# Journal Pre-proof
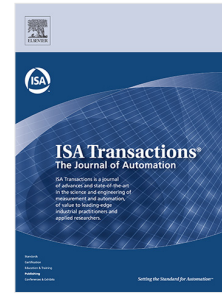
Early and extremely early multi-label fault diagnosis in induction motors

Mario Juez-Gil, Juan José Saucedo-Dorantes, Álvar Arnaiz-González,
Carlos López-Nozal, César García-Osorio, David Lowe

Please cite this article as: M. Juez-Gil, J.J. Saucedo-Dorantes, Á. Arnaiz-González et al., Early and extremely early multi-label fault diagnosis in induction motors. *ISA Transactions* (2020), doi: https://doi.org/10.1016/j.isatra.2020.07.002.

**Title:**

Early and extremely early multi-label fault diagnosis in induction motors

**Authors:**

Mario Juez-Gil[a], Juan José Saucedo-Dorantes[b], Álvar Arnaiz-González[a], Carlos López-Nozal[a], César García-Osorio[a,*], David Lowe[c]

[a] Universidad de Burgos, Burgos, Spain
[b] Autonomous University of Queretaro, Mexico
[c] Aston University, Birmingham, United Kingdom

*Corresponding author
Address: Avda. Cantabria, s/n; 09006 Burgos; Spain
Telephone: +34 947112457
E-mail: cgosorio@ubu.es

**Conflict of interest:**

There is no conflict of interest with the concerned persons or organizations.

**Graphical Abstract (for review)**

INDUCTION MOTOR    MEASUREMENTS    MULTI-FAULT

SENSOR #1    PCA #1

SENSOR #2    PCA #2

⋮    ⋮

SENSOR #12    PCA #12
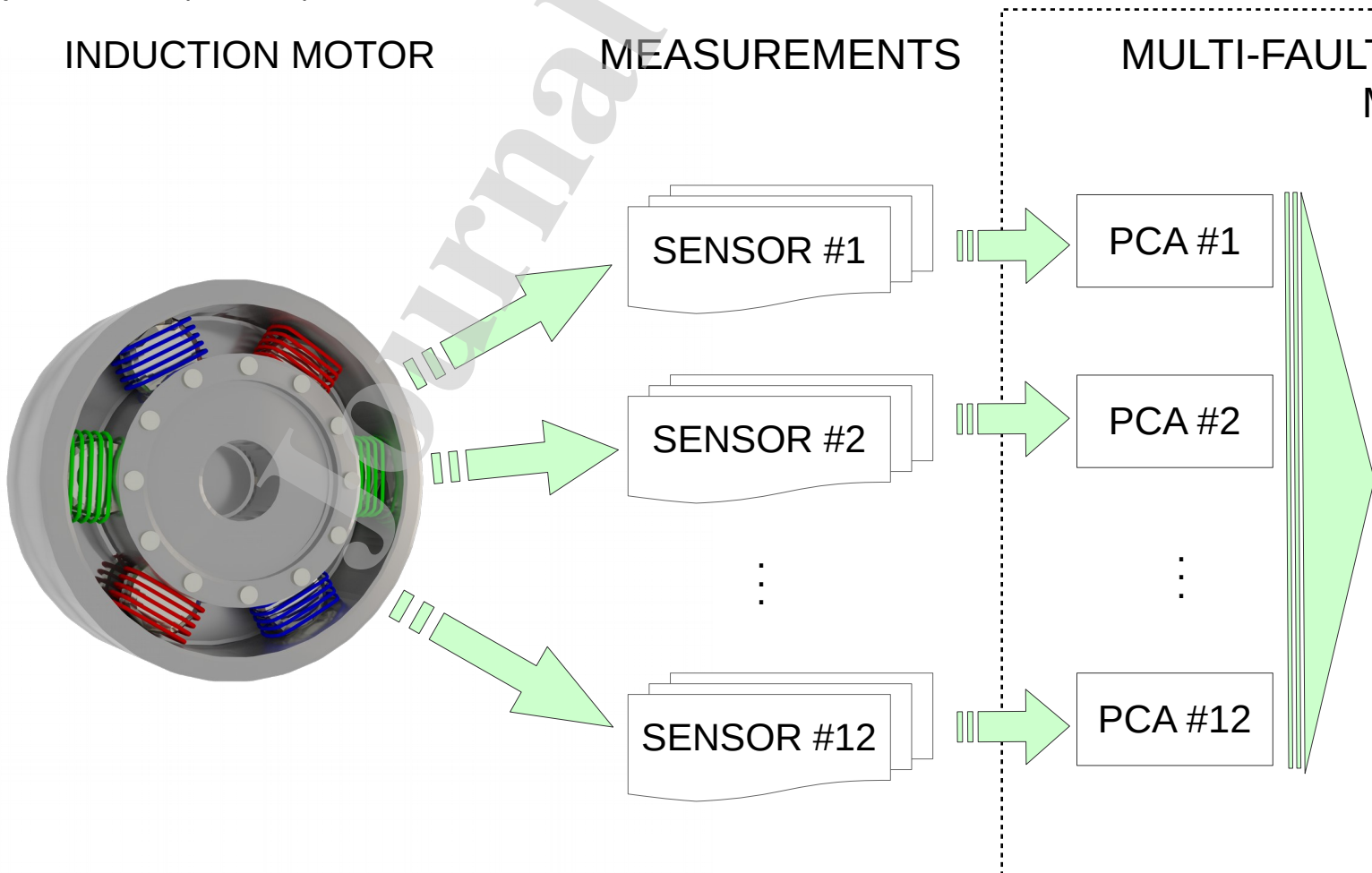
A method for multifault detection during the first half second is presented.
The approach combines the usage of PCA and multi-label decision trees.
The evaluation is performed using multi-label measures.
The proposed method is insensitive to the load and the frequency of the induction motor.
The proposed method gives good results even at frequencies as low as 3Hz.

