizadas:
 - **Amplitud en el instante de tiempo correspondiente a la periodicidad de los armónicos de fallo.**
 - **Instante de tiempo de la periodicidad de los armónicos de fallo.**

- Para la técnica de diagnóstico basada en el **vector de Park extendido** se ha utilizado como primera característica simple la amplitud de la componente armónica de fallo para $k=1$. Como segunda característica se ha utilizado la distancia existente desde el origen hasta la frecuencia de fallo para $k = 1$. Características utilizadas:
 - **Amplitud de la componente frecuencial $k=1$.**
 - **La distancia entre las componentes de fallo.**

Para cada técnica de diagnóstico utilizada se han desarrollado variantes en las que se ha aplicado un filtro de ventana de Hann (apartado 2.3.3.1) y/o se ha sustituido la FFT para realizar la transformación al dominio de frecuencias por una transformación basada en el desarrollo del periodograma de Welch (apartado 2.3.3.2).

Como se ha desarrollado en el apartado 2.3.3, la aplicación del filtro de Hann o la transformación mediante el periodograma de Welch, tendría en teoría que mejorar la representatividad de las características de fallo obtenidas con las técnicas de diagnóstico. Y, como se observa posteriormente en los resultados expuestos en este apartado, esta presuposición de mejora al utilizar el filtro de Hann o la transformada de Welch es correcta en la mayoría de casos.

Antes de comentar en detalle los resultados obtenidos se han de tener en cuenta dos detalles importantes:

- Los diversos coeficientes de ajustes de los sistemas expertos están optimizados de forma automática por el sistema automático de generador de sistemas de diagnóstico óptimos de esta tesis.

Se ha comprobado los **sistemas de diagnóstico óptimos** generados para las muestras de los motores del banco de ensayos (apartado 4.4) sobre otro conjunto de muestras obtenidas previamente con un motor “**M4BP 160 MLB**”

3GBP” de tipología distinta a los motores usados para generar el banco de ensayos. La detección del fallo de los sistemas de diagnóstico sobre motores de otra tipología ha sido errónea en la mitad de los casos.

De esta forma se ha corroborado que estos ajustes no son generales, pero sí óptimos respecto a la base de datos de muestras procesadas (apartado 4.4) de la tipología de motores utilizada.

Por tanto estos ajustes no son extrapolables, en principio, para el diagnóstico de motores de otra tipología distinta.

En tal caso, ya existe el sistema creado en esta tesis para volver a generar sistemas de diagnóstico óptimos con nuevos coeficientes para esos otros tipos de motores.

- Todos los tiempos que se han obtenidos son relativos, pero han sido obtenidos en el mismo equipo informático. Todos los tiempos que se muestran en las tablas de resultados dependen tanto de la implementación del algoritmo usado como del dispositivo donde se ejecuta (apartado 5.6). En consecuencia, estos tiempos se tienen que tomar solo como aproximaciones comparables con el resto de tiempos mostrados en los resultados de esta tesis.

La siguiente Tabla 5.2 muestra los tiempos obtenidos al ejecutar las diversas técnicas (y sus variantes) de obtención de características representativas de fallo a partir de las muestras obtenidas con cada técnica de diagnóstico de fallos. La primera columna de esta tabla muestra cada uno de las técnicas utilizadas en esta tesis. El resto de las columnas de la tabla representa:

- **Utilización de ventana de Hann.** Indica si en esa variación en concreto de la técnica de obtención de características se ha aplicado un filtro de ventana de Hann a la señal resultante. Con este parámetro se puede observar en esta tabla si el uso de ventanas de Hann influye significativamente en el tiempo de proceso.
- **Utilización de periodograma de Welch.** Indica si en esa variación en concreto de la técnica de obtención de características se ha reemplazado a la función FFT para obtener el espectro de frecuencias por la función Welch. Con este parámetro se puede observar la variación positiva o negativa del tiempo al sustituir la función FFT por el periodograma de Welch
- **Tiempo total para generar características <TTC>.** Muestra en segundos, para cada alternativa, el tiempo total necesario para generar las características representativas. Comprende el tiempo de carga de cada una de las muestras y el tiempo de generación de las características de todas las muestras.

- **Tiempo medio para procesar una muestra <TMm>**. Muestra en milisegundos, para cada alternativa, el tiempo medio calculado de procesar una muestra para obtener las características de dicha muestra. A diferencia del tiempo (TTc), en este caso no se tiene en cuenta el tiempo de carga de la muestra.

Tabla 5.2. Resultados de los tiempos obtenidos en la ejecución de las técnicas de diagnóstico en donde se han obtenido las características representativas de fallo a partir de las muestras.

	Variaciones		Tiempo de computo	
	Ventana de Hann	Periodograma Welch	Tiempo total para generar características < TTc > (s)	Tiempo medio para procesar una muestra < TMm > (ms)
FFT	No	No	2.134,86	3.630,21
	Si	No	2.140,41	3.814,02
	No	Si	2.374,42	4.706,31
	Si	Si	2.472,63	4.889,77
AC del módulo de la Señal Analítica	No	No	2.556,39	5.546,95
	Si	No	2.597,87	5.702,35
	No	Si	2.703,95	6.244,51
	Si	Si	2.740,94	6.395,85
HOTA estacionario	No	No	2.580,57	5.431,47
	Si	No	2.593,11	5.677,07
	No	Si	2.677,86	6.072,63
	Si	Si	2.729,43	6.256,86
Cepstrum	No	No	2.195,24	3.911,79
	Si	No	2.239,16	4.069,69
	No	Si	2.390,04	4.798,27
	Si	Si	2.424,65	4.949,63
EPVA	No	No	1.660,50	4.198,11
	Si	No	1.658,75	4.377,32
	No	Si	1.707,05	4.917,01
	Si	Si	1.719,27	5.073,35

En cada una de las columnas de tiempo de cómputo de la Tabla 5.2 se muestra en color rojo el peor tiempo (el de mayor valor) y de color verde el mejor tiempo (el de menor valor).

De esta misma tabla de tiempos 5.2 se observa que la obtención de características representativas de todo el conjunto de muestras (216) para cada técnica de diagnóstico con su correspondiente variación en torno de 27 a 46 minutos en el mejor y peor caso, dependiendo del número de muestras a procesar este tiempo variará.

En este caso es importante el tiempo obtenido para procesar una sola muestra, ya que es independiente del número de muestras de entrenamiento y muestra una aproximación al tiempo requerido en obtener las características representativas para una muestra desconocida que va a ser diagnosticada. El tiempo en este caso varía entre los 3 y 6 segundos, que es un tiempo suficientemente pequeño para realizar un diagnóstico en tiempo real.

En la técnica EPVA se muestra como, a pesar de que el tiempo de media para procesar una muestra es mayor que en otras técnicas, el tiempo total de entrenamiento es sensiblemente menor.

Esto se debe a que, para obtener las características representativas de fallo de una muestra con la técnica EPVA, esta tiene que procesar las tres corrientes de fase, por lo que el tiempo de cálculo es aproximadamente 3 veces mayor al resto de técnicas que solo procesan una sola fase. Sin embargo, al obtener solo 2 características por muestra en comparación del resto de técnicas que obtienen 6 características por muestra (2 características por 3 corrientes de fase), el tiempo de entrenamiento es menor ya que el número total de características a procesar es tres veces menor que el resto de técnicas de diagnóstico.

Al sustituir la transformada FFT por la transformada Welch, se observa un incremento del tiempo que no es demasiado significativo. Por supuesto, las variantes que añaden un filtrado de Hann añaden también el tiempo extra de dicho filtrado.

Todos estos tiempos pueden mejorarse optimizando por una parte la implementación de las técnicas de diagnóstico, de los sistemas expertos y en general de cualquier algoritmo implementado en el sistema de generación de sistemas de diagnósticos expertos óptimos; y por supuesto también se puede mejorar los tiempos obtenidos mediante el uso de un hardware más potente.

5.7.1 Resultados obtenidos con sistemas expertos de Máquinas de vectores de soporte (SVM).

Los resultados obtenidos respecto del sistema experto de tipo SVM se muestran a continuación en las tablas 5.3 y 5.4 respecto del tiempo requerido, y 5.6 y 5.7 respecto de los ajustes óptimos de parámetros y tasas de acierto.

Para todas las variantes de las técnicas en que se han aplicado sistemas expertos SVM, se ha verificado dentro de la fase de entrenamiento que dicho sistema experto no se encuentre nunca sobreentrenado.

Para obtener estos resultados, como ya se ha comentado en el apartado 5.4.2, en el proceso de generación del SVM óptimo se han utilizado **dos kernels** (Polinomial y de Base Radial Gaussiana), se ha usado un algoritmo de ajuste de coeficientes mediante **búsqueda en malla** y a cada nodo de la búsqueda en malla se ha aplicado una **validación cruzada de uno a muchos** para obtener la estimación de su tasa de aciertos.

5.7.1.1 Comparación de tiempos entre kernel polinomial y kernel RBF.

La tabla 5.3 para SVM con kernel polinomial y la tabla 5.4 para SVM con kernel RBF muestran en primera columna cada uno de las técnicas desarrolladas. El resto de las columnas de las tablas se corresponden con:

- **Utilización de ventana de Hann.** Indica si esta variación utiliza o no un filtro de ventana de Hann.
- **Utilización de periodograma de Welch.** Indica si esa variación reemplaza la función FFT por la función Welch.
- **Tiempo de cómputo del SVM - Entrenamiento < Te-1 >.** Muestra en segundos el tiempo total necesario para realizar el mejor ajuste de coeficientes y el mejor entrenamiento de forma que se obtiene la tasa de aciertos más alta
- **Tiempo de cómputo del SVM - Clasificación < Tc-1 >.** Muestra en milisegundos la media del tiempo requerido para clasificar las características representativas de una sola muestra.
- **Tiempo de cómputo total - Entrenamiento < Te-2 >.** Muestra en segundos el tiempo total de entrenamiento de todo el sistema de diagnóstico. Es decir el resultado de sumar el tiempo de procesar las muestras (TTc) con la variante de la técnica de obtención de características y el tiempo de entrenamiento óptimo posterior del sistema SVM (Te-1). Estos tiempos obtenidos pueden dar una idea del tiempo requerido para el entrenamiento del sistema experto SVM con cada variante de las técnicas de obtención de características.

- **Tiempo de cómputo total - Clasificación < Tc-2 >.** Muestra en milisegundos el tiempo total procesar una muestra de un motor (TMm) y clasificarlo con el sistema experto mejor SVM con esa variante de las técnicas de obtención de características (Tc-1). Estos tiempos permiten obtener una aproximación del tiempo final que requerirá el sistema de diagnóstico optimizado para diagnosticar el fallo en una muestra de motor desconocida.

Como se observa en estas tablas, los costes de tiempos en clasificación del SVM (Tc-1) son despreciables respecto del tiempo de diagnóstico final del sistema de diagnóstico (Tc-2), ya que el tiempo de obtención de características para una muestra (TMm) es mucho más significativo.

Aunque el tiempo final para el diagnóstico es algo excesivo para realizar un diagnóstico en tiempo real, sigue siendo lo bastante pequeño como para poder mejorarlo, utilizando un equipo informático adecuado, y así alcanzar un margen de tiempo menor para realizar un diagnóstico en tiempo real.

Respecto de los costes de tiempos en entrenamiento tanto del clasificador como respecto de todo el sistema de diagnóstico varían bastante más, pasando del orden de segundos al de miles de segundos. Aun así, siguen siendo tiempos asumibles si se tiene en cuenta que el entrenamiento solo se realiza una única vez para obtener el sistema de diagnóstico óptimo.

Tabla 5.3. Resultados de los tiempos obtenidos del sistema experto de clasificación SVM configurado con un kernel polinomial, y los tiempos totales de los diversos sistemas de diagnóstico.

SVM con kernel Polinomial	Variaciones		Tiempo de cómputo del SVM		Tiempo de cómputo total	
	Ventana de Hann	Periodograma Welch	Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-2 > (s) (TTc+Te-1)	Clasificación < Tc-2 > (ms) (TMm+Tc-1)
FFT	No	No	1.935,26	0,469	4.070,12	3.630,67
	Si	No	617,55	0,469	2.757,96	3.814,49
	No	Si	6.289,73	0,460	8.664,15	4.706,77
	Si	Si	4.683,08	0,458	7.155,71	4.890,23
AC del módulo de la Señal Analítica	No	No	1.517,30	0,466	4.073,69	5.547,42
	Si	No	14.420,65	0,462	17.018,52	5.702,82
	No	Si	432,63	0,458	3.136,58	6.244,97
	Si	Si	9,53	0,467	2.750,47	6.396,32
HOTA estacionario	No	No	1.742,81	0,643	4.323,38	5.432,12
	Si	No	2,63	0,462	2.595,74	5.677,53
	No	Si	853,82	0,466	3.531,68	6.073,10
	Si	Si	7.379,91	0,652	10.109,34	6.257,51
Cepstrum	No	No	3.173,68	0,464	5.368,92	3.912,25
	Si	No	5.801,92	0,466	8.041,08	4.070,16
	No	Si	3.168,20	0,462	5.558,24	4.798,73
	Si	Si	4.472,12	0,469	6.896,77	4.950,10
EPVA	No	No	272,12	0,402	1.932,62	4.198,51
	Si	No	1.788,49	0,402	3.447,24	4.377,72
	No	Si	563,35	0,402	2.270,40	4.917,41
	Si	Si	475,39	0,402	2.194,66	5.073,75

Tabla 5.4. Resultados de los tiempos obtenidos del sistema experto de clasificación SVM configurado con un kernel Gaussiano RBF, y los tiempos totales de los diversos sistemas de diagnóstico.

SVM con kernel RBF Gaussiano	Variaciones		Tiempo de cómputo del SVM		Tiempo de cómputo total	
	Ventana de Hann	Periodo- grama Welch	Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-2 > (s) (TTc+Te-1)	Clasificación < Tc-2 > (ms) (TMm+Tc-1)
FFT	No	No	3,71	0,511	2.138,57	3.630,72
	Si	No	3,64	0,508	2.144,05	3.814,53
	No	Si	2,68	0,521	2.377,10	4.706,83
	Si	Si	55,93	0,507	2.528,56	4.890,28
AC del módulo de la Señal Analítica	No	No	61,02	0,517	2.617,41	5.547,47
	Si	No	63,03	0,509	2.660,90	5.702,86
	No	Si	3,65	0,731	2.707,60	6.245,24
	Si	Si	61,01	0,507	2.801,95	6.396,36
HOTA estaciona- rio	No	No	4,44	0,514	2.585,01	5.431,98
	Si	No	61,01	0,516	2.654,12	5.677,59
	No	Si	233,23	0,513	2.911,09	6.073,14
	Si	Si	62,27	0,730	2.791,70	6.257,59
Cepstrum	No	No	88,40	0,517	2.283,64	3.912,31
	Si	No	6,71	0,518	2.245,87	4.070,21
	No	Si	6,80	0,519	2.396,84	4.798,79
	Si	Si	6,98	0,518	2.431,63	4.950,15
EPVA	No	No	214,56	0,447	1.875,06	4.198,56
	Si	No	38,32	0,450	1.697,07	4.377,77
	No	Si	0,63	0,448	1.707,68	4.917,46
	Si	Si	0,76	0,451	1.720,03	5.073,80

Tabla 5.5. Resultados máximo, mínimo y promedio de los tiempos obtenidos del sistema experto de clasificación SVM con kernel polinomial y con kernel Gaussiano RBF.

		Tiempo de cómputo del SVM		Tiempo de cómputo total	
		Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-2 > (s) (TTc+Te-1)	Clasificación < Tc-2 > (ms) (TMm+Tc-1)
SVM kernel Polinomial	Mínimo	2,63	0,402	1.932,62	3.630,67
	Máximo	14.420,65	0,652	17.018,52	6.396,32
	Promedio	2.980,01	0,460	5.294,86	5.033,63
SVM kernel RBF Gaussiano	Mínimo	0,63	0,447	1.697,07	3.630,72
	Máximo	233,23	0,731	2.911,09	6.396,36
	Promedio	48,94	0,520	2.363,79	5.033,68

La alta variación del entrenamiento del sistema se da porque al ser un proceso exploratorio en búsqueda del mejor ajuste de coeficientes con mayor ratio, el tiempo de entrenamiento en consecuencia está supeditado al coste de exploración requerido hasta llegar a la mejor solución. Dicho coste de exploración depende de las ramificaciones desarrolladas en la búsqueda de la optimización y en el número de niveles de profundidad explorados.

Lo más relevante de todos estos datos de tiempo se da, independientemente del tipo de kernel, en los sistemas de diagnóstico SVM que utilizan la técnica FFT con y sin filtro de Hann. Estos sistemas de diagnóstico promedian el tiempo de diagnóstico (Tc-2) más pequeño, en torno a los 3 segundos.

En contraposición, los sistemas de diagnóstico basados en sistemas expertos SVM que utilizan HOTA con la función Welch muestran el mayor requisito de tiempo medio para el diagnóstico. Este tiempo llega a duplicar incluso al tiempo mínimo de diagnóstico obtenido. Aun así, el tiempo mayor de diagnóstico se encuentra en torno a los 6 segundos, por lo que no se puede decir que sea demasiado elevado.

Respecto al tipo de kernel utilizado en el sistema experto SVM, se observa en la Tabla 5.5 como el kernel RBF es bastante más rápido en entrenamiento debido a que llega a sobreentrenamiento mucho antes. Sin embargo, en el caso de la clasificación de una muestra se observa como el kernel polinomial es ligeramente más rápido.

La selección del mejor “sistema óptimo” se evalúa comparando la mejor tasa de aciertos. En SVM estas tasas de aciertos se muestran en las tablas 5.6 y 5.7. No obstante, si existen limitaciones en el requisito del hardware donde se implementará el sistema experto, las tablas de tiempos (5.3, 5.4 y la comparativa de la tabla 5.5) pueden ayudar a elegir una alternativa aceptable de sistema de diagnóstico que cumpla con los requisitos de tiempo manteniendo una buena tasa de aciertos.

5.7.1.2 Comparación de resultados entre kernel polinomial y kernel RBF.

En las siguientes tablas 5.6 para kernel polinomial y 5.7 para kernel RBF, se muestran los parámetros óptimos y las mejores tasas de aciertos obtenidas con la combinación de un sistema experto SVM y una variante de las técnicas de obtención de características. En la primera columna se indica la técnica de obtención de características usada para ese conjunto de filas, mientras que el resto de las columnas de las tablas se corresponden con:

- **Utilización de ventana de Hann.** Indica si esta variación utiliza o no un filtro de ventana de Hann.
- **Utilización de periodograma de Welch.** Indica si esa variación reemplaza a la función FFT por una función Welch.
- **Nivel de exploración alcanzado <lvl>.** Como se ha comentado en el apartado 5.4.2.1, el ajuste de los coeficientes se realiza mediante un algoritmo de búsqueda en malla (ver imagen 5.15). Este valor de “Nivel de exploración alcanzado” muestra hasta qué nivel de profundidad ha llegado la búsqueda en malla antes de obtener el mejor ajuste de los coeficientes.
- **Límites de caja <C>.** Coeficiente del sistema experto SVM que limita la ventana de ajuste del hiperplano. Al modificarse este se regula el margen de distancia del hiperplano variando el error aceptado en la clasificación de clases.
- **Orden Polinomial <d> (solo para kernel polinomial).** Este parámetro afecta al kernel polinomial, indicando el orden polinomial y afectando a la curvatura del hiperplano.
- **Sigma (σ) (solo para kernel RBF).** Este parámetro afecta al kernel RBF usándose como factor de escala de la función de base radial.
- **Tasa de aciertos.** Muestra la tasa de aciertos estimada del sistema de diagnóstico óptimo formado por el método de obtención de características indicado y por el sistema experto SVM entrenado de forma óptima. Este parámetro es esencial en esta tesis, pues nos indica la calidad de diagnóstico del mejor sistema experto final generado.

En ambas tablas, para el sistema experto SVM con distinto kernel se muestran unos ratios bastante similares comparando cada técnica con su homónimo de la otra tabla. En las dos tablas se observa claramente que la técnica Cepstrum en cualquiera de sus variantes es la que se comporta de forma más desfavorable, mientras que el resto de técnicas de obtención de características optimizadas con SVM muestran en la mayoría de sus variantes unas tasas de acierto excelentes.

La técnica HOTA utilizando la transformada Welch es la que presenta mejores resultados. Este hecho es importante ya que en los estudios anteriores realizados con HOTA como por ejemplo en Sapena Baño [35]-[36] no se ha estudiado esta variante, siendo en este caso la que mejor ratio de aciertos ha obtenido independientemente del kernel utilizado. Lo cual representa una **aportación** de esta tesis, donde se han combinado el uso de la técnica HOTA aplicándole el periodograma de Welch como mejora de la resolución.

Con la única excepción de Cepstrum, las variantes de las técnicas en que se ha utilizado la transformada Welch han obtenido una tasa de aciertos igual o superior a las variantes de la misma técnica en donde se ha utilizado la transformada FFT.

Tanto en la tabla 5.6 como en la tabla 5.7, se observa que los resultados de las técnicas FFT y HOTA, en sus variantes sin aplicar una ventana de Hann o el periodograma de Welch, muestra una notable disminución en el ratio de aciertos respecto al resto de sus variantes. Este resultado con menor tasa de aciertos es debido a la perturbación provocada por la fuga espectral de la fundamental, pues tanto FFT como HOTA necesitan algún tipo de filtro o transformación (Hann o Welch) que disminuya esta perturbación.

En estos casos afectados por la fuga espectral (FFT sin Hann ni Welch y HOTA sin Hann ni Welch), cuando se trata el problema de la fuga espectral (leakage) con una ventana de Hann o transformación de Welch, el ratio de aciertos ha aumentado más del 10%, evidenciando que esta **fuga espectral de la frecuencia de fundamental** es un **problema serio que hay que evitar** al diagnosticar fallos en motores.

Tabla 5.6. Resultados obtenidos respecto del clasificador experto de tipo SVM configurado con un kernel polinomial.

SVM con kernel Polinomial	Variaciones			Mejor ajuste de parámetros		
	Ventana de Hann	Periodograma Welch	Nivel de exploración alcanzado < lvl >	Límites de caja < C >	Orden Polinomial < d >	Tasa de Aciertos (%)
FFT	No	No	1	1,000	4	86,11
	Si	No	1	32,000	6	99,53
	No	Si	3	2.435,496	3	98,61
	Si	Si	3	2.435,496	4	99,07
AC del módulo de SA con Hilbert	No	No	1	1,000	4	94,90
	Si	No	4	49,350	3	96,75
	No	Si	1	32,000	3	98,61
	Si	Si	0	1,000	3	96,75
HOTA estacionario	No	No	1	32,000	2	84,25
	Si	No	0	1,000	3	96,29
	No	Si	2	1.024,000	5	100,00
	Si	Si	4	117,376	3	100,00
Cepstrum	No	No	2	0,176	5	68,05
	Si	No	3	0,420	3	64,35
	No	Si	1	0,031	6	59,25
	Si	Si	4	0,003	6	60,64
EPVA	No	No	2	0,176	6	94,44
	Si	No	3	0,013	5	95,83
	No	Si	3	0,074	3	98,61
	Si	Si	3	0,074	3	98,61

Tabla 5.7. Resultados obtenidos respecto del clasificador experto de tipo SVM configurado con un kernel RBF Gaussiano.

SVM con kernel RBF Gaussiano	Variaciones		Mejor ajuste de parámetros			
	Ventana de Hann	Periodograma Welch	Nivel de exploración alcanzado <lvl>	Límites de caja <C>	Sigma (σ)	Tasa de Aciertos (%)
FFT	No	No	0	1,000	1,00	82,87
	Si	No	0	1,000	1,00	96,29
	No	Si	0	1,000	1,00	97,68
	Si	Si	1	32,000	1,00	97,22
AC del módulo de SA con Hilbert	No	No	1	32,000	1,00	94,44
	Si	No	1	32,000	1,00	95,83
	No	Si	0	1,000	1,00	95,83
	Si	Si	1	32,000	1,00	96,29
HOTA estacionario	No	No	0	1,000	1,00	83,33
	Si	No	1	32,000	1,00	95,83
	No	Si	2	32,000	0,17	100,00
	Si	Si	1	32,000	1,00	97,22
Cepstrum	No	No	1	32,000	1,00	65,27
	Si	No	0	1,000	1,00	59,72
	No	Si	0	1,000	1,00	29,62
	Si	Si	0	1,000	1,00	31,01
EPVA	No	No	4	430,538	1,00	91,66
	Si	No	2	5,656	1,00	95,83
	No	Si	0	1,000	1,00	95,83
	Si	Si	0	1,000	1,00	95,83

Aunque la técnica de HOTA ha sobresalido con un 100% de **tasa de aciertos**, las otras técnicas del análisis FFT, el módulo de la Señal Analítica y EPVA también han obtenido unas excelentes tasas de aciertos. En contraposición se encuentra la técnica Cepstrum que ha obtenido unos resultados de aciertos muy malos.

Comparando la tabla 5.6 (SVM con kernel polinomial) con la tabla 5.7 (SVM con kernel RBF), se observa una clara diferencia en el nivel de búsqueda que se desarrolla con un kernel y otro. Usando el kernel polinomial, el **nivel de búsqueda** del mejor ajuste de parámetros durante el entrenamiento varía entre 0 (coeficientes por defecto) hasta llegar en algunos casos al 4º nivel de profundidad en la búsqueda de ajuste de coeficientes. Sin embargo, en el caso del kernel RBF, el nivel máximo de búsqueda del mejor ajuste de coeficientes durante el entrenamiento rara vez se adentra hasta el nivel 4, siendo en la mayoría de casos lo adecuado tomar como coeficiente el inicial dado por defecto.

Esta diferencia en el nivel de búsqueda alcanzado entre un kernel y otro se puede explicar a partir del estudio de Hsu [91]. El kernel RBF es menos restrictivo que el kernel polinomial, permitiendo así una mayor deformación de espacio de características donde se ajusta el hiperplano. Al ser menos restrictivo el kernel RBF, el entrenamiento del sistema de diagnóstico tiende mucho más rápidamente al sobreentrenamiento, finalizándose dicho entrenamiento mucho antes ya que una de las condiciones de parada se da justo cuando se detecta el sobreentrenamiento del sistema experto.

Como se observa en el tiempo de entrenamiento del sistema SVM de la tabla 5.5, esta mayor versatilidad del kernel RBF hace que el aprendizaje de las diversas técnicas de diagnóstico finalice en un tiempo mucho menor al usar el kernel RBF en vez del kernel polinomial.

Por contrapartida, al tender más rápido a la solución el aprendizaje con kernel RBF es ligeramente menos preciso. Este fenómeno se puede apreciar en los ratios de acierto obtenidos para los dos kernels (tabla 5.8), donde el ratio de aciertos para el mismo método y misma variación con kernel RBF es ligeramente inferior al mismo método y misma variación con kernel polinomial.

Tabla 5.8. Resultados máximo, mínimo y promedio obtenidos del sistema experto de clasificación SVM con kernel polinomial y con kernel Gaussiano RBF.

Mejor ajuste de parámetros

		Nivel de exploración alcanzado $\langle lvl \rangle$	Límites de caja $\langle C \rangle$	Orden Polinomial $\langle d \rangle$	Tasa de Aciertos (%)
				Sigma (σ)	
SVM kernel Polinomial	Mínimo	0	0,013	2	59,25
	Máximo	4	2.435,496	6	100,00
	Promedio	2,1	308,134	4	89,53
SVM kernel RBF Gaussiano	Mínimo	0	1,000	0,17	29,62
	Máximo	4	430,538	1	100,00
	Promedio	0,75	36,90	0,96	84,30

Resumiendo, se obtienen tres casos óptimos con una tasa de aciertos del 100%:

- **HOTA estacionario con periodograma de Welch y sistema experto SVM con kernel polinomial.** 100% de aciertos con tiempo total de entrenamiento de 3532 segundos y un tiempo medio de clasificación de una muestra de 6073 milisegundos.
- **HOTA estacionario con ventana de Hann, periodograma de Welch y sistema experto SVM con kernel polinomial.** 100% de aciertos con tiempo total de entrenamiento de 10109 segundos y un tiempo medio de clasificación de una muestra de 6258 milisegundos.
- **HOTA estacionario con periodograma de Welch y sistema experto SVM con kernel de base radial Gaussiana.** 100% de aciertos con tiempo total de entrenamiento de 2911 segundos y un tiempo medio de clasificación de una muestra de 6073 milisegundos.

