

Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions

Piero Baraldi¹, Francesco Cannarile^{1,2}, Francesco Di Maio^{1*}, Enrico Zio^{1,2,3}

¹*Energy Department, Politecnico di Milano, Via la Masa 34, 20156 Milano, Italy*

francesco.dimaio@polimi.it

²*Aramis Srl, Milano, Italy*

³*Chair on System Science and the Energetic Challenge, European Foundation for New Energy-Electricité, Ecole Centrale Paris and Supelec, Paris, France*

Abstract

Electric traction motors in automotive applications work in operational conditions characterized by variable load, rotational speed and other external conditions: this complicates the task of diagnosing bearing defects. The objective of the present work is the development of a diagnostic system for detecting the onset of degradation, isolating the degrading bearing, classifying the type of defect. The developed diagnostic system is based on an hierarchical structure of K-Nearest Neighbours classifiers. The selection of the features from the measured vibrational signals to be used in input by the bearing diagnostic system is done by a wrapper approach based on a Multi-Objective (MO) optimization that integrates a Binary Differential Evolution (BDE) algorithm with the K-Nearest Neighbour (KNN) classifiers. The developed approach is applied to an experimental dataset. The satisfactory diagnostic performances obtain show the capability of the method, independently from the bearings operational conditions.

Key words: Bearing diagnostics; K-Nearest Neighbours (KNN) Classifier; Feature Selection; Wrapper Approach; Multi-Objective (MO) Optimization; Differential Evolution (DE); Wavelet Packet Transform (WPD)

1. INTRODUCTION

According to both the IEEE large machine survey (Zhang et al., 2011) and the Norwegian offshore and petrochemical machines data, bearing-related defects are responsible of more than 40% of the failure in industrial machines (O'Donnell et al., 1983). Then, in industrial practice it is of great interest to promptly detect the bearing degradation onset, to identify which bearing is degrading, to correctly classify the cause of the bearing degradation (type of defects) and to assess the bearing degradation level. The most critical bearing degradation modes involve the bearing inner race, outer race and balls (Rao et al., 2012) (Schoen et al., 1995). At the earliest stage of bearing degradation, information on the bearing health state, and, eventually, on the type of degradation can be obtained by observing the machine vibrational behavior. Thus, a typical approach to fault diagnosis in bearings is based on the extraction of features from the raw vibrational signals (accelerations) and on the use of classification models, such as Support Vector Machine (SVM) (Gryllias & Antoniadis, 2012) (Zhu et al., 2014), Relevance Vector Machines (Di Maio et al., 2012a), K-Nearest Neighbors (KNN) (Jiang et al., 2013), Artificial Neural Networks (ANN) (Li & Ma, 1997), neuro-fuzzy techniques (Zio et al., 2009) (Pan et al., 2014) and multi-symptom-domain consensus diagnosis techniques (He et al., 2001): input to the classifiers are the selected features, whereas the outputs are the detection of the onset of bearing degradation, the isolation of which bearing is degrading, the classification of the degradation mechanism and the assessment of the bearing degradation level.

Approaches to fault diagnosis in bearings have been developed considering the vibrational signals in the time domain, in the frequency domain and in both time and frequency domains. Time-domain approaches are based on the use of statistical indicators of the raw acceleration signals, such as mean, standard deviation, peak value, root mean square error, crest factor, kurtosis and skewness (Martin et al., 1995). Alternative time domain indicators have been developed (Tao et al., 2007) for dealing with incipient bearing faults, although the most critical shortcoming of all time-domain approaches is their inability to correctly diagnose bearing failures at the last stages of the degradation process, when the signal behaviors tend to be highly unpredictable and random (Ocak et al., 2007). In frequency-domain

approaches, the principal frequencies of the vibrational signals and their amplitudes are identified (Chebil et al., 2009). Most of the proposed approaches to fault diagnosis for bearings in the frequency domain assume a priori knowledge of the principal frequencies associated to the bearings faults (Chebil et al., 2009). This setting is not realistic in automotive applications where the environmental and operational conditions modify the frequency spectra of the vibrational signals. Furthermore, real bearing spectra are characterized by a large number of frequency components, which can be difficult to be managed (Ocak et al., 2007). Time-frequency approaches, which combine time and frequency domain information, have been reported to provide the most satisfactory performances (Georgoulas et al., 2013). Several time-frequency features have been proposed in literature, such as Short Time Fourier Transforms (STFT) (Kaewkongka et al., 2003), Wigner-Ville Distribution (WVD) (Hui et al., 2006), Wavelet Transform (WT) (Loutas et al., 2012) (Abbasion et al., 2007), and Empirical Mode Decomposition (EMD) (Huang et al., 1998) (Ben Ali et al., 2015). For example, a multilevel classification approach for bearing diagnosis based on WT has been proposed in (Chebil et al., 2009). Conversely, EMD is suitable and attractive in dealing with highly non-linear, non-stationary signals but can be computationally expensive due to the non-smooth behaviour of vibration signals. This limitation can be partially overcome using EMD and the Hilbert Huang transforms for the extraction of a compact set of features (Georgoulas et al., 2013).

A common characteristic of the frequency and time-frequency domain approaches is that they typically generate feature sets of very high dimensionality. Reducing the dimensionality of the feature set allows to remarkably reduce the computational burden. Furthermore, it has been shown that irrelevant and noisy features unnecessarily increase the complexity of the classification problem and can degrade modeling performance (Emmanouilidis et al., 1999). Thus, in this work, the development of classification algorithms for bearing diagnosis is accompanied by the application of feature extraction methods which map the n -dimensional data being classified onto an m -dimensional space, where $m < n$ (Dash & Liu, 1997). Examples of feature extraction methods are Kernel Principal Component Analysis (KPCA) (Schölkopf et al., 1998), Kernel Fisher Discriminant Analysis (KFDA)

(Mika et al., 1999) (Baudat & Anouar, 2000) or Semi-supervised Kernel Marginal Fisher Analysis (SKMFA) (Jiang, 2013), Linear Local Tangent Space Alignment (LLTSA) (Li, 2013), Self-Organizing feature Map (SOM) (Kohonen, 1982). A special case of feature extraction is feature selection, whereby $(n - m)$ irrelevant features are discarded. More specifically, the objective of feature selection is that of finding a subset of the original features such that the classification algorithm based on these features generates a classifier with the highest possible performance (Zio, 2006). In general, feature selection methods can be classified into two categories: filter and wrapper methods (Kohavi & John, 1997). In filter methods, the feature selector algorithm is used as a filter to discard irrelevant and/or redundant features a priori of the construction of the classification algorithm. A numerical evaluation function is used to compare the feature subsets with respect to their classification performance (Dash & Liu, 1997). On the contrary, in wrapper methods the feature selector behaves as a wrapper around the specific learning algorithm used to construct the classifier. The feature subsets are compared using as criterium the classification performance achieved by the classification algorithm itself (Zio et al., 2008).

This work is motivated by the interest of investigating the possibility of effectively performing in practice fault diagnostics of bearings installed on the powertrain of a Fully Electric Vehicle (FEV). The research is part of the European Union funded project Electrical power train Health Monitoring for Increased Safety of FEVs (HEMIS, www.hemis-eu.org) (Sedano et al., 2015), (Baraldi et al., 2013), which aims at the development of a Prognostics and Health Monitoring System (PHMS) for the most critical components of FEVs. The difficulty of the fault diagnostics task is that automotive motors differ from other industrial motors since they work in operational conditions characterized by variable load, rotational speed and other external conditions which can cause major modifications of the vibrational signal behaviour. Electrical machines and drive systems are subject to many different types of faults which include: 1) stator faults such as stator winding open or short circuited; 2) rotor electrical faults such as rotor winding open or short circuited for wound rotor machines and broken bar(s) or cracked end-ring for squirrel-cage machines; 3) rotor mechanical faults such as bearing

damage, eccentricity, bent shaft, and misalignment; and 4) failure of one or more power electronic components of the drive system (Bellini et al., 2008). (Bonnett & Yung, 2008) describes the distribution of induction motor faults and shows possible scenarios for after fault, detailing the repair-replace decision process. The distribution of induction motor faults is listed in (Bellini et al., 2008) as bearing (69%), rotor bar (7%), stator windings (21%), and shaft/coupling (3%). Fault diagnostics of bearing installed on the powertrain of electric machines is an attracting research field. In (Tian et al., 2016), different features are extracted from spectral kurtosis and then combined to build a health index based on PCA and a semi-supervised KNN distance measure to detect incipient faults and diagnose the locations of the bearings faults. In (Abed et al., 2015), DWT is used to extract features from stator current and lateral vibrations current measurements. The obtained features are further reduced via the applications of orthogonal fuzzy neighbourhood discriminant analysis. Finally, a Recurrent Neural Networks (RNN) is used to detect and classify the presence of bearing faults. In (Geramifard et al., 2013), a semi-nonparametric approach based on a hidden Markov model classifier is introduced for fault detection and diagnosis of bearings in synchronous motors. In (Zhang & Zhou, 2013), a procedure based on Ensemble Empirical Mode Decomposition (EEMD) and SVM for multi-fault diagnosis of bearings in induction motors is discussed. In (Dalvand et al., 2016) the kurtosis of instantaneous frequency of motor voltage is used for the identification of defective bearings. In (Jin et al., 2014), Trace Ratio Linear Discriminant Analysis (TRLDA) is used to deal with high dimensional non-Gaussian fault data for dimension reduction and fault classification of bearings in induction motors. Although the listed works have been reported to achieve satisfactory performance, the industrial applicability of these methods is limited by the fact that the features extracted to train the empirical model for the diagnosis are not independent from operational conditions: fault diagnostics is tacitly based on the hypothesis that the training patterns and the testing patterns are similar. As a result, if the diagnostic model is used in working conditions different from those considered to train the model, its performance may be unsatisfactory. To overcome this limitation, the main contribution of this work is the development of a novel feature selection approach to identify

features independent from operational conditions. This is expected to allow developing a diagnostic system that can be used independently from the operational and environmental conditions that the FEV is experiencing. A further novelty of the work is that the feature selection problem is embedded into a multi-classification problem, where several classifiers developed for different scopes (detection, isolation, degradation mode classification and degradation level assessment) are integrated. The proposed diagnostic system is based on an hierarchical model of K-Nearest Neighbor (KNN) (Jiang, 2013) classifiers. A multi-objective (MO) Binary Differential Evolution (BDE) optimization algorithm has been used for the identification of the feature set to be used. The optimization aims at the identification of a feature set, which allows to obtain a high classification performance by using a low number of features extracted from a low number of vibrational signals. Notice that the use of a low number of features allows reducing the computational burden and memory demand of the diagnostic system, whereas the use of a limited number of vibration signals allows minimizing the cost of the installation of the measurement system. The proposed approach is verified with respect to the Western Reserve Case University Bearing dataset (CWRUBD).

The paper is organized as follows: in Section 2 the hierarchical model for bearing degradation detection, isolation, diagnosis and degradation level assessment is proposed; in Section 3, a wrapper approach for optimal feature selection based on the use of a BDE-based MO optimization algorithm is discussed; the application to the Western Reserve Case University Bearing dataset is described in Section 4, whereas in Section 5 conclusions are drawn.