# Early and extremely early multi-label fault diagnosis in induction motors

## Abstract

The detection of faulty machinery and its automated diagnosis is an industrial priority because efficient fault diagnosis implies efficient management of the maintenance times, reduction of energy consumption, reduction in overall costs and, most importantly, the availability of the machinery is ensured. Thus, this paper presents a new intelligent multi-fault diagnosis method based on multiple sensor information for assessing the occurrence of single, combined, and simultaneous faulty conditions in an induction motor. The contribution and novelty of the proposed method include the consideration of different physical magnitudes such as vibrations, stator currents, voltages, and rotational speed as a meaningful source of information of the machine condition. Moreover, for each available physical magnitude, the reduction of the original number of attributes through the Principal Component Analysis leads to retain a reduced number of significant features that allows achieving the final diagnosis outcome by a multi-label classification tree. The effectiveness of the method was validated by using a complete set of experimental data acquired from a laboratory electromechanical system, where a healthy and seven faulty scenarios were assessed. Also, the interpretation of the results do not require any prior expert knowledge and the robustness of this proposal allows its application in industrial applications, since it may deal with different operating conditions such as different loads and operating frequencies. Finally, the performance was evaluated using multi-label measures, which to the best of our knowledge, is an innovative development in the field condition monitoring and fault identification.

*Keywords:* multi-fault detection, early detection, multi-label classification, principal component analysis, load insensitive model, prediction at low operating frequencies

## 1. Introduction

The operation of most industrial applications is provided by rotating machinery powered by electricity. In that context, the Induction Motor (IM) represents the most used electrical machine, due to its robustness, efficiency, and safety. Despite its effectiveness and reliability, the occurrence of sudden and unexpected faults can affect the IM operation, provoking undesirable interruptions that may lead to crucial stoppages in industrial processes. Thereby, the continuous monitoring for assessing the operating condition to identify faults in rotating machines, that are powered by electricity, is a key safeguard for ensuring the machine reliability and safety in industrial production systems [1, 2]. In that sense, most of the condition monitoring strategies comprise three main tasks: fault detection, fault isolation, and fault identification [3]. Specifically, fault detection provides information relating to the existence of a malfunction; fault isolation locates the faulty components; and, fault identification labels the fault type that has been identified [4].

Recently, the use of Artificial Intelligence (AI) and classification techniques have been included as a part of the condition monitoring strategies. These techniques usually consist of several steps: data collection (gathering fault-related measurements); feature calculation, extraction and selection (determining fault-related features and measurements); and, training an algorithm (building a classification/regression model with the selected features) [5, 6, 7]. In that regard, AI techniques such as Neural Networks (NN) and Fuzzy-based inference systems have been developed as classification algorithms. Consequently, automatic Fault Detection and Diagnosis (FDD), based on AI algorithms using a trained AI model that is capable of identifying the relationship between both input data (information provided by sensors) and output data (machine-assessed condition), can replace the human intervention.

The training process of an AI algorithm usually involves the modeling of several inputs that specifically are the set of features estimated from the available physical magnitudes that characterize the working condition of the machine under observation [8]. Normally, the output is a single value (a 'label') that can be numeric (for regression problems) or discrete/categorical (for classification problems) but represents a specific condition/operation

2

of the assessed system. If there are only two classes, e.g. fault or no fault, then the problem involves binary classification, and if there are more labels (two or more fault types), it is a multi-class problem. Nevertheless, more than one label can be assigned whenever multiple faults occur simultaneously, which is also possible and a critical issue that must be addressed. This task, known as multi-label classification, has recently been gaining ground [9, 10].

Therefore, the main contribution of this paper lies in the proposal of a new intelligent multi-fault diagnosis method based on multiple sensor information for assessing the occurrence of single, combined, and simultaneous faulty conditions in an IM. The condition assessment and fault identification is performed by analyzing the available physical magnitudes, vibrations, stator currents, voltages, and rotational speed, through the Principal Component Analysis a multi-label classification tree. Additionally, the contribution and novelties of the proposed multi-fault detection and identification approach include:

1. Multi-label classification: failure prediction of multiple faults that occur simultaneously.

2. Insensitivity to load and frequency: capable of achieving good performance regardless of whether the motor is running under variable load or no-load conditions. Likewise, training is possible without taking into account the operating frequency.

3. High-performance results in both transient and in steady state regimen: system predictions are generated both in the transient state and in the easier steady mode.