Aunque la técnica HOTA con variante de periodograma de Welch, ventana de Hann y sistema experto SVM con kernel polinomial muestra el sistema de diagnóstico óptimo (tabla 5.6), por contrapartida esta misma combinación requiere de un tiempo mayor para realizar el diagnóstico de nuevas muestras desconocidas (tabla 5.3).

5.7.2 Resultados obtenidos con sistemas expertos de Redes Neuronales Artificiales (ANN).

Como se ha comentado en el apartado 5.5.2.1. en el caso del algoritmo de aprendizaje de redes neuronales, este algoritmo de aprendizaje solo explora posibles soluciones con sistema expertos ANN de dos capas ocultas cuando el rendimiento del entrenamiento con una sola capa oculta es muy bajo. No obstante, para mostrar la diferencia de resultados al entrenar una o varias capas ocultas, se ha forzado el sistema de entrenamiento para explorar también **en cualquier caso** los sistemas expertos ANN con dos capas ocultas.

5.7.2.1 Comparación de tiempos entre ANN de 1 y de 2 capas ocultas.

En las siguientes tablas 5.9 (para una sola capa oculta) y 5.10 (para dos capas ocultas), se muestran los tiempos requeridos para el entrenamiento y los tiempos medios requeridos para la clasificación de las muestras. En la primera columna de estas tablas se muestran las técnicas desarrolladas en la tesis, en el resto de las columnas de las tablas se muestran:

- **Utilización de ventana de Hann.** Indica si la variación usa o no un filtro de ventana de Hann.
- **Utilización de periodograma de Welch.** Indica si la variación sustituye la función FFT por una función Welch.
- **Tiempo de cómputo del ANN - Entrenamiento < Te-1 >.** Muestra en segundos el tiempo total de entrenamiento para la obtención de los mejores parámetros de ajuste del clasificador con los que se obtiene la tasa de aciertos más alta.
- **Tiempo de cómputo del ANN - Clasificación < Tc-1 >.** Muestra en segundos la media del tiempo requerido para clasificar las características representativas obtenidas de una sola muestra.
- **Tiempo de cómputo total - Entrenamiento < Te-2 >.** Muestra en segundos el tiempo total de entrenamiento de cada sistema de diagnóstico generado en cada fila. Este tiempo es la combinación de los tiempos mostrados en la tabla 5.2 para generar todas características representativas de las muestras (TTC) más el tiempo necesario para desarrollar el sistema experto ANN óptimo respecto de las características representativas obtenidas (Te-1).
- **Tiempo de cómputo total - Clasificación < Tc-2 >.** Muestra en segundos el tiempo total necesario para realizar el diagnóstico mediante la extracción de las características representativas de fallo de una muestra (TMm) y realizando su clasificación (Tc-1).

Tabla 5.9. Resultados de los tiempos obtenidos del sistema experto de clasificación ANN de una sola capa oculta, y los tiempos totales de los sistemas de diagnóstico.

ANN con 1 capa oculta	Variaciones		Tiempo de cómputo del ANN		Tiempo de cómputo total	
	Ventana de Hann	Periodograma Welch	Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-2 > (s) (TTc+Te-1)	Clasificación < Tc-2 > (ms) (TMm+Tc-1)
FFT	No	No	4.632,56	9,80	6.767,42	3.640,01
	Si	No	2.469,10	9,90	4.609,51	3.823,92
	No	Si	2.870,46	9,82	5.244,88	4.716,13
	Si	Si	2.502,58	9,81	4.975,21	4.899,58
AC del módulo de la Señal Analítica	No	No	5.898,46	9,81	8.454,85	5.556,76
	Si	No	2.965,05	9,83	5.562,92	5.712,18
	No	Si	2.890,22	9,81	5.594,17	6.254,32
	Si	Si	2.896,78	9,82	5.637,72	6.405,67
HOTA estacionario	No	No	2.524,79	9,78	5.105,36	5.441,25
	Si	No	4.753,80	9,90	7.346,91	5.686,97
	No	Si	3.989,56	9,86	6.667,42	6.082,49
	Si	Si	3.599,93	9,80	6.329,36	6.266,66
Cepstrum	No	No	3.826,01	9,82	6.021,25	3.921,61
	Si	No	3.869,44	9,82	6.108,60	4.079,51
	No	Si	5.596,32	9,82	7.986,36	4.808,09
	Si	Si	2.456,41	9,81	4.881,06	4.959,44
EPVA	No	No	2.872,01	9,83	4.532,51	4.207,94
	Si	No	2.894,80	9,82	4.553,55	4.387,14
	No	Si	2.857,48	9,82	4.564,53	4.926,83
	Si	Si	2.876,05	9,83	4.595,32	5.083,18

Tabla 5.10 Resultados de los tiempos obtenidos del sistema experto de clasificación ANN de dos capas ocultas, y los tiempos totales de los sistemas de diagnóstico.

ANN con 2 capas ocultas	Variaciones		Tiempo de cómputo del ANN		Tiempo de cómputo total	
	Ventana de Hann	Periodo- grama Welch	Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-2 > (s) (TTc+Te-1)	Clasificación < Tc-2 > (ms) (TMm+Tc-1)
FFT	No	No	96.344,42	11,63	98.479,28	3.641,84
	Si	No	41.297,55	11,63	43.437,96	3.825,65
	No	Si	27.144,93	9,83	29.519,35	4.716,14
	Si	Si	28.092,66	11,63	30.565,29	4.901,40
AC del mó- dulo de la Señal Analí- tica	No	No	42.842,72	11,62	45.399,11	5.558,57
	Si	No	38.080,01	11,59	40.677,88	5.713,94
	No	Si	46.557,43	11,61	49.261,38	6.256,12
	Si	Si	33.453,03	9,81	36.193,97	6.405,66
HOTA estacionario	No	No	55.174,96	11,62	57.755,53	5.443,09
	Si	No	43.027,15	9,85	45.620,26	5.686,92
	No	Si	39.771,23	9,83	42.449,09	6.082,46
	Si	Si	39.051,83	9,86	41.781,26	6.266,72
Cepstrum	No	No	64.204,75	11,63	66.399,99	3.923,42
	Si	No	101.551,19	11,61	103.790,35	4.081,30
	No	Si	104.735,85	11,59	107.125,89	4.809,86
	Si	Si	48.257,78	11,57	50.682,43	4.961,20
EPVA	No	No	27.652,53	9,86	29.313,03	4.207,97
	Si	No	30.992,69	9,76	32.651,44	4.387,08
	No	Si	28.306,47	9,79	30.013,52	4.926,80
	Si	Si	43.761,34	11,54	45.480,61	5.084,89

Respecto de los costes de tiempos en clasificación en sistema experto ANN (Tc-1), aunque estos son del orden de 10 veces mayores a los obtenidos con un sistema experto SVM (tablas 5.3 y 5.4), estos siguen siendo lo bastante reducido (centésimas de segundo) como para ser considerados no relevantes. Por tanto al igual que en SVM, los tiempos realmente relevantes para reducir el tiempo de diagnóstico final del sistema de diagnóstico sigue siendo el tiempo necesario para la obtención de características representativas de una muestra (Tc-2).

Tabla 5.11. Resultados máximos, mínimos y promedio de los tiempos obtenidos del sistema experto de clasificación ANN para una sola capa oculta y para dos capas ocultas.

		Tiempo de cómputo del SVM		Tiempo de cómputo total	
		Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-2 > (s) (TTc+Te-1)	Clasificación < Tc-2 > (ms) (TMm+Tc-1)
ANN con 1 capa oculta	Mínimo	2.456,41	9,78	4.532,51	3.640,01
	Máximo	5.596,32	9,90	8.454,85	6.405,67
	Promedio	3.462,09	9,83	5.776,95	5.042,98
ANN con 2 capas ocultas	Mínimo	27.144,93	9,76	29.313,03	3.641,84
	Máximo	104.735,85	11,63	107.125,89	6.405,66
	Promedio	49.015,03	10,89	51.329,88	5.044,05

Comparando los costes de tiempos en entrenamiento (tabla 5.11) entre el entrenamiento con una capa oculta (tabla 5.9) y dos capas ocultas (tabla 5.10), el coste temporal con 2 capas ocultas crece del orden mayor a 10 veces. Donde es necesaria poco más de una hora para entrenar de forma óptima una red neuronal de una capa oculta, cuando se incrementa el número de capas ocultas a 2 en algunos casos el tiempo necesario de entrenamiento óptimo llega a superar incluso el día.

Comparando con los tiempos obtenidos para el entrenamiento completo (tabla 5.11) con el sistema experto SVM (tabla 5.5), se observa que el sistema experto ANN (promedio de 5776 segundos para una capa y de 51329 segundos para dos capas ocultas) requiere de un coste de tiempo en promedio superior para realizar el entrenamiento completo que cualquiera de las dos versiones del SVM (promedio de 5294 segundos para SVM con kernel polinomial y de 2363 segundos con kernel Gaussiano RBF).

El tiempo realmente importante en todo caso se corresponde con el **tiempo de cómputo total** (tiempo de la técnica de diagnóstico más el tiempo del sistema experto) para la clasificación de un elemento (**Tc-2**). Ya que este tiempo es una aproximación media del tiempo que tardaría el sistema de diagnóstico óptimo en procesar una muestra y obtener la respuesta.

Sin embargo, al igual que ocurría con el entrenamiento de los sistemas expertos SVM, la fase de entrenamiento (número de capas ocultas, ajuste de neuronas por capa, desarrollo de entropía cruzada y el entrenamiento local para valores aleatorios iniciales) no tiene un coste constante, sino que es relativo a la aproximación de la mejor solución de sistema experto ANN, por lo que ese tiempo de entrenamiento depende del coste de encontrar la configuración óptima del experto óptimo con la tasa de aciertos mayor.

Como se observa en la tabla 5.11, el tiempo de entrenamiento para 2 capas ocultas siempre va a ser superior a su equivalente de una sola capa oculta, llegando a multiplicarse por un factor de 10 o superior. Sin embargo, como se muestra al comparar las tablas 5.12 y 5.13, la mejora del ratio de acierto al pasar de una a dos capas cuando el ratio es superior al 90% es como máximo del 2%.

5.7.2.2 Comparación de resultados entre ANN de 1 y 2 capas ocultas.

En las siguientes tablas 5.12 para ANN de una sola capa oculta y 5.13 para ANN de dos capas ocultas, se muestran el número de neuronas óptimo por capa oculta y las mejores tasas de aciertos obtenidas con la combinación del sistema experto ANN y una variante de unos de las técnicas de obtención de características. En la primera columna se indica la técnica de obtención de características utilizado, mientras que el resto de las columnas representan a:

- **Utilización de ventana de Hann.** Indica si la variación utiliza o no un filtro de ventana de Hann.
- **Utilización de periodograma de Welch.** Indica si la variación reemplaza a la función FFT por una función Welch.
- **Intervalo de búsqueda de capa oculta.** Muestra el intervalo máximo del número de neuronas por capa oculta que ha sido explorado. Para cada capa oculta explorada de la red neuronal se muestra este intervalo.
- **Mejor ajuste del número de neuronas en capa oculta.** Se especifica en esta variable el número óptimo de neuronas en la o las capas ocultas que obtiene el mayor ratio de aciertos. Aunque la estructura de la red neuronal influye en su optimización también se ha de tener en cuenta otros parámetros como la selección de una buena función de transferencia y el entrenamiento correcto de la red sin sobreentrenamiento.

- **Tasa de aciertos.** Muestra la tasa de aciertos estimada por del sistema de diagnóstico óptimo formado por la técnica de obtención de características indicado y por el sistema experto ANN entrenado de forma óptima.

Tabla 5.12. Resultados obtenidos respecto del clasificador experto de tipo ANN con una sola capa oculta.

ANN con 1 capa oculta	Variaciones		Mejor ajuste		
	Ventana de Hann	Periodograma Welch	Intervalo de búsqueda en capa oculta	Numero de neuronas en capa oculta 1	Tasa de Aciertos (%)
FFT	No	No	[1 - 13]	7	87,04
	Si	No	[1 - 7]	1	97,69
	No	Si	[1 - 8]	2	98,15
	Si	Si	[1 - 7]	1	97,22
AC del módulo de SA con Hilbert	No	No	[1 - 16]	10	95,37
	Si	No	[1 - 8]	2	97,22
	No	Si	[1 - 8]	2	98,61
	Si	Si	[1 - 8]	2	98,61
HOTA estacionario	No	No	[1 - 7]	1	84,72
	Si	No	[1 - 13]	7	96,76
	No	Si	[1 - 11]	5	100,00
	Si	Si	[1 - 10]	4	100,00
Cepstrum	No	No	[1 - 11]	5	64,81
	Si	No	[1 - 11]	5	64,35
	No	Si	[1 - 16]	10	61,57
	Si	Si	[1 - 7]	1	54,17
EPVA	No	No	[1 - 8]	2	95,83
	Si	No	[1 - 8]	2	97,22
	No	Si	[1 - 8]	2	98,61
	Si	Si	[1 - 8]	2	98,61

Tabla 5.13. Resultados obtenidos respecto del clasificador experto de tipo ANN con dos capas ocultas.

ANN con 2 capas ocultas	Variaciones		Intervalo de búsqueda		Mejor ajuste de neuronas		Tasa de Aciertos (%)
	Ventana de Hann	Periodograma Welch	Búsqueda en capa oculta 1	Búsqueda en capa oculta 2	Neuronas en capa oculta 1	Neuronas en capa oculta 2	
FFT	No	No	[0 - 17]	[1 - 29]	11	23	89,35
	Si	No	[0 - 9]	[1 - 22]	3	16	99,54
	No	Si	[0 - 7]	[1 - 12]	0	5	98,15
	Si	Si	[0 - 7]	[1 - 13]	1	3	97,69
AC del módulo de SA con Hilbert	No	No	[0 - 9]	[1 - 16]	3	2	96,30
	Si	No	[0 - 9]	[1 - 14]	3	3	97,69
	No	Si	[0 - 12]	[1 - 13]	6	7	100,00
	Si	Si	[0 - 8]	[1 - 13]	0	2	98,61
HOTA estaciona- rio	No	No	[0 - 14]	[1 - 14]	8	5	87,04
	Si	No	[0 - 9]	[1 - 14]	0	7	96,76
	No	Si	[0 - 9]	[1 - 12]	0	5	100,00
	Si	Si	[0 - 9]	[1 - 13]	0	4	100,00
Cepstrum	No	No	[0 - 12]	[1 - 18]	6	7	68,98
	Si	No	[0 - 17]	[1 - 24]	11	3	69,44
	No	Si	[0 - 19]	[1 - 22]	13	10	68,98
	Si	Si	[0 - 8]	[1 - 22]	2	5	62,50
EPVA	No	No	[0 - 8]	[1 - 10]	0	2	95,83
	Si	No	[0 - 8]	[1 - 11]	0	2	97,22
	No	Si	[0 - 8]	[1 - 9]	0	2	98,61
	Si	Si	[0 - 13]	[1 - 11]	7	4	100,00

Como cabía de esperar, **la tasa de aciertos** cuando se utiliza un sistema ANN de dos capas ocultas (tabla 5.13) es igual o mayor a su respectiva variante con una sola capa oculta (tabla 5.12).

Se ha comentado en la investigación de Bishop [124] con anterioridad que, cuando se obtiene un ratio de aciertos con 1 capa oculta superior al 90%, el beneficio entre el ratio de aciertos y el coste de tiempo al aumentar a 2 capas ocultas es minúsculo (un importante incremento del tiempo de cálculo para obtener una pequeña mejora en la tasa de aciertos).

En esta investigación se ha realizado la comparación del ratio de aciertos y coste de tiempo necesario para entrenar un ANN de 1 y 2 capas ocultas, de forma que se ha estudiado esta afirmación expuesta por Bishop [124]. Conforme a los resultados obtenidos, se observa claramente que el tiempo para entrenar 2 capas ocultas es significativamente mayor al tiempo necesario para entrenar un ANN de 1 capa oculta (tabla 5.11), pero en los casos en que con 1 capa oculta se ha obtenido una tasa de fallos superior al 90%, la mejora en el ratio de aciertos al entrenar con 2 capas ocultas ha sido como máximo del 2.36% (tabla 5.14). Por tanto se ha validado tal afirmación, mostrando que es poco significativa y no compensa el tiempo requerido para su entrenamiento.

Tabla 5.14. Resultados máximo, mínimo y promedio obtenidos del sistema experto de clasificación ANN para una sola capa oculta y para dos capas ocultas.

Mejor ajuste de parámetros

		Búsqueda en capa oculta 1	Búsqueda en capa oculta 2	Neuronas en capa oculta 1	Neuronas en capa oculta 2	Tasa de Aciertos (%)
ANN con 1 capa oculta	Mínimo	[1 - 7]	-	1	-	54,17
	Máximo	[1 - 16]	-	10	-	100,00
	Promedio	-	-	3,65	-	89,32
ANN con 2 capas ocultas	Mínimo	[0 - 7]	[1 - 9]	0	2	62,50
	Máximo	[0 - 19]	[1 - 29]	13	23	100,00
	Promedio	-	-	3,7	5,85	91,13

Las tasas de aciertos resultantes obtenidas con el uso de un sistema experto ANN (tablas 5.12 y 5.13) se muestra muy similar al obtenido con el uso de un sistema experto SVM (tablas 5.6 y 5.7).

Que estas tasas de aciertos sean tan parecidas al utilizar sistemas expertos tan diferentes indica que el ajuste en ambos casos se encuentra cerca del óptimo sin llegar a darse un sobreentrenamiento. Ya que en caso contrario, los ratios de acierto respecto de una misma técnica y distinto sistema experto serían mucho más dispares, dado al falso aumento de la tasa de fallos potenciada por el sobreentrenamiento.

Al igual que ocurría en el caso de SVM, se observa claramente como peor caso a la técnica Cepstrum en cualquiera de sus variantes, mientras que el resto de técnicas de obtención de características optimizados con ANN muestran en la mayoría de sus variantes unas tasas de acierto bastante elevadas.

Descartando los resultados obtenidos por Cepstrum, en las tablas 5.12 y 5.13 los resultados obtenidos con las técnicas FFT y HOTA sin el uso de filtros ni transformadas obtienen una tasa de aciertos menor, dada la **fuga espectral** (leakage) de la frecuencia fundamental que dificulta la clasificación correcta de las características representativas de fallo.

Como en los resultados obtenidos en SVM, la técnica HOTA con transformada Welch destaca sobre el resto, incluso en los casos en que se utiliza una sola capa oculta en la red neuronal.

Aunque el entrenamiento de las redes neuronales de dos capas ocultas muestra un tiempo excesivo, también son las que mejoran en conjunto el ratio de aciertos respecto las redes neuronales de una sola capa oculta y respecto del sistema experto SVM para ambos tipos de kernel. En este caso, el número de variantes que alcanza el 100% de aciertos son los siguientes:

- **Componente alterna del módulo de la señal analítica con ventana de Hann, periodograma de Welch y sistema experto ANN con dos capas ocultas.** 100% de aciertos con tiempo total de entrenamiento de 36193 segundos y un tiempo medio de clasificación de una muestra de 6405 milisegundos.
- **HOTA estacionario con periodograma de Welch y sistema experto ANN con una capa oculta.** 100% de aciertos con tiempo total de entrenamiento de 6667 segundos y un tiempo medio de clasificación de una muestra de 6082 milisegundos.
- **HOTA estacionario con periodograma de Welch y sistema experto ANN con dos capas ocultas.** 100% de aciertos con tiempo total de entrenamiento de 42449 segundos y un tiempo medio de clasificación de una muestra de 6082 milisegundos.

- **HOTA estacionario con ventana de Hann, periodograma de Welch y sistema experto ANN con una capa oculta.** 100% de aciertos con tiempo total de entrenamiento de 6329 segundos y un tiempo medio de clasificación de una muestra de 6266 milisegundos.
- **HOTA estacionario con ventana de Hann, periodograma de Welch y sistema experto ANN con dos capas ocultas.** 100% de aciertos con tiempo total de entrenamiento de 41781 segundos y un tiempo medio de clasificación de una muestra de 6266 milisegundos.
- **EPVA con ventana de Hann, periodograma de Welch y sistema experto ANN con dos capas ocultas.** 100% de aciertos con tiempo total de entrenamiento de 45480 segundos y un tiempo medio de clasificación de una muestra de 5084 milisegundos.

También se ha de recalcar que, independientemente del sistema experto, todas las variantes que han llegado al 100% de aciertos reemplazan la transformada FFT por el espectrograma de la transformada Welch.

De los resultados obtenidos se deduce que, en el caso de que el fin último del sistema de diagnóstico sea el diagnóstico con la mayor probabilidad de aciertos independientemente del tiempo requerido, la técnica HOTA con su variante de transformada Welch es una apuesta segura.

Si además de un excelente ratio de aciertos es necesario realizar el diagnóstico de nuevas muestras en el menor tiempo posible en este caso, según los datos obtenidos, una buena alternativa como sistema de diagnóstico se da con la técnica FFT aplicando un filtrado de Hann (para evitar la fuga espectral) y un sistema experto SVM con kernel polinomial o un sistema experto ANN de dos capas ocultas. En este caso se obtiene de promedio un tiempo de diagnóstico de 3.8 segundos manteniendo una tasa de aciertos estimada, tanto en SVM con kernel polinomial como con ANN con dos capas ocultas, superior al 99%.

5.8 Conclusiones del capítulo

Como se ha podido observar de las tasas de acierto obtenidas en el sistema experto basado en SVM (tabla 5.6 y tabla 5.7 del apartado 5.7.1.2) y en el sistema experto basado en ANN (tabla 5.12 y tabla 5.13 del apartado 5.7.2.2), a excepción de Cepstrum los resultados obtenidos han sido bastante altos llegando a ser en el caso de la técnica de diagnóstico HOTA excelentes (100% de aciertos) para ambos sistemas expertos.

Considerando que las tasas de acierto, conseguidas con dos sistemas expertos tan diferentes entre sí, son tan similares; se puede decir que la implementación de los algoritmos automáticos que se han desarrollado para los ajustes de parámetros, de estructuras, de coeficientes, algoritmos de entrenamiento, validación de los sistemas expertos y sistemas de detección y anulación del sobreentrenamiento, desarrollados respectivamente para el sistema experto SVM (apartado 5.3) y para el sistema experto ANN (apartado 5.4) son totalmente correctos y válidos.

Además, la obtención de tasas de acierto tan altas con la mayoría de técnicas de diagnóstico sugiere que el proceso de obtener sistemas expertos optimizados adaptados a las características obtenidas dentro del sistema de generación automático de sistemas óptimos de diagnóstico de fallos funciona perfectamente.

También se han podido comprobar los costes relativos de tiempo que requiere cada una de las técnicas de diagnóstico y de los sistemas expertos para realizar su función. Y, aunque ambos sistemas expertos obtienen unas tasas de acierto muy parecidas, es notable el tiempo de cálculo extra que requieren los sistemas expertos de tipo ANN en comparación con los sistemas expertos de tipo SVM.

Independientemente del sistema experto y técnica de diagnóstico usado, también se ha podido comprobar como el espectrograma de Welch mejora ligeramente la detección de fallos respecto al uso de **su homónimo** la transformada rápida de Fourier (FFT).

Igualmente, ha sido interesante comprobar cómo, aunque la técnica de diagnóstico FFT es de las técnicas más utilizadas en diagnóstico (entre las consideradas en la tesis), con el filtro (Ventana de Hann) o transformación del dominio (periodograma de Welch) se obtienen unos resultados notables que no tienen nada que envidiar a los resultados obtenidos por técnicas más nuevas.

Capítulo 6

Generador de detectores de fallo en máquinas de inducción en régimen transitorio

6.1 Introducción

En el caso de motores de inducción operando en régimen transitorio se sigue un desarrollo similar al correspondiente en régimen estacionario para la creación del generador de diagnóstico de fallos a partir de la corriente de fase. Al igual que en el capítulo 5, este sistema de diagnóstico de fallos está formado por una técnica de obtención de las características relevantes para la detección de fallo (apartado 2.3) y un sistema inteligente entrenado para clasificar y discernir a partir de esas características (Capítulo 3).

Como se puede observar comparando las técnicas de obtención de características del apartado 2.3.1 para régimen estacionario y del apartado 2.3.2 para régimen transitorio, existe una notable diferencia en la complejidad de estas y de sus resultados.

En el caso de diagnóstico para régimen estacionario, observando las técnicas que aparecen en el estado del arte del capítulo 2, se ve claramente en la mayoría de estas técnicas tienen los siguientes aspectos comunes:

- La técnica FFT, la técnica de componente alterna del módulo de la señal analítica, la técnica HOTA, la técnica Cepstrum y la técnica del vector extendido de Park desarrollan transformaciones sobre las corrientes de fase para llegar a un dominio de datos de un único vector de información de donde se extraen las características relevantes.

- El número de características obtenido en todos los casos es bastante reducido.
- En los casos claros de fallo es fácil detectar las características de fallo, incluso puede ser detectado por un operador humano observando la representación del vector de información.
- Las técnicas de detección de fallos utilizadas en el capítulo 5 para régimen estacionario no requieren de coeficientes de ajuste para mejorar la detección de las características representativas de los fallos mientras que en el régimen transitorio sí, como se verá más adelante.

En contraposición a las técnicas y métodos en régimen de trabajo estacionario, las técnicas desarrolladas para la obtención de características relevantes para el diagnóstico en régimen transitorio muestran la característica común de desarrollar un espacio tiempo-frecuencia de donde se obtienen las características representativas de fallo.

Como la característica significativa en régimen transitorio está formada por una figura bidimensional compleja (por ejemplo una recta), la detección del fallo es más robusto y preciso, ya que es más difícil que una perturbación pueda enmascarar completamente la figura correspondiente al fallo.

Las características que comparten mayoritariamente las técnicas de detección de fallos en régimen transitorio son:

- Todas ellas desarrollan transformaciones sobre las corrientes de fase para llegar a conjuntos de información más complejos que en régimen estacionario; bien a múltiples vectores de información (DWT, Hilbert-Huang) o a un único conjunto de datos de dos dimensiones como mínimo (CWT, CCWT, Gabor, HOTA transitorio, Wigner-Ville).
- Al ser el conjunto de información más complejo también lo es la precisión en la detección de fallos, ya que es más fácil reconocer el patrón.
- En algunas técnicas es necesario ajustar parámetros para mejorar la detección de las características de fallos. Por ejemplo, en la técnica DWT se puede variar la ondícula madre para mejorar la resolución de la característica de fallo, mientras que en la transformada STFT se debe ajustar lo mejor posible la dimensión del filtro de ventana para resaltar lo máximo posible las características de fallos.

Esta mejora en la precisión de las técnicas de diagnóstico para régimen transitorio, las hace preferentes sobre las técnicas de diagnóstico para régimen estacionario. Además,

solo se usan técnicas de diagnóstico en régimen transitorio a nivel industrial, ya que es muy difícil encontrar motores que siempre estén trabajando en régimen estacionario perfecto.

Actualmente el campo de investigación sobre el diagnóstico de fallos para motores funcionando en régimen transitorio ha avanzado más, es el más utilizado y en el que más artículos de investigación se desarrollan en comparación con las técnicas orientadas para régimen estacionario.

Por tanto, el objetivo de este capítulo ha sido el de desarrollar un sistema autónomo de generación de sistemas de diagnóstico óptimos de motores en régimen de trabajo transitorio.

Como se ha visto en el apartado 2.3.2, existe un amplio abanico de técnicas de diagnóstico en régimen transitorio. En este caso para régimen transitorio, a diferencia del capítulo 5, se ha optado por seleccionar e implementar una única técnica de diagnóstico que permita obtener unas características de fallo lo bastante representativas a la par que simples para no complicar demasiado el diseño del sistema de generación de sistemas de diagnóstico. Se deja para un futuro estudio implementar un mayor número de técnicas de diagnóstico de fallos.

Además, como se ha podido observar en los resultados obtenidos en el capítulo 5, el algoritmo HOTA para régimen de trabajo estacionario ha obtenido el mejor ratio de aciertos. En base a este criterio y otros que se exponen en el siguiente apartado de este capítulo, esta técnica HOTA de análisis en régimen transitorio (apartado 2.3.2.6) ha sido seleccionada, en este trabajo.

Durante la implementación del sistema se ha comprobado que esta técnica de obtención de características en régimen transitorio requiere de grandes recursos tanto de cálculo como de memoria, por tanto, se ha modificado el algoritmo original propuesto en Sapeña Bañó [35] para obtener unas características representativas de fallo cercanas a las originales del algoritmo HOTA, pero con un coste mucho menor de los recursos demandados. Esto constituye una contribución original en esta tesis.