2. THE HIERARCHICAL DIAGNOSTIC MODEL

In this work, a motor system containing two bearings, one installed at the drive end (DE) and one at the Fan End (FE) of the powertrain, is considered. The main objective of the work is the development of a diagnostic system for the identification of: *i*) the onset of the degradation (fault detection), *ii*) which bearing is degrading (fault isolation), *iii*) the failure mode (failure mode identification) and *iv*) the degradation level (degradation level assessment). Fault detection (task *i*) is typically based on the

quantification of inconsistencies between the actual and the expected behavior in nominal conditions and can be tackled resorting to anomaly detection techniques (Pichler et al., 2016) (Serdio et al., 2014), Auto Associative Kernel Regression (AAKR) (Baraldi et al., 2014) or one-class classifier. These methods are typically based on the use of data collected from the component operation in nominal conditions, before the onset of the degradation and do not require the availability of data collected in abnormal conditions, after the onset of the degradation process. The task of fault isolation, failure mode identification and failure size assessment (task *ii*, *iii* and *iv*) are, then, typically tackled in cascade, as shown in Figure 1.

Notice that the problem of fault isolation and failure mode identification requires the development of a classification model whose inputs are the measured signals and whose output is a label indicating the location and type of the fault and cannot be addressed by means of unsupervised techniques. With respect to the assessment of the degradation level, it can be tackled as a classification problem where the degradation level is discretized into a predefined set of levels. Finally, since data of abnormal conditions are available and will be exploited for the tasks of fault isolation, failure mode identification and degradation level assessment (level 2, 3 and 4 of Figure 1), the fault detection task (level 1) is also framed as a binary classification problem. To this aim, we have developed a hierarchical model based on a set of classifiers (Figure 2). The first classifier identifies the onset of the bearing degradation (stage 0, classifier C_0), the second the location of the degradation, i.e. which bearing is degrading (stage 1, classifier C_1), the third the degradation mode (stage 2, classifiers C_2^b , $b=1, 2$) and the last one the degradation intensity of the failure (stage 3, classifier $C_3^{b,i}$, $b=1, 2$, $i=1, \dots, N_c$ with N_c indicating the number of possible bearing degradation modes). Notice that for each bearing a different classifier, C_2^b , $b = 1, 2$ of the degradation mode is developed, and for each bearing and each degradation mode a different classifier, $C_3^{b,i}$, of the intensity, $b = 1, 2$ and $i = 1, \dots, N_{dm}$ is developed (Figure 2).

All the classifiers are fed with information extracted from vibrational signals correlated to the degradation process of the bearings. In particular, in this work we consider the possibility of installing up to S accelerometers in different locations of the motor housing and motor supporting base plate, and the possibility of extracting from each vibrational signal, K features, including statistical indicators (Di Maio et al., 2012b), Discrete Wavelet Transform (DWT) (Baraldi et al., 2012) and Wavelet Packet Transform (WPT) (Chebil et al., 2009). These different types of features have been considered since they have been already used in bearing diagnostic problems and they have been shown to contain information correlated with the bearing degradation.

3. THE FEATURE SELECTION PROBLEM

Each classifier of the hierarchical structure can receive in input up to $n = K \cdot S$ features. In this work, the problem of selecting the most performing features for the classifiers is addressed considering only the classifiers at stages 2 and 3 of the hierarchical model (identification of the degradation mode and assessment of the degradation level, respectively). The input features used by classifiers C_0 and C_1 for the detection of the onset of the degradation and the identification of which bearing is degrading will be identified in a second phase considering only the features identified for the classifiers at the second and third stages. This simplification of the problem is justified by the fact that the classifiers for the detection of the degradation (C_0) and the isolation of the degrading bearing (C_1) will be shown to achieve high performance using the same features selected for the fault diagnosis (stage 2).

The overall objectives of the feature selection process are to identify a set of features which guarantees:

- I. high classification performance in each stage of the classification (diagnosis of the degradation mode, assessment of the degradation level);
- II. low cost for the development of the overall diagnostic system. The cost should take into account: the number of vibrational sensors required, the computational burden and memory

demand for processing of the vibrational signals, the training of the classification algorithms and the storage of the training examples.

With respect to I), notice that the selected features should be able to provide good classification performances independently from the operational conditions experienced from the automotive vehicle. In practice, the first objective that is considered is the minimization of the misclassification rates of the two classifiers C_2^1 and C_2^2 at stage 2 of the hierarchical model, dedicated to the identification of the degradation mode at the DE and FE bearings, respectively. With respect to a feature set represented by a n -dimensional vector $\mathbf{x} \in \{0,1\}^n$, where $x(k) = 1$ denotes that feature k is selected whereas $x(k) = 0$ that it is not selected, the objective function F_1 , i.e. the average misclassification rate at stage 2, is defined by:

$$F_1(\mathbf{x}) = \frac{1}{2} \sum_{b=1}^2 R_2^b(\mathbf{x}) \quad (1)$$

where $R_2^b(\mathbf{x})$ is the misclassification rate of classifier C_2^b on a set of test patterns. In order to verify the capability of the classifiers to provide good performances independently from the operational conditions, we propose the following “leave all but one” cross validation method inspired to the leave one out cross validation method (Polikar, 2007). Let l be the number of possible operational condition that can be experienced by the bearing. We train l different classifiers: the j^{th} , $j = 1, \dots, l$, classifier is built using a training set containing patterns taken at operational condition j , and tested on patterns taken at operational condition different from j .

Finally, $R_2^b(\mathbf{x})$ is computed as average of the misclassification rates $R_2^{b,j}(\mathbf{x})$ of the j^{th} , $j = 1, \dots, l$ classifier:

$$R_2^b(\mathbf{x}) = \frac{1}{l} \sum_{j=1}^l R_2^{b,j}(\mathbf{x}) \quad (2)$$

Figure 3 shows the “leave all but one” cross validation procedure for $l = 4$.

The name “leave all but one” derives from the fact that when developing the j^{th} classifier, we leave from the training set all patterns which have been taken at operational condition different from j but

those taken at operational condition j . By so doing, the different classifiers are tested using patterns in operational conditions different from those considered to train the model. Notice that this procedure requires the availability of condition monitoring data collected in different, but stationary operating conditions. Thus, the selected features are not expected to provide satisfactory performances when the component works in variable operating conditions, which can lead to complex dynamics in the measured signals. In practical applications characterized by continuously changing operating conditions, this drawback of the method can be overtaken by applying the fault diagnostic system only when the operating conditions are constant for a given period of time (e.g., the vehicle is used on an highway at constant speed).

The second objective takes into account the performance of the classifiers $C_3^{b,i}$ for the assessment of the degradation level, with $b = 1, 2$, and $i = 1, \dots, N_{dm}$. With respect to a feature set represented by the n -dimensional vector $\mathbf{y} \in \{0,1\}^n$, where $y(k) = 1$ indicates that feature k is selected as input of the classifier, whereas $y(k) = 0$ is not selected, the average misclassification rate at stage 3, F_2 , is defined by:

$$F_2(\mathbf{y}) = \frac{1}{2N_{dm}} \sum_{b=1}^2 \sum_{i=1}^{N_{dm}} R_3^{b,i}(\mathbf{y}) \quad (3)$$

where $R_3^{b,i}(\mathbf{y})$ is the misclassification rate of classifier $C_3^{b,i}$ obtained applying the same procedure followed in eq (2) to guarantee independence from the operational conditions.

With respect to the objectives in II), we consider two different cost indicators: the net number of features employed by the overall hierarchical model, F_3 , and the number of accelerometers to be used, F_4 . For a given feature set $\mathbf{z} = (\mathbf{x}, \mathbf{y})$, F_3 is given by:

$$F_3(\mathbf{z}) = \sum_{k=1}^n x_k + \sum_{k=1}^n y_k - \sum_{k=1}^n 1_{\{x_k=y_k\}} \quad \mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \{0,1\}^{2 \cdot n} \quad (4)$$

where n is the total number of features which can be extracted.

The number of accelerometers to be used, F_4 , is given by:

$$F_4(\mathbf{z}) = \sum_{s=0}^{S-1} R_{4,s}(\mathbf{z}) \quad (5)$$

where $S - 1$ is the total number of accelerometers which can be installed and $R_{4,s}(\mathbf{z})$ is equal to 1 if at least one feature extracted from the acceleration signal measured by accelerometer s is selected.

According to the proposed wrapper approach (Figure 4), the search engine builds a candidate group of features set $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ whose performance is evaluated with respect to a fitness function \mathbf{F} that is defined as:

$$\mathbf{F}(\mathbf{z}) = [F_1(\mathbf{x}), F_2(\mathbf{y}), F_3(\mathbf{z}), F_4(\mathbf{z})] \quad \mathbf{x}, \mathbf{y} \in \{0,1\}^n \quad \mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \{0,1\}^{2 \cdot n} \quad (6)$$

Dealing with a MO optimization problem (in our specific case a MO minimization), we introduce the definition of *Pareto Optimal Set* $\mathcal{P}^* = \{\mathbf{z} \in \mathcal{F} : \mathbf{z} \text{ is Pareto-optimal}\}$, that is a set of optimal solutions among which we select the preferred solution \mathbf{z}_{opt} . A vector of decision variable $\mathbf{z}^* \in \mathcal{F}$ is *Pareto Optimal* if it is non-dominated with respect to \mathcal{F} , i.e., it does not exist another solution $\mathbf{z}' \in \mathcal{F}$ such that $\mathbf{F}(\mathbf{z}')$ dominates $\mathbf{F}(\mathbf{z}^*)$:

$$\forall \alpha \in \{1, \dots, 4\}, F_\alpha(\mathbf{z}') \leq F_\alpha(\mathbf{z}^*), \text{ and } \exists \tilde{\alpha} \in \{1, \dots, 4\}, \text{ such that } F_{\tilde{\alpha}}(\mathbf{z}') < F_{\tilde{\alpha}}(\mathbf{z}^*) \quad (7)$$

3.1 Binary Differential Evolution for feature selection

Performing an exhaustive search of the best solution among all the possible $2^{2 \cdot n}$ solutions is typically impracticable unless $2 \cdot n$ is very small (Dong & Kothari, 2003). For this reason, different combination of optimization heuristics such as Ant Colony (Al-ani, 2005), Genetic Algorithm (Sikora & Piramuthu, 2007), Particle Swarm Optimization (PSO) (Samanta & Nataraj, 2009) (Firpi & Goodman, 2005), Binary Genetic Algorithms (Zio et al., 2006), and Binary Differential Evolution BDE (He, 2009) (Kushaba et al., 2011) have been used within wrapper approaches for feature selection. In this work, we resort to a Binary Differential Evolution (BDE) algorithm to address the MO feature selection problem, since BDE has been shown to explore the decision space more efficiently than other multi-objective evolutionary algorithms (Tušar & Filipič, 2007) such as Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al., 2002), Strength Pareto Evolutionary Algorithms (SPEA2) (Zitzler et al., 2001) and Indicator Based Evolutionary Algorithm (IBEA) (Zitzler & Künzli, 2004).