4. The method is able to identify faults without the need of expert knowledge of each failure.

5. Extremely-early detection: the classification performance of the method is remarkable, predicting from signals within only 0.5 of a second.

The rest of the paper is organized as follows: a brief description of related works is be provided in Section 2; in Section 3, a general introduction to the relevant Data Analysis procedure is given; then, an explanation of the data set and the process of parameter setting is offered in Section 4; the proposed diagnostic methodology is presented in Section 5; in Section 6, the results of the experimental study for the performance assessment of the method

is discussed; and, finally, in Section 7, the main conclusions of the paper is summarized, and some lines of future research is discussed.

## 2. Related works

The development of strategies for condition monitoring and fault identification in industrial processes is important to reduce unexpected stoppages avoiding economic losses and also for ensuring the continuous machining operations. As stated earlier, undesirable faulty conditions may occur repeatedly during the continuous working cycle of an IM. This issue has normally been addressed as a single-label problem in most condition monitoring strategies that, one by one, identify each single fault mode.

Commonly, the application of condition monitoring strategies for fault identification in IMs has been divided into two widely studied groups: mechanical and electrical problems [11]. In this regard, most of mechanical failures presented in IM are usually associated with problems to the bearing components, axis eccentricities, misalignments, and couplings, among others; in contrast, electrical problems are associated with faults in the stator and rotor windings, such as broken rotor bars, insulation defects, and short circuits [12, 13]. On the other hand, different physical magnitudes, such as voltage, stator currents, temperatures, load torque, and mechanical vibrations are usually monitored in most industrial processes in order to assess its current condition. Indeed, a big deal of the industrial processes usually prefer the application of condition monitoring strategies that are based on the installation of non-invasive sensors. Additionally, the analysis of a unique physical measurement remains as the most preferred approach to condition assessment. Hence, vibration-based and stator current-based analyses have been widely addressed by using signal processing for condition monitoring [14]. At present, the classical fault identification in electrical rotating machines with industrial applications involves the analysis of vibration or stator current signals, separately. In fact, prior to the application of condition monitoring strategies in industrial applications, the development of new proposals can also be based on theoretical approaches that consider equivalent mechanical, electrical or magnetic models that aim to simulate different faulty conditions in a specific and real system [13]. Likewise, different

4

data-driven condition monitoring strategies have been proposed to detect faults and their sudden appearance in IMs and, although significant and advantageous results have emerged, most of these proposals have been focused on the analysis of single fault-mode [4, 15]. Thus, the critical issue of multiple faults and the adaptation of condition monitoring strategies to detect them, in view of their harmful consequences during the regular operation of an IM, therefore suggests that the measurement of different physical magnitudes represent a coherent approach for being considered during condition monitoring applied to electrical machines.

The rest of the section follows the structure of the review by Liu & Bazzi [1]. According to those authors, Fault Detection and Diagnosis (FDD) methods can be grouped into four groups: time-domain methods, frequency-domain methods, time-frequency-domain methods, and artificial-intelligence-based methods. Traditionally, signal processing techniques based on time domain, frequency domain, and time-frequency domain approaches are the most common and well-known processing strategies [16]. As this paper is not a complete state-of-the-art review, it is recommended that interested readers consult the following papers [1, 2].

### 2.1. Time-domain FDD methods

These methods are used to identify defective machine functions. The use of time-domain methods may be preferred due to one of its advantages is the simplicity in computation burden and implementation, although their low sensitivity, especially in noisy environments, might also be mentioned. Some relevant methods in this category are [17, 18, 19].

### 2.2. Frequency-domain FDD methods

The intrinsic mechanics of all IM mean that their characteristic fault-related patterns are within the frequency domain. Whenever a fault is present, it therefore provoke changes to the spectra, thus, the Fast Fourier Transform (FFT) is among the simplest of mathematical methods of classifying signals from spectrum measurements. Nevertheless, the overlapping of several faults within the same frequency band and the difficulties associated with low loads

and frequencies are some of the main drawbacks of the frequency-domain methods. These methods can directly process both raw and preprocessed signals [20, 21]. Other methods subtract the fault-independent components from the faulty signal, to obtain a residual that can be analyzed [22].

## 2.3. Time-frequency-domain FDD methods

As is known, non-stationary signals are difficult to process and time-frequency domain methods therefore yield results of greater accuracy. Time-frequency methods use a moving time window on which spectral analysis is performed, with the result that the non-stationary components can be treated as constants. These analyses require more complex implementation and computational complexities (log-linear or higher). Some examples of these techniques are the wavelet transform [23] and the short-time Fourier-transform [24].

## 2.4. Artificial Intelligence-based FDD methods

The consideration of AI techniques in condition monitoring strategies has been towards automatic fault diagnosis. Over the years, different variants of neural networks have been one of the most frequently used techniques: the multilayer perceptron [25] and radial-basis functions [26], among others. On the other hand, evolutionary methods (such as genetic algorithms and particle swarm optimization) are widely used to compute the best algorithm parameters. Whereas, Support Vector Machines (SVMs), another popular family of algorithms in this context, have likewise demonstrated good classification performance in hierarchical structures [27]. Recently, deep learning has been used for FDD [6], although its main drawback is the huge amount of data that is needed for training FDD systems. The recent work of Liu et al. [2] is strongly recommended for a detailed review of AI-based methods.