Con la técnica de generación de características completado, se ha configurado de nuevo el uso de los sistemas expertos vistos en el capítulo 5 (SVM y ANN). En todo caso, la metodología de desarrollo de los sistemas expertos desarrollada en el apartado 5.4 (SVM) y 5.5 (ANN) es totalmente compatible con las nuevas características obtenidas mediante HOTA transitorio.

6.2 Obtención de características relevantes en régimen transitorio.

En las siguientes subsecciones se desarrollan optimizaciones de la técnica HOTA para régimen de trabajo transitorio. **Estas son aportaciones novedosas de esta tesis al campo del diagnóstico de fallos.**

Teniendo en consideración las técnicas de obtención de características representativas de fallo para régimen transitorio (apartado 2.3.2), en vez de aplicar diversas técnicas, se ha optado por seleccionar y optimizar aquella que obtiene unos buenos resultados de tasa de aciertos en diversas situaciones de transición dados por la variación de la velocidad, debido a la actuación del regulador de velocidad o a la variación de la carga. De esta forma se implementa y se optimiza un buen sistema de diagnóstico para el trabajo del motor en régimen transitorio.

6.2.1 Análisis del Orden Armónico para régimen transitorio (HOTA transitorio).

Como se ha descrito en el apartado 6.1, todas las técnicas de obtención de características de fallo para régimen transitorio tienen sus ventajas y desventajas.

Sin embargo, al analizar los resultados obtenidos de cada una de las técnicas para régimen transitorio y que hay que interpretar para obtener la característica representativa de fallo es de destacar la información que proporciona la técnica HOTA [35]. Como se desarrolla en el apartado 2.3.2.6, el vector de datos resultante obtenido (denominado resultado comprimido de HOTA transitorio) es un vector de amplitudes respecto de un dominio del tipo *k-orden armónico*, donde cada armónico de fallo se muestra en su respectiva posición dentro de este dominio de *k-orden armónico*.

Por tanto, esta técnica ofrece una mayor simplificación a la hora de obtener posibles características de fallo así como una mayor significancia en dichas características. Como prueba de la claridad de esta técnica, al igual que ocurría en la mayoría de técnicas para régimen estacionario, en los casos en que el fallo es muy evidente este puede ser diagnosticado incluso por un operario humano analizando los datos de una forma totalmente objetiva.

Como **características representativas de fallo**, ya que la propia técnica en estacionario como en transitorio ofrecen el mismo tipo de características representativas, se ha optado por utilizar las mismas características representativas de fallo que se han seleccionado en el caso de HOTA para régimen estacionario del apartado 5.2.3. Es decir, se ha utilizado como características para los sistemas expertos las características las amplitudes obtenidas de las dos frecuencias de fallo para la posición de *k-orden armónico* para $k=\pm 1$.

Además, gracias a que se utiliza el mismo tipo de características es posible usar sin necesidad de un ajuste previo los sistemas expertos de tipo SVM y ANN desarrollados en el apartado 5.4 y 5.5.

No obstante, a diferencia del HOTA para régimen estacionario, la técnica HOTA para régimen transitorio utiliza de forma interna la técnica STFT (apartado 2.3.2.1). Esta técnica STFT requiere de varias selecciones de configuración para optimizar la obtención de las características representativas de fallo respecto de cada muestra procesada.

En este caso, la mayor definición de ancho de banda que muestre dicho espacio tanto en la dimensión de tiempo como en la dimensión de frecuencias depende del tipo y ajustes de parámetros de la ventana de filtro usada. Por tanto, para poder extraer finalmente unas buenas características de fallo significativas del espacio tiempo frecuencia generado es necesario tener en cuenta los siguientes puntos sobre la ventana de filtro:

- **Hay que utilizar una ventana que obtenga la mayor concentración de energía posible en el espacio tiempo frecuencia.**

De este modo se intenta maximizar la diferenciación de las amplitudes de las características representativas de fallo del resto de frecuencias. En el apartado 2.3.2.1 para STFT se muestra como la ventana Gaussiana es óptima en concentrar la mayor energía posible cuando la longitud de la señal es infinita.

Por ello, como se comenta en el apartado 2.3.2.1, esta ventana es utilizada en la mayoría de investigaciones sobre diagnóstico de fallos basadas en la técnica STFT.

- **Es necesario ajustar los parámetros de la función de ventana para maximizar la resolución frecuencial de las componentes características de fallo.**

En todo caso no es posible mejorar la resolución en una dimensión del filtro de ventana sin llegar a afectar a la resolución en la otra dimensión. En la mayoría de casos se intenta ajustar el mejor balance entre la resolución frecuencial y los periodos de tiempo que reflejen mejor las características deseadas.

Hay que usar un sistema de ajuste que encuentre la configuración óptima del filtro de ventana de cara a maximizar la resolución de las componentes de fallo dentro de este espacio tiempo frecuencia.

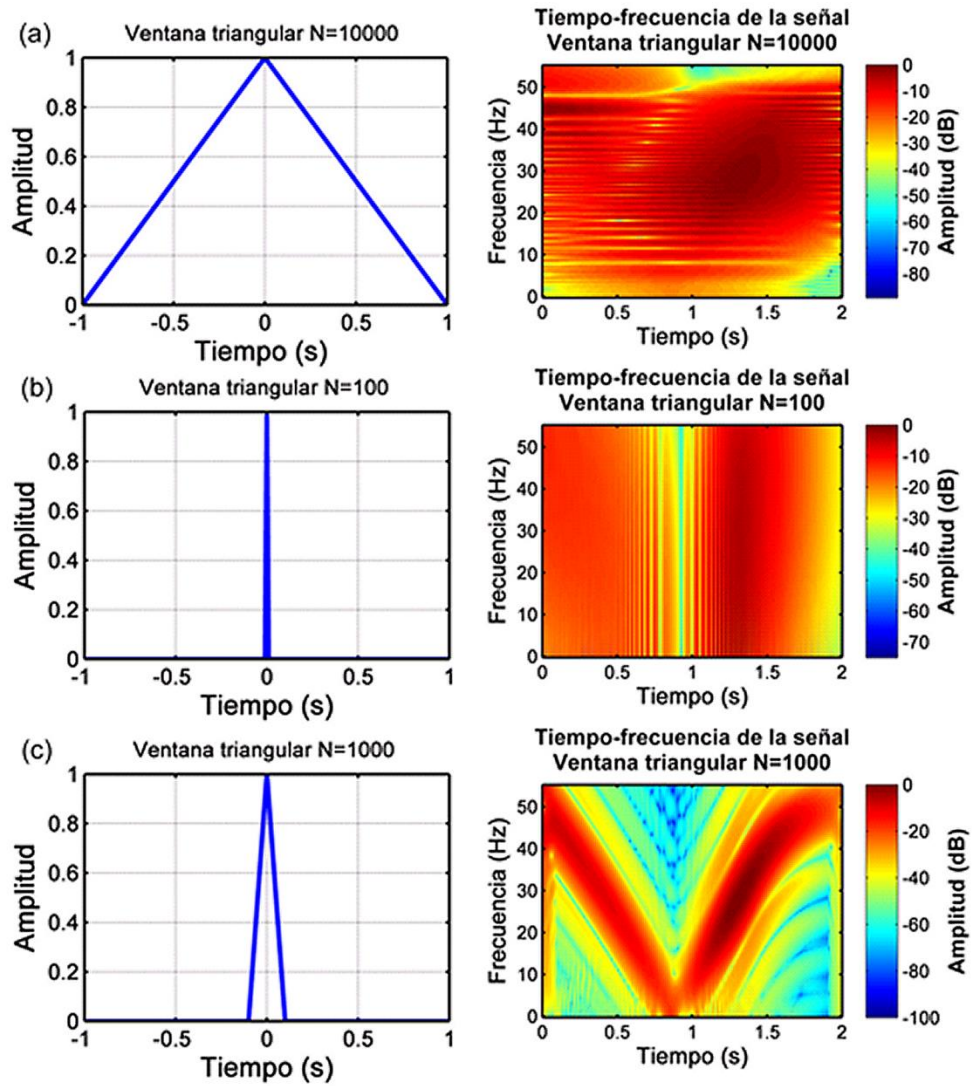


Figura 6.1. Ejemplo de la generación de un espacio tiempo/frecuencia para una señal de arranque de un motor utilizando una ventana triangular con las siguientes proporciones: a) longitud máxima en tiempo, b) longitud mínima en tiempo y c) ventana ajustada a la rampa de fallo del espacio tiempo/frecuencia.

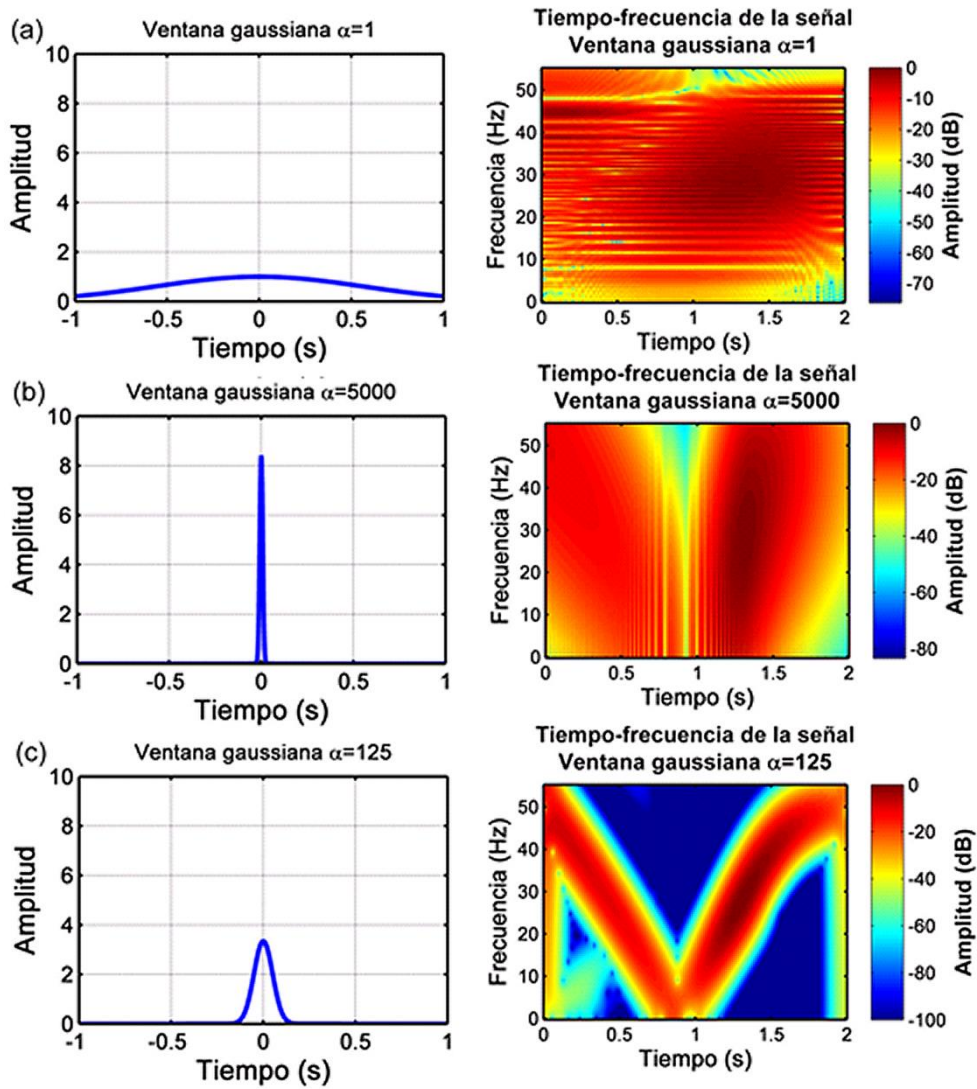


Figura 6.2. Ejemplo de la generación de un espacio tiempo/frecuencia para una señal de arranque de un motor utilizando una ventana Gaussiana con las siguientes proporciones: a) longitud máxima en tiempo, b) longitud mínima en tiempo y c) ventana ajustada a la rampa de fallo del espacio tiempo/frecuencia.

Como ejemplo de estos dos puntos anteriores, se muestra en la Figura 6.1 un ejemplo de generar un espacio tiempo/frecuencia (de señal de arranque de un motor con fallo de barra rota) utilizando una ventana triangular con diversas proporciones. En el caso a) de la imagen se utiliza una ventana triangular de longitud máxima en tiempo (mínima en frecuencia), en el caso b) de la imagen se utiliza una ventana de longitud mínima en tiempo (máxima en frecuencia) y en el caso c) se utiliza una ventana triangular ajustada a la rampa de fallo mostrada en el espacio tiempo/frecuencia. Como se ve claramente, el ajuste de la proporción de la ventana de filtro es esencial para obtener un buen vector de amplitudes correspondientes al fallo dentro del espacio tiempo/frecuencia (características representativas de fallo).

Como alternativa a la ventana triangular de la Figura 6.1, en la Figura 6.2 se muestra otro ejemplo de generación de un espacio tiempo/frecuencia (de señal de arranque de un motor con fallo de barra rota) pero en este caso se ha utilizado una ventana Gaussiana igualmente con diversas proporciones de ventana. En el caso a) se utiliza una ventana Gaussiana de longitud máxima en tiempo (mínima en frecuencia), en el caso b) se utiliza una ventana de longitud mínima en tiempo (máxima en frecuencia) y en el caso c) se utiliza una ventana ajustada a la rampa de fallo de barra rota representada en el espacio tiempo/frecuencia.

En este caso de ventana Gaussiana (Figura 6.2), el ajuste de la proporción de la ventana de filtro mejora la detección de la característica de fallo, además, comparando la ventana Gaussiana con la ventana Triangular de la imagen 6.1 se observa como la ventana Gaussiana define mejor el vector de amplitudes de fallo cuando las dos ventanas tienen la proporción correcta ((c) en ambas imágenes) para maximizar la detección del vector de amplitudes del fallo en forma de “V”.

Por tanto, en esta tesis se ha utilizado la ventana Gaussiana como ventana por defecto para utilizar en la técnica STFT que implementa internamente HOTA.

En el algoritmo de ejecución de HOTA transitorio mostrado en Figura 6.3, idéntico al apartado 2.3.2.6, se observa como la función que se encarga de aplicar la ventana de filtro (técnica STFT) sobre el vector de corrientes para generar un vector de frecuencias respecto de un incremento de tiempo concreto (marcada en rojo), es una función crítica que debe desarrollar HOTA necesaria para poder obtener finalmente el resultado simplificado de k-orden armónico, o el resultado reducido, de donde obtener las características representativas de fallo.

No obstante y como se ha comentado anteriormente, al igual que el resto de técnicas para régimen transitorio basados en espacios tiempo/frecuencia, la técnica de análisis STFT (de la que hace uso HOTA) consume muchos recursos tanto de tiempo de cálculo como de almacenamiento de memoria.

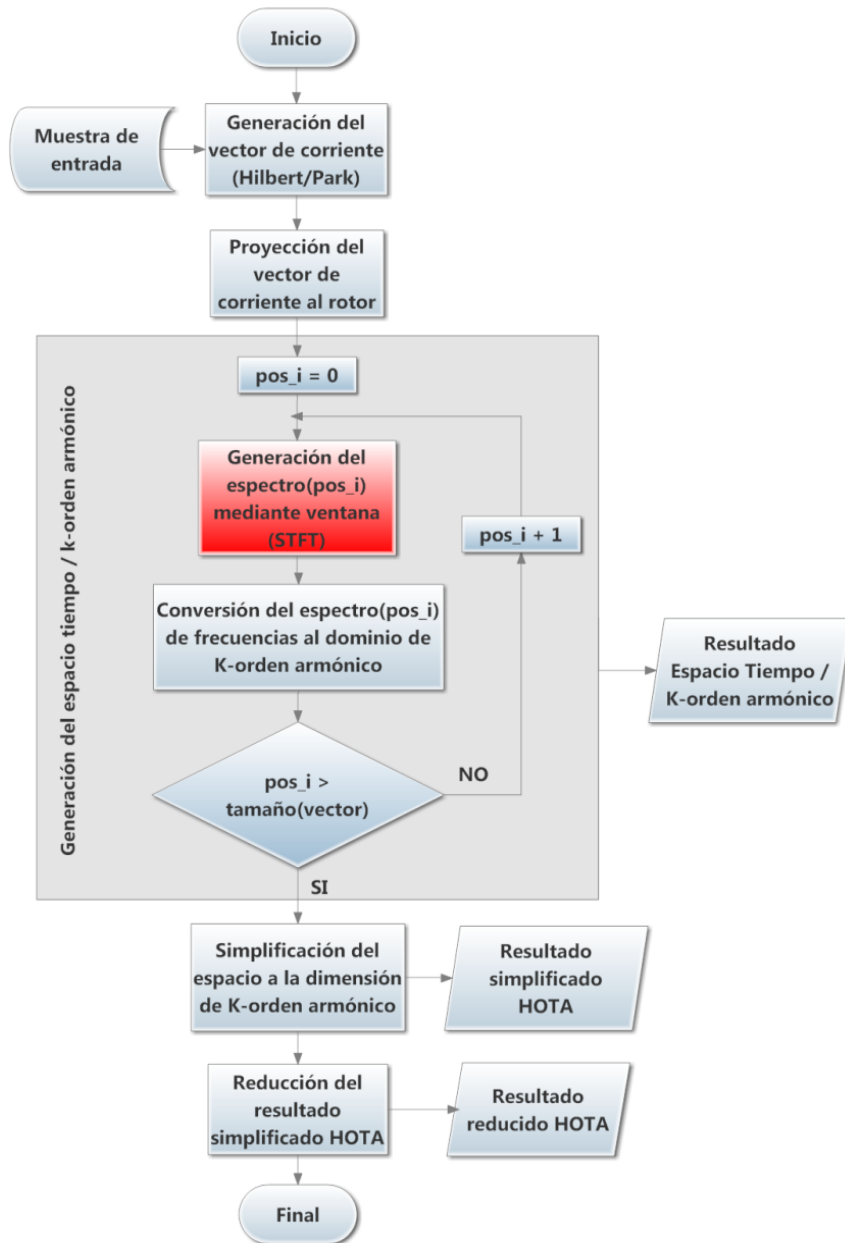


Figura 6.3. Esquema de flujo del funcionamiento de la técnica HOTA para régimen de trabajo transitorio.

En las pruebas desarrolladas de forma experimental con las muestras obtenidas del banco de ensayo (apartado 4.4) se han observado varios problemas tanto por el coste de cálculo como el coste de almacenamiento de memoria. Estos problemas se han dado al implementar la técnica STFT descrita en el estado del arte (apartado 2.3.2.1) dentro de HOTA transitorio, utilizando como hardware el ordenador personal destinado a la ejecución del sistema (especificado en el apartado 5.6).

- **Tiempo excesivo de cálculo.** Se ha observado que la STFT dentro de la técnica de diagnóstico HOTA requiere de un tiempo de cálculo para obtener el espacio tiempo frecuencia demasiado elevado.

Por ejemplo, para generar un espacio tiempo frecuencia para una proporción específica de ventana Gaussiana, se ha contabilizado en pruebas una media aproximada para generar el espacio tiempo frecuencia de 9 minutos.

La aplicación del algoritmo de ajuste de la ventana de filtro mediante la búsqueda del espacio de mínima entropía (apartado 2.3.2.1.1), con el objetivo de obtener el espacio de mayor energía donde extraer las características representativas de fallo, requiere de varias decenas de procesos de generación de espacios de tiempo frecuencia variando la proporción de la ventana para obtener la curva de entropías. Por tanto, si se desarrollan en teoría 20 espacios tiempo/frecuencia (un número bastante conservador para hacer una curva entrópica) el tiempo total para obtener la característica de fallo para una única corriente de fase de una muestra sería aproximadamente de 3 horas.

Y por tanto, si se tiene en cuenta el total de muestras solo para régimen de trabajo transitorio obtenidas en el banco de ensayos (apartado 4.4.2). Existen 173 muestras de motor, donde cada una tiene tres muestras de corrientes de fase que procesar (735 muestras corrientes de fase). Considerando la media hipotética de tres horas por muestra de corriente de fase, sería necesario esperar alrededor de **tres meses** para obtener las características representativas del conjunto de fallo solo para proseguir con el entrenamiento de los sistemas expertos (apartados 5.4 y 5.5).

- **Almacenamiento insuficiente de memoria.** Se ha tenido que utilizar la versión del entorno de desarrollo de MATLAB R2014a de 64 bits para ejecutar la STFT dentro de la técnica HOTA transitoria, ya que el límite de 4GB de memoria RAM para programas de 32 bits no ha sido suficiente para poder procesar una muestra con esta técnica.

Además, aun disponiendo de 12GB de RAM, el sistema ha requerido puntualmente un número mayor de capacidad de almacenamiento, obligando al sistema operativo a utilizar memoria virtual, ralentizando todavía más el proceso de ejecución de la técnica HOTA para una muestra.

En consecuencia, se han planteado e implementado diversas novedades sobre el algoritmo HOTA transitorio para aumentar la precisión del diagnóstico y disminuir al mismo tiempo los requisitos de tiempo de cálculo y de almacenamiento en memoria.

6.2.1.1 Etapa de filtrado paso banda [Mejora en el tiempo de cálculo y almacenamiento de memoria].

Como innovación se ha implementado una etapa de filtrado que se aplica sobre las corrientes estáticas proyectadas en el rotor. Al reducir la muestra de corriente al rango de frecuencias “interesantes” para la obtención de fallos, se está minimizando el tiempo de cálculo y el almacenamiento de memoria necesario.

Un filtro paso bajo es aquel que deja pasar las señales con una frecuencia inferior a cierta frecuencia de corte y atenúa o elimina las señales con frecuencias superiores a dicha frecuencia de corte. El nivel de atenuación de cada señal depende del propio diseño del filtro. El caso contrario consiste en el uso de filtros paso-alto, es decir, permiten el paso de frecuencias superiores a la frecuencia de corte y atenúan las señales con frecuencias inferiores. La combinación de ambos filtros, da lugar a lo que se denomina como filtros paso-banda, que permiten estudiar las frecuencias de la señal en un ancho de banda concreto. Existen numerosos tipos de filtros que incluyen circuitos electrónicos, filtros anti-aliasing, filtros digitales, barreras acústicas, etc.

Los filtros paso-bajo, paso-alto y paso-banda no solo se implementan de manera física (circuitos electrónicos, etc.) sino que se pueden implementar en una etapa posterior en el propio procesado de la señal utilizando alguna transformada matemática precisando menores requerimientos de hardware y permitiendo mayores ajustes de la frecuencia de corte. En la literatura científica aplicada al campo del diagnóstico de máquinas eléctricas se han utilizado numerosas transformadas matemáticas como filtros, no obstante, entre todas ellas prevalecen las transformadas discretas de wavelet [146] [147]. Estas, se basan en descomponer la señal en una serie de bandas utilizando una serie de filtros paso banda. No obstante, tal como se ha visto en el capítulo del estado del arte, este banco de filtrado tiene una serie de limitaciones. Por un lado, el filtrado no es ideal, es decir, existe una cierta superposición entre las bandas de descomposición lo que puede dar lugar a interferencias en la descomposición. Por otro lado, la descomposición en bandas de frecuencia sigue una progresión diádica, de modo, que el ancho de cada banda queda determinado por la frecuencia de muestreo y por el tipo de función wavelet madre elegido. Expresado en otras palabras, el usuario no puede especificar la frecuencia de corte deseada, sino que esta dependerá específicamente de la frecuencia de muestreo utilizada para capturar la señal. Para tratar de dar solución a este problema, permitiendo la descomposición en anchos de banda especificados por el usuario se ha propuesto el uso de la transformada discreta “packet” de wavelet [148]. No obstante, este filtrado, aunque permite la descomposición en anchos de banda variables, sigue

teniendo el inconveniente de que el filtrado no es ideal, es decir, sigue existiendo superposición entre las bandas de descomposición que dan lugar a interferencias.

Con el objetivo de dar solución a ambos problemas, recientemente se ha propuesto el uso de la transformada discreta armónica wavelet (DHWT) [149] que permite, por un lado, una descomposición en anchos de banda arbitrarios, determinados por el usuario, y, por otro lado, un filtrado ideal, es decir, no existe superposición entre bandas eliminando los indeseables efectos de interferencias entre bandas. Por otro lado, su algoritmo de implementación basado en la FFT es rápido y consume menores recursos que cualquier otro tipo de filtrado citado anteriormente.

Por tanto, con el objetivo de mejorar la etapa en la que se realiza la STFT en HOTA, se ha **procedido a realizar un filtrado utilizando la DHWT eliminando las frecuencias que no son de interés** para el diagnóstico de la máquina y, por ende, reduciendo el volumen de datos a tratar en las etapas posteriores. Dicho filtrado se debe ajustar dependiendo del tipo de fallo a detectar.

Tal como se ha visto en el capítulo del estado del arte, cada tipo de fallo genera o amplifica una serie de componentes armónicas cuya frecuencia se ha determinado de manera teórica. Por ejemplo, en el caso de rotura de barras se ha observado que, al convertir la corriente estática en corriente rotórica, las componentes de fallo aparecen cercanas a la componente fundamental (la cual se encuentra igualmente cerca de los 0 Hz). Por tanto, con un estudio entre -50 y 50Hz sería más que suficiente para detectar las componentes debidas a este tipo de fallo. Por tanto, en el trabajo presentado en esta tesis que está siendo validado para la detección de asimetrías rotóricas se establecerá una banda de filtrado entre -50 y 50 Hz, es decir, se emplea un filtro paso banda ideal con frecuencia de corte de -50 y 50 Hz.

Respecto de las muestras experimentales utilizadas en esta tesis, al pasar de un ancho de banda de 100kHz a un ancho de banda de 100Hz la diferencia tanto en coste de tiempo de cálculo como de almacenamiento de memoria ha sido bastante significativa.

6.2.1.2 Implementación de la función elemental de onda esferoidal Prolate como nueva ventana de filtro (PSWF) [Mejora en la precisión del diagnóstico y en el tiempo de cálculo].

Como se ha analizado en el estado del arte para sistemas de diagnóstico (apartado 2.3.2.1) para la técnica de diagnóstico STFT, el filtro de ventana Gaussiana es capaz teóricamente de obtener la mayor concentración de energía (periodos mínimos de tiempo y un ancho de banda frecuencial pequeño) para señales de duración infinita. Sin embargo, en la realidad cualquier tipo de señal tiene una duración determinada, por lo que nunca se alcanzaría por completo la mayor concentración de energía con una ventana Gaussiana.

Adentrándose en la literatura científica de otros campos de estudio donde se utiliza la técnica STFT, se propone como alternativa de ventana para la STFT (Short Time Fourier Transform) una función elemental de onda esferoidal Prolate (PSWF) [150]. Esta ventana de filtro es usada con éxito en diversos campos de estudio, como por ejemplo el tratamiento de imágenes médicas, sistemas de transmisión inalámbrica o tratamiento acústico.

A diferencia de la función elemental Gaussiana, el estudio de Landau y Pollak [151] demuestra teóricamente que la función elemental de onda esferoidal Prolate es capaz de obtener la mayor concentración de energía para una señal de duración delimitada en tiempo T y frecuencia B , y por tanto, en la detección de fallos, la definición de las características representativas de fallo teóricamente sería mayor respecto al uso del filtro de ventana Gaussiana.

Aunque anteriormente en las funciones de ventana se ha definido la letra Ω para referirse a la dimensión de frecuencia, en el caso de la ventana Prolate se utiliza la letra B para referirse a la dimensión de frecuencia, de forma que se mantiene la coherencia con los estudios previos.

Este tipo de filtro de ventana nunca ha sido usado en un problema de espacio tiempo frecuencia para la detección de fallos en motores de inducción, por lo que se puede considerar un **elemento innovador** dentro de esta área.

A continuación, se muestra la función de ventana PSWF ($\varphi(t)$) (6.1). La función $\varphi(t)$ está limitada por la variable T correspondiente a la mitad de la duración del tiempo, y la variable B correspondiente al ancho de banda. La variable λ se corresponde con una serie infinita de autovalores (λ_n) de tipo real, positivo y menores a 1 (factor de escala).

$$\int_{-T}^T \varphi(x) \frac{\text{sen}(B(t-x))}{\pi(t-x)} dx = \lambda \varphi(t) \quad (6.1)$$

Al igual que en el caso de la función elemental Gaussiana, la función elemental de onda esferoidal Prolate obtiene la mejor definición de las características representativas de fallo dentro del espacio tiempo frecuencia cuando la relación de aspecto T/B de la función de ventana mantiene el mayor solapamiento posible con la trayectoria de las características de fallo.