In BDE, each candidate solution $\mathbf{z}_{p,G}$, called *target vector*, of the G^{th} population is encoded by a binary sequence (*chromosome*) of $2 \cdot n$ bits (*genes*) for $2 \cdot n$ decision variables, where each bit indicates whether a feature is present (1) or discarded (0) in the candidate solution $\mathbf{z}_{p,G}$. Each gene, $z_{p,k,G}$, $p = 1:NP$, $k = 1:2 \cdot n$ of each chromosome of the G -th population is conveniently mapped into a continuous variable $\tilde{z}_{p,k,G}$. In practice, the interval $[0,1]$ is partitioned into two equal subintervals $[0,0.5)$ and $[0.5,1]$, such that if the gene $z_{p,k,G} = 0$, $\tilde{z}_{p,k,G}$ belongs to the first subinterval, whereas if $z_{p,k,G} = 1$, $\tilde{z}_{p,k,G}$ it belongs to the second interval. The mapping operator

$$\tilde{z}_{p,k,G} = \begin{cases} 0.5 * rand & \text{if } z_{p,k,G} = 0 \\ 0.5 + 0.5 * rand & \text{if } z_{p,k,G} = 1 \end{cases} \quad (8)$$

is used for this purpose, where *rand* is a random number in $[0,1)$.

1. Mutation

For each vector $\tilde{\mathbf{z}}_{p,G}$ in the population, a noisy vector $\tilde{\mathbf{v}}_p$ is generated randomly choosing three mutually different vector indices $r_1, r_2, r_3 \in \{1, \dots, NP\}$ with $p \neq \{r_1, r_2, r_3\}$

$$\tilde{\mathbf{v}}_{p,G} = \tilde{\mathbf{z}}_{r_1,G} + SF(\tilde{\mathbf{z}}_{r_2,G} - \tilde{\mathbf{z}}_{r_3,G}) \quad (9)$$

where the *scaling factor* $SF \in (0,2]$ (Khushaba et al., 2011).

A sigmoid function is applied to entry $\tilde{v}_{p,k,G}$ to ensure that the result generated by the mutation operator falls into the interval $[0,1]$:

$$\tilde{v}_{p,k,G} = \frac{1}{1 + e^{-\tilde{v}_{p,k,G}}} \quad (10)$$

An inverse operator is then used:

$$v_{p,k,G} = \begin{cases} 0 & \text{if } \tilde{v}_{p,k,G} \in [0, 0.5) \\ 1 & \text{if } \tilde{v}_{p,k,G} \in [0.5, 1] \end{cases} \quad (11)$$

2. Crossover

In order to increase diversity of the perturbed parameter vectors, crossover can be introduced. This procedure is typically referred to as *recombination*. To this aim, the *trial vector* $\mathbf{u}_{p,G} = (u_{p,1,G}, \dots, u_{p,k,G}, \dots, u_{p,2 \cdot n,G})$ is defined by:

$$\mathbf{u}_{p,k,G} = \begin{cases} \mathbf{z}_{p,k,G} & \text{if } \mathcal{U}(0,1] \leq CR \text{ or } k = \text{irand}(NP) \\ \mathbf{v}_{p,k,G} & \text{if } \mathcal{U}(0,1] > CR \text{ and } k \neq \text{irand}(NP) \end{cases} \quad (12)$$

where $\mathcal{U}(0,1]$ is a uniform continuous random value $[0,1]$, whereas $\text{irand}(NP)$ is a discrete random number in the set $\{1,2, \dots, NP\}$ sampled from a uniform distribution. The *crossover parameter* $CR \in [0,1]$ influences the probability that the noisy vector's variables are selected for the mutation process (Wang et al., 2011).

3. Selection

In order to avoid stagnation of population in local minima due to the impoverishment of the population, selection strategies have been deeply investigated in literature (Mezura-Montes et al., 2008) (Salman et al., 2007). According to the MODE-III selection technique (Wang et al., 2011), each trial vector generated at each iteration by mutation and crossover operations, $\mathbf{u}_{p,G}$, is compared only with its target vector $\mathbf{z}_{p,G}$ from which it inherits some variables: if $\mathbf{u}_{p,G}$ dominates $\mathbf{z}_{p,G}$, it takes its place in the population for the next generation, otherwise $\mathbf{z}_{p,G}$, survives (Wang et al., 2011). Notice, however, that, this approach suffers of a low level of elitism since each trial vector is compared only with its own target vector.

In the present work, we have applied a different technique, called Non-Dominated Sorting Binary Differential Evolution (NSDBE), which combines the robust and effective BDE strategy with the fast non-dominated sorting and ranking selection scheme of NSGA-II (Deb, 2002). In practice, at the G^{th} generation the combined population of size $2NP$ comprising all $\mathbf{u}_{p,G}$ and $\mathbf{z}_{p,G}$ is ranked using a fast non-dominated sorting algorithm that identifies the ranked non-dominated solutions of the Pareto optimal set, Σ . Then, the first NP candidate solutions are selected according to the *crowding distance* (Deb, 2002).

3.3. The classification algorithm

The feature selection algorithm requires the availability of a classification algorithm characterized by 1) few parameters to be tuned and 2) very fast computational times for the classification model training and executions. Among the possible choices of classification algorithms, we have considered Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees (DT) and K-Nearest Neighbours (KNN) classifier. The SVM have been shown to produce satisfactory performances when applied to datasets characterized by many classes (Gryllias & Antoniadis, 2012), even if few labelled examples are available for the SVM training process. Unfortunately, SVMs require the tuning of the kernel, the kernel parameter optimization and the setting of an appropriate soft margin parameter and computational efforts in the training phase than other classification algorithms. ANN based classifiers allow dealing with nonlinear and multi-class classification problems, but they typically require a large amount of data for the ANN training and are computationally time consuming in the training phase (Li & Ma, 1997). Decision Trees (DTs) based classifiers are simple to understand and to interpret and can handle multi-class classification problems (Parvin et al., 2015). However, DT learners can create over-complex trees that do not generalise the data well, i.e., they suffer from overfitting and can be unstable because small variations in the data might result in a completely different tree being generated. This problem is mitigated by training multiple trees in an ensemble learner which is computationally consuming both in the training phase and in the execution phase. The KNN algorithm is automatically non-linear and requires the setting the number \tilde{K} of near neighbours to be considered for the classification and of the metric used to quantify the distance between patterns in the feature space. The KNN classification of a test pattern o is based on the computation of its distance with the T labelled patterns of a training set, $T_r = \{(o_t, c_t)\}$, $t = 1:T$, $c_t \in \{1, \dots, Cl\}$, with c_t indicating the class of the t -th pattern and Cl the total number of classes. In practice, the KNN algorithm (Hellman, 1970):

- a) finds the \tilde{K} closest training patterns to the test pattern, according to an opportune distance (e.g. Euclidean distance, Mahalanobis distance etc.), where \tilde{K} is a user-defined nonnegative integer;
- b) assigns the test pattern o to the class with most representatives among those of its \tilde{K} neighbors.”

Although KNN classification performance has been reported to be less satisfactory in some applications than that of SVM, ANN and DT classifiers, KNN classifiers have been employed in this work for their simplicity and low computational requirements (Compare et al., 2016). Furthermore, notice that, once the time consuming feature selection step has been performed using the KNN classifier, one according to the available computational and memory resources, can potentially choose any other classifier for the development of the diagnostic system.

4. CASE STUDY: THE CASE WESTERN RESERVE UNIVERSITY BEARING DATASET

The Case Western Reserve University bearing dataset contains the results of 72 experiments consisting in the measurement of 3 acceleration signals. The acceleration signals are measured using $S = 3$ accelerometers placed at the 12 o'clock position at the drive end and at the fan end of the motor housing and on the motor supporting base plate. Data are collected at frequencies of 12000 samples per second for time lengths of about 10 seconds. Two ball bearings are installed at the drive end and at the fan end of the motor, respectively. For both bearing, $N_{dm} = 3$ degradation mode are considered affecting the inner race, outer race and ball, respectively. For each failure mode, 12 experiments have been performed, considering all the possible combinations of $N_{dl} = 3$ different degradation levels (i.e., $f = 7, 14, 21$, mils (mil inches) long defects) and $j = 4$ different operation conditions represented by motor loads from 0 to 3 horsepower. Bearings in normal conditions have also been tested at the $l = 4$ different loads. The vibration time series have been verified to be stationary by applying the Kwiatkowski, Phillips, Schmidt, and Shin's test (KPSS test) (Kwiatkowski et al., 1992).

4.1 Feature extraction

Each vibration signal has been segmented using a fixed time window of approximately 1.4 seconds, overlapping of about 0.37 seconds. Each time window contains 2^{14} acceleration measures from each sensor. Therefore, from each time series, we have extracted 10 different time windows, hereafter called records. From each record, we have extracted $K = 29$ different features: these include statistical indicators (1 to 9) (Di Maio et al., 2012b), Discrete Wavelet Transform (DWT) using Haar basis (10 and 11) (Baraldi et al., 2012), DWT using Daubechies3 basis (12 to 15) and Wavelet Packet Transform (WPT) using Symlet6 basis (16 to 29) (Chebil et al., 2009), as listed in Appendix A. Since these features have been extracted from $S = 3$ vibrational signals measured by $S = 3$ different accelerometers, the total number of features extracted is $n = 87$. Thus, the available data-set consists of 720 87-dimensional patterns (Table 1). Notice that for each pattern we know whether it corresponds to a motor with a degraded or healthy bearing and in the former case, the occurring degradation mode and the degradation level. Thus, the patterns are labelled with respect to all the classifiers of the hierarchical model.

All the available 720 labelled data are partitioned into a set used for the feature selection task formed by 80% of the total number of patterns and obtained by randomly sampling 8 patterns among the 10 at a given load in each row of Table 1, and a validation set formed by the remaining patterns, which will be used for validating the performance of the diagnostic model after the optimal features subset selection.

4.2 Validation of the feature selection algorithm

In this subsection, we compare the results obtained by the proposed feature selection algorithm with those obtained in literature considering the same dataset (Zhang et al., 2013) (Jiang et al., 2013) (Zhu et al., 2014) (Li et al., 2013). To this aim, in order to have the same test conditions used in the literature works, the feature selection task has been performed considering only the failure of the drive-end bearing and the vibrational signal registered at the drive-end of the motor housing. Furthermore, in

accordance with the literature works, a direct, one-stage classification of the fault type and intensity has been performed. In practice, we have considered a 10 classes classification problem, where the classes correspond to the normal state and all the possible 9 combinations of the 3 failure types and 3 failure intensities. The only objective of the feature selection is the minimization of the misclassification rate, i.e. the fraction of test patterns not assigned to the correct class. The best solution identified by the DE algorithm is reported in Table 2, the performance in terms of misclassification rates obtained by adopting a 50-fold cross-validation approach on validation data not used for the feature selection is 0.0059 with a standard deviation 0.0008. In other words, for 50 times we have randomly chosen among the validation set, 75% of the patterns for the training set and 25% for the test set, ensuring that at least 3 patterns of each class are present in the training set. Table 3 compares the obtained results with those of other literature works. Notice that the performance obtained using the selected features is more satisfactory than those obtained in (Jiang et al., 2013) (KNN classifier), (Zhang et al., 2013) and are comparable to those obtained in (Zhu et al., 2014) and (Li et al., 2013) (SVM) which are based on a more refined classification model. It is, however, worth noting that our approach is the only one which is tackling the problem of independence from operational conditions, which complicates the classification problem since it reduces the amount of data available for training the classifier and the similarity between the training and test data.