## 2.5. Unresolved Issues

Even though the literature on FDD methods is very copious, there are still some unresolved issues that must be addressed, the most critical issues that have to be faced are:

- Simultaneous fault detection [1]: most methods analyze a single or, at best, a couple of faults, and are not designed for the simultaneous identification of multiple faults.

- Non-stationary conditions [1]: the authors make significant progress examining fault detection under non-stationary conditions.

- Integrated methodology/system [1, 2]: this point can be analyzed from two perspectives. Firstly, there are four main faults in induction motors (broken bar, misalignment, unbalance, and bearing failure); different methods approach different kind of faults, underlining the need for an integrated system that is capable of identifying all four faults [1]. Secondly, FDD are usually built as a combination of different parts instead of a whole diagnostic system. The incorporation of both feature extraction and "intelligent" (Machine Learning) FDD in a single system would represent an integrated method [2] that would be easier to use in industrial production.

### 2.6. Our proposal vs. related works

Although several works focused on the fault detection and identification in IM have been proposed, the important advantages of the diagnosis method presented in this paper are outlined below.

Most condition monitoring approaches that have been proposed are classically based on the analysis of stator current and vibration signals and, the fault diagnosis is usually performed through the estimation of characteristic fault-related frequency patterns. However, although the occurrence of different faulty modes is among the aims of such proposals, the fault identification has been widely addressed as a single-fault identification problem [28]; thus, the advantage of the method proposed in this paper is that it can be applied to assess the occurrence single, combined, and simultaneous faulty conditions.

In fact, the implementation of complex condition monitoring structures including stages relating to calculation, extraction and selection of feature have also been proposed. Those stages are specifically implemented to improve and highlight the characterization of a large set of features estimated through time-domain techniques such as CWT (Continuous Wavelet

7

Transform), HHT (Hilbert Huang Transform) and EMD (Empirical Mode Decomposition) [29, 30, 7]. Although such optimization processes are mostly performed with genetic algorithms, the use of those techniques requires additional knowledge for their proper implementation. Whereas, in the proposed method, the need for expert knowledge on each type of failure is not necessary, as the novel implementation of the PCA and decision trees means that there is no need to configure the condition monitoring approach with specific parameters, depending on the system to be assessed.

Additionally, the condition monitoring strategies that can detect the occurrence of multiple and combined faults are available in only very few studies. Indeed, such proposals are based on complex time-frequency domain techniques, such as the high-resolution spectral method of multiple signal classification (MUSIC) [31], which compromises the computational burden. The strategies are also limited to the identification of multiple faults under the start-up or steady-state regime [31]. Hence, in this work, the assessment of multiple combined faults is performed during the start-up and steady-state regime under variable load conditions and, in addition, the low computational burden of this tool means that it can be used for real-time condition monitoring.

## 3. Theoretical background

If sensor data and records of mechanical operations are not filtered, organized and analyzed, to extract information that may be of use to understand a process and to solve a problem, they are of no inherent utility. The challenge is how to perform the early diagnosis of multiple malfunction conditions in an IM. The techniques used to perform the process of extracting information from raw data, range from the simplest, such as data summary and visualization, to the most sophisticated, such as the application of multivariate statistical methods, Machine Learning and Data Mining. Thus, in this study, a statistical technique, Principal Component Analysis (Subsection 3.1), is combined with a Machine Learning technique (Subsection 3.2), Decision Trees, to solve a multiple prediction problem, commonly known as multi-label classification (Subsection 3.3).

8

## 3.1. Principal Component Analysis

The Principal Component Analysis (PCA) [32] is commonly used for data exploratory analysis (in data visualization) and it is still one of the most well-known and popular methods for dimensionality reduction [33, 34]. The main objective of the PCA attempts to establish the directions, in the original high-dimensional space, with the highest variance (directions that, assuming Gaussian noise distributions, usually correlate with a more informative content). As a result, the projection of the multivariate measurement data onto the most significant orthogonal principal component directions is the data and the noise reduction step in the process with the lowest loss of overall data variance. The assumption is that the primary information loss is the noise subspace, which is assumed not to be informative for the multi-label classification problem.

More formally, if $X$ is an $n \times m$ matrix with the data ($m$ attributes for $n$ instances), the principal components are obtained by solving an eigenvalue problem:

$$\mathrm{cov}(X)\mathbf{c}_k = \lambda_k \mathbf{c}_k, \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m,$$

$$\mathrm{cov}(X) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T,$$

where, $\mathrm{cov}(X)$ is the covariance matrix of $X$, $\mathbf{c}_k$ is a principal component (and an eigenvector), $\mathbf{x}_i$ is a row of $X$ (data instances), and $\overline{\mathbf{x}} = (1/n)\sum_{i=1}^{n}\mathbf{x}_i$ is the average of the $\mathbf{x}_i$. In practice, an explicit calculation of the covariance matrix is not necessary, as other matrix decompositions, such as singular value decomposition (SVD), can be used to obtain the same eigenvectors with greater efficiency, especially if only a few principal components are needed.