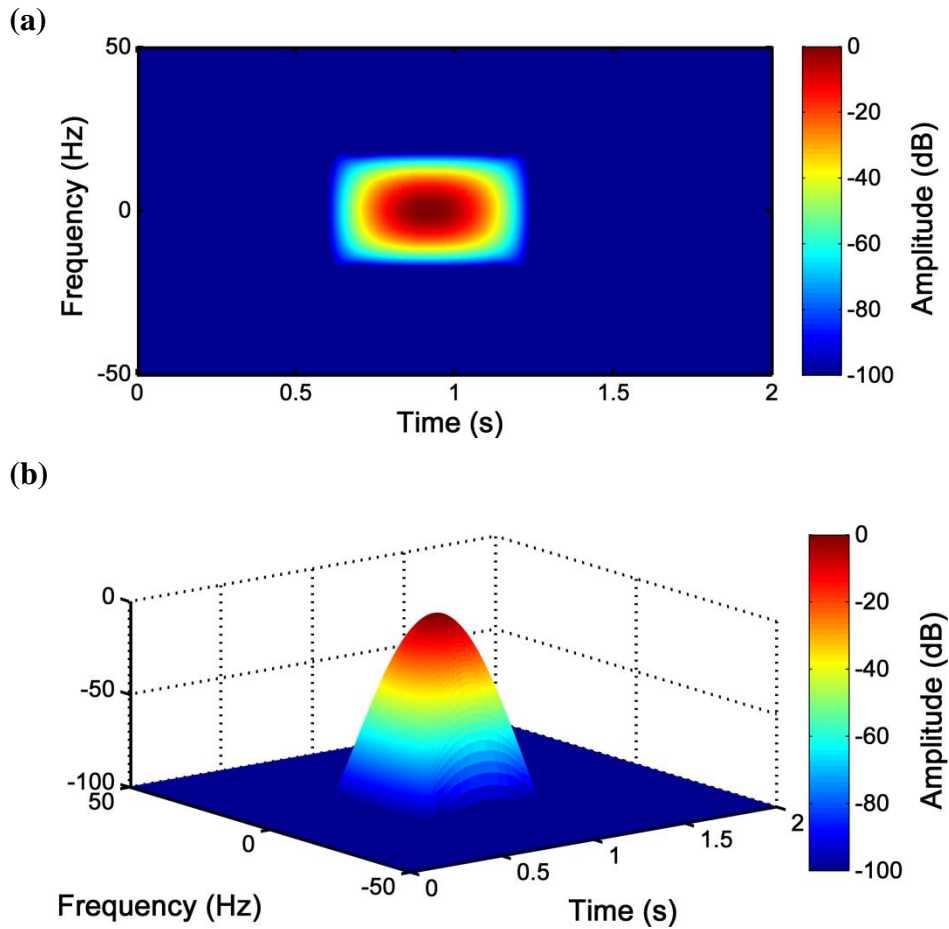


Figura 6.4 Ejemplo de función de onda esferoidal Prolate representada en (a) 2D y (b) 3D.

En la Figura 6.4 se muestra un ejemplo de la función de ventana de onda esferoidal Prolate optimizada para la detección de fallo de barra rota en 2D 6.4(a) y 3D 6.4(b).

Por otro lado, la función de ventana Gaussiana [Figura 6.5 (a)] se muestra como una onda de campana de Gauss donde la máxima energía de la ventana se encuentra en la copa de esta campana de Gauss. Por tanto, como se explicó en el apartado 2.3.2.1, para obtener un espacio de tiempo frecuencia representativo se desplaza la ventana Gaussiana en incrementos de tiempo con un desplazamientos igual o menor a la longitud de la ventana.

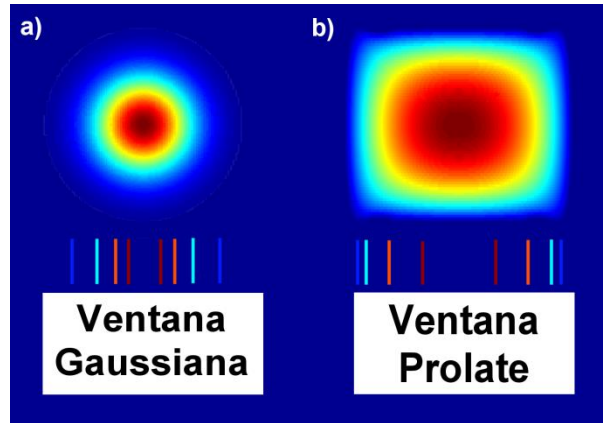


Figura 6.5. Comparativa gráfica en 2D entre una ventana Gaussiana y una ventana Prolate.

Sin embargo, como se observa en Figura 6.5 (b), la función de ventana Prolate es una función de onda cercana a una onda cuadrada, donde la máxima energía de esta ventana se centra en un intervalo “cuadrado” mayor que el disponible en la ventana gaussiana. Por tanto, si se utiliza una ventana Prolate (cuya energía es mayor para la misma longitud) se puede obtener un espacio tiempo frecuencia bastante representativo realizando un desplazamiento de la ventana Prolate más largo que con la ventana Gaussiana.

Al realizar desplazamientos más largos con la ventana Prolate (Figura 6.6), es necesario calcular un número menor de incrementos de tiempo para generar el espacio tiempo frecuencia, disminuyendo así el tiempo de cálculo necesario.

En conclusión, al haber sustituido la ventana Gaussiana por una ventana Prolate que concentra una mayor energía para un intervalo T/B definido, se ha mejorado la resolución de las características significativas de fallo. Además, al requerir esta ventana Prolate un número menor de desplazamientos para generar el espacio tiempo frecuencia también se ha mejorado el tiempo de cálculo necesario.

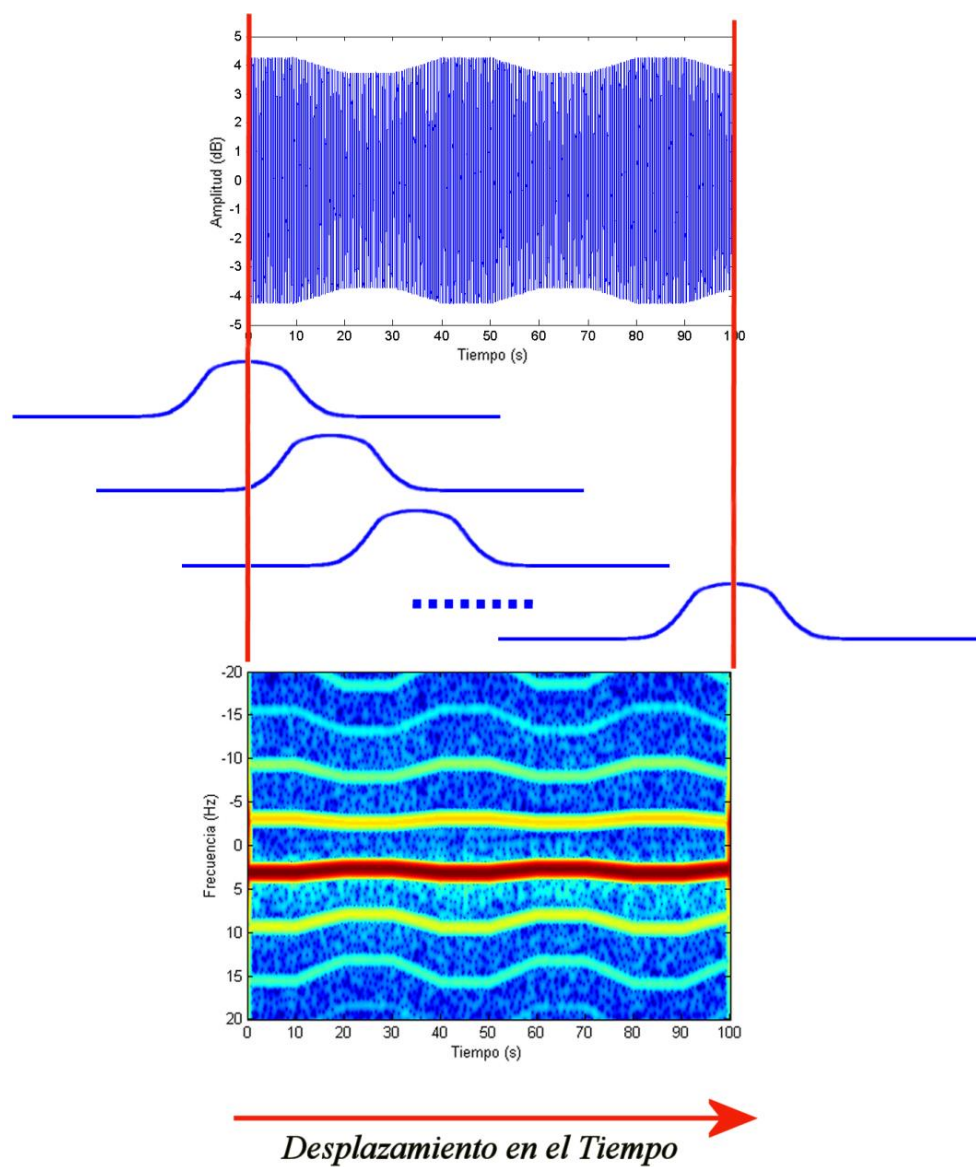


Figura 6.6. Desplazamiento de la función de ventana Prolate sobre la señal de corriente obteniendo así el espacio de tiempo frecuencia STFT.

6.2.1.3 Limitación del recorrido de la ventana para minimizar el efecto de borde [Mejora en la precisión del diagnóstico].

El efecto de borde es un problema bien conocido en la generación de espacios tiempo/frecuencia mediante la técnica STFT. Como se muestra en la Figura 6.7, este efecto se visualiza como una deformación de los valores de amplitud en los límites del espacio tiempo/frecuencia, generados a causa de aplicar la ventana en el límite de la señal.

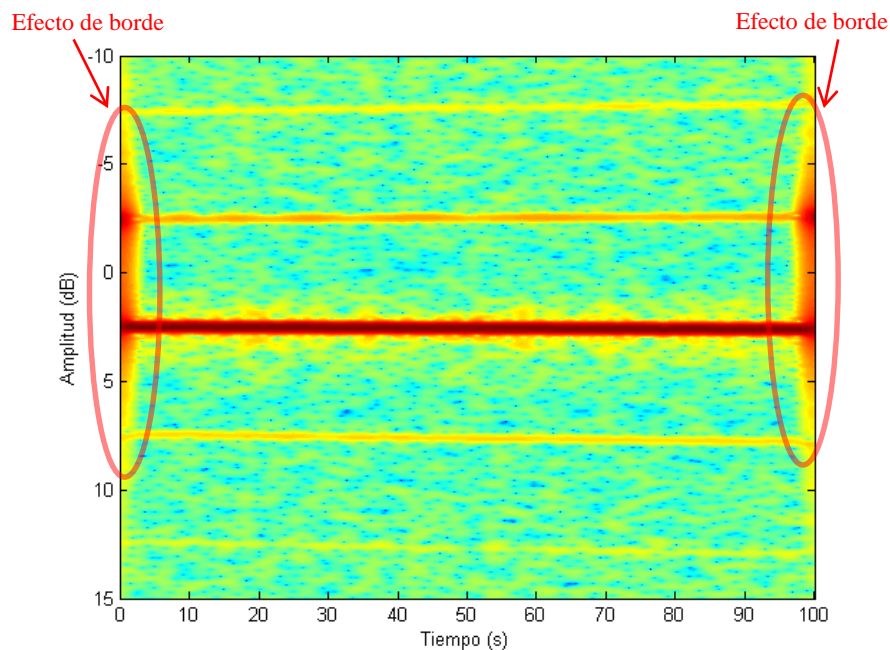


Figura 6.7 Espacio tiempo frecuencia de una corriente de fase generado con la técnica STFT, donde se observa un acusado efecto de borde.

En la Figura 6.8 se muestra el resultado simplificado de HOTA transitorio obtenido del espacio tiempo/frecuencia mostrado en la anterior Figura 6.7. En comparación con la imagen Figura 6.8, se muestra la Figura 6.9 correspondiente al resultado simplificado de HOTA transitorio creado a partir del mismo espacio tiempo/frecuencia de la Figura 6.7 pero eliminando con anterioridad el efecto de borde.

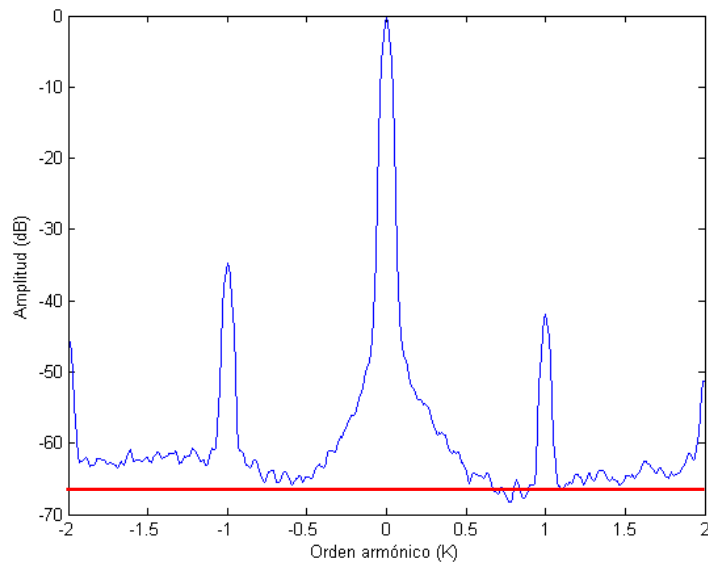


Figura 6.8. Resultado simplificado de HOTA transitorio obtenido a partir de un espacio tiempo frecuencia con efecto de borde.

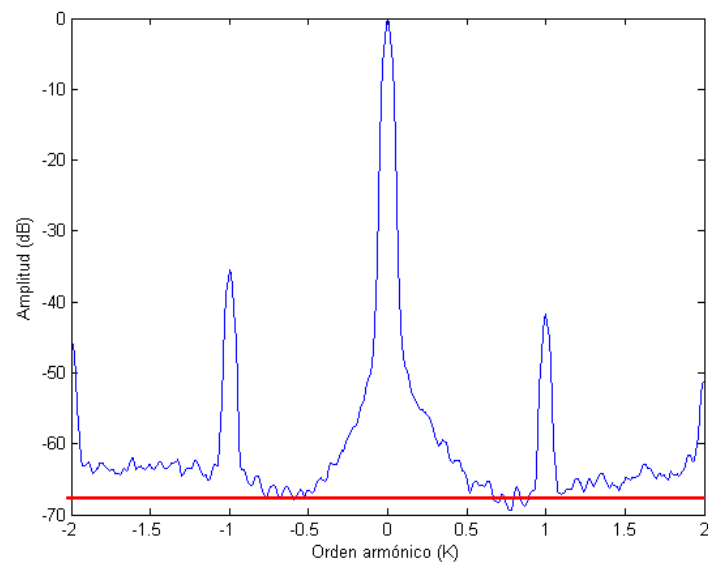


Figura 6.9. Resultado simplificado de HOTA transitorio obtenido a partir de un espacio tiempo frecuencia sin efecto de borde.

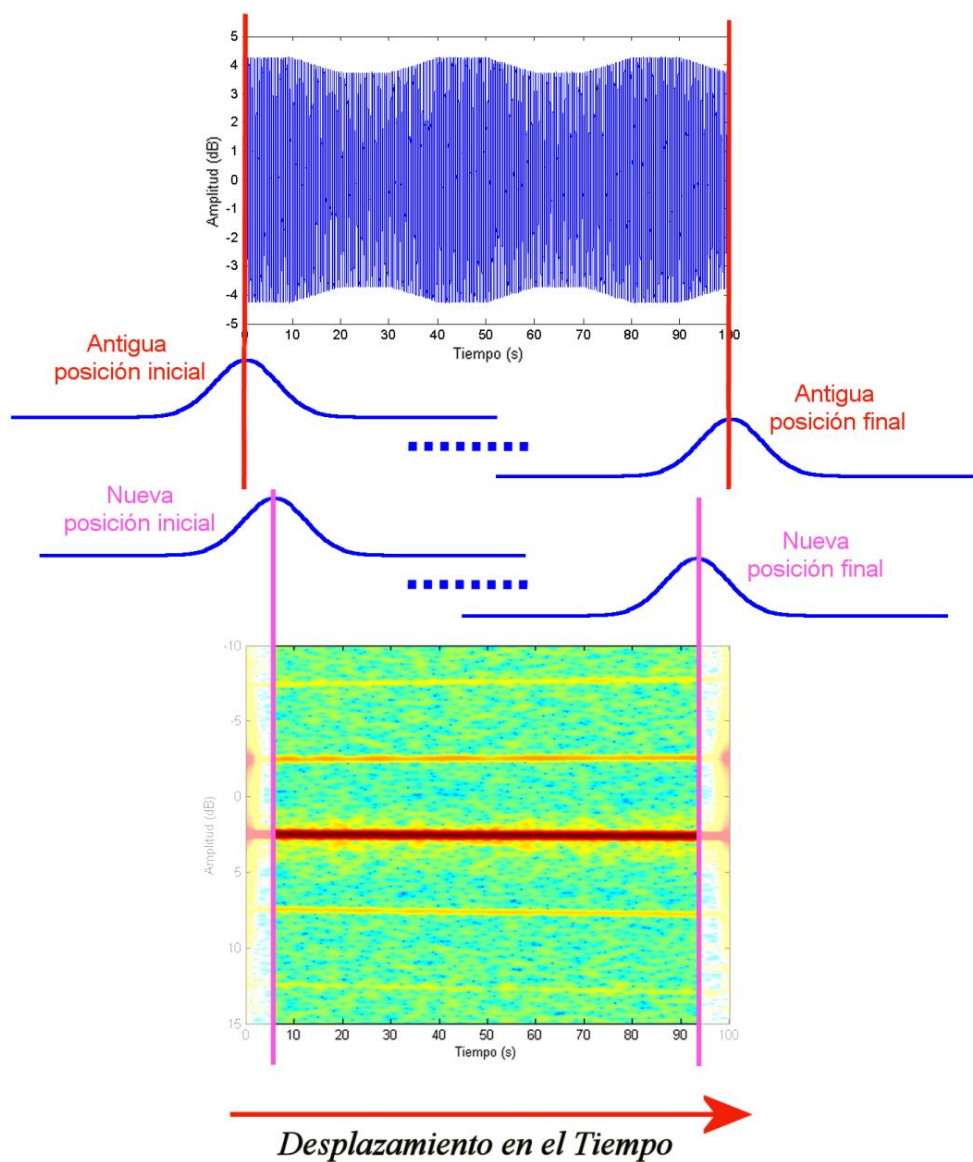


Figura 6.10. Desplazamiento de la función de ventana Gaussiana sobre la señal de corriente para generar el espacio de tiempo frecuencia STFT, omitiendo la zona en los límites de la señal donde se produce el efecto de borde.

Como se observa en la comparativa entre la Figura 6.8 y la Figura 6.9, el efecto de borde actúa sobre el resultado de HOTA amplificando de forma artificial las componentes características de fallo para $k=\pm 1$, produciendo un efecto negativo igual de problemático al obtenido para el régimen estacionario con la fuga espectral.

Como innovación para evitar el no deseado efecto de borde, se ha decidido modificar el desplazamiento inicial y final de la ventana sobre la corriente de fase proyectada en el rotor.

Como se muestra en la representación de la Figura 6.10, se ha reconfigurado la posición inicial del recorrido de la ventana en la primera posición de la corriente donde se procesa por completo la zona de la ventana con una energía significativa (la parte de la campana Gaussiana con una amplitud mayor al 0.5% respecto de su valor máximo). De este modo se asegura que la parte más significativa de la ventana se procesa adecuadamente y en consecuencia el efecto de borde es mínimo.

De forma análoga, se ha configurado la posición final del recorrido de la ventana en la última posición de la corriente de fase donde es posible procesar por completo la zona de la ventana con una energía significativa mayor al 0.5% respecto de su valor máximo.

Es evidente que al acortar el recorrido de la ventana se pierde algo de resolución, pero se puede considerar la parte de resolución que se pierde como una parte inútil, ya que contiene el problemático efecto de borde. En consecuencia, el resultado simplificado de HOTA al eliminar el efecto de borde muestra unas características significativas de mayor precisión.

Esta limitación del recorrido de la ventana es independiente del tipo de ventana, por lo que se puede aplicar sin ningún problema tanto a la ventana Gaussiana como a la ventana Prolate.

6.2.1.4 Método de ajuste de la proporción de la función de ventana para maximizar la frecuencia de fallo [Mejora en la precisión del diagnóstico, en el tiempo de cálculo y en almacenamiento en memoria].

A continuación se exponen varios problemas sobre el algoritmo de ajuste de la proporcionalidad de las dimensiones de la ventana de la técnica STFT propuesto en el estado del arte (apartado 2.3.2.1.1), basado en la búsqueda del espacio tiempo frecuencia de mínima entropía:

- Para obtener el vector de entropías se requiere generar un número mínimo de espacios tiempo frecuencia con diferentes proporciones de la función de ventana para obtener un vector de entropías de donde obtener la proporción de

ventana idónea para generar un espacio de mínima entropía (espacio de máxima energía).

Como ya se ha comentado en el apartado 6.2.1, el coste tanto de cálculo como de espacio de memoria de generar un espacio de tiempo frecuencia con STFT es alto para el sistema de hardware disponible, por lo que generar el total de espacios tiempo frecuencia necesarios para obtener un vector de entropías con buena resolución llega a ser prohibitivo tanto en tiempo como en espacio de memoria.

- El espacio de tiempo frecuencia generado con esta técnica obtiene la mayor concentración de energía (mínima entropía) en conjunto del espacio. Sin embargo lo ideal en la detección de fallos es obtener las componentes de fallo con una mayor concentración de energía independientemente de la energía concentrada en el resto de frecuencias del espacio tiempo frecuencia.

Por tanto, aunque el ajuste de la proporcionalidad de la función de ventana obtiene un espacio de mínima entropía, pueden existir otras configuraciones de ventana con una entropía mayor pero con una concentración de energía para las componentes características de fallo también mayor, y por tanto más idóneas para su uso como características de fallo.

- En algunos casos en que se han procesado señales de motores de la base de datos (apartado 4.4), la zona de mínima entropía del vector de entropías no ha sido demasiado diferenciable del resto del vector de entropías, por lo que la selección de la mejor proporción de la función de ventana se vuelve más complicada al haber varias soluciones mínimas locales pero no hay ninguna que destaque notablemente del resto de soluciones.

Por tanto, como método de ajuste de la proporción de la función de ventana de la técnica STFT (apartado 2.3.2.1) dentro de la técnica HOTA transitorio (apartado 2.3.2.6), se ha optado por implementar un nuevo método de ajuste **innovador**, centrado en maximizar la amplitud de las componentes frecuenciales características de fallo manteniendo al mismo tiempo un coste de cálculo y de almacenamiento de memoria asumibles.

En la Figura 6.11 se muestran varios espacios de tiempo frecuencia generados a partir de la misma muestra de fase de corriente, pero variando la proporción de la ventana de filtro Gaussiana con la que se genera cada espacio tiempo frecuencia.

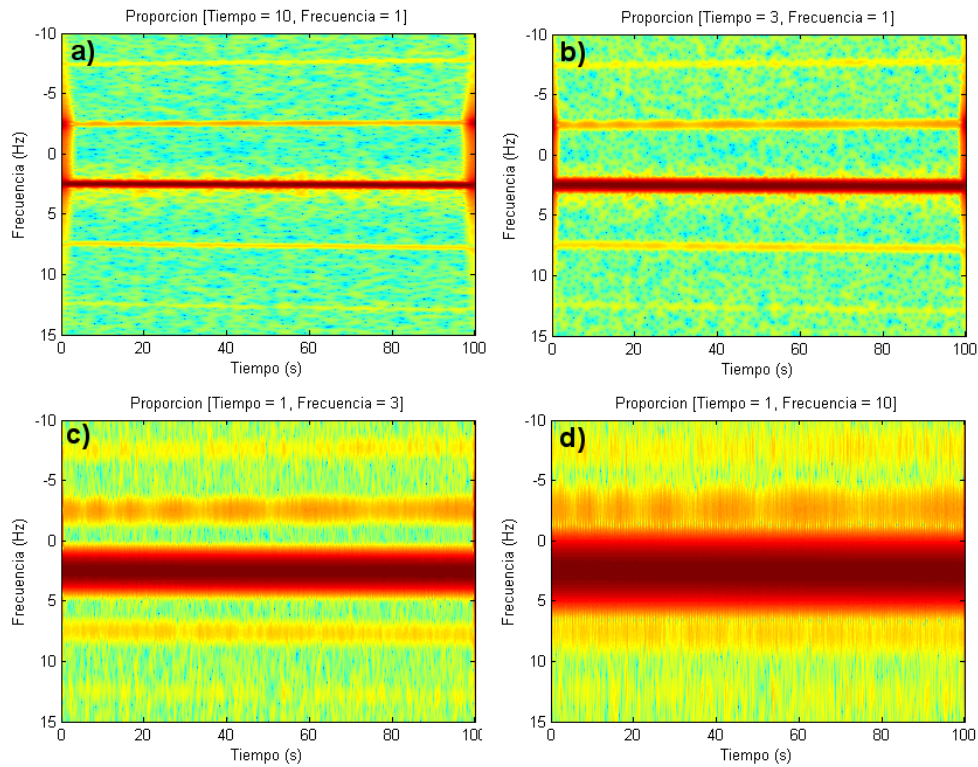


Figura 6.11. Comparativa del espacio tiempo frecuencia para una muestra de corriente de fase variando la proporción de ventana entre a) 10 unidades en tiempo por 1 en frecuencia, b) 3 unidades en tiempo por 1 en frecuencia, c) 1 unidad en tiempo por 3 en frecuencia y d) 1 unidad en tiempo por 10 en frecuencia.

Como se observa en esta Figura 6.11, conforme aumenta la proporción en longitud en una de las dimensiones (de tiempo o frecuencia) de la ventana de filtro (y disminuye en consecuencia la proporción del otro dominio), aumenta la resolución de la dimensión del espacio tiempo frecuencia cuya proporción.

Como ya se comentó en el apartado 2.3.2.1 y como se observa en esta Figura 6.11, cuando se mejora la resolución en uno de las dimensiones (tiempo o frecuencia) del espacio de tiempo frecuencia inevitablemente la resolución de la otra dimensión empeora.

Se puede decir que esta variación de la resolución para diferentes proporciones de ventana actúa como un filtro de borrosidad para la dimensión del espacio afectada, donde las frecuencias de mayor amplitud solapan a las frecuencias contiguas de menor ampli-

tud. En el caso de las componentes frecuenciales de fallo, la frecuencia de control es la que mayormente afecta a las componentes frecuenciales de fallo más importantes ($k=1$ y $k=-1$).

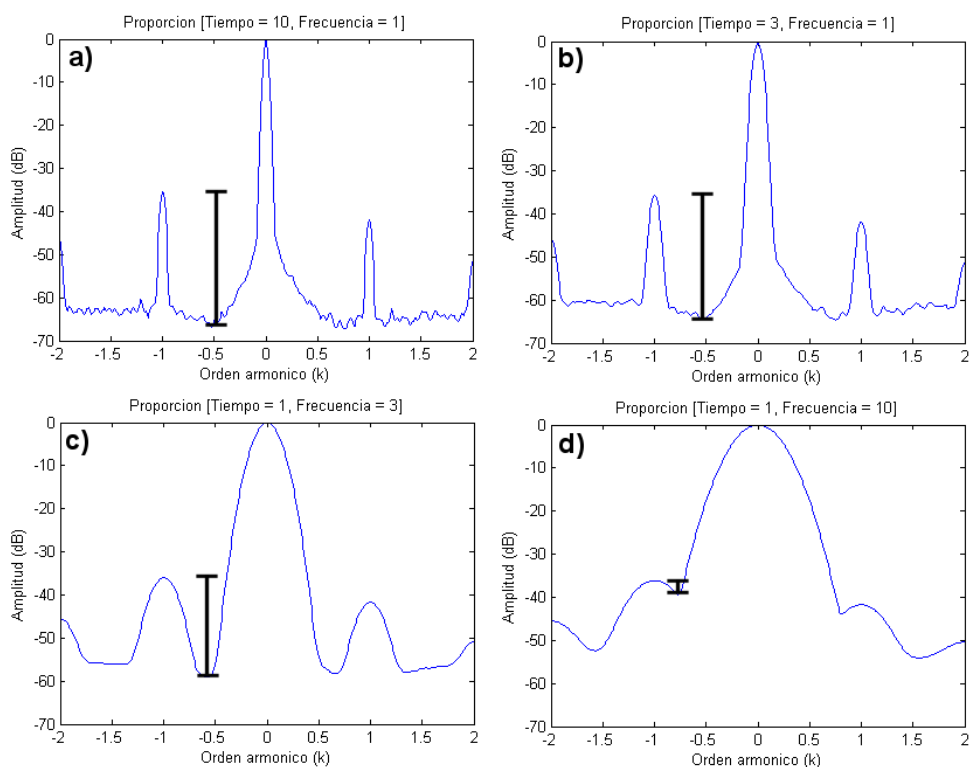


Figura 6.12. Comparativa del resultado reducido para HOTA transitorio para una muestra de corriente de fase variando la proporción de ventana entre a) 10 unidades en tiempo por 1 en frecuencia, b) 3 unidades en tiempo por 1 en frecuencia, c) 1 unidad en tiempo por 3 en frecuencia y c) 1 unidad en tiempo por 10 en frecuencia.