4.3 The overall hierarchical classification model

According to Figure 2, the overall hierarchical model is formed by:

1. one classifier for identifying the onset of the degradation, C_0
2. one classifier for identifying the location of the degradation, C_1
3. two classifiers for identifying the degradation mode, C_2^b with $b = 1,2$ indicating which bearing is degrading, where $b = 1$ and $b = 2$ refer to drive end and fan end bearing, respectively

4. six classifiers for identifying the degradation level, $C_3^{b,i}$ with $i = 1,2,3$ indicating the degradation mode (1 refers to inner race defects, 2 to balls defects and 3 to outer race defects)

In order to obtain independence from the operational conditions, the training sets used to build the classifiers are always formed by patterns extracted from signals collected from a motor operating at a load different from that from which the patterns of the test sets have been obtained. The test is repeated considering classifiers trained with patterns collected from motor operating at different loads, until all the loads have been considered. The number of patterns used for the training and test of the different classifiers are summarized in Table 4.

4.4 Feature selection results

A MOBDE-based approach has been applied using the MO fitness functions $F(\mathbf{z})$ in (7) as criteria for the selection of the relevant features. Each candidate solution \mathbf{z} is a binary string of 174 bits (*genes*), the first 87 genes represent the input features to the classifiers at stage 2, whereas genes from 88 to 174 represent the input features to the classifiers at stage 3. The parameters CR , SF and NP of the BDE have been set to 0.30, 0.5 and 350, respectively. The choice of the value of 0.30 for the crossover parameter, CR , is motivated by the necessity of maintaining diversity in the population and it has been set according to the suggestions in (Gong et al., 2014), where it is shown that low CR values can lead to a gradual and successful exploration of a complex search space. The scale factor parameter, SF , has been set to 0.5 according to the suggestion of (Ali, 2005). Finally, a large population, formed by 350 chromosomes has been used in order to allow a deep exploration of the multidimensional search space (Mallipeddi & Suganthan, 2008).

The performance of the MO optimization can be quantified in terms of the *diversity* of the solutions and the *convergence* to the Pareto optimal front (Deb et al., 2002). Since in a MO optimization problem, it is typically not possible to simultaneously improve the values of two or more objective functions without causing the deterioration of some other objectives (Azevedo & Arujo, 2011), diversity is a fundamental requirement in a MO evolutionary optimization. In practice, diversity in

the population allows improving the coverage of the search space and exploring different evolutionary paths. An indicator of the diversity of a Pareto optimal set is the hyper-volume over the non-dominated set, which has been defined as the Lebesgue-measure of the hyper-volume with respect to a lower reference bound (normally, the ideal worst values of each objective function) (Zitzler, 2003): when two Pareto fronts are compared, higher is the value of such indicator, better is the performance in terms of objective function evaluations and wider is the exploration of the search space. In our case, we set as upper reference point, the point (1,1,87,3) i.e., the feature set characterized by the worst possible performances i.e. all the patterns are misclassified and 87 features extracted from $S = 3$ sensors are used. Figure 5 shows the Pareto fronts obtained after $G=1500$ generations applying the NSBDE and the MODE III selection strategies, and Table 5 reports the statistics of the corresponding hyper-volumes.

Notice that NSBDE performs better than MODE III in terms of diversity and performance of the solutions. This has justified the application of the NSBDE strategy with a high number of generations in order to identify the optimal Pareto set. Figure 6 shows that the optimal Pareto set hyper-volume is increasing until generation 15500 and then it tends to remain constant. This indicates that the Pareto set becomes stable and no improvement of the solutions is expected to be found by further increasing the number of generations.

In order to select the solution to be actually used for the development of the bearing diagnostic system, we have considered the following information provided by experts:

- a) the computational cost of memory pre-allocation depends on the number of slots to be used. A slot typically allows to use from one to eight features, thus the computational cost is the same if the number of features is between 1 and 8, and it increases when the number of features exceeds 9. Since solutions with more than 16 features have not been identified, the computational cost can be that of 1 or 2 slots.

b) the monetary cost for sensors (i.e., measurement devices and data collection system) is directly proportional to the number of sensors to be installed.

In order to select the best compromise solution \mathbf{z}_{opt}^* , we firstly normalize the four objective functions in a scale from 0 to 1, where 0 corresponds to the minimum value of the objective function in the Pareto optimal front and 1 to the maximum value. With respect to the objective function 3, in order to take into account the information provided by the expert, we have assigned a normalized value of 0 to all the solutions of the Pareto optimal set characterized by less than 9 features (all characterized by the same cost) and of 1 to all the solutions with more than 8 features (all characterized by the same cost).

Then, we resort to the TOPSIS method (Technique for Order Preference by Similarity to an Ideal Solution) (Opricovic & Tzeng, 2004), which is a multiple criteria decision making method whose basic principle is that the chosen solution should have the shortest distance from the ideal solution and the farthest distance from the negative-ideal solution (Appendix B). Table 6 reports the features in the best compromise solution, \mathbf{z}_{opt}^* , whereas its performance is reported in Table 7.

It can be observed that only one statistical indicator, the peak value, has been selected for both classifiers, whereas all the other features, except a minimum wavelet coefficient, are norms computed at different levels of the WPD. This result confirms the superiority of the WPD feature for diagnostics in bearings with respect to DWT, as already pointed out in (Chebil et al., 2009) where, however, the problem of the independence from operational conditions is not addressed and the possibility of building classifiers based on a mixture of DWT and WPD features is not considered. It is also interesting to notice that 6 features are extracted from the DE sensor and just 2 from the FE sensor. Thus, it seems that the drive-end features are more informative with respect to bearing degradation than the fan-end features. This is also confirmed by the analysis of the solution of the Pareto optimal front with features extracted from only one sensor (circles in Figure 7): in all these solutions the DE sensor is selected. Finally, according to the identified optimal compromise solution, the classifiers

for the identification of the degradation mode require more features than those for the identification of the degradation level. This can be interpreted by observing that the task of the degradation mode classifiers is more complex since it has to consider a large set of patterns characterized by all the types of degradation modes, whereas the degradation level classifiers have to consider only a subset of those patterns, i.e. those characterized by a specific degradation mode (Figure 2) and thus a more limited training space.

4.5 Classification results

Once the feature selection task has been performed, the bearing diagnostic system has been developed using as input features for classifiers at stages 2 and 3 the features in Table 6 (first and second columns, respectively). With respect to the classifiers, C_0 and C_1 , we have performed an exhaustive search among all the possible combinations of the 8 selected features for the classifiers at stage 2 and we have obtained the best performance using the feature sets in Tables 8 and 9, respectively. It is interesting to notice that one feature is sufficient for the bearing detection task.

The overall performance of the hierarchical classification model has been verified on the data of the validation set, not previously used during the feature selection. The percentage of patterns for which the classification is correct in all the 4 stages of the diagnostic system is 79.78% in the case in which the training set is forced to contain only patterns collected from an operational condition (load) different from that of the test patterns (hereafter referred to as Case 1) and 97.61% in the case in which the training set contains patterns collected at any load (hereafter referred to as Case 2). Table 10, second column, reports the performances of the single classifiers of the hierarchical structure in Case 1 and column 4 in Case 2.

The less satisfactory performance is obtained by classifier $C_2^{2,3}$ which is devoted to the identification of the intensity of the FE bearing degradation due to outer raceway defects. It is interesting to notice

that according to (CWRU), the Case Western Reserve University Bearing data referring to the outer raceway defects with an intensity level of 21 mils and at load 0 have been collected considering bearing with defects located at the 3 o'clock direction (directly in the load zone), whereas, for the other degradation levels, the defects are located at a 6 o'clock direction (orthogonal to the load zone). Thus, the misclassifications are due to the different ways in which the defects are induced, as it can be seen in Table 11 which reports the misclassifications of the patterns at the different intensity levels. Misclassifications of C_2^2 are also due to the same cause.

With respect to the analysis of the other misclassification causes, it is interesting to observe that the proposed feature selection approach is constraining all the classifiers of a given level of the hierarchical model to be based on the same set of features, i.e. the 2 classifiers (C_2^1 and C_2^2) at level 2 are all based on the features in Table 6, first column and all the 6 classifiers at level 3 ($C_3^{b,i}$, $b = 1,2$ and $i = 1,2,3$) on the features, Table 6, second column. This choice allows obtaining, for each level of the hierarchical model, a set of features which provides a good compromise between the performance of the different classifiers of the level, but is not optimal for the single classifier. Considering, for example, classifier C_2^1 which is devoted to the classification of the degradation mode for the drive-end bearing, its performance can be remarkably increased by considering a subset of the selected features which does not contain the features measured by the FE sensor. In particular, the misclassification rate of C_2^1 reduces from 4.17 to 0.02 when only features 2A, 2B and 2E (Table 6) are used as input of the classifier. Thus, the fan-end features (2F and 2G in Table 6) have been selected by the MOBDE algorithm only for the information that they provide for the classification of defects at the fan-end bearing, but they cause a decrease in the performance of the drive-end bearing fault classifier. This can be graphically seen in Figures 8 and 9: the patterns representative of the different degradation modes are clearly separated when the drive-end features 2A, 2B and 2E are used (Figure 8), whereas they become more confused when the fan-end feature 2G is taken into account (Figure 9)

5. CONCLUSIONS

In this work, we have developed a diagnostic approach for the identification and characterization of defeats in automotive bearings based on a hierarchical architecture of K-Nearest Neighbour classifiers. Different features extracted from acceleration signals in the time and frequency domains have been considered, and an optimal feature set has been identified by resorting to a wrapper approach based on the use of a binary differential evolution algorithm. Multiple objectives of the search have been the maximization of the diagnostic system performance, and the minimization of the cost associated to the development of the diagnostic system and the measurement of the acceleration signals. Since the external conditions experienced by automotive bearings remarkably influence the acceleration signal data and, thus, may cause unsatisfactory performance in application, a further requirement is the independence of the extracted features from the external conditions. The developed method has been applied with success to the data of the Western Reserve Case University Bearing dataset which contains real vibrational data collected in experimental tests performed on degraded bearings.

The practical deployment and validation of the proposed diagnostic approach for automotive bearings requires the design and execution of further tests reproducing bearing degradation in automotive vehicles under realistic external conditions. Further research work will be devoted to the application of the developed method to real data collected during FEV operation.

This activity is being performed within the European Project HEMIS (www.hemis-eu.org), whose objective is the development of prognostics and health monitoring systems for the most critical components of Fully Electric Vehicles.

5.1 Acknowledgements

The research leading to these results has received funding from the European Community's Framework Programme (FP7/2007-2013) under grant agreement n° 314609. The authors are grateful for the support and contributions from other members of the HEMIS project consortium, from CEIT (Spain), IDIADA (Spain), Jema (Spain), MIRA (UK), Politecnico di Milano (Italy), VTT (Finland),

and York EMC Services (UK). Further information can be found on the project website (www.hemis-eu.org).