In this paper, the value of $n$ is the number of experiments performed in the experimental test bench and the value of $m$ depends on the measurement sampling frequency, e.g. 12 kHz or 3 kHz (see Section 4.1). In this paper, 12 matrices were used, one for each of the measurements.

When PCA is used as a dimensionality reduction method, the determination of the number of principal components to use is one of the most important decisions. The wrong

9

choice could reduce the impact of the results. Retaining too few components could lead to a loss of important information [35]. Whereas, retaining too many components could produce high dimensional data, which might still embody much of the measurement noise that is detrimental to the performance of subsequent AI models.

There are empirical prescriptions for determining the number of components, some of which were compared in [35]. Kaiser's eigenvalue greater than one rule (K1 rule) [36] is one of the most popular methods used for the selection of the number of retained principal components [35]. The rule states that if the eigenvalue of a component is greater than 1.0, then the component is significant. Catell's scree test [37] is a graphical method that consists of plotting the eigenvalues in descending order (scree plot) and looking for an "elbow" in the plot, which implies a steep slope from large to small eigenvalues [38].

### 3.2. Classification trees

A decision tree is an algorithm in which a set of tests, organized in a hierarchical way, is used to guide the process of class assignment or output value calculation. The process begins at the root node, where the value of one of the attributes of the instances to be classified is compared (or whose output value is needed to be determined). Depending on the result of this comparison the process is directed to one or several branches (typically, two in a binary decision tree), where nodes apply new tests to conditionally branch further in the tree. The process continues until a leaf node (a node without further branches) is reached, at which point a class is assigned to the instance, or there is a function to calculate the output value for that instance.

The process of building a decision tree also starts at the root node. In each node, the training set is divided into subsets, for which it is necessary to determine the best division attribute and the value for the division into subsets. If only one threshold value is used, the division will be into two subsets, but it would be possible to use several ranges of values permitting a node to have more than two branches. The criterion for the selection of attributes and values is the optimization of a merit function. Some of these functions are: the Gini Index [39], Information Gain, and Gain Ratio [40]. Once the attribute and the

10

values of the node have been selected, they are used to determine through which branches each instance proceeds, and the process is repeated for the corresponding arrival nodes. The process terminates when the number of instances is less than a certain value, or when other stop criteria are satisfied. The leaf nodes are assigned the majority class of the instances that have reached that node (or those instances are used to calculate a function that gives the output value when the tree is used for prediction).

Decision trees are popular in Data Mining and Machine Learning for several reasons: they are quick to build, they are interpretable, and they are unstable (that is, small changes in the training set results in very different trees). This last property has made them suitable for the construction of ensembles,[1] both of classifiers [41, 42] and of regressors [43, 44].

In this research, decision trees were used on account of their speed of construction and because decision trees adapted to multi-label problems already exist. Further details on these types of problems are given in the next section.

### 3.3. Multi-label and multi-output learning

Supervised Machine Learning is used primarily to solve two types of problems:

- The determination of the relationship between the attributes of an instance and its class, in order to be able to label new instances, that is, so that one class may be assigned to them from a finite number of existing classes; usually single-label classification problems.

- The prediction of the value of a continuous attribute, from those attributes that are already available; known as regression problems.

However, an instance can belong to several classes simultaneously, or, several attributes must be predicted at the same time, which are known as multi-label (classification) and multi-output (regression) problems, respectively. Although the prediction of each of the

---

[1]An ensemble is an algorithm that makes predictions by combining several models, but these models should be diverse, and the instability of the trees makes them ideal in this context.

11

attributes could be addressed as a problem of single-label or single-output prediction, it has been shown that simultaneous prediction can be beneficial [45], since learning algorithms can exploit the relationships between the attributes to be predicted. So in recent years, research has increasingly focused on adapting existing techniques to these new paradigms, or presenting new techniques to resolve these multi-label problems. Papers have been published on classification [9, 10], regression [46, 47], and pre-processing techniques [48, 49, 50].

In this paper, a decision tree with multi-label learning was designed to identify motor malfunctions, even for simultaneously occurring faults. The model was also evaluated using multi-label performance measures, which is discussed in Subsection 4.2.

## 4. Experimental study

In this section, the method of data acquisition obtained from an experimental test-bench based on a laboratory electromechanical system is described (Subsection 4.1). The multi-label evaluation measures is then be introduced (Subsection 4.2), and the method for determining the optimal number of principal components is also subsequently described (Subsection 4.3).

The proposed condition monitoring method was programmed in Python using the Scikit-learn library [51]. And, the experimental data was analyzed by using multi-label measures and $10 \times 10$-fold cross-validation (cross-validation instead of train-test splits were used, as it has been demonstrated that the cross-validation test better exploits the available data [52]). For the purpose of achieving load-insensitive models, the training data contained examples of the IM functioning with three types of load conditions: no-load, half-load, and full-load.

All source code is publicly available on Github: `https://github.com/mjuez/early-im-fault-diagnosis`.