En consecuencia, como se observa en la Figura 6.12 (para los resultados reducidos con HOTA transitorio) una buena aproximación al valor real de las componentes de fallo será aquella cuya componentes de fallo más importantes ($k=1$ y $k=-1$) muestren una mayor diferencia con el ruido de fondo entre estas y la frecuencia de control. En caso contrario, cuanto menor sea esta diferencia sugiere que la frecuencia fundamental ejerce una mayor influencia sobre las amplitudes de las frecuencias de fallo falseando la amplitud correcta del fallo.

Desafortunadamente, este ajuste de las componentes de fallo para la técnica STFT no puede usarse de forma genérica para las muestras de corriente con transitorios, donde las frecuencias correspondientes a los fallos varían.

Sin embargo, en este caso en concreto la técnica STFT se utiliza como técnica para obtener el espacio tiempo frecuencia requerido por la técnica HOTA transitorio (Figura 6.3 del apartado 6.2.1), donde esta técnica HOTA tiene la ventaja de transformar el espacio tiempo frecuencia en un espacio tiempo k-orden armónico donde en este caso todas las componentes de fallo sí se muestran siempre en su posición de k-orden armónico independientemente de la transitoriedad de la muestra de corriente original.

Por tanto, y como se ha mostrado anteriormente en la Figura 6.12, la selección de la mejor proporción de ventana para obtener las amplitudes de las frecuencias de fallo más representativas puede obtenerse analizando la diferencia entre las frecuencias de fallo más significativas y el valor medio del ruido de fondo entre estas y la componente frecuencial de control; datos obtenidos de los resultados reducidos extraídos con HOTA transitorio de sus respectivos espacios de tiempo k-orden armónico.

Por otro lado, como se observa en la Figura 6.12, existe una relación directa y proporcional entre la variación en la proporción en longitud de la función de la ventana y la variación en la distancia entre las frecuencias de fallo más significativas y el valor medio del ruido de fondo entre estas y la componente frecuencial de control.

Por consiguiente, con el anterior método comentado para el mejor ajuste de la proporción de ventana basado en la búsqueda del espacio de mínima entropía (apartado 2.3.2.1.1) era necesario procesar un cierto número mínimo de espacios tiempo frecuencia para obtener una aproximación de la curva de entropías. Sin embargo, con este nuevo método se puede obtener una buena aproximación a la solución ideal de ajuste de la proporción de ventana mediante una búsqueda binaria partiendo del rango inicial de búsqueda entre una ventana con una proporción máxima en tiempo respecto de frecuencia (en esta tesis definida por una relación de 1 en tiempo por cada 1000 en frecuencia) y una ventana con una proporción máxima en frecuencia respecto del tiempo (definida por una relación de 1 en frecuencia por cada 1000 en tiempo).

En este caso, la búsqueda binaria garantiza una buena solución de ajuste de la proporción de ventana en pocos pasos de búsqueda, es decir, generando unos pocos espacios de tiempo k-orden armónico, por lo que el coste de cálculo necesario para obtener el mejor ajuste de la proporción de ventana también es notablemente menor en comparación con el método de ajuste de la proporción de ventana basado en la búsqueda del espacio de mínima entropía (apartado 2.3.2.1.1).

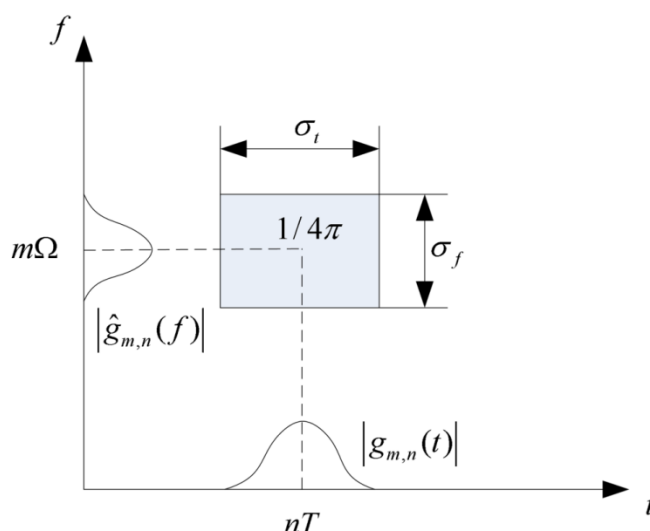
Finalmente, como se ha desarrollado en este punto, con este novedoso método de ajuste de la proporción de ventana se optimiza la amplitud de las componentes frecuenciales

de fallo (características de fallo) y se mejora el tiempo de cálculo necesario para su obtención.

6.2.1.5 Transformada de Fourier corta en frecuencia (SFFT) con bandas adaptativas como alternativa a la transformada STFT [Mejora en el tiempo de cálculo y en almacenamiento en memoria].

Como **última novedad** para mejorar el coste requerido por la técnica HOTA para el diagnóstico de motores en régimen transitorio, se ha optado por sustituir dentro de esta técnica la transformada STFT destinada a generar el espacio tiempo frecuencia, por una transformada de Fourier corta en frecuencia (SFFT)..

Al igual que la transformada STFT, la transformada SFFT es una técnica de transformación de una muestra en un espacio tiempo-frecuencia [42] (Figura 6.13) donde se utiliza una función de ventana $\hat{g}(f)$ (expresada en función de la frecuencia) para generar ese espacio tiempo/frecuencia con la mayor resolución posible.



Fuente: Riera-Guasp et al. (2012) [42]

Figura 6.13. Representación paramétrica de la función de ventana en el espacio de tiempo-frecuencia.

Sin embargo, a diferencia de la transformada STFT, esta función de ventana $\hat{g}(f)$ genera el espacio tiempo frecuencia desplazándose en la dimensión de frecuencias (Figura 6.14).

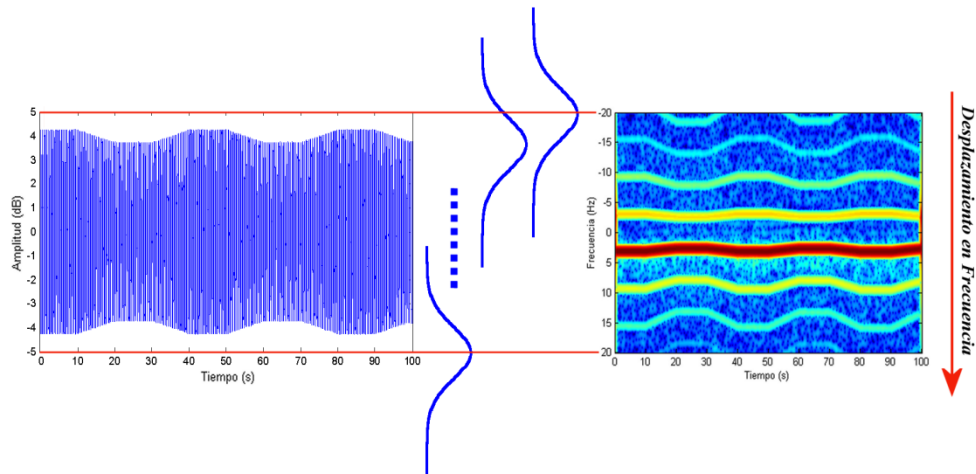


Figura 6.14. Desplazamiento de la función de ventana Gaussiana sobre la señal de corriente para generar el espacio tiempo frecuencia mediante la técnica SFFT.

Aparentemente, la generación del espacio tiempo frecuencia mediante la transformada STFT o la transformada SFFT generan prácticamente la misma representación del espacio de tiempo frecuencia para una corriente de fase proyectada en el rotor, con un coste del tiempo de cálculo y de espacio necesario de memoria muy similares.

En consecuencia, la sustitución de la transformada STFT por la transformada SFFT dentro de la técnica HOTA transitorio no produce ningún tipo de mejora.

No obstante, el sistema de generación del espacio tiempo frecuencia de la transformada SFFT recorriendo la ventana en la dimensión de frecuencia permite procesar únicamente, en el caso de que sean conocidas, los intervalos de frecuencias necesarios.

Para el caso del diagnóstico de fallos de esta tesis, solo es necesario conocer dentro de la dimensión de frecuencias tres tipos de frecuencias distintas para cada incremento de tiempo dentro del espacio tiempo frecuencia:

- **La frecuencia de control**, necesaria para conocer la posición frecuencial de las componentes de fallo.
- **Las frecuencias de fallo de la componente armónica para $k=\pm 1$** , siendo estas las más componentes frecuenciales de fallo más representativas.
- **El valor medio del ruido de fondo entre la frecuencias de fallo $k=\pm 1$ y la frecuencia de control.** Este valor medio es necesario para desarrollar el méto-

do de ajuste de la proporción de la función de ventana para maximizar las frecuencias de fallo, desarrollado en el apartado 6.2.1.4.

Por tanto, se ha desarrollado un **novedoso algoritmo** de búsqueda para la obtención de estas frecuencias mediante bandas adaptativas (Figura 6.15)

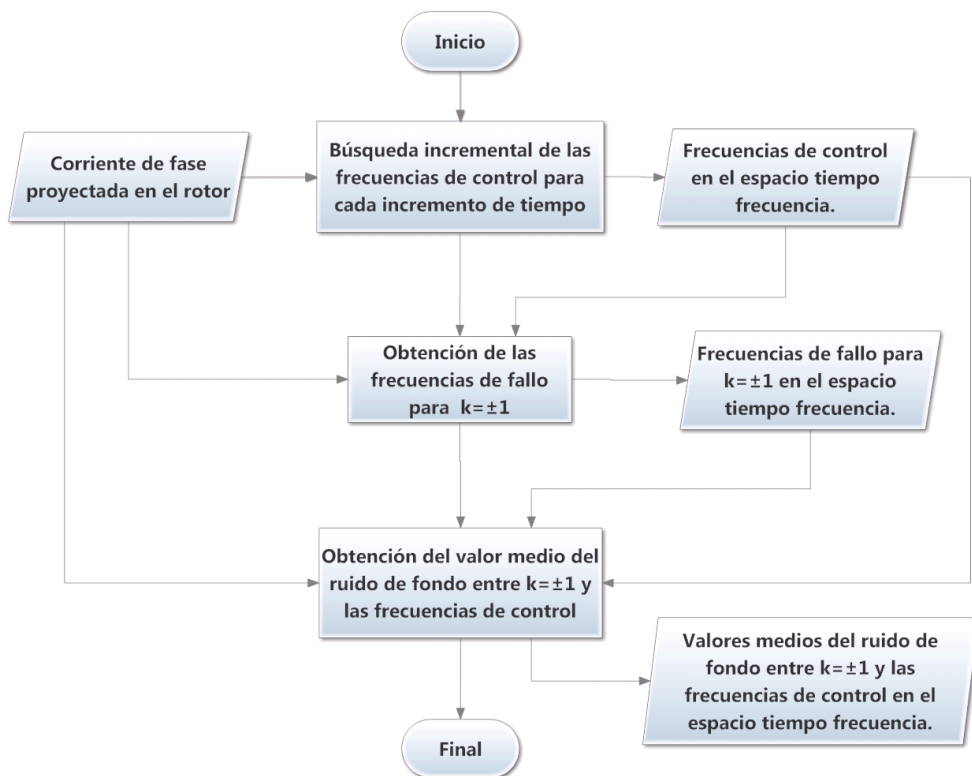


Figura 6.15. Algoritmo de obtención de las frecuencias representativas de fallo y los valores medios del ruido de fondo basado en bandas adaptativas.

Como se muestra en la Figura 6.15, este algoritmo se desarrolla en tres pasos:

- 1) Se realiza una búsqueda incremental de las frecuencias de control para cada incremento de tiempo dentro del espacio tiempo frecuencia mediante la transformada SFFT para esas frecuencias de búsqueda.

Al analizarse una corriente de fase proyectada al sistema de referencia del rotor, se conoce que la frecuencia de control es próxima al origen 0Hz de fre-

cuencias. Por consiguiente, se genera la búsqueda de las frecuencias de control a partir de un intervalo inicial de confianza respecto del origen de 0Hz. En este caso el intervalo inicial se ha definido como el intervalo de [-5, 5] Hz.

Partiendo de los límites de este intervalo inicial se realiza una exploración incremental a partir de la parte positiva del intervalo ($> 5\text{Hz}$) y decremental a partir de la parte negativa del intervalo ($< -5\text{Hz}$) de forma que la búsqueda termina cuando se obtiene la frecuencia de control (la amplitud máxima).

- 2) A partir del vector de frecuencias de control obtenido en el punto 1) para cada incremento de tiempo dentro del espacio, se calculan las posiciones correspondientes a las componentes frecuenciales de fallo para $k=\pm 1$.

Conocida la posición en frecuencia de estas componentes de fallo, se obtiene su amplitud (característica representativa de fallo) para cada incremento de tiempo del espacio de tiempo frecuencia mediante la transformada SFFT aplicada sobre esa posición concreta de frecuencia.

- 3) Como último paso, se obtiene el valor medio del ruido de fondo entre las componentes frecuenciales de fallo para $k=\pm 1$ y la frecuencia de control para cada incremento de tiempo dentro del espacio de tiempo frecuencia.

En este caso se define el intervalo intermedio de frecuencias igual al 10% (definido en el algoritmo de ajuste de ventana del apartado 6.2.1.4) de la distancia frecuencial entre la frecuencia de control y la respectiva componente de fallo $k=1$ o $k=-1$. Mediante la transformada SFFT se obtiene las amplitudes de este intervalo de frecuencia seleccionado y se calcula su valor medio para cada incremento de tiempo dentro del espacio tiempo frecuencia. .

Como se puede observar, el algoritmo se adapta para generar únicamente las bandas de frecuencia necesarias del espacio tiempo frecuencia para el desarrollo del diagnóstico de fallos mediante la técnica HOTA transitorio. Por tanto, al evitar mediante la transformada SFFT tener que procesar el resto de frecuencias no necesarias, el coste de tiempo de cálculo y de espacio de memoria requerido es sensiblemente menor.

En la Figura 6.16 se muestra un ejemplo de espacio tiempo frecuencia, donde se ha marcado en color únicamente las frecuencias que serían procesadas en la técnica HOTA transitorio haciendo uso de la transformada SFFT y del algoritmo de bandas adaptativas.

En este ejemplo el ancho de banda necesario para obtener las frecuencias relevantes es inferior a 12Hz. Si se compara con la técnica anterior STFT, donde siempre se mantiene un ancho de banda de frecuencia de 100Hz (por la reducción de frecuencia de la señal mediante el filtro paso banda del apartado 6.2.1.1), se ha reducido el coste del

proceso de obtención de las características de fallo para el espacio tiempo frecuencia en un 88%.

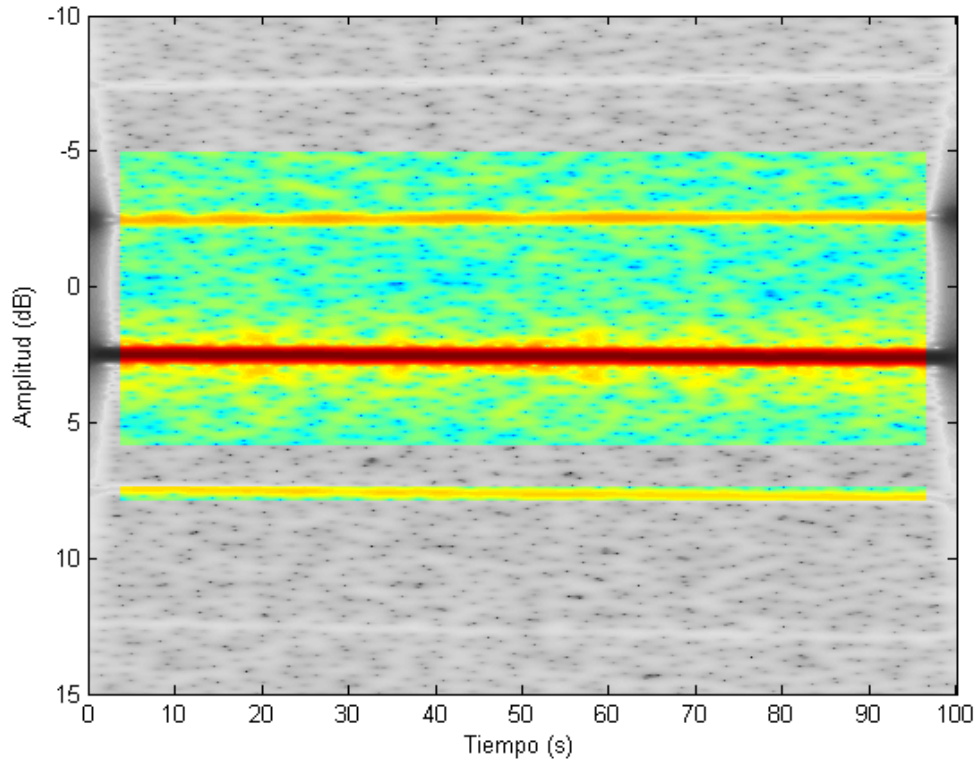


Figura 6.16. Ejemplo de espacio tiempo frecuencia, donde se ha marcado en color las frecuencias procesadas por el algoritmo de bandas adaptativas.

6.2.2 Resumen de variantes optimizadas de HOTA transitorio.

En este apartado se resumen las diferentes variantes de HOTA transitorio optimizadas con las mejoras desarrolladas en el anterior punto 6.2.1.

Para esta tesis se ha diferenciado 4 variantes de la técnica HOTA transitorio, basadas en el tipo de transformación del espacio tiempo frecuencia utilizado (STFT o SFFT) y en el tipo de ventana utilizada para la generación de ese espacio tiempo frecuencia (ventana Gaussiana o ventana Prolate):

- **HOTA Transitorio con transformada STFT y ventana Gaussiana.** Internamente se han aplicado las mejoras de la etapa de filtrado paso bajo (apartado 6.2.1.1), de limitación del recorrido de la ventana para minimizar el efecto de borde (apartado 6.2.1.3) y el método de ajuste de la proporción de la función de ventana Gaussiana (apartado 6.2.1.4).
- **HOTA Transitorio con transformada STFT y ventana Prolate.** Internamente se han aplicado las mejoras de la etapa de filtrado paso bajo (apartado 6.2.1.1), de implementación de la función de ventana Prolate (apartado 6.2.1.2), de limitación del recorrido de la ventana para minimizar el efecto de borde (apartado 6.2.1.3) y el método de ajuste de la proporción de la función de ventana Prolate (apartado 6.2.1.4).
- **HOTA Transitorio con transformada SFFT y ventana Gaussiana.** Internamente se han aplicado las mejoras de la etapa de filtrado paso bajo (apartado 6.2.1.1), de limitación del recorrido de la ventana para minimizar el efecto de borde (apartado 6.2.1.3), el método de ajuste de la proporción de la función de ventana Gaussiana (apartado 6.2.1.4) y la transformada SFFT con bandas adaptativas (apartado 6.2.1.5).
- **HOTA Transitorio con transformada SFFT y ventana Prolate.** Internamente se han aplicado las mejoras de la etapa de filtrado paso bajo (apartado 6.2.1.1), de implementación de la función de ventana Prolate (apartado 6.2.1.2), de limitación del recorrido de la ventana para minimizar el efecto de borde (apartado 6.2.1.3), el método de ajuste de la proporción de la función de ventana Prolate (apartado 6.2.1.4) y la transformada SFFT con bandas adaptativas (apartado 6.2.1.5).

6.3 Gestión de fallos mediante sistemas expertos.

Como ya se ha comentado en el apartado 6.2.1 referente al método HOTA de obtención de características representativas de fallo para régimen transitorio, las características de fallo obtenidas mediante esta técnica son equivalentes a las características de fallo obtenidas con el método HOTA para régimen estacionario del apartado 5.2.3.

Gracias a esta cualidad de HOTA de la que carecen los otros métodos de obtención de características para régimen transitorio (apartado 2.3.2.6), se pueden usar para desarrollar el mejor sistema final de diagnóstico de fallos en régimen transitorio los mismos sistemas expertos de máquinas de vectores de soporte y sistemas expertos por redes neuronales artificiales, explicados en el apartado 5.4 y 5.5 respectivamente, sin necesidad de variar la técnica de ajuste de parámetros, desarrollo estructural, metodología de entrenamiento o validación de los sistemas de diagnóstico generados.

6.4 Validación práctica.

Se ha usado el mismo equipo presentado en el apartado 5.6 para desarrollar la implementación del sistema completo de desarrollo de sistemas de diagnóstico mediante el uso de sistema experto de tipo SVM y ANN para régimen transitorio:

La implementación del nuevo sistema para régimen transitorio se ha desarrollado igualmente dentro del entorno de desarrollo de software matemático *MATLAB* versión R2014a de 64bits. Las *Toolbox* utilizadas en este caso de régimen transitorio son:

- *Signal Processing Toolbox*. Toolbox utilizada para el procesamiento de señales requerido dentro del método HOTA para régimen transitorio.
- *Statistics and Machine Learning Toolbox*. Toolbox utilizada para la implementación del conjunto de sistemas expertos SVM.
- *Neural Network Toolbox*. Toolbox utilizada para la implementación del conjunto de sistemas expertos ANN.

Al ejecutarse ambos sistemas (estacionario y transitorio) en el mismo equipo y mismo entorno de desarrollo se ha podido realizar de forma objetiva la comparación de los tiempos de ejecución y entrenamiento obtenidos para ambos tipos de régimen de trabajo.

6.5 Resultados obtenidos del sistema autónomo para régimen transitorio.

En este apartado se exponen y comentan los mejores sistemas de diagnóstico mediante las distintas variantes de la técnica HOTA para régimen transitorio (apartado 6.2.2) y sus respectivos sistemas expertos optimizados. De este modo se representan los diversos sistemas de diagnóstico óptimos para el caso de motores de inducción trabajando en régimen transitorio.

Para obtener los resultados se ha utilizado para todas las variantes de la técnica HOTA para régimen transitorio el mismo conjunto de características representativas de fallo, el cual es además el mismo conjunto de características representativas de fallo utilizadas en la técnica HOTA para régimen estacionario expuesto en el apartado 5.2.3.

Es decir, se ha utilizado como **primera y segunda característica simple la amplitud de la componente armónica de fallo para $k = 1$ y $k = -1$** respectivamente.

Un sistema experto desarrolla un mejor aprendizaje cuando el conjunto de características de entrenamiento es más representativo, es decir, cuando las características repre-

sentan todo el intervalo de valores posibles que pueden obtenerse respecto del problema de clasificación. Por tanto, para este caso de sistemas expertos orientados al diagnóstico de motores operando en régimen transitorio se ha utilizado todas las muestras generadas en régimen de trabajo transitorio (apartado 4.4.2) y todas las muestras generadas en régimen de trabajo estacionario (apartado 4.4.1) que también pueden ser procesadas con la técnica de diagnóstico HOTA transitorio.

En conjunto, se han utilizado en total 245 muestras de motores de inducción sanos y de barra rota. Teniendo en cuenta que cada muestra de motor contiene 3 muestras de fase de corriente y que el sistema de diagnóstico HOTA transitorio procesa cada muestra por separado. Considerando cada muestra de corriente de fase por separado se ha utilizado un total de 735 muestras de corrientes de fase para la generación del sistema de diagnóstico óptimo.

En este caso de sistemas de diagnóstico para régimen transitorio, antes de analizar los datos se han de tener en cuenta los siguientes puntos:

- Los parámetros de ajuste de las ventanas utilizadas en los espacios tiempo frecuencia internamente dentro de cualquiera de las variantes de la técnica HOTA para régimen transitorio, han sido optimizados mediante el algoritmo desarrollado en el apartado 6.2.1.4. Este proceso se realiza de forma automática por el sistema automático de generador de sistemas de diagnóstico óptimos de esta tesis y solo son óptimos respecto a la base de datos de muestras procesadas (apartado 4.4).
- Los coeficientes de ajustes de los sistemas expertos también están optimizados de forma automática por el sistema automático de generador de sistemas de diagnóstico óptimos de esta tesis. Por lo que igualmente son óptimos respecto a la base de datos de muestras (apartado 4.4)

En el caso de requerir un nuevo sistema de diagnóstico óptimo para las muestras de otro tipo de motor distinto, es necesario volver a ejecutar el sistema de optimización de sistemas de diagnóstico de forma que vuelva a ajustar de forma óptima los parámetros internos de la técnica HOTA transitorio (parámetros de ventana) y los coeficientes de los sistemas expertos respecto de ese nuevo tipo de motor.

- Al igual que ocurre en los resultados de tiempo obtenidos para el régimen de trabajo estacionario (apartado 5.7), todos los tiempos que se han obtenidos en este caso son relativos y dependen de la implementación del algoritmo y del dispositivo donde se ejecuta. Por tanto, los tiempos que se exponen en este apartado se tienen que tomar solo como valores locales válidos para comparar con el resto de tiempos obtenidos en los resultados de esta tesis.

La siguiente Tabla 6.1 muestra los tiempos obtenidos al ejecutar las variantes del sistema de diagnóstico HOTA transitorio para obtener las características representativas de fallo. Cada columna de la tabla representa:

- **Variación del sistema de diagnóstico HOTA transitorio.** Indica el tipo de técnica que ha sido utilizada internamente (STFT o SFFT) dentro de HOTA transitorio para generar el espacio tiempo frecuencia, y el tipo de función de ventana utilizada en cada caso.
- **Tiempo total para generar características <TTc>.** Muestra en segundos, para cada alternativa, el tiempo total necesario para generar las características representativas. Comprende el tiempo de carga de cada una de las muestras y el tiempo de generación de las características de todas las muestras.
- **Tiempo medio para procesar una muestra <TMm>.** Muestra en milisegundos, para cada alternativa, el tiempo medio calculado de procesar una muestra para obtener las características de dicha muestra. A diferencia del tiempo (TTc), en este caso no se tiene en cuenta el tiempo de carga de la muestra

Tabla 6.1. Resultados de los tiempos obtenidos en la ejecución de las variantes de la técnica de diagnóstico HOTA transitorio en donde se han obtenido las características representativas de fallo a partir de las muestras.

HOTA para régimen transitorio	Tiempo de cómputo	
	Tiempo total para generar características < TTc > (s)	Tiempo medio para procesar una muestra < TMm > (ms)
Técnica STFT y ventana Gaussiana	45.236,35	36.641,65
Técnica STFT y ventana Prolate	44.454,17	30.915,03
Técnica SFFT y ventana Gaussiana	6.550,53	5.597,42
Técnica SFFT y ventana Prolate	6.192,53	5.483,18

En cada una de las columnas de tiempo de cómputo de la Tabla 6.1 se muestra en color rojo el peor tiempo (el de mayor valor) y de color verde el mejor tiempo (el de menor valor).

Como se observa en la tabla de tiempos 6.1, la optimización (no común a todas las variantes de HOTA transitorio) más influyente se centra en la sustitución de la técnica STFT por la técnica SFFT con bandas dinámicas. Aunque la optimización basada en la sustitución de la función de ventana Gaussiana por una ventana Prolate también muestra cierta mejoría, esta no es demasiado notable.

En todo caso es importante el tiempo obtenido para procesar una sola muestra, ya que ofrece una aproximación al tiempo requerido en obtener las características representativas para el diagnóstico de una nueva muestra desconocida. El tiempo en este caso varía bastante dependiendo del tipo de transformada de espacio frecuencia utilizada (STFT o SFFT) con sus respectivas optimizaciones.