The participation of Enrico Zio to this research is partially supported by the China NSFC under grant number 71231001. The participation of Piero Baraldi and Francesco Di Maio to this research is partially supported by the European Union Project INNOvation through Human Factors in risk analysis and management (INNHF, www.innhf.eu) funded by the 7th framework program FP7-PEOPLE-2011- Initial Training Network: Marie-Curie Action.

References

- Abbasion, S., Rafsanjani, A., Farshidianfar, A., Irani, N., “Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine”, *Mechanical Systems and Signal Processing*, 21 (7), pp. 2933-2945, 2007.
- Abed, W., Sharma, S., Sutton, R., Motwani, A., “A Robust Bearing Fault Detection and Diagnosis Technique for Brushless DC Motors Under Non-stationary Operating Conditions”, *Journal of Control, Automation and Electrical Systems*, 26 (3), pp. 241-254, 2015.
- Al-ani, A., “Feature subset selection using ant colony optimization”, *International Journal of Computational Intelligence*, pp. 53-58, 2005.
- Ali, M.M., Törn, A., “Population set-based global optimization algorithms: Some modifications and numerical studies”, *Computers and Operations Research*, 31 (10), pp. 1703-1725, 2004.
- Azevedo, C.R.B., Araujo, A.F.R., “Correlation between diversity and hypervolume in evolutionary multiobjective optimization”, *IEEE Congress of Evolutionary Computation, CEC 2011*, art. no. 5949962, pp. 2743-2750, 2011.
- Baraldi, P., Di Maio, F., Pappaglione, L., Zio, E., Seraoui, R., “Condition Monitoring of Electrical Power Plant Components During Operational Transients”, *Proceedings of the Institution of Mechanical Engineers, Part O, Journal of Risk and Reliability*, 226(6) 568–583, 2012.
- Baraldi, P., Barbisotti, F., Capanna, R., Colombo, S., Rigamonti, M., Zio, E., “Assessment of the Performance of a Fully Electric Vehicle Subsystem in Presence of a Prognostic and Health Monitoring System”, *Chemical Engineering Transactions*, 33, pp. 787-792, 2013.
- Baraldi, P., Zio, E., Mangili, F., Gola, G., Nystad, B.H., “Ensemble of Kernel Regression Models for Assessing the Health State of Choke Valves in Offshore Oil Platforms”, *International Journal of Computational Intelligence Systems*, 7 (2), pp. 225-241, 2014.
- Baudat, G., Anouar, F., “Generalized discriminant analysis using a kernel approach”, *Neural Computation*, 12 (10), pp. 2385-2404, 2000.
- Bellini, A., Filippetti, F., Tassoni, C., Capolino, G.-A., “Advances in diagnostic techniques for induction machines”, *IEEE Transactions on Industrial Electronics*, 55 (12), pp. 4109-4126, 2008.

- Ben Ali, J., Saidi, L., Mouelhi, A., Chebel-Morello, B., Fnaiech, F., “Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally progressing degradations”, *Engineering Applications of Artificial Intelligence*, 42, pp. 67-81, 2015.
- Bonnett, A.H., Yung, C., “Increased efficiency versus increased reliability”, *IEEE Industry Applications Magazine*, 14 (1), pp. 29-36, 2008.
- Chebil, G.Noel, M. Mesbah, M.Deriche, “Wavelet decomposition for the detection of faults in rolling bearing elements”, *Jordan Journal of Mechanical and Industrial Engineering*, pp. 260-267, 2009.
- Compare, M., Baraldi, P., Cannarile, F., Di Maio, F., Zio, E., “Homogeneous finite-time, finite-state, semi-Markov modelling for enhancing Empirical Classification System diagnostics of industrial components”, *Probabilistic Engineering Mechanics*, under review, 2016.
- CWRU, Case Western Reserve University website <http://csegroups.case.edu/bearingdatacenter/home> .
- Dash, M., Liu, H., “Feature selection for classification”, *Intelligent Data Analysis*, 1 (3), pp. 131-156, 1997.
- Dalvand, F., Kalantar, A., Safizadeh, M.S., “A novel bearing condition monitoring method in induction motors based on instantaneous frequency of motor voltage”, *IEEE Transactions on Industrial Electronics*, 63 (1), art. no. 7181676, pp. 364-376, 2016.
- Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., “A fast and elitist multiobjective genetic algorithm: NSGA-II”, *IEEE Transactions on Evolutionary Computation*, 6 (2), pp. 182-197, 2002.
- Di Maio, F., Hu, J., Tse, P., Pecht, M., Tsui, K., Zio, E., “Ensemble-approaches for clustering health status of oil sand pumps”, *Expert Systems with Applications*, 39 (5), pp. 4847-4859, 2012a.
- Di Maio, F., Tsui, K.L., Zio, E., “Combining Relevance Vector Machines and Exponential Regression for Bearing RUL estimation”, *Mechanical Systems and Signal Processing*, 31, 405–427, 2012b.
- Dong, M., Kothari, R., “Feature subset selection using a new definition of classifiability”, *Pattern Recognition Letters*, 24 (9-10), pp. 1215-1225, 2003.
- Emmanouilidis, C., Hunter, A., MacIntyre, J., Cox, C., “Selecting features in neurofuzzy modelling by multiobjective genetic algorithms”, *Conference Publication*, 2 (470), pp. 749-754, 1999.
- Firpi, H.A., Goodman, E., “Swarmed feature selection”, *Proceedings - Applied Imagery Pattern Recognition Workshop*, pp. 112-118, 2005.
- Geramifard, O., Xu, J.-X., Kumar Panda, S., “Fault detection and diagnosis in synchronous motors using hidden Markov model-based semi-nonparametric approach”, *Engineering Applications of Artificial Intelligence*, 26 (8), pp. 1919-1929, 2013.
- Georgoulas, G., Loutas, T., Stylios, C.D., Kostopoulos, V. “Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition”, *Mechanical Systems and Signal Processing*, 41 (1-2), pp. 510-525, 2013.
- Gong, W., Cai, Z., Wang, Y., “Repairing the crossover rate in adaptive differential evolution”, *Applied Soft Computing Journal*, 15, pp. 149-168, 2014.
- Gryllias, K.C., Antoniadis, I.A., “A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments”, *Engineering Applications of Artificial Intelligence*, 25 (2), pp. 326-344, 2012.

- He, Y., Chu, F., Zhong, B., "A study on group decision-making based fault multi-symptom-domain consensus diagnosis", *Reliability Engineering and System Safety*, 74(1), pp. 43-52, 2001.
- He, X., Zhang, Q., Sun, N., Dong, Y., "Feature selection with discrete binary differential evolution" 2009 International Conference on Artificial Intelligence and Computational Intelligence, AICI 2009, 4, art. no. 5376334, pp. 327-330, 2009.
- Hellman, M.E., "The Nearest Neighbor Classification Rule with a Reject Option", *IEEE Transactions on Systems Science and Cybernetics*, 6 (3), pp. 179-185, 1970.
- Huang, N.H., et al., "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proc. R. Soc. Lond. A.*, 1998.
- Hui, L., Haiqi, Z., Liwei, T., "Wigner-Ville distribution based on EMD for faults diagnosis of bearing", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4223 LNAI, pp. 803-812, 2006.
- Jiang, L., Xuan, J., Shi, T., "Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis", *Mechanical Systems and Signal Processing*, 41 (1-2), pp. 113-126, 2013.
- Jin, X., Zhao, M., Chow, T.W.S., Pecht, M., "Motor bearing fault diagnosis using trace ratio linear discriminant analysis", *IEEE Transactions on Industrial Electronics*, 61 (5), art. no. 6583974, pp. 2441-2451, 2014.
- Kaewkongka, T., Au, Y.H.J., Rakowski, R.T., Jones, B.E., "A comparative study of short time fourier transform and Continuous Wavelet Transform for bearing condition monitoring", *International Journal of COMADEM*, 6 (1), pp. 41-48, 2003.
- Khushaba, R.N., Al-Ani, A., Al-Jumaily, A., "Feature subset selection using differential evolution and a statistical repair mechanism", *Expert Systems with Applications*, 38 (9), pp. 11515-11526, 2011.
- Kohavi, R., John, G.H., "Wrappers for feature subset selection", *Artificial Intelligence*, 97 (1-2), pp. 273-324, 1997.
- Kohonen, T., "Self-organized formation of topologically correct feature maps", *Biological Cybernetics*, 43 (1), pp. 59-69, 1982.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y., "Testing the null hypothesis of stationarity against the alternative of a unit root". *Journal of Econometrics* 54 (1-3), 159-178, 1992.
- Li, F., Tang, B., Yang, R., "Rotating machine fault diagnosis using dimension reduction with linear local tangent space alignment", *Measurement: Journal of the International Measurement Confederation*, 46 (8), pp. 2525-2539, 2013.
- Li, J, Ma, J, "Wavelet decomposition of vibrations for detection of bearing-localized defects", *NDT & E International*, Volume 30, Issue 3, Pages 143-149, June 1997.
- Mallipeddi, R., Suganthan, P.N., "Empirical study on the effect of population size on differential evolution algorithm", 2008 IEEE Congress on Evolutionary Computation, CEC 2008, art. no. 4631294, pp. 3663-3670, 2008.
- Martin, H.R., Honarvar, F., "Application of statistical moments to bearing failure detection", *Applied Acoustics*, 44 (1), pp. 67-77, 1995.
- Mezura-Montes, E., Reyes-Sierra, M., Coello, C.A., "Multi-objective optimization using differential evolution: A survey of the state-of-the-art", *Studies in Computational Intelligence*, 143, pp. 173-196, 2008.

- Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Muller, K., "Fisher discriminant analysis with kernels", *Neural Networks for Signal Processing - Proceedings of the IEEE Workshop*, pp. 41-48, 1999.
- Ocak, H., Loparo, K.A., Discenzo, F.M., "Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics", *Journal of Sound and Vibration*, 302 (4-5), pp. 951-961, 2007.
- O'Donnell, Pat, Bell, R.N., McWilliams, D.W., Singh, C., Wells, S.J., "Report of large motor reliability survey of industrial and commercial installations", *IEEE Conference Record of Industrial and Commercial Power Systems Technical Conference*, pp. 58-70, 1983.
- Opricovic, S., Tzeng, G.-H. "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS", *European Journal of Operational Research*, 156 (2), pp. 445-455, 2004.
- Pan, Y., Er, M.J., Li, X., Yu, H., Gouriveau, R., "Machine health condition prediction via online dynamic fuzzy neural networks" *Engineering Applications of Artificial Intelligence*, 35, pp. 105-113, 2014.
- Parvin, H., Mirnabibaboli, M., Alinejad-Rokny, H., "Proposing a classifier ensemble framework based on classifier selection and decision tree", *Engineering Applications of Artificial Intelligence*, 37, pp. 34-42, 2015.
- Pichler, K., Lughofer, E., Pichler, M., Buchegger, T., Klement, E.P., Huschenbett, M., "Fault detection in reciprocating compressor valves under varying load conditions", *Mechanical Systems and Signal Processing*, 70-71, pp. 104-119, 2016.
- Polikar, R., "Bootstrap-inspired techniques in computational intelligence", *IEEE Signal Processing Magazine* 59, 59-72, 2007.
- Rao, B.K.N., Srinivasa Pai, P., Nagabhushana, T.N., "Failure diagnosis and prognosis of rolling - Element bearings using artificial neural networks: A critical overview", *Journal of Physics: Conference Series*, 364 (1), 2012.
- Salman, A., Engelbrecht, A.P., Omran, M.G.H., "Empirical analysis of self-adaptive differential evolution", *European Journal of Operational Research*, 183 (2), pp. 785-804, 2007.
- Samanta, B., Nataraj, C., "Use of particle swarm optimization for machinery fault detection", *Engineering Applications of Artificial Intelligence*, 22 (2), pp. 308-316, 2009.
- Schoen, Randy R., Habetler, Thomas G., Kamran, Farrukh, Bartheld, Robert G., "Motor bearing damage detection using stator current monitoring", *IEEE Transactions on Industry Applications*, 31 (6), pp. 1274-1279, 1995.
- Schölkopf, B., Smola, A., Müller, K.-R., "Nonlinear Component Analysis as a Kernel Eigenvalue Problem", *Neural Computation*, 10 (5), pp. 1299-1319, 1998.
- Sedano, B., Astigarraga, D., Baraldi, P., Rigamonti, M., Zio, E., "Assessment of the improvement achieved in RAMS by a FEV embedding a powertrain PHMS", *Safety and Reliability of Complex Engineered Systems - Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015*, pp. 1833-1838, 2015.
- Serdio, F., Lughofer, E., Pichler, K., Buchegger, T., Pichler, M., Efendic, H., "Fault detection in multi-sensor networks based on multivariate time-series models and orthogonal transformations", *Information Fusion*, 20 (1), pp. 272-291, 2014.
- Sikora, R., Piramuthu, S., "Framework for efficient feature selection in genetic algorithm based data mining", *European Journal of Operational Research*, 180 (2), pp. 723-737, 2007.

- Tao, B., Zhu, L., Ding, H., Xiong, Y. “An alternative time-domain index for condition monitoring of rolling element bearings-A comparison study”, *Reliability Engineering and System Safety*, 92(5), pp. 660-670, 2007.
- Tian, J., Morillo, C., Azarian, M.H., Pecht, M., “Motor Bearing Fault Detection Using Spectral Kurtosis-Based Feature Extraction Coupled with K-Nearest Neighbor Distance Analysis”, *IEEE Transactions on Industrial Electronics*, 63 (3), art. no. 7360161, pp. 1793-1803, 2016.
- Tušar, T., Filipič, B. “Differential evolution versus genetic algorithms in multiobjective optimization”, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4403 LNCS, pp. 257-271, 2007.
- Wang, Y., Cai, Z., Zhang, Q., “Differential evolution with composite trial vector generation strategies and control parameters”, *IEEE Transactions on Evolutionary Computation*, 15 (1), art. no. 5688232, pp. 55-66, 2011.
- Zhang, Y., Zuo, H., Bai, F., “Classification of fault location and performance degradation of a roller bearing”, *Measurement: Journal of the International Measurement Confederation*, 46 (3), pp. 1178-1189, 2013.
- Zhang, P., Du, Y., Habetler, T.G., Lu, B., “A survey of condition monitoring and protection methods for medium-voltage induction motors”, *IEEE Transactions on Industry Applications*, 47 (1), art. no. 5620974, pp. 34-46, 2011.
- Zhang, X., Zhou, J., “Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines”, *Mechanical Systems and Signal Processing*, 41 (1-2), pp. 127-140, 2013.
- Zhu, K., Song, X., Xue, D., “A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm”, *Measurement*, Volume 47, Pages 669-675, January 2014.
- Zio, E., Baraldi, P., Pedroni, N. “Selecting features for nuclear transients classification by means of genetic algorithms”, *IEEE Transactions on Nuclear Science*, 53 (3), pp. 1479-1493, 2006.
- Zio, E., Baraldi, P., Gola, G., “Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery”. *Applied Soft Computing*, Vol. 8 (4), pp. 1365–1380, 2008.
- Zio, E., Gola, G., “A neuro-fuzzy technique for fault diagnosis and its application to rotating machinery”, *Reliability Engineering and System Safety*, 94(1), pp. 78-88, 2009.
- Zitzler, E., Laumanns, M., Thiele, L., “SPEA2: Improving the Performance of the Strength Pareto Evolutionary Algorithm”, *Technical Report 103*, Computer Engineering and Communication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH) Zurich, 2001
- Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Da Fonseca, V.G., “Performance assessment of multiobjective optimizers: An analysis and review”, *IEEE Transactions on Evolutionary Computation*, 7 (2), pp. 117-132, 2003.
- Zitzler, E., Künzli, S., “Indicator-based selection in multiobjective search”, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3242, pp. 832-842, 2004.

Appendix A: List of features

1. Mean value
2. Kurtosis
3. Skew value
4. Standard Deviation
5. Crest indicator
6. Clearance indicator
7. Shape indicator
8. Impulse indicator
9. Peak value
10. Minimum Haar Wavelet coefficient
11. Maximum Haar Wavelet coefficient
12. Norm level A3 Daubechies Wavelet transform
13. Norm level D3 Daubechies Wavelet transform
14. Norm level D2 Daubechies Wavelet transform
15. Norm level D1 Daubechies Wavelet Transform
16. Norm Node 1 Symlet6 Wavelet
17. Norm Node 2 Symlet6 Wavelet
- ...
28. Norm Node 13 Symlet6 Wavelet
29. Norm Node 14 Symlet6 Wavelet

Appendix B: the TOPSIS method for the selection of the best compromise solution

The basic principle of this technique is that the chosen alternative should have the shortest distance from the ideal solution and the farthest distance from the negative-ideal solution. The TOPSIS method is described in (Opricovic & Tzeng, 2004):

1. Compute for each solution in the Pareto optimal set the values $\tilde{z}_{p,\alpha}$:

$$\tilde{z}_{p,\alpha} = F_p(z_p) \frac{F_p(z_p)}{\sqrt{\sum_{s=1}^P F_\alpha(z_s)^2}} \quad p = 1:P, \alpha = 1:4 \quad (13)$$

2. Calculate the weighted values $\xi_{p,\alpha}$:

$$\xi_{p,\alpha} = \tilde{z}_{p,\alpha} \cdot \omega_\alpha \quad p = 1:P, \alpha = 1:4 \quad (14)$$

$$\sum_{\zeta=1}^4 \omega_\zeta = 1 \quad (15)$$

where ω_α indicates the relative importance of the i -th objective and is here taken equal to 0.25 for all the objectives.

3. Determine the ideal and negative-ideal solution:

$$A^* = \{\xi_1^*, \xi_2^*, \xi_3^*, \xi_4^*\} = \min_{\alpha} (\xi_{p,\alpha} | \alpha = 1:4) \quad (16)$$

$$A^- = \{\xi_1^-, \xi_2^-, \xi_3^-, \xi_4^-\} = \max_{p=1:P} (\xi_{p,\alpha} | \alpha = 1:4) \quad (17)$$

4. Compute the separation measures, according to the Euclidean distance. The separation of each candidate solution from the ideal solution is given by

$$D_p^* = \sqrt{\sum_{\alpha=1}^4 (\xi_{p,\alpha} - \xi_{\alpha}^*)^2} \quad p = 1:P \quad (18)$$

Analogously, from the negative-ideal solution is given by

$$D_p^- = \sqrt{\sum_{\alpha=1}^4 (\xi_{p,\alpha} - \xi_{\alpha}^-)^2} \quad p = 1:P \quad (19)$$

5. Calculate the relative closeness to the ideal solution. For each candidate solution \mathbf{z}_p the relative closeness with respect to A^* is defined as

$$C_p^* = \frac{D_p^-}{D_p^* + D_p^-} \quad p = 1:P \quad (20)$$

6. Ranking the solutions in increasing order.

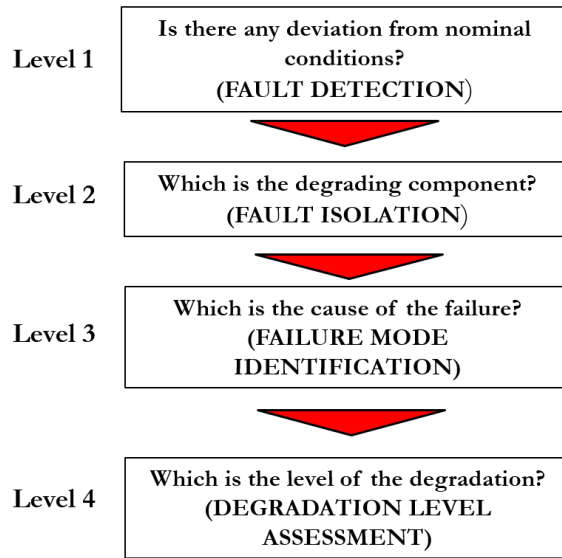


Figure 1: Four level representation of fault detection, fault isolation, failure mode identification and degradation level assessment.

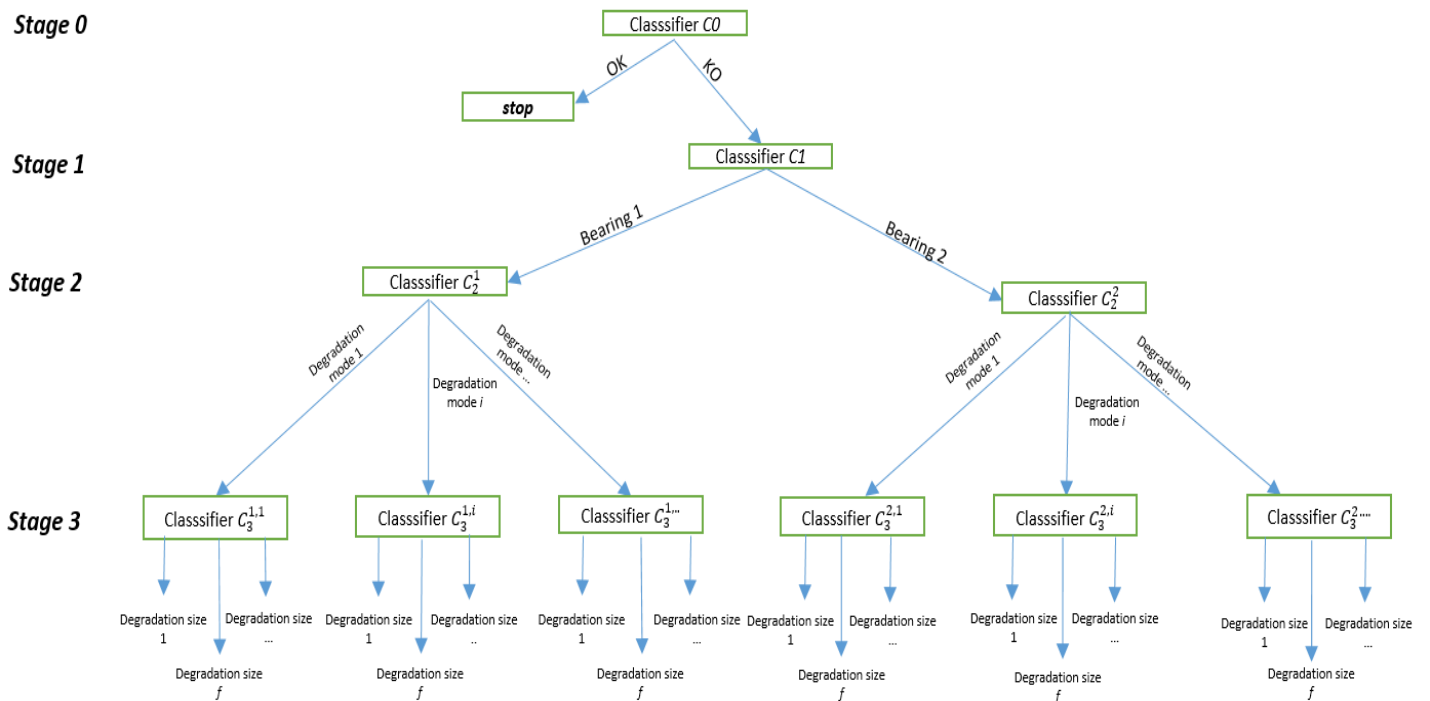


Figure 2: The hierarchical model of the bearing diagnostic system.