### 4.1. Data set description

The experimental test bench used to validate the proposed condition monitoring approach consists of a pulley-belt electromechanical system driven by a 971-W three-phase IM (model WEG 00136APE48T) and coupled to an automotive alternator. The IM had 1 pair
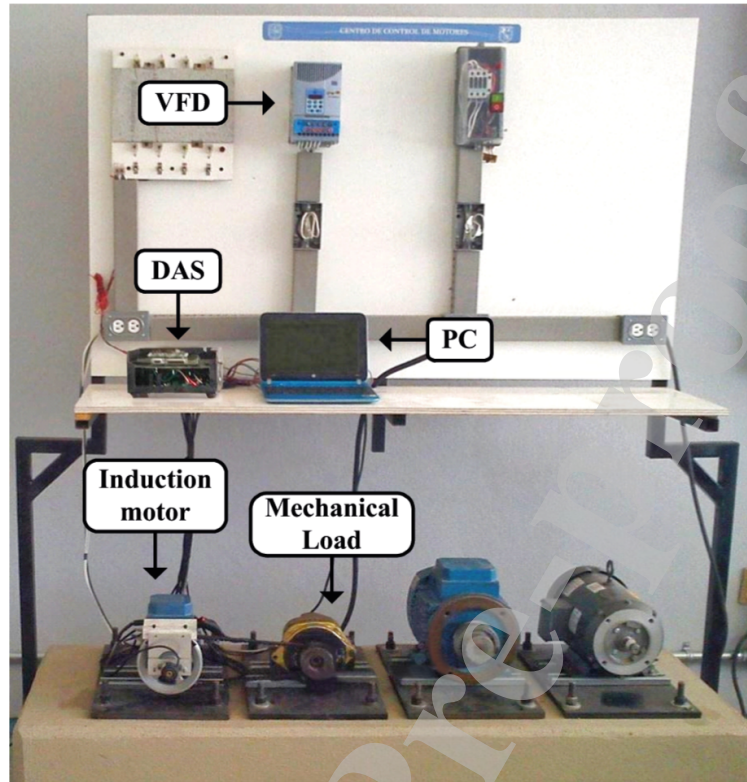
Figure 1: Experimental test bench used to validate the proposed condition monitoring method.

of poles, a power supply of 220V AC, and its full load current was 3.4 amperes. A variable frequency driver (VFD) (model WEG CFW08) was used to feed and drive the rotational speed of the IM. The automotive alternator was therefore driven by the IM through mechanical coupling based on a pulley-belt system. The alternator was used as the mechanical load and functioned, during the experiments, at three different load capacities: unloaded, half-load and full-load. The experimental test bench is shown in Figure 1.

Different physical magnitudes were acquired during the experiments. The effect of mechanical vibrations was recorded by a triaxial accelerometer, model LIS3L02AS4, fixed on top of the IM. Additionally, the voltage and the stator current consumption were continuously monitored and acquired by means of a set of Hall-effect sensors, model L08P050D15, from Tamura corporation. Thus, there were two groups of measurements: 7 related to currents and voltages; voltage A, voltage B, voltage C, current A, current B, current C, current N; and 3 relating to the measurement or vibrations; accelerometer X, accelerom-
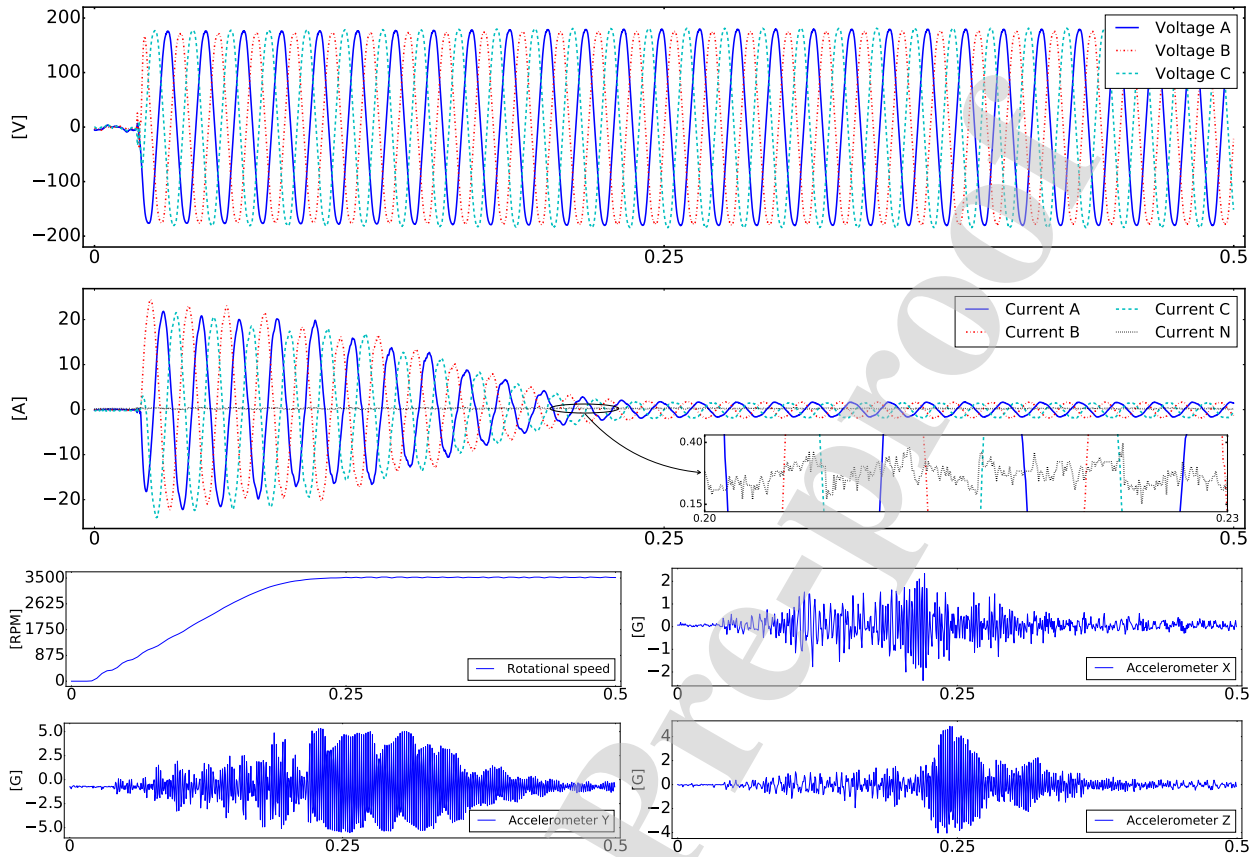
13