Aunque el tiempo de procesar una muestra es aceptable en cualquier caso, en los casos concretos donde se utiliza internamente la transformada SFFT dentro de la técnica HOTA, se muestra un tiempo medio de diagnóstico muy bueno. Este tiempo incluso se encuentra dentro del intervalo obtenido en el anterior capítulo 5 respecto del tiempo obtenido para procesar una muestra en régimen de trabajo estacionario (tabla 5.2 del apartado 5.7).

6.5.1 Resultados obtenidos con sistemas expertos de Máquinas de vectores de soporte (SVM).

Los resultados obtenidos respecto del sistema experto SVM se muestran en las tablas 6.2 y 6.3 para valores de tiempo, y 6.5 y 6.6 para ajustes óptimos de parámetros y tasas de acierto.

Como ya se expuso en el apartado 5.4, en el proceso de generación del SVM óptimo se ha utilizado un kernel Polinomial y un **kernel de Base Radial Gaussiana**, se ha usado un algoritmo de ajuste de coeficientes mediante **búsqueda en malla** y a cada nodo de la malla de búsqueda se ha aplicado una **validación cruzada de uno a muchos** para estimar tasa de aciertos.

6.5.1.1 Comparación de tiempos entre kernel polinomial y kernel RBF.

La tabla 6.2 para SVM con kernel polinomial y la tabla 6.3 para SVM con kernel RBF muestra los tiempos necesarios para optimizar el mejor sistema de diagnóstico basado en sistemas expertos SVM. Cada columna de la tabla representa:

- **Variación del sistema de diagnóstico HOTA transitorio.** Indica el tipo de técnica que ha sido utilizada internamente (STFT o SFFT) dentro de HOTA transitorio para generar el espacio tiempo frecuencia, y el tipo de función de ventana utilizada en cada caso.

- **Tiempo de cómputo del SVM - Entrenamiento < Te-1 >.** Muestra en segundos el tiempo total necesario para realizar el mejor ajuste de coeficientes y el mejor entrenamiento de forma que se obtiene la tasa de aciertos más alta
- **Tiempo de cómputo del SVM - Clasificación < Tc-1 >.** Muestra en segundos la media del tiempo requerido para clasificar las características representativas de una sola muestra.
- **Tiempo de cómputo total - Entrenamiento < Te-2 >.** Muestra en segundos el tiempo total de entrenamiento de todo el sistema de diagnóstico. Es decir el resultado de sumar el tiempo de procesar las muestras (TTc) con la variante de la técnica HOTA transitorio y el tiempo de entrenamiento óptimo posterior del sistema SVM (Te-1). Estos tiempos obtenidos pueden dar una idea del tiempo requerido para el entrenamiento del sistema experto SVM con cada variante de las técnicas de obtención de características.
- **Tiempo de cómputo total - Clasificación < Tc-2 >.** Muestra en milisegundos el tiempo total procesar una muestra de un motor (TMm) y clasificarlo con el sistema experto mejor SVM con esa variante de HOTA transitorio (Tc-1). Estos tiempos permiten obtener una aproximación del tiempo final que requerirá el sistema de diagnóstico optimizado para diagnosticar el fallo en una muestra de motor desconocida.

Tabla 6.2. Resultados de los tiempos obtenidos del sistema experto de clasificación SVM configurado con un kernel polinomial, y los tiempos totales de las variantes del sistema de diagnóstico HOTA transitorio.

SVM con kernel Polinomial	Tiempo de cómputo del SVM		Tiempo de cómputo total	
	Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-2 > (s) (TTc+Te-1)	Clasificación < Tc-2 > (ms) (TMm+Tc-1)
HOTA (STFT y ventana Gaussiana)	4.499,41	0,671	49.735,76	36.642,32
HOTA (STFT y ventana Prolate)	3.555,66	0,688	48.009,83	30.915,72
HOTA (SFFT y ventana Gaussiana)	3.189,25	0,680	9.739,78	5.598,10
HOTA (SFFT y ventana Prolate)	4.936,78	0,703	11.129,31	5.483,88

Tabla 6.3. Resultados de los tiempos obtenidos del sistema experto de clasificación SVM configurado con un kernel RBF, y los tiempos totales de las variantes del sistema de diagnóstico HOTA transitorio.

SVM con Kernel RBF	Tiempo de cómputo del SVM		Tiempo de cómputo total	
	Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-2 > (s) (TTc+Te-1)	Clasificación < Tc-2 > (ms) (TMm+Tc-1)
HOTA (STFT y ventana Gaussiana)	24,95	0,723	45.261,30	36.642,37
HOTA (STFT y ventana Prolate)	25,11	0,726	44.479,28	30.915,76
HOTA (SFFT y ventana Gaussiana)	23,24	0,723	6.573,77	5.598,14
HOTA (SFFT y ventana Prolate)	25,10	0,755	6.217,63	5.483,45

Como se observa en ambas tablas (y al igual que ocurría para régimen estacionario), los costes de tiempos en clasificación del SVM son tan bajos que son irrelevantes respecto del tiempo de diagnóstico final del sistema de diagnóstico, que es mucho más significativo.

Respecto de los costes de tiempos entre el entrenamiento, se muestra como el entrenamiento del SVM con kernel RBF (tabla 6.3 y tabla 6.4) es de dos órdenes de magnitud más rápido que el entrenamiento con un SVM con kernel polinomial (tabla 6.2 y tabla 6.4).

Este fenómeno ya se observó en los resultados obtenidos con el sistema experto SVM para el diagnóstico en régimen estacionario (apartado 5.7.1) y tiene como explicación el rápido sobreentrenamiento del sistema cuando se utiliza un kernel RBF. Esta rápida tendencia al sobreentrenamiento hace finalizar al sistema de ajuste de coeficientes mucho más rápido; pero como se verá posteriormente, en la tabla 6.6 de resultados, a costa de obtener una pequeña penalización de la tasa de aciertos.

Tabla 6.4. Resultados máximo, mínimo y promedio de los tiempos obtenidos del sistema experto de clasificación SVM con kernel polinomial y con kernel Gaussiano RBF.

		Tiempo de cómputo del SVM		Tiempo de cómputo total	
		Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-2 > (s) (TTc+Te-1)	Clasificación < Tc-2 > (ms) (TMm+Tc-1)
SVM kernel Polinomial	Mínimo	3.189,25	0,671	9.739,78	5.483,88
	Máximo	4.936,78	0,703	49.735,76	36.642,32
	Promedio	4.045,28	0,686	29.653,67	19.660,00
SVM kernel RBF Gaussiano	Mínimo	23,24	0,723	6.217,63	5.483,94
	Máximo	25,11	0,755	45.261,30	36.642,37
	Promedio	24,60	0,730	25.633,00	19.660,05

En general, no existe un tiempo para una respectiva variante de HOTA transitorio que destaque notablemente del resto, tanto en tiempos de entrenamiento como el tiempo medio de clasificación del sistema experto SVM. Respecto del tiempo de clasificación mostrados en la tabla 6.2 y 6.3, tanto para un kernel polinomial como un kernel RBF, la variante de HOTA transitorio con transformación STFT y función de ventana Gaus-

siana obtiene el mejor tiempo de clasificación del clasificador SVM, no obstante, este valor de tiempo se ve ensombrecido por el tiempo de cómputo para clasificar una muestra, convirtiéndose en el peor de todos.

Justo ocurre lo contrario con la variante de HOTA transitorio con transformación SFFT y función de ventana Prolate, la cual, independientemente del kernel utilizado obtiene el peor tiempo de clasificación de SVM, pero sin embargo obtiene el mejor tiempo total de cómputo en clasificación.

6.5.1.2 Comparación de resultados entre kernel polinomial y kernel RBF.

En las siguientes tablas 6.5 para SVM con kernel polinomial y 6.6 para SVM con kernel RBF, se muestran los parámetros óptimos y las mayores tasas de aciertos obtenidas con el sistema experto SVM y una variante de la técnica HOTA para régimen de trabajo transitorio. Cada columna de la tabla representa:

- **Variación del sistema de diagnóstico HOTA transitorio.** Indica el tipo de técnica que ha sido utilizada internamente (STFT o SFFT) dentro de HOTA transitorio para generar el espacio tiempo frecuencia, y el tipo de función de ventana utilizada en cada caso.
- **Nivel de exploración alcanzado <lvl>.** El ajuste de los coeficientes se realiza mediante un algoritmo de búsqueda en malla. Este valor de “Nivel de exploración alcanzado” muestra hasta qué nivel de profundidad ha llegado la búsqueda en malla antes de obtener el mejor ajuste de los coeficientes.
- **Límites de caja <C>.** Coeficiente del sistema experto SVM que limita el número de características clasificadas como erróneas en la ventana de ajuste del hiperplano.
- **Orden Polinomial <d> (solo para kernel polinomial).** Este parámetro afecta al kernel polinomial, indicando el orden polinomial óptimo.
- **Sigma (σ) (solo para kernel RBF).** Este parámetro afecta al kernel RBF usándose como factor de escala de la función de base radial.
- **Tasa de aciertos.** Muestra la tasa de aciertos estimada del sistema de diagnóstico óptimo formado por la variación de HOTA transitorio para la obtención de características y por el sistema experto SVM entrenado de forma óptima. Este parámetro es esencial en esta tesis, pues muestra una aproximación la tasa de aciertos real en diagnóstico de fallos que es capaz de realizar este sistema experto final generado.

Tabla 6.5. Resultados obtenidos respecto del clasificador experto de tipo SVM configurado con un kernel polinomial.

Mejor ajuste de parámetros

SVM con kernel Polinomial	Nivel de exploración alcanzado < lvl >	Límites de caja < C >	Orden Polinomial < d >	Tasa de Aciertos (%)
HOTA (STFT y ventana Gaussiana)	1	32,000	3,0	98,62
HOTA (STFT y ventana Prolate)	1	0,031	6,0	98,34
HOTA (SFFT y ventana Gaussiana)	1	0,500	6,0	98,89
HOTA (SFFT y ventana Prolate)	1	32,000	5,0	98,76

Tabla 6.6. Resultados obtenidos respecto del clasificador experto de tipo SVM configurado con un kernel RBF Gaussiano.

Mejor ajuste de parámetros

SVM con Kernel RBF	Nivel de exploración alcanzado < lvl >	Límites de caja < C >	Sigma (σ)	Tasa de Aciertos (%)
HOTA (STFT y ventana Gaussiana)	0	1	1	97,24
HOTA (STFT y ventana Prolate)	0	1	1	96,96
HOTA (SFFT y ventana Gaussiana)	0	1	1	97,38
HOTA (SFFT y ventana Prolate)	0	1	1	97,24

Ambas tablas muestran unas tasas de acierto muy elevadas, por lo que se puede considerar que todos los sistemas de diagnóstico están muy bien optimizados.

La tabla 6.6 para kernel RBF Gaussiano muestra el nivel de ajuste de coeficientes mínimo {0} por defecto. Como ya se ha comentado anteriormente, esto se debe a su rápido sobreaprendizaje, esto afecta de forma negativa a las tasas de aciertos que, como se ve en la tabla 6.6 para kernel RBF son algo menores a las tasas de aciertos mostradas en la tabla 6.5 para kernel RBF.

De forma curiosa, se observa en ambas tablas que la tasa de aciertos cuando se utiliza una ventana Prolate es ligeramente inferior a cuando se utiliza su variación de HOTA equivalente pero con una ventana Gaussiana. Aunque en teoría la ventana Prolate concentra una mayor energía, no parece ser lo bastante relevante como para mostrar una mejora significativa de la tasa de aciertos.

Como se observa en la tabla 6.7, el sistema experto SVM con kernel polinomial muestra una tasa de aciertos ligeramente superior al sistema experto SVM con kernel RBF.

Tabla 6.7. Resultados máximo, mínimo y promedio obtenidos del sistema experto de clasificación SVM con kernel polinomial y con kernel Gaussiano RBF.

Mejor ajuste de parámetros

		Nivel de exploración alcanzado < lvl >	Límites de caja < C >	Orden Polinomial < d >	Tasa de Aciertos (%)
				Sigma (σ)	
SVM kernel Polinomial	Mínimo	1	0,031	3,0	98,34
	Máximo	1	32,000	6,0	98,89
	Promedio	1	16,133	5,0	98,65
SVM kernel RBF Gaussiano	Mínimo	0	1,000	1,0	96,96
	Máximo	0	1,000	1,0	97,38
	Promedio	0	1,000	1,0	97,20

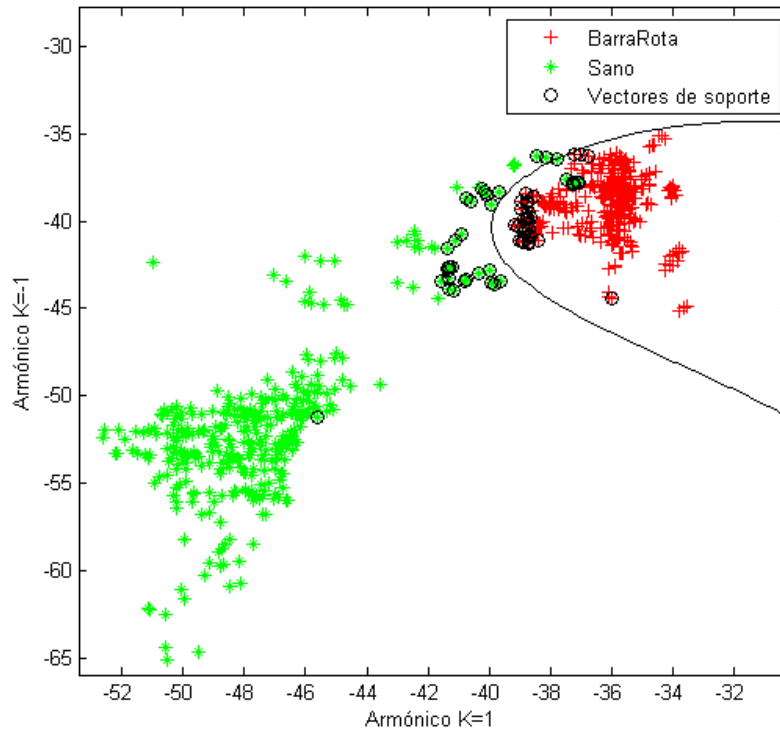


Figura 6.17. Espacio de características del sistema experto SVM con kernel polinomial, y la variante de HOTA para régimen transitorio con transformada SFFT y ventana Gaussiana.

Si el coste de tiempo de cálculo para clasificar las muestras no es relevante, el **mejor sistema de diagnóstico** obtenido por el sistema generador **respecto de un sistema experto SVM** con un ratio de aciertos del 98.89%, se centra en la **variante de HOTA con transformada SFFT y ventana Gaussiana** (tabla 6.5), y un **kernel polinomial**.

En la Figura 6.17, se muestra el espacio de características de este sistema de diagnóstico (**transformada SFFT, ventana Gaussiana y kernel polinomial**) que ha obtenido la mayor tasa de aciertos. En este caso se puede ver claramente como el hiperplano discriminante es capaz de separar casi sin problema los dos conjuntos de elementos en sus respectivas clases.

Si el coste de tiempo es relevante, se puede considerar como alternativa de **buen sistema de diagnóstico** con una **relación bastante alta de tasa de aciertos y coste de cálculo**, al sistema de diagnóstico óptimo basado en la **variante de HOTA con transformada SFFT, ventana Prolate** (tabla 6.5) y **kernel polinomial**. Este sistema

de diagnóstico ha mostrado una alta tasa de aciertos del 98.76% con un tiempo medio de clasificación para una muestra de 5.48 segundos.

6.5.2 Resultados obtenidos con sistemas expertos de Redes Neuronales Artificiales (ANN).

Al igual que se ha realizado en el apartado 5.7.2 para un régimen de trabajo estacionario, en este caso también se ha forzado el algoritmo del sistema principal para generar los resultados tanto para la red neuronal de 1 como de 2 capas ocultas, independientemente de la tasa de aciertos obtenida con la red de 1 capa oculta.

6.5.2.1 Comparación de tiempos entre ANN de 1 y de 2 capas ocultas.

En las siguientes tablas 6.8 (para una sola capa oculta) y 6.9 (para dos capas ocultas), se muestran los tiempos obtenidos durante el entrenamiento y los tiempos medios calculados para la clasificación de una muestra por cada variante del sistema de diagnóstico HOTA transitorio. Cada columna de la tabla representa:

- **Variación del sistema de diagnóstico HOTA transitorio.** Indica el tipo de técnica que ha sido utilizada internamente (STFT o SFFT) en HOTA transitorio y el tipo de función de ventana (Gaussiana o Prolate).
- **Tiempo de cómputo del ANN - Entrenamiento < Te-1 >.** Muestra en segundos el tiempo total de entrenamiento para la obtención de los mejores parámetros de ajuste del clasificador con los que se obtiene la tasa de aciertos más alta.
- **Tiempo de cómputo del ANN - Clasificación < Tc-1 >.** Muestra en segundos la media del tiempo requerido para clasificar las características representativas obtenidas de una sola muestra.
- **Tiempo de cómputo total - Entrenamiento < Te-2 >.** Muestra en segundos el tiempo total de entrenamiento de cada sistema de diagnóstico generado en cada fila. Este tiempo es la combinación de los tiempos mostrados en la tabla 5.2 para generar todas características representativas de las muestras (TTc) más el tiempo necesario para desarrollar el sistema experto ANN óptimo respecto de las características representativas obtenidas (Te-1).
- **Tiempo de cómputo total - Clasificación < Tc-2 >.** Muestra en milisegundos el tiempo total necesario para realizar el diagnóstico mediante la extracción de las características representativas de fallo de una muestra (TMm) y realizando su clasificación (Tc-1) con el sistema experto ANN.

Tabla 6.8. Resultados de los tiempos obtenidos del sistema experto de clasificación ANN de una sola capa oculta, y los tiempos totales de los sistemas de diagnóstico.

ANN con 1 capa oculta	Tiempo de cómputo del ANN		Tiempo de cómputo total	
	Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)
HOTA (STFT y ventana Gaussiana)	3.679,70	9,83	48.916,05	36.651,48
HOTA (STFT y ventana Prolate)	3.746,10	9,90	48.200,27	30.924,93
HOTA (SFFT y ventana Gaussiana)	4.340,63	9,90	10.891,16	5.607,32
HOTA (SFFT y ventana Prolate)	4.383,65	9,80	10.576,18	5.492,98

Tabla 6.9. Resultados de los tiempos obtenidos del sistema experto de clasificación ANN de dos capas ocultas, y los tiempos totales de los sistemas de diagnóstico.

ANN con 2 capas ocultas	Tiempo de cómputo del ANN		Tiempo de cómputo total	
	Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)
HOTA (STFT y ventana Gaussiana)	77.242,54	11,70	122.478,89	36.653,35
HOTA (STFT y ventana Prolate)	71.886,20	11,70	116.340,37	30.926,73
HOTA (SFFT y ventana Gaussiana)	66.034,40	10,00	72.584,93	5.607,42
HOTA (SFFT y ventana Prolate)	57.185,89	11,60	63.378,42	5.494,78

Respecto de los costes de tiempos en clasificación en sistema experto ANN (tablas 6.8 y 6.9), el tiempo promedio de clasificación es de 1 orden de magnitud mayor al mostrado por los sistemas expertos SVM (tablas 6.2 y 6.3). Sin embargo, estos tiempos continúan siendo tan poco significativos que pueden ser despreciados del tiempo total de clasificación.

Tabla 6.10. Resultados máximo, mínimo y promedio de los tiempos obtenidos del sistema experto de clasificación ANN para una sola capa oculta y para dos capas ocultas.

		Tiempo de cómputo del ANN		Tiempo de cómputo total	
		Entrenamiento < Te-1 > (s)	Clasificación < Tc-1 > (ms)	Entrenamiento < Te-2 > (s) (TTc+Te-1)	Clasificación < Tc-2 > (ms) (TMm+Tc-1)
ANN con 1 capa oculta	Mínimo	3.679,70	9,80	10.576,18	5.492,98
	Máximo	4.383,65	9,90	48.916,05	36.651,48
	Promedio	4.037,52	9,85	29.645,91	19.669,17
ANN con 2 capas ocultas	Mínimo	57.185,89	10,00	63.378,42	5.494,78
	Máximo	77.242,54	11,70	122.478,89	36.653,35
	Promedio	68.087,25	11,25	93.695,65	19.670,57

Comparando el tiempo en entrenamiento (tabla 6.10) entre el entrenamiento con una capa oculta (tabla 6.8) y dos capas ocultas (tabla 6.9), el coste temporal de entrenamiento con 2 capas ocultas aumenta en 1 orden de magnitud.

Comparando los tiempos del entrenamiento (tabla 6.10) con los tiempos de entrenamiento del sistema experto SVM (tabla 6.4), el sistema ANN es de 1 capa oculta desarrolla un tiempo similar a un sistema SVM con kernel polinomial. En todo caso, cuando el sistema ANN es de 2 capas ocultas el tiempo de entrenamiento es sensiblemente superior

El tiempo de cómputo total para la clasificación <Tc-2> es realmente el tiempo más importante a tener en cuenta, ya que este tiempo se representa como una aproximación al tiempo necesario para clasificar una nueva muestra desconocida para la variación de HOTA transitorio asociado a ese tiempo de clasificación.

Como ya se ha comentado anteriormente, el entrenamiento de un sistema experto ANN (ajuste del número de capas ocultas, de neuronas por capa, desarrollo de entropía cru-

zada y el entrenamiento local para valores aleatorios iniciales) no tiene un coste constante por lo que el tiempo de entrenamiento puede variar significativamente dependiendo del ajuste que realice el sistema principal para obtener el mejor sistema de diagnóstico.

6.5.2.2 Comparación de resultados entre ANN de 1 y 2 capas ocultas.

En las tablas 6.11 para ANN de una capa oculta y 6.12 para ANN de dos capas ocultas, se muestran los valores óptimos obtenidos, concernientes al número de neuronas óptimo por capa oculta y las mejores tasas de aciertos. Cada columna de la tabla representa:

- **Variación del sistema de diagnóstico HOTA transitorio.** Indica el tipo de técnica que ha sido utilizada internamente (STFT o SFFT) en HOTA transitorio y el tipo de función de ventana (Gaussiana o Prolate).
- **Intervalo de búsqueda de capa oculta.** Muestra el intervalo máximo del número de neuronas por capa oculta que ha sido explorado. Para cada capa oculta explorada de la red neuronal se muestra este intervalo.
- **Mejor ajuste del número de neuronas en capa oculta.** Se especifica en esta variable el número óptimo de neuronas en la o las capas ocultas que obtiene el mayor ratio de aciertos. Aunque la estructura de la red neuronal influye en su optimización también se ha de tener en cuenta otros parámetros como la selección de una buena función de transferencia y el entrenamiento correcto de la red sin sobreentrenamiento.
- **Tasa de aciertos.** Muestra la tasa de aciertos estimada por del sistema de diagnóstico óptimo formado por la variante de la técnica HOTA transitorio y por el sistema experto ANN entrenado de forma óptima.

Tabla 6.11. Resultados obtenidos respecto del clasificador experto de tipo ANN con una sola capa oculta.

ANN con 1 capa oculta	Mejor ajuste		
	Intervalo de búsqueda en capa oculta	Numero de neuronas en capa oculta 1	Tasa de Aciertos (%)
HOTA (STFT y ventana Gaussiana)	[1 - 7]	1	97,10
HOTA (STFT y ventana Prolate)	[1 - 7]	1	97,10
HOTA (SFFT y ventana Gaussiana)	[1 - 7]	2	97,38
HOTA (SFFT y ventana Prolate)	[1 - 7]	2	97,38

Tabla 6.12. Resultados obtenidos respecto del clasificador experto de tipo ANN con dos capas ocultas.

ANN con 2 capas ocultas	Intervalo de búsqueda		Mejor ajuste de neuronas		Tasa de Aciertos (%)
	Búsqueda en capa oculta 1	Búsqueda en capa oculta 2	Neuronas en capa oculta 1	Neuronas en capa oculta 2	
HOTA (STFT y ventana Gaussiana)	[0 - 11]	[1 - 16]	5	9	98,89
HOTA (STFT y ventana Prolate)	[0 - 11]	[1 - 16]	5	8	98,76
HOTA (SFFT y ventana Gaussiana)	[0 - 9]	[1 - 14]	0	8	98,89
HOTA (SFFT y ventana Prolate)	[0 - 8]	[1 - 14]	2	3	98,89

Como ya se ha mostrado anteriormente para el caso de ANN de dos capas ocultas, **la tasa de aciertos** cuando se genera un sistema ANN de dos capas ocultas (tabla 6.12) la tasa de aciertos es igual o superior a la correspondiente variación de HOTA transitorio con un sistema ANN de una sola capa oculta (tabla 6.11).

Al igual que también se ha verificado la exposición en Bishop [30] respecto al ajuste del número de capas de las redes neuronales. Como recordatorio, según Bishop cuando se obtiene con un sistema ANN de una capa oculta un rendimiento superior al 90%, la búsqueda de obtener una tasa mayor añadiendo una nueva capa oculta no compensa el coste de tiempo añadido respecto de la mejora en la tasa de aciertos que se va a conseguir (menor a un 2% de mejoría en la tasa de aciertos).

Como se puede observar en la tabla 6.13, las tasas de aciertos obtenidos son algo peculiares si las comparamos con la tabla 6.7 de los sistemas expertos de tipo SVM.

Tabla 6.13. Resultados máximo, mínimo y promedio obtenidos del sistema experto de clasificación ANN para una sola capa oculta y para dos capas ocultas.

Mejor ajuste de parámetros

		Búsqueda en capa oculta 1	Búsqueda en capa oculta 2	Neuronas en capa oculta 1	Neuronas en capa oculta 2	Tasa de Aciertos (%)
ANN con 1 capa oculta	Mínimo	[1 - 7]	-	1	-	97,10
	Máximo	[1 - 7]	-	2	-	97,38
	Promedio	-	-	1,5	-	97,25
ANN con 2 capas ocultas	Mínimo	[0 - 8]	[1 - 14]	0	3	98,76
	Máximo	[0 - 11]	[1 - 16]	5	9	98,89
	Promedio	-	-	3	7	98,86

Cuando el sistema experto ANN es de 1 capa oculta no parece afectar demasiado en la tasa de aciertos el tipo de función de ventana utilizado. Y la influencia de la variación entre una transformada STFT y SFFT en este caso también es mínima.

Cuando el sistema experto ANN es de 2 capas ocultas las diferencias de tasas de aciertos entre distintas variantes del HOTA transitorio parece que se desvanecen todavía más, igualándose la misma tasa de aciertos en la mayoría de casos.

Al igual que ocurría con los resultados obtenidos en el apartado 5.7 para régimen estacionario, la tasas de aciertos obtenidas tanto con un sistema experto SVM como con un sistema experto ANN son muy similares.

Aunque el coste de entrenamiento de los sistemas expertos ANN de dos capas ocultas es bastante alto, esta clase de sistema experto ha obtenido la tasa de aciertos más alta en casi todas las variantes de la técnica HOTA transitorio. Únicamente esta tasa de aciertos ha sido igualada solo con la configuración del sistema experto SVM con kernel polinomial para la variante de HOTA transitorio con transformada SFFT y ventana Gaussiana (tabla 6.5).

De los resultados obtenidos se deduce que, en el caso de que el fin último del sistema de diagnóstico sea el diagnóstico con la mayor probabilidad de aciertos independientemente del tiempo requerido, el uso de un sistema experto ANN de 2 capas ocultas es el más recomendable.

En el caso de que se busque un sistema de diagnóstico que esté bien balanceado entre la tasa de aciertos y el coste de cómputo, según los datos no existe un sistema idóneo sin perder en tasa de aciertos o ganar coste de cómputo. Solo se puede disminuir notablemente el coste de cómputo utilizando un sistema experto ANN de 1 capa oculta, pero igualmente en todos los casos la penalización en la tasa de aciertos va a ser mayor al 1.5%, aunque tampoco es una pérdida excesiva.

6.6 Conclusiones del capítulo

En todos los casos, tanto con sistemas expertos de tipo SVM como con sistemas expertos de tipo ANN, se han obtenido unas tasas de aciertos muy altas.

Además, como se puede observar en la anterior Figura 6.17 del espacio de características SVM de la variante del HOTA transitorio, que ha obtenido la mayor tasa de aciertos para este tipo de sistema experto; no existe ningún sobreentrenamiento que pueda afectar a la tasa de aciertos incrementando esta de forma artificial.

Teniendo en cuenta estas las tasas de acierto, se sigue manteniendo la premisa de que estas tasas de aciertos tan similares obtenidas de dos sistemas expertos tan diferentes verifican nuevamente que la implementación de los algoritmos automáticos que se han desarrollado para los ajustes de parámetros, de estructuras, de coeficientes, algoritmos de entrenamiento, validación de los sistemas expertos y sistemas de detección y anulación del sobreentrenamiento, desarrollados respectivamente para el sistema experto SVM (apartado 5.4) y para el sistema experto ANN (apartado 5.5) son totalmente correctos.