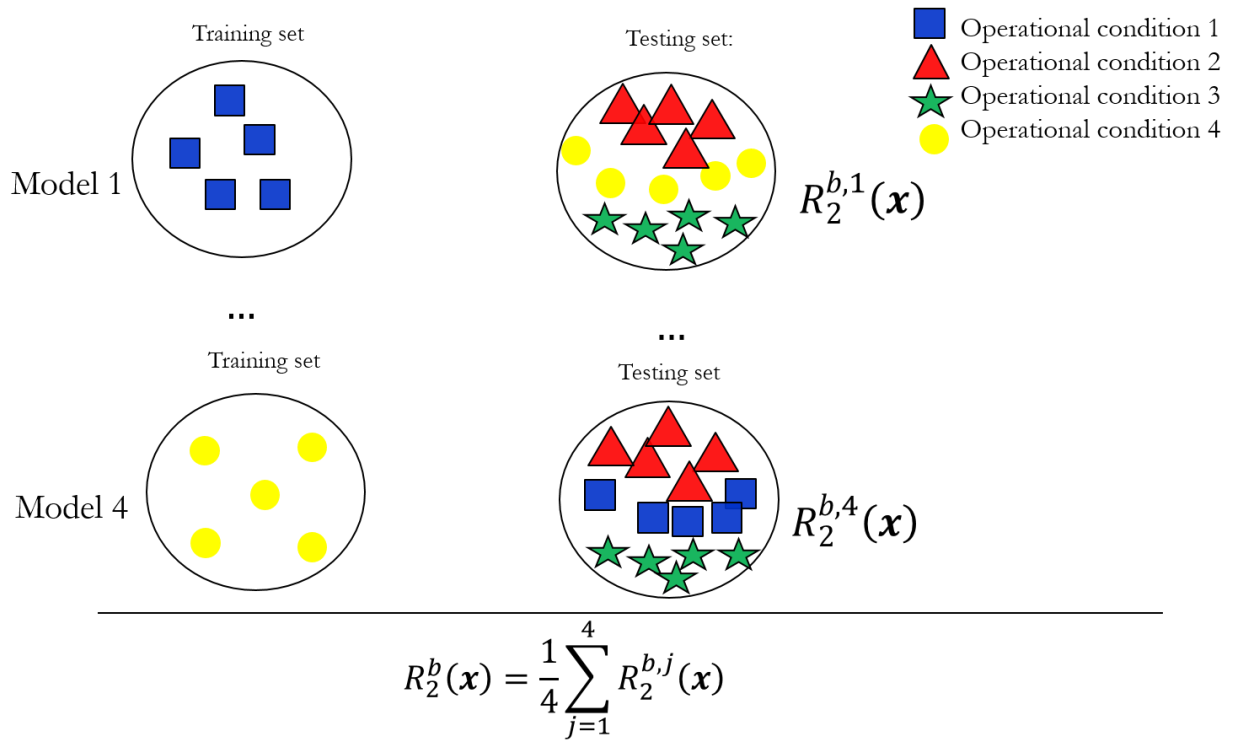


Figure 3: Leave all but one cross validation with $l = 4$.

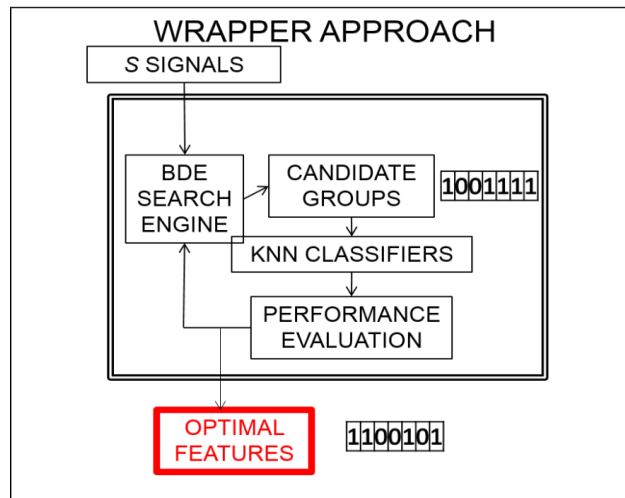


Figure 4: Wrapper approach for optimal feature subset selection based on BDE optimization algorithm.

<i>Degradation i</i>	<i>Failure intensity f</i>	<i>Number of patterns (all loads)</i>	<i>Number of patterns for each load</i>
Inner race (DE)	7 mils	40	10
Inner race (DE)	14 mils	40	10
Inner race (DE)	21 mils	40	10
Balls (DE)	7 mils	40	10

Balls (DE)	14 mils	40	10
Balls (DE)	21 mils	40	10
Outer race (DE)	7 mils	40	10
Outer race (DE)	14 mils	40	10
Outer race (DE)	21 mils	40	10
Inner race (FE)	7 mils	40	10
Inner race (FE)	14 mils	40	10
Inner race (FE)	21 mils	40	10
Balls (FE)	7 mils	40	10
Balls (FE)	14 mils	40	10
Balls (FE)	21 mils	40	10
Outer race (FE)	7 mils	40	10
Outer race (FE)	14 mils	40	10
Outer race (FE)	21 mils	40	10

Table 1: Type of degradation mode and intensity in all the available patterns. DE=Drive End bearing and FE=Fan End bearing

<i>Selected features</i>
Shape Indicator (DE)
Peak Value (DE)
Norm Node 5 Symlet6 wavelet (DE)
Norm Node 11 Symlet6 wavelet (DE)
Norm Node 14 Symlet6 wavelet (DE)

Table 2: Selected features (DE= Drive End sensor)

<i>Work</i>	<i>Number of features extracted</i>	<i>Feature selection approach</i>	<i>Number of features after reduction</i>	<i>Number of classes considered</i>	<i>Bearing considered</i>	<i>Classifier</i>	<i>Misclassification rate</i>
Zhang et al., 2013	21	Kernel Principal Component Analysis	3	7	Drive End	SVM	0.47%
Jiang et al., 2013	16	Semi-supervised kernel Marginal Fisher Analysis	5	10	Drive End	SVM KNN	0.00% 1.50%
Zhu et al., 2014	8	None	8	10	Drive End	SVM	0.00%
Li et al., 2013	14	Linear Local Tangent Space Alignment	3	7	Drive End	Littlewoods-Paley SVM	5.71%
Ours	29	Wrapper search	5	10	Drive End	KNN	0.01%

Table 3: Comparison of our work with other literature works

	<i>No. of patterns in the training set</i>	<i>No. of patterns in the test set</i>
$R_2^{b,j}$	144	432
$R_3^{b,i,j}$	48	144

Table 4: Number of patterns in the training and test set.

<i>Selection strategy</i>	<i>Hyper Volume Median</i>	<i>Hyper Volume Mean value</i>	<i>Hyper Volume standard deviation</i>
NSBDE	148.4094	148.4146	0.1913
MODE III	64.2715	64.4124	1.3780

Table 5: Statistics on the hyper-volume over the non-dominated set obtained by applying the NSDE and the MODE III selection strategies.

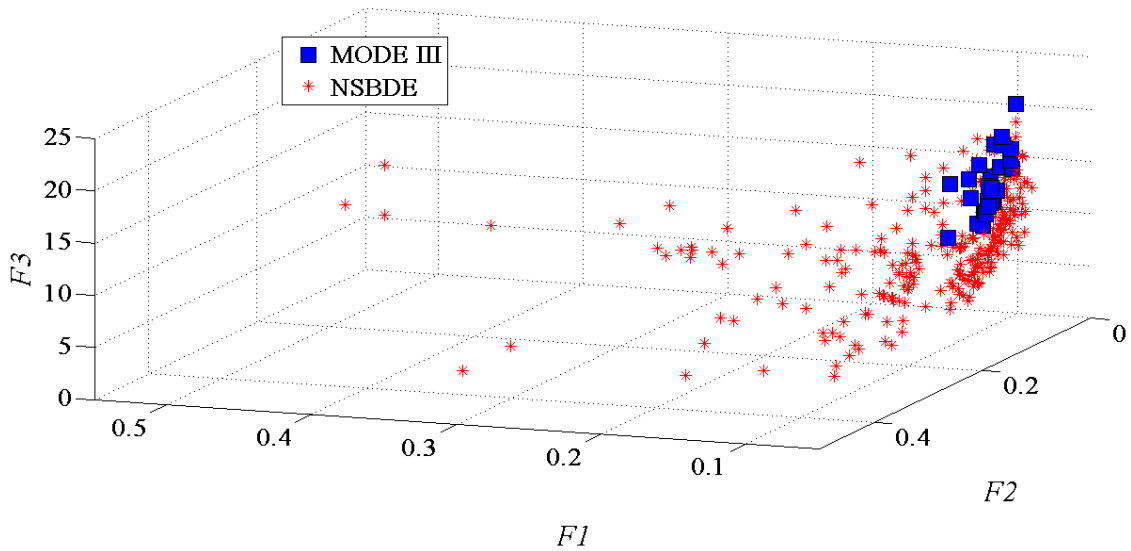


Figure 5: Pareto optimal front, after $G = 1500$ generations (stars NSBDE strategy, squares MODE III strategy).