Figure 2: First half second of sensor measurements used for predicting faults. On the $X$-axis, the time in seconds. The $Y$-axis refers to the sensor value (different scale for each one). The figure shows the measurements of a healthy engine operating at direct supply.

eter Y, accelerometer Z. Additionally, there was a last measure, the IM rotational speed that was acquired through a digital encoder. To give an idea of the nature of the signals used for the prediction of failures, Figure 2 shows the graphic representation of the signals (currents, voltages, speed, and accelerometers) during the first half second of an engine in good condition operating at direct supply.

The installed sensors were individually mounted on a board with its corresponding signal conditioning and anti-alias filtering. The data acquisition was performed by means of two 12-bit 4-channel serial-output sampling analog-to-digital converters, model ADS7841, from Texas Instruments, which were used as the Data Acquisition System (DAS); a proprietary low-cost DAS design based on a field programmable gate array technology.
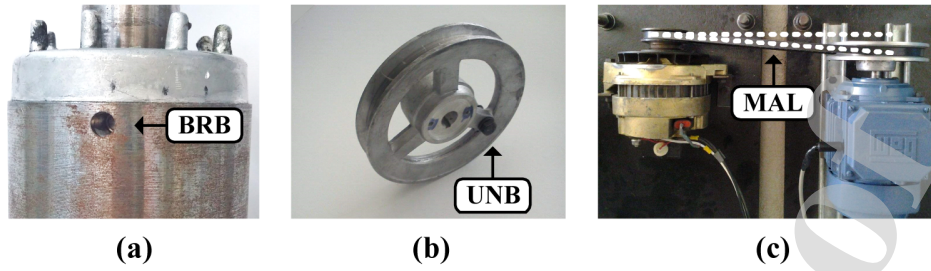
14

Figure 3: Set of experimental faults: a) broken rotor bar (BRB); b) unbalance (UNB); and, c) misalignment (MAL).

Four different operating conditions were considered for evaluation in this study: healthy (HLT), broken rotor bar (BRB), unbalanced (UNB), misalignment (MAL), as well as their combinations. The conditions under consideration were produced artificially, thus, the BRB condition was produced by drilling a rough-hole of 2 mm diameter, at a depth of 14 mm, into a bar of the rotor, without harming the rotor shaft, in such a way that the hole was drilled through the complete section of the rotor bar. The UNB condition was reproduced with a bolt in the rotor pulley that provoked a non-uniform load distribution that imbalanced the center of mass of the IM shaft. The MAL condition was caused by shifting the belt in the alternator pulley forward, thereby misaligning the transverse axial rotation of the motor and its load. Figures 3-a to Figure 3-c depict each of the conditions under consideration.

Moreover, the IM was driven at different operating frequencies during the experiments, so the data set contained the operating frequencies of 3 Hz, 30 Hz, and direct supply.

Each experiment was carried out during 10 seconds, the first 5 seconds corresponded to the IM start-up and were considered as a transient state. The next 5 seconds were considered as steady state (these last 5 seconds were not used in the experiments)[2]. The transient state was considered harder to predict, due to its non-stationary and inherently changing nature [53]. Nevertheless, as the experimental study addresses an early diagnosis problem, hereinafter only the transient state is considered.

In summary, an almost completely balanced data set was recorded of 2 521 instances: 831

---

[2]As the IM is powered by an inverter, the IM start-up was controlled, so that it always lasted 5 seconds, regardless of its frequency of use.

for 3 Hz, 833 for 30 Hz, and 857 for direct supply. As there were eight measures sampled at 12 kHz and four measures sampled at 3 kHz, a single instance consisted of 540 000 attributes ($8 \times 12\,000 \times 5 + 4 \times 3\,000 \times 5$). As a preliminary pre-processing step, all the features were normalized to have values between 0 and 1.