Además, al igual que ya se vio en los resultados del capítulo 5, la obtención de tasas de acierto tan altas con la mayoría de variantes de la técnica HOTA transitorio sugiere que el proceso de obtener sistemas expertos óptimos para un régimen de trabajo transitorio funciona correctamente.

Muchas de las conclusiones que se han podido extraer de los resultados obtenidos en este capítulo 6 de sistemas de diagnóstico de fallo óptimos para régimen de trabajo transitorio ya se habían observado en los resultados obtenidos en el capítulo 5 de sistemas de diagnóstico de fallos óptimos para régimen de trabajo estacionario.

Sin embargo, en este caso el número de muestras de corrientes de fase utilizadas ha sido en total de 735, muy superior a las 216 muestras usadas para régimen estacionario en el capítulo 5. Por lo que aparte de corroborar estas premisas, esta nueva validación se ha realizado además utilizando un conjunto de características representativas del problema mucho más amplio.

En conclusión, comparando los resultados de este capítulo 6 con los resultados obtenidos en el capítulo 5, **se deduce la posibilidad de utilizar únicamente el sistema de diagnóstico óptimo para régimen transitorio (tanto para régimen transitorio como para régimen estacionario) basado en la técnica HOTA transitorio en su variante de transformada SFFT con una ventana Prolate o Gaussiana y un sistema experto ANN de 2 capas ocultas.** Ya que, cualquiera de estas dos variantes muestra una tasa de aciertos excepcionalmente alta con un coste temporal de diagnóstico de nuevas muestras igual al obtenido para los sistemas de diagnóstico para régimen de trabajo estacionario del capítulo 5.

Capítulo 7

Conclusiones, aportaciones y futuras líneas de investigación

7.1 Conclusiones y aportaciones

En esta tesis doctoral se ha realizado una revisión exhaustiva acerca del diagnóstico de las máquinas eléctricas, y más concretamente de las máquinas asíncronas de inducción; analizándose los principales fallos que se producen, las principales magnitudes medibles de las mismas, las técnicas y/o métodos de diagnóstico basadas en la medida de la corriente en regímenes de funcionamiento estacionario y transitorio de la máquina. También se ha hecho una revisión de los principales sistemas expertos de clasificación.

Partiendo de la base de este conocimiento previo se ha desarrollado un suprasistema que es capaz de generar sistemas optimizados de diagnóstico de motores de inducción. Este suprasistema es capaz de seleccionar la combinación más óptima de técnica de diagnóstico y sistema experto para un régimen de trabajo de un tipo de motor asíncrono de inducción, tal y como se recoge en los resultados mostrados en el apartado 5.7 y apartado 6.5.

En la presente tesis se ha aplicado sólo a la detección de motores con barra rota; principalmente, porque de forma experimental se puede reproducir este fallo de una forma sencilla. La ampliación a otros fallos es inmediata si se conoce la característica de fallo (Tabla 2.2 del apartado 2.2.3) y se plantea su extensión para próximos trabajos.

Este suprasistema presenta unos resultados excelentes para el análisis en **régimen estacionario** de un 100% de tasa de acierto aplicando la técnica de diagnóstico HOTA y con clasificadores de tipo SVM y ANN. Es resaltable que mediante dos sistemas de inteligencia artificial tan dispares entre sí como son el SVM y las ANN se obtengan los mismos resultados, sólo justificables desde el punto de vista de un **correcto desarrollo de los algoritmos automáticos de ajustes y optimización**.

Se han medido diversos tiempos, obteniendo las mediciones equivalentes respecto al tiempo de entrenamiento necesario y el tiempo requerido para realizar un diagnóstico de fallos. En base a los resultados es notable la reducción de tiempos obtenidos en los sistemas expertos de tipo SVM frente a las ANN.

Además de los resultados principales obtenidos en esta tesis (tasas de aciertos de sistemas de diagnóstico, tiempos de computo, etc.), se han obtenido otros datos no relacionados directamente con el tema principal de la tesis, pero que pueden ser utilizados en otras investigaciones en el futuro para mejorar otras metodologías de diagnóstico de fallos.

Una de las mayores novedades presentadas en esta tesis es la aplicación de sistemas clasificatorios en **régimen de trabajo transitorio**. Este suprasistema también presenta unos resultados muy satisfactorios para los análisis en régimen transitorio para ambos clasificadores ANN y SVM, volviéndose a comprobar que el algoritmo de optimización empleado funciona perfectamente. Para el entrenamiento de estos sistemas se han empleado tanto las muestras provenientes de los ensayos en régimen estacionario como de los transitorios, ampliando mucho, para estos últimos, las características representativas, y por consiguiente aumentándose los tiempos debido a que el número de muestras ha aumentado considerablemente.

Se han propuesto varias **mejoras a la técnica de diagnóstico** empleada en el análisis de fallos en régimen transitorio, HOTA.

- La primera es el uso de una ventana Prolate, ya que esta concentra un mayor nivel de energía frente al uso tradicional de la ventana Gaussiana; lo cual, hace que la resolución sea mayor si cabe. Además, desde un punto de vista de optimización del algoritmo en su implementación, agiliza los tiempos. Aunque experimentalmente no se ha observado una mejora en la tasa de aciertos gracias a la mayor resolución de esta ventana, sí que se ha observado una disminución del tiempo de cómputo.
- La segunda es la aplicación de la transformada en frecuencia en vez de en tiempo, es decir el uso de la SFFT en vez de la STFT. Esta transformación SFFT ha aportado principalmente una notable reducción en el tiempo de cálculo necesario, así como de la memoria de almacenamiento de las variables.

En base a los resultados, **se propone** la posibilidad de utilizar exclusivamente el sistema de diagnóstico óptimo para régimen transitorio basado en la técnica HOTA transitorio en su variante de transformada SFFT con una ventana Prolate o Gaussiana y un sistema experto ANN de 2 capas ocultas. Ya que, cualquiera de estas dos variantes muestra una tasa de aciertos excepcionalmente alta con un coste temporal de diagnóstico de nuevas muestras similar al tiempo de diagnóstico obtenido para los sistemas de diagnóstico generados para régimen de trabajo estacionario del capítulo 5.

Finalmente, con este suprasistema generador y con todo el conjunto de resultados obtenidos en las pruebas experimentales se puede confirmar fehacientemente que se han alcanzado los objetivos propuestos en esta investigación de tesis. Además, con esta investigación se ha abierto nuevas puertas que posibilitan desarrollar nuevas mejoras en el futuro para el diagnóstico de maquinaria rotatoria (apartado 7.2).

Artículos en Revistas Internacionales:

- *Optimized Spheroidal Prolate Wavelet Function As STFT Window for Transient Motor Current Signature Analysis.* Enviado a **IEEE Transactions on Industrial Electronics** el 29 Octubre 2015. Revisado el 27 de enero de 2016 y Segunda revisión el 17 de Mayo de 2016
- *The Harmonic Order Tracking Analysis Method for the Fault Diagnosis in Induction Motors Under Time-Varying Conditions.* Enviado a **IEEE Transactions on Energy Conversion** el 9 Marzo 2016. Revisado el 13 de junio de 2016
- *Short Frequency Fourier Transform for Fault Diagnosis of Induction Machines Working in Transient Regime.* Enviado a **IEEE Transactions on Instrumentation and Measurement** el 11 de abril de 2016. Revisado el 1 de julio de 2016

Artículos en Congresos Internacionales:

- *Support Vector Machines optimization for Steady State diagnosis methods of Induction Motors. A comparative study.* Aceptado para su publicación en IEEE XXIIth International Conference on electrical Machines (ICEM) 4-7, Septiembre 2016
- *Multilayer Park's Vector Approach, a method for fault detection on induction motors* Publicado en IEEE International Conference on Industrial Technology (ICIT) 17-19, Marzo 2015
- *Study of Performance of Several Techniques of Fault Diagnosis for Induction Motors in Steady-State with SVM Learning Algorithms* Publicado en IEEE 2014 2nd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)
- *A new multilevel approach for Programmable Logical Controller (PLCs).* Publicado en IEEE International Conference on Industrial Technology (ICIT) 17-19, Marzo 2015
- *Motor current signal analysis on programmable logic controller study* Publicado en IEEE XXIth International Conference on electrical Machines (ICEM) 3-6, Septiembre 2014
- *DC Internal Inductance for a Conductor of Rectangular Cross Section computed with the Proper Generalized Decomposition* Publicado en 17th International Symposium on electromagnetic Fields (ISEF) 10-12, Septiembre 2015

7.2 Futuras líneas de investigación.

La tesis ha dejado varias líneas de investigación abiertas para que sean exploradas en un futuro:

- Aplicación a otros tipos de fallos o averías de máquinas eléctricas, como por ejemplo excentricidad, fallos en cojinetes, cortocircuitos entre espiras, etc.
- Aplicación a otras magnitudes o incluso combinación de las magnitudes, tales como vibraciones, potencia, etc.
- Extensión del sistema aplicado al desarrollo de sistemas de diagnóstico óptimos en las que el motor trabaje en modo generador.
- Extensión a otras técnicas de diagnóstico de fallos aplicadas al transitorio, como pudieran ser la frecuencia instantánea, la transformada de Gabor, etc
- Extensión de los sistemas expertos empleados, como pudieran ser redes bayesianas, redes neuronales difusas, etc.
- Implementación en equipos autónomos de bajo coste (DSP, FPGA,...) destinados al diagnóstico de máquinas eléctricas.
- Integración de estos nuevos sistemas en los equipos electrónicos de control, como puede ser un convertidor de frecuencia.

Capítulo 8

Referencias

- [1] Matić, D., Kulić, F., Pineda-Sánchez, M., & Kamenko, I. Support vector machine classifier for diagnosis in electrical machines: Application to broken bar. *Expert Systems with Applications*, 2012, vol. 39, no 10, p. 8681-8689.
- [2] Nandi, S., Toliyat, H. A., & Li, X. Condition monitoring and fault diagnosis of electrical motors-a review. *Energy Conversion, IEEE Transactions on*, 2005, vol. 20, no 4, p. 719-729.
- [3] Thorsen, O. V., & Dalva, M. (1999). Failure identification and analysis for high-voltage induction motors in the petrochemical industry. *Industry Applications, IEEE Transactions on*, 35(4), 810-818.
- [4] Bell, R. N., McWilliams, D. W., O'DONNELL, P., Singh, C., & Wells, S. J.. Report of large motor reliability survey of industrial and commercial installations. I. *IEEE Transactions on Industry applications*, 1985, vol. 21, no 4, p. 853-864.
- [5] Clay, H. R., & Wanless, J. W. Life expectancy of motors in mild nuclear plant environments. *Final report. NUS Corp.*, Menlo Park, CA (USA); Brancato (EL), Clarksville, MD (USA), 1985.
- [6] Puche, R. Nuevos Métodos de Diagnósis de Excentricidad y otras Asimetrías Rotóricas en Máquinas Eléctricas de Inducción a través del Análisis de la Corriente Estatística. *Tesis Doctoral, Universidad Politécnica de Valencia*, 2008.

- [7] Jung, J. H., Jong-Jae, L., & Kwon, B. H. Online diagnosis of induction motors using MCSA. *Industrial Electronics, IEEE Transactions on*, 2006, vol. 53, no 6, p. 1842-1852.
- [8] Xu, M., & Marangoni, R. D. (1994). Vibration analysis of a motor-flexible coupling-rotor system subject to misalignment and unbalance, Part I: theoretical model and analysis. *Journal of Sound and Vibration*, 176(5), 663-679.
- [9] Sapena Bañó, Á. Desarrollo de técnicas avanzadas de diagnóstico de máquinas eléctricas válidas para cualquier régimen de funcionamiento. 2014. *Tesis Doctoral. Universidad Politécnica de Valencia*.
- [10] Bellini, A., Filippetti, F., Franceschini, G., Tassoni, C., & Kliman, G. B. Quantitative evaluation of induction motor broken bars by means of electrical signature analysis. *Industry Applications, IEEE Transactions on*, 2001, vol. 37, no 5, p. 1248-1255.
- [11] Antonino-Daviu, J. A., Riera-Guasp, M., Folch, J. R., & Palomares, M. Validation of a new method for the diagnosis of rotor bar failures via wavelet transform in industrial induction machines. *Industry Applications, IEEE Transactions on*, 2006, vol. 42, no 4, p. 990-996.
- [12] Deleroi, W. Squirrel cage motor with broken bar in the rotor—physical phenomena and their experimental assessment. *En Proc. of Int. Conf. On Electrical Machines*. 1982. p. 767-770.
- [13] Castelli, M. Desarrollo de un equipo para la realización de mantenimiento predictivo en motores asíncronos de gran porte. *Available from internet: <URL: http://www.uruman.org/TrabajosTec/Marcelo_Castelli.pdf>*, 2006.
- [14] Kliman, G. B., Koegl, R. A., Stein, J., Endicott, R. D., & Madden, M. W. Noninvasive detection of broken rotor bars in operating induction motors. *Energy Conversion, IEEE Transactions on*, 1988, vol. 3, no 4, p. 873-879.
- [15] Schoen, R. R., Habetler, T. G., Kamran, F., & Bartfield, R. G. Motor bearing damage detection using stator current monitoring. *Industry Applications, IEEE Transactions on*, 1995, vol. 31, no 6, p. 1274-1279.
- [16] Prieto, M. D., Cirrincione, G., Espinosa, A. G., Ortega, J. A., & Henao, H. Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. *Industrial Electronics, IEEE Transactions on*, 2013, vol. 60, no 8, p. 3398-3407.
- [17] Blödt, M., Granjon, P., Raison, B., & Regnier, J. Mechanical fault detection in induction motor drives through stator current monitoring-theory and application examples. *Fault Detection*, 2010, p. 451-488.

- [18] Kral, C., Habetler, T. G., & Harley, R. G. Detection of mechanical imbalances of induction machines without spectral analysis of time-domain signals. *Industry Applications, IEEE Transactions on*, 2004, vol. 40, no 4, p. 1101-1106.
- [19] Nandi, S., Ahmed, S., & Toliyat, H. A. Detection of rotor slot and other eccentricity related harmonics in a three phase induction motor with different rotor cages. *Energy Conversion, IEEE Transactions on*, 2001, vol. 16, no 3, p. 253-260.
- [20] Dorrell, D. G., Thomson, W. T., & Roach, S. Analysis of airgap flux, current, and vibration signals as a function of the combination of static and dynamic airgap eccentricity in 3-phase induction motors. *Industry Applications, IEEE Transactions on*, 1997, vol. 33, no 1, p. 24-34.
- [21] Zhang, P., Du, Y., Habetler, T. G., & Lu, B. A survey of condition monitoring and protection methods for medium-voltage induction motors. *Industry Applications, IEEE Transactions on*, 2011, vol. 47, no 1, p. 34-46.
- [22] Elawgali, S. Effect of combined eccentricities on the spectrum of induction machine currents. *In XL International Symposium on Electrical Machines, SME*. 2004. p. 15-18.
- [23] Nandi, S., Bharadwaj, R. M., & Toliyat, H. A. Performance analysis of a three-phase induction motor under mixed eccentricity condition. *Energy Conversion, IEEE Transactions on*, 2002, vol. 17, no 3, p. 392-399.
- [24] Li, X., Wu, Q., & Nandi, S. Performance analysis of a three-phase induction machine with inclined static eccentricity. *Industry Applications, IEEE Transactions on*, 2007, vol. 43, no 2, p. 531-541.
- [25] Ying Xie, Chenglin Gu, & Wenping Cao. Study of broken bars in three phase squirrel cage induction motors at standstill. *International Transactions on Electrical Energy Systems*, 2013, vol. 23, no 7, p. 1124-1138.
- [26] Williamson, S., & Smith, A. C. Steady-state analysis of 3-phase cage motors with rotor-bar and end-ring faults. *In IEE Proceedings B (Electric Power Applications). IET Digital Library*, 1982. p. 93-100.
- [27] Cruz, S. M., & Cardoso, A. M. Stator winding fault diagnosis in three-phase synchronous and asynchronous motors, by the extended Park's vector approach. *Industry applications, IEEE Transactions on*, 2001, vol. 37, no 5, p. 1227-1233.
- [28] Díaz Sánchez, D. Diagnóstico de fallas en motores de inducción tipo jaula de ardilla mediante la aplicación de métodos híbridos. *Doctoral dissertation. Faculty of Engineering. Del Valle University (Santiago de Cali)*. 2011.

- [29] Burriel-Valencia, J., Sapena-Bano, A., Pineda-Sanchez, M., & Martinez-Roman, J. Multilayer Park's vector approach, a method for fault detection on induction motors. In *Industrial Technology (ICIT), 2015 IEEE International Conference on*. IEEE, 2015. p. 775-780.
- [30] Zarei, J., & Poshtan, J. An advanced Park's vectors approach for bearing fault detection. *Tribology International*, 2009, vol. 42, no 2, p. 213-219.
- [31] Cruz, S. M., Toliyat, H. A., & Cardoso, A. M. DSP implementation of the multiple reference frames theory for the diagnosis of stator faults in a DTC induction motor drive. *Energy Conversion, IEEE Transactions on*, 2005, vol. 20, no 2, p. 329-335.
- [32] El Hachemi Benbouzid, M. (2000). A review of induction motors signature analysis as a medium for faults detection. *Industrial Electronics, IEEE Transactions on*, 47(5):984-993.
- [33] Mehrjou, M., Mariun, N., Hamiruce Marhaban, M., & Misron, N. Rotor fault condition monitoring techniques for squirrel-cage induction machine—A review. *Mechanical Systems and Signal Processing*, 2011, vol. 25, no 8, p. 2827-2848.
- [34] Puche-Panadero, R., Pineda-Sanchez, M., Riera-Guasp, M., Roger-Folch, J., Hurtado-Perez, E., & Perez-Cruz, J. Improved resolution of the MCSA method via Hilbert transform, enabling the diagnosis of rotor asymmetries at very low slip. *Energy Conversion, IEEE Transactions on*, 2009, vol. 24, no 1, p. 52-59.
- [35] Sapena-Bano, A., Martinez-Roman, J., Perez-Cruz, J., Pineda-Sanchez, M., Roger-Folch, J., Riera-Guasp, M., & Puche-Panadero, R.. Harmonic order tracking analysis: A novel method for the diagnosis of induction generators. In *Electrical Machines (ICEM), 2014 International Conference on*. IEEE, 2014. p. 1765-1771.
- [36] Sapena-Bano, A., Pineda-Sanchez, M., Puche-Panadero, R., Perez-Cruz, J., Roger-Folch, J., Riera-Guasp, M., & Martinez-Roman, J. Harmonic Order Tracking Analysis: A Novel Method for Fault Diagnosis in Induction Machines. *Energy Conversion, IEEE Transactions on*, 2015, vol. 30, no 3, p. 833-841.
- [37] Liu, Z., Yin, X., Zhang, Z., Chen, D., & Chen, W. Online rotor mixed fault diagnosis way based on spectrum analysis of instantaneous power in squirrel cage induction motors. *Energy Conversion, IEEE Transactions on*, 2004, vol. 19, no 3, p. 485-490.
- [38] Khezzar, A., El Kamel Oumaamar, M., Hadjami, M., Boucherma, M., & Razik, H. Induction motor diagnosis using line neutral voltage signatures. *Industrial Electronics, IEEE Transactions on*, 2009, vol. 56, no 11, p. 4581-4591.

- [39] Abhisek Ukil, Andrea Andenna, & Shuo Chen. Detection of stator short circuit faults in three-phase induction motors using motor current zero crossing instants. *Electric Power Systems Research*, 2011, vol. 81, no 4, p. 1036-1044.
- [40] Çalış, H., & Çakır, A. Rotor bar fault diagnosis in three phase induction motors by monitoring fluctuations of motor current zero crossing instants. *Electric Power Systems Research*, 2007, vol. 77, no 5, p. 385-392.
- [41] Arabacı, H., & Bilgin, O. Neural network classification and diagnosis of broken rotor bar faults by means of short time Fourier transform. *In Proceedings of the International MultiConference of Engineers and Computer Scientists*. 2009.
- [42] Riera-Guasp, M., Pineda-Sanchez, M., Perez-Cruz, J., Puche-Panadero, R., Roger-Folch, J., & Antonino-Daviu, J. A. Diagnosis of induction motor faults via Gabor analysis of the current in transient regime. *Instrumentation and Measurement, IEEE Transactions on*, 2012, vol. 61, no 6, p. 1583-1596.
- [43] Friedlander, B., & Porat, B. (1989). Detection of transient signals by the Gabor representation. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 1989, vol. 37, no 2, p. 169-180.
- [44] Cui, J., & Wong, W. Optimal window length in the windowed adaptive chirplet analysis of visual evoked potentials. *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE. IEEE*, 2006. p. 4580-4583.
- [45] Zielinski, T. P. Joint time-frequency resolution of signal analysis using Gabor transform. *Instrumentation and Measurement, IEEE Transactions on*, 2001, vol. 50, no 5, p. 1436-1444.
- [46] Szyper, M. (1995). Time domain window. *Electronics Letters*, 1995. vol. 31, no 9, p. 707-708.
- [47] Gabor, D. Theory of communication. Part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of*, 1946, vol. 93, no 26, p. 429-441.
- [48] Jones, D. L., & Parks, T. W. (1992). A resolution comparison of several time-frequency representations. *Signal Processing, IEEE Transactions on*, 40(2), 413-420.
- [49] Meyer, Y. Wavelets-algorithms and applications. *Wavelets-Algorithms and applications Society for Industrial and Applied Mathematics Translation.*, 142 p., 1993, vol. 1.

- [50] Mallat, S. A wavelet tour of signal processing. *Book. Academic press*, 1999.
- [51] GRAPS, Amara. An introduction to wavelets. *Computational Science & Engineering, IEEE*, 1995, vol. 2, no 2, p. 50-61.
- [52] Antonino-Daviu, J. A., Lee, S. B., & Wiedenbrug, E. Reliable detection of rotor bar failures in induction motors operating in petrochemical plants. *In Petroleum and Chemical Industry Conference Europe*, 2014. IEEE, 2014. p. 1-9.
- [53] Giri, F (ed.). AC electric motors control: Advanced design techniques and applications. *Book, John Wiley & Sons*, 2013.
- [54] Pineda-Sanchez, M., Riera-Guasp, M., Perez-Cruz, J., & Puche-Panadero, R. Transient motor current signature analysis via modulus of the continuous complex wavelet: A pattern approach. *Energy Conversion and Management*, 2013, vol. 73, p. 26-36.
- [55] Huang, N. E., & Shen, S. S. Hilbert-Huang transform and its applications. *World Scientific*, 2014.
- [56] Antonino-Daviu, J., Rodriguez, P. J., Riera-Guasp, M., Arkkio, A., Roger-Folch, J., & Perez, R. B. Transient detection of eccentricity-related components in induction motors through the Hilbert–Huang Transform. *Energy Conversion and Management*, 2009, vol. 50, no 7, p. 1810-1820.
- [57] Peng, Z. K., Peter, W. T., & Chu, F. L. A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. *Mechanical systems and signal processing*, 2005, vol. 19, no 5, p. 974-988.
- [58] Hlawatsch, F., & Boudreaux-Bartels, G. F. (1992). Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 9(2), 21-67.
- [59] Cohen, L. Time-frequency distributions-a review. *Proceedings of the IEEE*, 1989, vol. 77, no 7, p. 941-981.
- [60] Climente-Alarcon, V., Antonino-Daviu, J. A., Riera-Guasp, M., Puche-Panadero, R., & Escobar, L. Application of the Wigner–Ville distribution for the detection of rotor asymmetries and eccentricity through high-order harmonics. *Electric Power Systems Research*, 2012, vol. 91, p. 28-36.
- [61] Pineda-Sanchez, M., Riera-Guasp, M., Antonino-Daviu, J. A., Roger-Folch, J., Perez-Cruz, J., & Puche-Panadero, R. Instantaneous frequency of the left side-band harmonic during the start-up transient: A new method for diagnosis of

- broken bars. *Industrial Electronics, IEEE Transactions on*, 2009, vol. 56, no 11, p. 4557-4570.
- [62] Pineda-Sanchez, M., Riera-Guasp, M., Roger-Folch, J., Antonino-Daviu, J. A., Perez-Cruz, J., & Puche-Panadero, R. Diagnosis of induction motor faults in time-varying conditions using the polynomial-phase transform of the current. *Industrial Electronics, IEEE Transactions on*, 2011, vol. 58, no 4, p. 1428-1439.
- [63] Pineda-Sanchez, M., Riera-Guasp, M., Antonino-Daviu, J. A., Roger-Folch, J., Perez-Cruz, J., & Puche-Panadero, R. Diagnosis of induction motor faults in the fractional Fourier domain. *Instrumentation and Measurement, IEEE Transactions on*, 2010, vol. 59, no 8, p. 2065-2075.
- [64] Benbouzid, M. E. H., Vieira, M., & Theys, C. Induction motors' faults detection and localization using stator current advanced signal processing techniques. *Power Electronics, IEEE Transactions on*, 1999, vol. 14, no 1, p. 14-22.
- [65] Nandi, S., Ilamparithi, T. C., Lee, S. B., & Hyun, D. Detection of eccentricity faults in induction machines based on nameplate parameters. *Industrial Electronics, IEEE Transactions on*, 2011, vol. 58, no 5, p. 1673-1683.
- [66] Dorrell, D. G. Sources and characteristics of unbalanced magnetic pull in three-phase cage induction motors with axial-varying rotor eccentricity. *Industry Applications, IEEE Transactions on*, 2011, vol. 47, no 1, p. 12-24.
- [67] Féki, N., Clerc, G., & Vexex, P. Gear and motor fault modeling and detection based on motor current analysis. *Electric power systems research*, 2013, vol. 95, p. 28-37.
- [68] Herrera Perez, E. Comunicaciones II: comunicación digital y ruido: una introducción a la teoría de la comunicación digital y el ruido. *Editorial Limusa*, 2002.
- [69] Kar, C., & Mohanty, A. R. Monitoring gear vibrations through motor current signature analysis and wavelet transform. *Mechanical systems and signal processing*, 2006, vol. 20, no 1, p. 158-187.
- [70] Stoica, P., & Moses, R. L. Introduction to spectral analysis. *Upper Saddle River: Prentice hall*, 1997.
- [71] Hekkala, A., Harjula, I., Panaitopol, D., Rautio, T., & Pacalet, R. Cooperative spectrum sensing study using welch periodogram. *In Telecommunications (ConTEL), Proceedings of the 2011 11th international conference on*. IEEE, 2011. p. 67-74.

- [72] Cupertino, F., De Vanna, E., Salvatore, L., & Stasi, S. Comparison of spectral estimation techniques applied to induction motor broken bars detection. In *Diagnostics for Electric Machines, Power Electronics and Drives, 2003. SDEMPED 2003. 4th IEEE International Symposium on. IEEE*, 2003. p. 129-134.
- [73] Ayhan, B., Chow, M. Y., & Song, M. H. Multiple signature processing-based fault detection schemes for broken rotor bar in induction motors. *Energy Conversion, IEEE Transactions on*, 2005, vol. 20, no 2, p. 336-343.
- [74] Rabiner, L., & Juang, B. H.. Fundamentals of speech recognition. *Book, Prentice Hall*, 1993.
- [75] Magnotta, V. A., Heckel, D., Andreasen, N. C., Cizadlo, T., Corson, P. W., Ehrhardt, J. C., & Yuh, W. T. Measurement of brain structures with artificial neural networks: two-and three-dimensional applications 1. *Radiology*, 1999, vol. 211, no 3, p. 781-790.
- [76] Newell, A., & Simon, H. A.. GPS, a program that simulates human thought. *RAND CORP SANTA MONICA CALIF*, 1961.
- [77] Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A., & Lederberg, J. DENDRAL: a case study of the first expert system for scientific hypothesis formation. *Artificial intelligence*, 1993, vol. 61, no 2, p. 209-261.
- [78] Colmerauer, A., & Roussel, P. The birth of Prolog. En History of programming languages-II. *ACM*, 1996. p. 331-367.
- [79] O'Brien, D. P. Conditionals and disjunctions in mental-logic theory: A response to Liu and Chou (2012) and to López-Astorga (2013). *Universum*, 2014, vol. 29, no 2, p. 221-235.
- [80] Friedl, M. A., & Brodley, C. E. Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment*, 1997, vol. 61, no 3, p. 399-409.
- [81] Silveyman, B. W., Jones, M. C., Fix, E., & Hodges, J. L. An important contribution to nonparametric discriminant analysis and density estimation. *International Statistical Review*, 1951, vol. 57, no 3, p. 233-247.
- [82] Chang, E. I., & Lippmann, R. P. Using genetic algorithms to improve pattern classification performance. In *Advances in neural information processing systems*. 1991. p. 797-803.
- [83] Cortes, C., & Vapnik, V. Support-vector networks. *Machine learning*, 1995, vol. 20, no 3, p. 273-297.