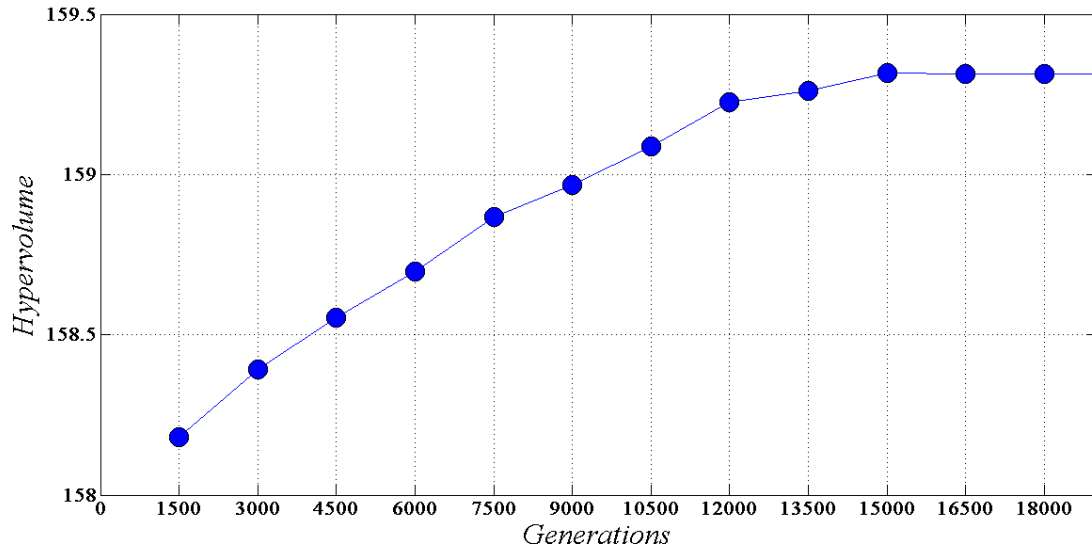


Figure 6: Hyper-volume values every 1500 generations.

The NSBDE based Pareto optimal front consists of $P = 211$ solutions, \mathbf{z}^* (Figure 6).

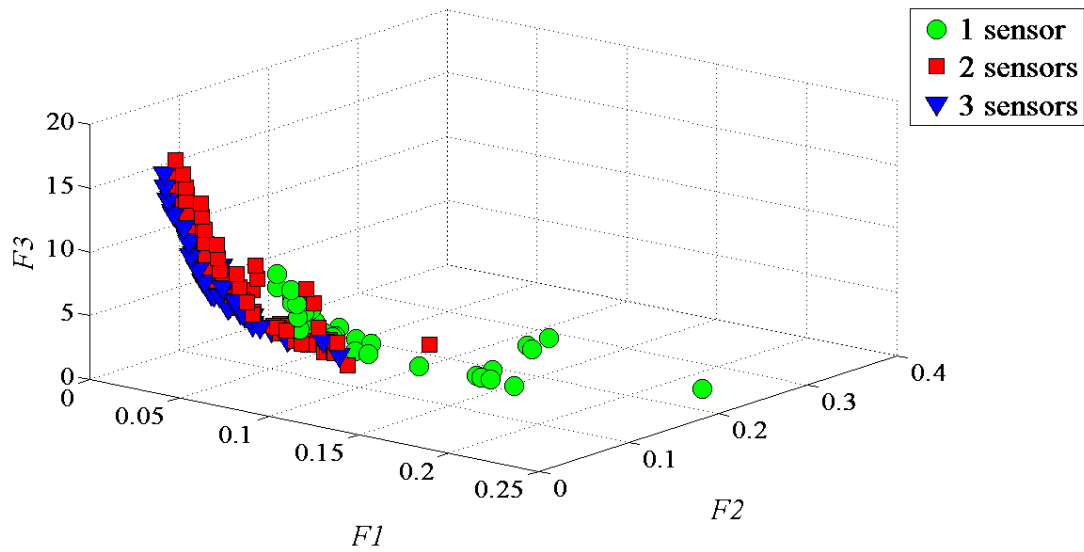


Figure 7: Pareto optimal front after 19500 generations.

<i>Input features selected for the classifiers of the degradation mode (C_2^1 and C_2^2)</i>	<i>Input features selected for the classifiers of the degradation level ($C_3^{1,1}, C_3^{1,2}, C_3^{1,3}, C_3^{2,1}, C_3^{2,2}, C_3^{2,3}$)</i>
2A = Peak Value (DE)	3A = Peak Value (DE)
2B = Norm Node 5 Symlet6 wavelet (DE)	3B = Minimum Haar wavelet coefficient (DE)
2C = Norm Node 7 Symlet6 wavelet (DE)	3C = Norm Node 5 Symlet6 wavelet (DE)

2D = Norm Node 12 Symlet6 wavelet (DE)	3D = Norm Node 12 Symlet6 wavelet (DE)
2E = Norm Node 14 Symlet6 wavelet (DE)	3E = Norm Node 11 Symlet6 wavelet (FE)
2F = Minimum Haar wavelet coefficient (FE)	
2G = Norm Node 11 Symlet6 wavelet (FE)	

Table 6: Features in the optimal compromise solution z_{opt}^* : the column on the left contains the features selected for the classifiers of the degradation mode (hereafter indicated by 2A, 2B,..., 2G), the column on the right that for the classifiers of the degradation level (hereafter indicated by 3A, 3B,..., 3E); DE refers to features extracted from the Drive End sensor, FE from the Fan End sensor.

F_1	F_2	F_3	F_4
0.0463	0.0495	8	2

Table 7: Objective function values in the optimal solution z_{opt}^* .

<i>Detection of the degradation</i>
Norm Node 7 Symlet6 wavelet (DE)

Table 8: Input feature of the bearing d the solution z_{opt}^* for the detection of the degradation classifier (DE= Drive End sensor).

<i>Isolation of the degrading bearing</i>
Peak Value (DE)
Minimum Haar wavelet coefficient (DE)
Norm Node 12 Symlet6 wavelet (DE)
Minimum Haar wavelet coefficient (FE)

Table 9: Features of the solution z_{opt}^* for the degradation isolation of the degrading bearing classifier (DE= Drive End sensor, FE=Fan End sensor).

<i>Classifiers</i>	<i>Misclassification rate Case 1</i>	<i>Misclassification rate Case 2</i>	<i>Standard deviation Case2</i>
C_0	0.00%	0.00%	$\pm 0.00\%$
C_1	1.61%	0.22%	$\pm 0.01\%$
C_2^1	4.17%	1.11%	$\pm 3.33\%$
C_2^2	8.33%	4.44%	$\pm 5.07\%$
$C_3^{1,1}$	0.00%	0.00%	$\pm 0.00\%$
$C_3^{1,2}$	12.50%	0.00%	$\pm 0.00\%$
$C_3^{1,3}$	0.00%	0.00%	$\pm 0.00\%$
$C_3^{2,1}$	12.50%	0.00%	$\pm 0.00\%$

$C_3^{2,2}$	8.33%	2.67%	$\pm 6.24\%$
$C_3^{2,3}$	16.67%	6.67%	$\pm 8.21\%$

Table 10: Misclassification rates of the KNN-classifiers for the solution Z_{opt}^* .

<i>True class of the test pattern (intensity of the degradation level)</i>	<i>Misclassification rate</i>
7	0
14	0
21	0.50
<i>Mean</i>	0.1667

Table 11: Analysis of the misclassification rate of the $C_3^{2,3}$ classifier.

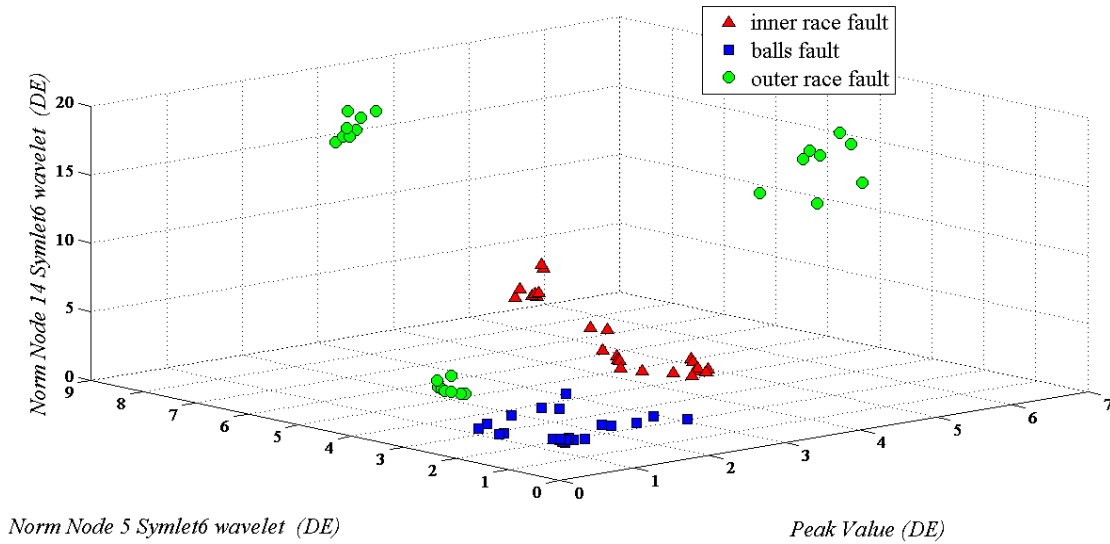


Figure 8: Representation of the patterns used to train classifier C_2^1 in the space of the Peak Value (DE), Norm Node 5 Symlet6 (DE) and Norm Node 14 Symlet6 (DE) features.

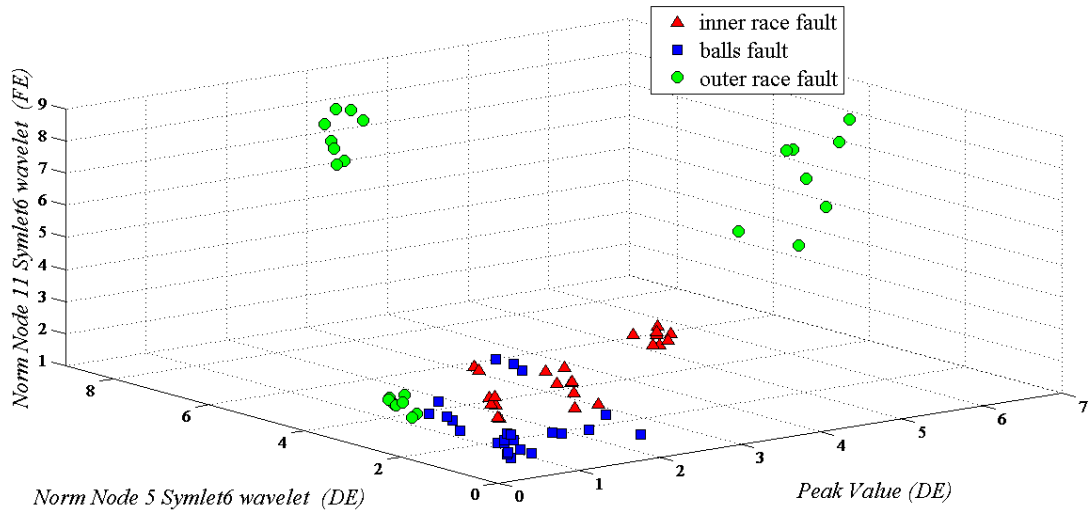


Figure 9: Representation of the patterns used to train classifier C_2^1 in the space of Peak Value (DE), Norm Node 5 Symlet6 (DE) and Norm Node 11 Symlet6 (FE).