- [84] McCulloch, W. S., & Pitts, W. A. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 1943, vol. 5, no 4, p. 115-133.
- [85] Demuth, H., Beale, M., & Hagan, M. Neural network toolbox™ 6. *User's guide*, 2008.
- [86] Antonino-Daviu, J. A., Riera-Guasp, M., Pineda-Sanchez, M., & Perez, R. B. A critical comparison between DWT and Hilbert–Huang-based methods for the diagnosis of rotor bar failures in induction machines. *Industry Applications, IEEE Transactions on*, 2009, vol. 45, no 5, p. 1794-1803.
- [87] Hassen Keskes, Ahmed Braham, & Zied Lachiri. Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM. *Electric Power Systems Research*, 2013, vol. 97, p. 151-157.
- [88] Sotomayor, G., & Esteban, P. Aportaciones al mantenimiento predictivo de motores de inducción mediante una metodología de diagnóstico basada en el uso combinado de técnicas estadísticas y redes neuronales artificiales. *Doctoral dissertation. Electrical engineering department. Valladolid University (Spain)*. 2013.
- [89] Hammo, R. Faults Identification in Three-Phase Induction Motors Using Support Vector Machines. *PhD. Thesis. Master of Technology Management-Engineering Technology*. 2014.
- [90] García, A. G. & Muñoz-Bouzo, M. J. Espacios de Hilbert y Análisis de Fourier: los primeros pasos (2ª edición). *Ed. Sanz y Torres, Madrid*. 2014.
- [91] Hsu, C. W., Chang, C. C., & Lin, C. J. A practical guide to support vector classification. *Department of Computer Science, National Taiwan University*, 2003, p. 1-16.
- [92] Henao, R., & Gómez, J. E. H. Selección de hiperparámetros en máquinas de soporte vectorial. *Universidad Nacional de Colombia*, 2004.
- [93] Chapelle, O., Vapnik, V., Bousquet, O., & Mukherjee, S. Choosing multiple parameters for support vector machines. *Machine learning*, 2002, vol. 46, no 1-3, p. 131-159.
- [94] Vapnik, V. N., & Vapnik, V. Statistical learning theory. *New York: Wiley*, 1998.
- [95] Rechenberg, I. The evolution strategy. a mathematical model of darwinian evolution. *In Synergetics—from microscopic to macroscopic order. Springer Berlin Heidelberg*, 1984. p. 122-132.

- [96] Zien, A., Rätsch, G., Mika, S., Schölkopf, B., Lengauer, T., & Müller, K. R. Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics*, 2000, vol. 16, no 9, p. 799-807.
- [97] Anderson, T. W., & Bahadur, R. R. Classification into two multivariate normal distributions with different covariance matrices. *Annals of Mathematical Statistics*, 1962, vol. 33, no 2, p. 420-431.
- [98] Manning, C. D., Raghavan, P., & Schütze, H. Introduction to information retrieval. *Cambridge: Cambridge university press*, 2008.
- [99] Boser, B. E., Guyon, I. M., & Vapnik, V. N. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory. ACM*, 1992. p. 144-152.
- [100] Schölkopf, B., Burgest, C., & Vapnik, V. Extracting support data for a given task. *no. x*, 1995.
- [101] Lin, H. T., & Lin, C. J. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. *Neural Computation*, 2003, p. 1-32.
- [102] Martínez Ruedas, C. Detección multiusuario para DS-CDMA basado en SVM. *Degree thesis. Technical college of engineering. Sevilla University (Spain)*. 2006.
- [103] Machado, P., Wade, J., & McGinnity, T. M. Si elegans: FPGA hardware emulation of *C. elegans* nematode nervous system. In *Nature and Biologically Inspired Computing (NaBIC), 2014 Sixth World Congress on. IEEE*, 2014. p. 65-71.
- [104] Rich, E., & Knight, K. Inteligencia artificial. *McGraw-Hill*. 1994.
- [105] Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 1958, vol. 65, no 6, p. 386.
- [106] Widrow, B., & Angell, J. B. Reliable, trainable networks for computing and control. *Institute of the Aerospace Sciences*, 1962.
- [107] Minsky, M., & Papert, S. Perceptron: an introduction to computational geometry. *The MIT Press, Cambridge, expanded edition*, 1969, vol. 19, no 88, p. 2.
- [108] Werbos, P.J. Beyond regression: New tools for prediction and analysis in the behavioral sciences. *PhD thesis, Harvard University*. 1974.
- [109] Isasi, P., & Galván, I. Redes neuronales artificiales—un enfoque práctico. *Pearson-Prentice Hall*. 2004.

- [110] Martín-del-Brio, B., & Sanz-Molina, A. Redes neuronales y sistemas borrosos. *Ra-Ma, Madrid*, 2001
- [111] Hagan, M. T., Demuth, H. B., & Beale, M. H. Neural network design. Boston: PWS publishing company, 1996.
- [112] Sharma, B., & Venugopalan, K. Comparison of neural network training functions for Hematoma classification in brain CT images. *Int J Comput Sci Eng*, 2014, vol. 16, no 1, p. 31-35.
- [113] Bertona, F. Entrenamiento de Redes Neuronales basado en Algoritmos Evolutivos. Buenos Aires, *Argentina: Universidad de Buenos Aires*, 2005.
- [114] Beale, M. H., Hagan, M. T., & Demuth, H. B. Neural network toolbox 7. *User's Guide, MathWorks*, 2010.
- [115] Anastasiadis, A. D., Magoulas, G. D., & Vrahatis, M. N. New globally convergent training scheme based on the resilient propagation algorithm. *Neurocomputing*, 2005, vol. 64, p. 253-270.
- [116] Møller, M. F. A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, 1993, vol. 6, no 4, p. 525-533.
- [117] Fletcher, R., & Reeves, C. M. Function minimization by conjugate gradients. *The computer journal*, 1964, vol. 7, no 2, p. 149-154.
- [118] Sandhu, P. S., & Chhabra, S. A comparative analysis of Conjugate Gradient Algorithms & PSO Based Neural Network approaches for reusability evaluation of procedure based software systems. *Chiang Mai j. Sci*, 2011, vol. 38, p. 123-135.
- [119] Pham, D. T., & Sagioglu, S. Training multilayered perceptrons for pattern recognition: a comparative study of four training algorithms. *International Journal of Machine Tools and Manufacture*, 2001, vol. 41, no 3, p. 419-430.
- [120] Sharifi, A., & Mohebbi, A. Introducing a new formula based on an artificial neural network for prediction of droplet size in venturi scrubbers. *Brazilian Journal of Chemical Engineering*, 2012, vol. 29, no 3, p. 549-558.
- [121] Chacon, A., Hiremath, S., & Ding, D. Evaluation of the RT3 Tri-axial Accelerometer to Measure Physical Activity in Manual Wheelchair Users with Spinal Cord Injury. In *Proceedings of the RESNA Annual Conference*. 2010. p. 26-30.
- [122] Nielsen, M.A. Neural Networks and Deep Learning. *Determination Press (Draft)*. 2015.

- [123] Hassan-Esfahani, L., Torres-Rua, A., Jensen, A., & McKee, M. Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks. *Remote Sensing*, 2015, vol. 7, no 3, p. 2627-2646.
- [124] Bishop, C. M. Neural networks for pattern recognition. *Oxford university press*, 1995.
- [125] Blum, A. L., & Rivest, R. L. Training a 3-node neural network is NP-complete. *Neural Networks*, 1992, vol. 5, no 1, p. 117-127.
- [126] Díez, R. P., Gómez, A. G., & de Abajo Martínez, N. Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva. *Universidad de Oviedo*, 2001.
- [127] SIMOTICS Low-Voltage Motors. Distribution offering [en línea]. *Siemens AG* 2012. <http://www.motology.co.th/download/motors/SIMOTICS_Low-Voltage_Motors_complete_English_2012.pdf> [Consulta: 28 de noviembre de 2015]
- [128] VERILEC. CATALEG VARIACS. [Catalogo]. *VERILEC S.A.* <http://www.verilec.com/pdf/CATALEG_VARIACS.pdf> [Consulta: 28 de noviembre de 2015]
- [129] MICROMASTER 440 [Documentación de usuario]. *Siemens*. <https://cache.industry.siemens.com/dl/files/362/11887362/att_30316/v1/MM440_PLList_Span_B1.pdf> [Consulta: 28 de noviembre de 2015]
- [130] Lascu, C., Jafarzadeh, S., Fadali, S. M., & Blaabjerg, F. Direct torque control with feedback linearization for induction motor drives. In *Power Electronics and Applications (EPE'15 ECCE-Europe), 2015 17th European Conference on. IEEE*, 2015. p. 1-10.
- [131] ACSM1 Hardware Manual [Documentación de usuario]. *ABB*. <https://library.e.abb.com/public/e77c7dbb873624cbc12572f8002dd548/EN_ACSM1_04_HWMan_C.pdf> [Consulta: 28 de noviembre de 2015]
- [132] ABB servo motors. MS series for ABB high performance machinery drives. [Catalogo]. *ABB*. <https://library.e.abb.com/public/4e8dbaddac6d252fc12575f200461a02/MSseries_servomotors%20catalog_EN_REVCpdf.pdf> [Consulta: 28 de noviembre de 2015]
- [133] Detailed information for: PM583-ETH [en línea]. *ABB*. <<http://new.abb.com/products/1SAP140300R0271/pm583-ethac500-programmable-logic-controller-1mb-24vdc-ethernet-2xrs232-485-fbp>> [Consulta: 28 de noviembre de 2015]

- [134] DL750/DL750P ScopeCoder. User's Manual. [Documentación de usuario]. *Yokogawa Electric Corporation. 4th Edition.* <http://www.yokogawa.co.jp/ftp/dist/ks/eusers/dl/odehj/im/IM701210-04E_041.pdf> [Consulta: 28 de noviembre de 2015]
- [135] MN60. [en línea]. *Chauvin Arnoux Group.* <<http://www.chauvin-arnoux.com/fr/produit/mn60.html>> [Consulta: 28 de noviembre de 2015]
- [136] Samira Ben Salem, Khmais Bacha, & Abdelkader Chaari Sengupta r. Support vector machine based decision for mechanical fault condition monitoring in induction motor using an advanced Hilbert-Park transform. *ISA transactions*, 2012, vol. 51, no 5, p. 566-572.
- [137] Cristianini, N., & Shawe-Taylor, J. An introduction to support vector machines and other kernel-based learning methods. Cambridge university press, 2000.
- [138] Ramesh, A. N., Kambhampati, C., Monson, J. R. T., & Drew, P. J. Artificial intelligence in medicine. *Annals of The Royal College of Surgeons of England*, 2004, vol. 86, no 5, p. 334.
- [139] Badler, N. I. Artificial Intelligence, Natural Language. State-of-the-art in Computer Animation: *Proceedings of Computer Animation '89, 19.* 2012.
- [140] Schmidt, C. F., Sridharan, N. S., & Goodson, J. L. The plan recognition problem: An intersection of psychology and artificial intelligence. *Artificial Intelligence*, 1978, vol. 11, no 1-2, p. 45-83.
- [141] MacKay, D. J. Bayesian methods for adaptive models. 1992. *Tesis Doctoral. California Institute of Technology.*
- [142] Golik, P., Doetsch, P., & Ney, H. Cross-entropy vs. squared error training: a theoretical and experimental comparison. En INTERSPEECH. 2013. p. 1756-1760.
- [143] Mustafidah, H., Hartati, S., Wardoyo, R., & Harjoko, A. Selection of Most Appropriate Backpropagation Training Algorithm in Data Pattern Recognition. *arXiv preprint arXiv:1409.4727*, 2014.
- [144] Lippmann, R. P. An introduction to computing with neural nets. *ASSP Magazine, IEEE*, 1987, vol. 4, no 2, p. 4-22.
- [145] Braun, J., & Griebel, M. (2009). On a constructive proof of Kolmogorov's superposition theorem. *Constructive approximation*, 30(3), 653-675.
- [146] Zhang, T. Y. Ji, M. S. Li, Q. H. Wu and Q. H. Wu, "Application of discrete wavelet transform for identification of induction motor stator inter-turn short

- circuit,"*Smart Grid Technologies - Asia (ISGT ASIA), 2015 IEEE Innovative*, Bangkok, 2015, pp. 1-5.
- [147] Yahia, A. J. Cardoso, A. Ghoggal, and S. E. Zouzou, "Induction motors airgap-eccentricity detection through the discrete wavelet transform of the apparent power signal under non-stationary operating conditions," *ISA Trans.*, vol. 53, no. 2, pp. 603–611, 2014.
- [148] Wang, Guanghua Xu, Lin Liang, Kuosheng Jiang, Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis, *Mechanical Systems and Signal Processing*, Volumes 54–55, March 2015, Pages 259-276, ISSN 0888-3270.
- [149] Sapena-Bañó, M. Pineda-Sanchez, R. Puche-Panadero, J. Martinez-Roman and D. Matic, "Fault Diagnosis of Rotating Electrical Machines in Transient Regime Using a Single Stator Current's FFT," in *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 11, pp. 3137-3146, Nov. 2015.
- [150] Walter, G. G., & Shen, X. Wavelets based on Prolate spheroidal wave functions. *Journal of Fourier Analysis and Applications*, 2004, 10(1), 1-26.
- [151] Landau, H. J., & Pollak, H. O. Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—III: The Dimension of the Space of Essentially Time-and Band-Limited Signals. *Bell System Technical Journal*, 1962. 41(4), 1295-1336.
- [152] Hansen, L. K., & Salamon, P. Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, 1990, 12(10), 993-1001.
- [153] Fernández, J. C., Carbonero, M., Gutiérrez, P. A., & Hervás, C. Ensembles de redes neuronales construidos mediante algoritmos híbridos multiobjetivo para optimizar la precisión y la sensibilidad. In *proceedings de: VI Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados*. 2009.
- [154] Whitepaper. NVIDIA GeForce GTX 750 Ti. *NVIDIA Corporation*.

Anexos

Anexo 1. Clúster de redes neuronales.

Del apartado 3.3.4.3 se deduce que para obtener la red neuronal óptima para una misma estructura se realiza un proceso en que se desarrollan múltiples redes neuronales entrenadas. No obstante, al descartar prácticamente todas las redes neuronales entrenadas y quedarse solo con la mejor no se aprovecha el coste computacional derivado de entrenar todas ellas.

Por ello, como se expone en Hansen [152] y en Fernández [153] con mejoras, se puede construir un clúster o ensamble con el conjunto de todas las redes neuronales formando un único sistema experto de clasificación. En este caso, además de aprovechar el coste consumido por el entrenamiento, aumenta la generalización al reducirse el efecto de la componente de ruido en los datos de cada conjunto de validación utilizado para cada entrenamiento por separado.

En consecuencia, la tasa de error en conjunto se reduce debido al promedio de las soluciones, de forma que se puede llegar a obtener un mayor índice de aciertos. Como demuestra Bishop [124], este sistema experto en clúster no aumentará en ningún caso la tasa de error si se compara con la red neuronal óptima extraída de dicho conjunto de redes.

En este nuevo sistema experto compuesto por el conjunto de redes neuronales es necesario definir un método de evaluación de los resultados parciales de cada red para determinar el resultado final del sistema. En el artículo de Fernández [153] se presentan algunos de los métodos comúnmente usados en un clúster de redes neuronales con salidas discretas para la clasificación de elementos:

- **Voto por mayoría (Majority Voting).** En este método, la solución global del sistema experto para un patrón de entrada corresponde a la solución parcial con mayor número de apariciones dentro del clúster para dicho patrón de datos, es decir, su moda. En (A1.1) se muestra dicha fórmula donde $C(x, g_i)$ corresponde a la clase aportada como solución por la red neuronal g_i para el patrón de entrada x .

$$C(x) = \text{Moda} \{ C(x, g_i) \} \quad (\text{A1.1})$$

Fuente: Fernández et al (2009) [153]

- **Media aritmética (Simple averaging).** En el siguiente método (A1.2) se calcula la media aritmética de la probabilidad $P(x, g_i, C_k)$ de cada clase C_k de mostrarse como solución de un mismo patrón x para cada red neuronal g_i del clúster. El resultado del sistema experto se corresponde con la clase con la mayor probabilidad media de acierto de todo el conjunto de probabilidades T correspondiente al conjunto de todas las redes neuronales artificiales.

$$C(x) = \frac{\arg \max_k \sum_{i=1}^T P(x, g_i, C_k)}{T} \quad (\text{A1.2})$$

Fuente: Fernández et al (2009) [153]

- **“El ganador se lo lleva todo” (Winner-Takes-All).** En este caso (A1.3) el resultado del sistema experto se asigna a la red neuronal g_i que tiene la mayor probabilidad de asignación para el patrón de entrada x respecto de la clase C_k .

$$C(x) = \arg_k \max_{i,k} \{ P(x, g_i, C_k) \} \quad (\text{A1.3})$$

Fuente: Fernández et al (2009) [153]

Aunque la tasa de error de este sistema experto formado por el clúster de redes neuronales en general puede mejorar, hay que tener en cuenta que también existen desventajas al usar este clúster de redes neuronales, como por ejemplo el mayor coste de cómputo de nuevos datos y el mayor almacenamiento necesario para guardar todo el clúster.

Anexo 2. Algoritmo de entrenamiento por retropropagación del gradiente conjugado escalado.

Este algoritmo de entrenamiento para un sistema experto del tipo red neuronal artificial es un algoritmo iterativo que realiza el ajuste de pesos de la red neuronal de forma heurística.

En el artículo de Møller [116], se define este algoritmo mediante los siguientes pasos que se desarrollan de forma iterativa:

- 1) Al inicio del algoritmo se elige el vector de pesos inicial (w_1) y los parámetros iniciales σ_1 y λ_1 ambos mayores que 0.

Se calcula el valor de los conjugados de la base p_1 y r_1 (A2.1), se inicializa el número de ciclo k a 1 (A2.2) y se asigna a la variable *success* el valor de cierto (A2.3).

$$p_1 = r_1 = -E'(w_1) \quad (\text{A2.1})$$

$$k = 1 \quad (\text{A2.2})$$

$$\text{success} = \text{true} \quad (\text{A2.3})$$

- 2) Si *success* = *true* se calcula la información de segundo orden.

$$\sigma_k = \frac{\sigma}{|p_k|} \quad (\text{A2.4})$$

$$s_k = -\frac{E'(w_k + \sigma_k \cdot p_k) - E'(w_k)}{\sigma_k} \quad (\text{A2.5})$$

$$\delta_k = p_k^T \cdot s_k \quad (\text{A2.6})$$

3) Se escala s_k .

$$s_k = s_k + (\lambda_k - \bar{\lambda}_k) p_k \quad (\text{A2.7})$$

$$\delta_k = \delta_k + (\lambda_k - \bar{\lambda}_k) |p_k|^2 \quad (\text{A2.8})$$

4) Si $\delta_k \leq 0$ se hace que la matriz Hessiana se defina positiva.

$$s_k = s_k + \left(\lambda_k - 2 \frac{\delta_k}{|p_k|^2} \right) p_k \quad (\text{A2.9})$$

$$\bar{\lambda}_k = 2 \left(\lambda_k - \frac{\delta_k}{|p_k|^2} \right) \quad (\text{A2.10})$$

$$\delta_k = -\delta_k + \lambda_k |p_k|^2 \quad (\text{A2.11})$$

$$\lambda_k = \bar{\lambda}_k \quad (\text{A2.12})$$

5) Se calcula el tamaño del siguiente paso.

$$\mu_k = p_k^T r_k \quad (\text{A2.13})$$

$$\alpha_k = \frac{\mu_k}{\delta_k} \quad (\text{A2.14})$$

6) Se calcula el parámetro de comparación.

$$\Delta_k = \frac{2\delta_k [E(w_k) - E(w_k + \alpha_k p_k)]}{\mu_k^2} \quad (\text{A2.15})$$

7) Si el incremento $\Delta_k \geq 0$ entonces se puede realizar una reducción del error de forma satisfactoria.

$$w_{k+1} = w_k + \alpha_k p_k \quad (\text{A2.16})$$

$$r_{k+1} = -E'(w_{k+1}) \quad (\text{A2.17})$$

$$\bar{\lambda}_k = 0 \quad (\text{A2.18})$$

$$\text{success} = \text{true} \quad (\text{A2.19})$$

a) Si $\text{mod}(k, N) = 0$ entonces se reinicia el algoritmo.

$$p_{k+1} = r_{k+1} \quad (\text{A2.20})$$

Si $\text{mod}(k, N) \neq 0$ se crea una nueva dirección conjugada.

$$\beta_k = \frac{|r_{k+1}|^2 - r_{k+1} \cdot r_k}{\mu_k} \quad (\text{A2.21})$$

$$p_{k+1} = r_{k+1} + \beta_k p_k \quad (\text{A2.22})$$

b) Si el incremento $\Delta_k \geq 0.75$ se reduce el parámetro de escalado λ a la mitad.

$$\lambda_k = \frac{1}{2} \lambda_k \quad (\text{A2.23})$$

Si el incremento $\Delta_k < 0$ no es posible reducir el error.

$$\bar{\lambda}_k = \lambda_k \quad (\text{A2.24})$$

$$\text{success} = \text{false} \quad (\text{A2.25})$$

- 8) Si el incremento $\Delta_k < 0.25$, se incrementa el parámetro de escalado λ cuatro veces.

$$\lambda_k = 4\lambda_k \quad (\text{A2.26})$$

- 9) Si la dirección más acusada es distinta de 0 ($r_k \neq 0$) entonces se incrementa el ciclo $k=k+1$ y se vuelve a iterar al paso número 2.

Si la dirección más acusada es igual a 0 ($r_k = 0$) el algoritmo termina y se devuelve el vector de pesos w_{k+1} como solución.

Anexo 3. Aceleración computacional mediante hardware CUDA.

En el mundo de los sistemas expertos se suele requerir una gran cantidad de cómputo, sobretodo en la fase de entrenamiento el cual puede durar entre horas y días según la capacidad de cómputo. Cuanto mayor es la dificultad de entrenamiento los requisitos de cómputo son mayores.

Aprovechando la capacidad de paralelización computacional de la *Parallel Computing Toolbox* en el entorno de desarrollo *MATLAB*, se ha intentado mejorar la velocidad de cálculo en la fase de entrenamiento y ejecución de los sistemas.

Esta mejora se da al dividir el sistema en subfunciones que puedan ser ejecutadas independientemente en paralelo, disminuyendo así el tiempo necesario para completar la tarea.

Por ejemplo, en el caso de las redes neuronales estas se calculan como parte de un problema de multiplicación de matrices, el cual puede ser dividido en subproblemas que pueden ser resueltos en paralelo.

Existen diversos tipos de hardware donde ejecutar la paralelización de aplicaciones, tanto en procesadores multinúcleo (CPU) como en procesadores gráficos con multinúcleos vectoriales (GPU). Cuando la paralelización del problema es muy alta (como en este caso) se suele utilizar la potencia de cálculo de los núcleos vectoriales de un procesador gráfico, ya que, aunque la velocidad de funcionamiento de los núcleos vectoriales suele ser menor a la de un núcleo de CPU y su funcionamiento interno más básico, estos suelen ser de una cantidad significativa, lo que compensa su menor velocidad y mayor simplicidad.

Como hardware para desarrollar la computación en paralelo en este caso se ha utilizado la tarjeta gráfica NVIDIA GeForce GTX 750 Ti con procesador CUDA Maxwell GM107 de 640 núcleos a 1GHz y 2GB de memoria RAM.

En la figura A3.1 se muestra el diagrama del procesador gráfico de NVIDIA tipo Maxwell GM107 con varios clústeres de núcleos vectoriales sumando en conjunto el total de 640 núcleos, un número significativamente mayor a los 4 núcleos que se pueden encontrar actualmente en una CPU de gama alta como por ejemplo el modelo Core i7 de Intel.

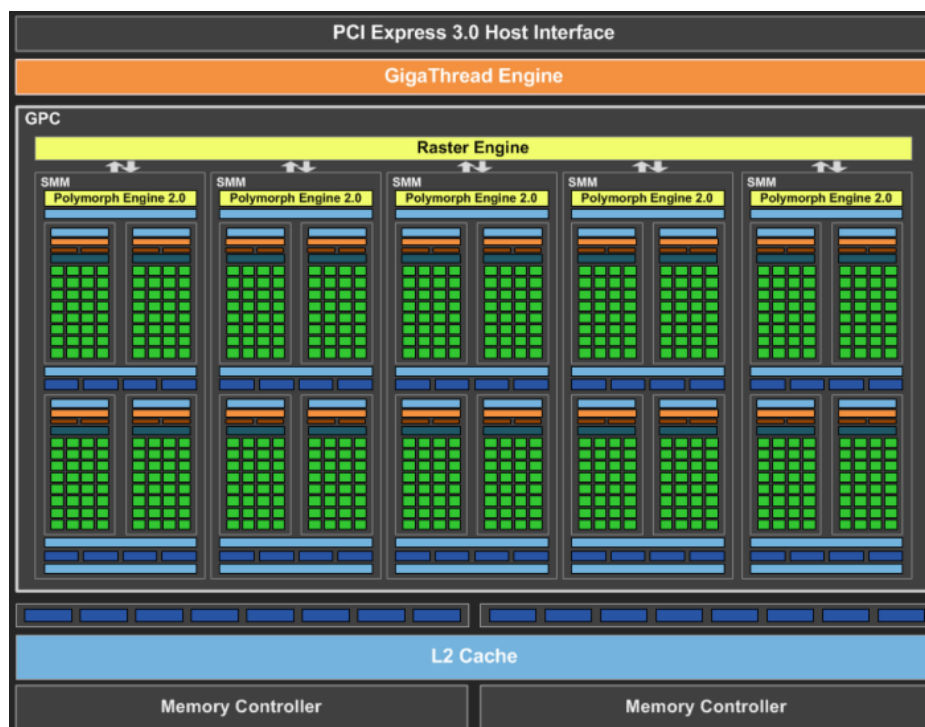


Figura A3.1. Diagrama de bloques de una GPU Maxwell GM107.

Sin embargo, durante el desarrollo del sistema se ha observado claramente que el intento de mejora de los costes computacionales paralelizando los sistemas expertos únicamente ha elevado el tiempo de cómputo necesario.

Aunque la paralelización computacional se aplica en muchos casos de forma satisfactoria, en este trabajo se ha descartado finalmente por sus penosos resultados.

Existen varios factores en la paralelización de programas que pueden mermar el rendimiento temporal de cómputo, como por ejemplo:

- El tiempo de preparación del entorno paralelo y carga en memoria de los datos es superior a la ejecución de las funciones paralelizadas con una pequeña duración. Si la duración de la ejecución de dichas funciones es dependiente del número de datos a procesar; el uso de un mayor número de datos puede superar el umbral mínimo necesario para obtener un mínimo beneficio.

Por ejemplo, como se ha comentado en el apartado 5.5.2.2.1, para una única estructura de red neuronal se realizan 1000 entrenamientos distintos, se requie-

re en este caso 1000 preparaciones del entorno de paralelización y 1000 volcados de datos para desarrollar todos esos entrenamientos con paralelización.

- En este caso el hardware usado no es óptimo para los requisitos del problema. Se ha utilizado una tarjeta gráfica NVIDIA GeForce GTX 750 Ti para la paralelización de los sistemas expertos, el cual requiere el uso de datos en coma flotante de doble precisión para maximizar la precisión de los cálculos.

No obstante, según las especificaciones aportadas por el fabricante

[154], la potencia de cálculo de este procesador GPU en doble precisión es de unos 60 GFlops, aproximadamente 32 veces más lenta que su velocidad de cálculo en precisión simple (1920 GFlops).

Un hardware más adecuado a este problema hubiera sido por ejemplo la tarjeta gráfica NVIDIA GeForce GTX Titan con una potencia de cálculo de doble precisión operando hasta un máximo de 1520 GFlops.

En consecuencia, no se descarta que, aplicar al mismo sistema una base de muestras mayor que eleve sustancialmente la necesidad de cómputo y/o usando un hardware más idóneo, la paralelización de su ejecución pueda derivar finalmente en un acortamiento del tiempo requerido de procesamiento.