## 4.2. Evaluation metrics for multi-label problems

In single-label classification, accuracy and confusion matrix-based metrics are commonly used. Nevertheless, the performance of multi-label classification methods are more complex, because of the presence of several labels for each instance, and because the combined performance for each label can be approached in different ways. Hence, different evaluation metrics for multi-label classification and taxonomies have even been proposed to group those measures [54], although the simplest and most accepted division corresponds to measures based on either predictions or rankings. In the experimental evaluation, the measures in use in the first and the second groups were, accuracy, $F_1$ and Hamming loss and, in turn, one-error and rank loss, respectively. The definition of all these measurements are given in equations 1 to 6. The up/down arrow besides the measurement label indicates whether the higher or the lower value is best.

$$\uparrow \text{Accuracy} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\omega_i \cap \hat{\omega}_i|}{|\omega_i \cup \hat{\omega}_i|}. \tag{1}$$

$$\uparrow \text{Macro F}_1 = \frac{1}{|\Omega|} \sum_{l \in \Omega} F_1(TP_l, FP_l, TN_l, FN_l). \tag{2}$$

$$\uparrow \text{Micro F}_1 = F_1\left(\sum_{l \in \Omega} TP_l, \sum_{l \in \Omega} FP_l, \sum_{l \in \Omega} TN_l, \sum_{l \in \Omega} FN_l\right). \tag{3}$$

$$\downarrow \mathcal{H} \text{ Loss} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\omega_i \,\triangle\, \hat{\omega}_i|}{|\Omega|}. \tag{4}$$

16

$$\downarrow \text{One Error} = \frac{1}{n} \sum_{i=1}^{n} \delta\big(\arg\min_{l \in \Omega} r_i(l)\big), \delta(l) = \begin{cases} 1 & \text{if } l \notin \omega_i \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$\downarrow \mathcal{R} \text{ Loss} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{|\omega_i||\bar{\omega}_i|} \Big| (l_q, l_r) : r_i(l_q) > r_i(l_r), (l_q, l_r) \in \omega_i \times \bar{\omega}_i \Big|. \tag{6}$$

The notation for the equations is as follows: $n$ is the number of instances; $\Omega$ is the set of labels; $\omega_i \subseteq \Omega$ is the actual labelset of instance $\mathbf{x}_i$; $\hat{\omega}_i$ is the predicted labelset of instance $\mathbf{x}_i$; and $\bar{\omega}_i$ is the complementary labelset of $\omega_i$. Then, $r_i(l_j)$ is the predicted rank of class label $l_j$ for instance $\mathbf{x}_i$ and $\triangle$ is the symmetric difference between two labelsets. Finally, $TP$, $TN$, $FP$, and $FN$, are the number of true positives, true negatives, false positives, and false negatives, respectively. All of these, $TP$, $FP$, $TN$, and $FN$, are computed by comparing the predicted and the actual labelsets.

### 4.3. Determining the number of principal components

As explained in Subsection 3.1, the number of the selected principal components is essential, if accurate models are subsequently to be trained. Two classical approaches were used to determine that number: the K1 rule and the Scree test. Parallel analysis [55] was also used, as it is more reliable [35], although the computational resources it requires can be prohibitive, due to the high dimensionality of the data set.

The models trained with the number of principal components selected with the K1 rule performed slightly better than those that used the scree test results. Also, for all sensors except those monitoring voltages, the number of retained components was lower following the K1 rule, meaning that a higher dimensionality reduction had occurred. Although Kaiser's rule states that only those components with an eigenvalue greater than 1 should be retained, whenever there was no eigenvalue greater than 1, then at least the first principal component was selected in this study, as it would be of no interest to either discard a sensor or completely disregard any measurement.

17

As described in the Section 5, the proposed predictive model is a combination of the results of applying twelve PCA transformations (one for each input sensor) and a multi-label classifier. It means that the model has twelve input parameters, which correspond to the number of principal components that will be retained for each signal. Therefore, in this scenario, the determination of the optimal value for those parameters, can be treated as a hyperparameter optimization problem [56]. When optimizing hyperparameters, a set of parameter combinations is defined, then the model performance is evaluated for each combination. Thus, in this particular case, good model evaluations will indicate an optimal number of principal components to be retained.

Considering the K1 rule and the scree test results, 50 was selected as the maximum number of principal components to retain, and 1 as the minimum. 12 parameters with values between 1 and 50 will produce a set of $2.44 \times 10^{20}$ combinations of parameters. Any evaluation of so many models is extremely costly in terms of time and computation effort, so the Hyperopt [57] library was used for the hyperparameter tuning task, which uses only two algorithms to optimize the parameter search space. In this study, the probabilistic method called the Tree of Parzen Estimators (TPE) was used [56], rather than the only other possible alternative, Random Search. Optimization of the hyperparameters also implies the minimization of an objective function, which in this study, was the inverse of model accuracy $(1 - accuracy)$. The maximum number of evaluations was set to 500, which leads to $150\,000$ models evaluated in total (500 parameter combinations $\times$ 3 frequencies $\times$ 10 repetitions of 10-fold cross-validation).

Table 1 shows the number of principal components selected according to the different methods (K1 rule, Scree test, and Hyperopt) and frequencies. In terms of model accuracy, the number of principal components selected by Hyperopt produced significantly better results than those selected by the Scree and the K1 methods. Furthermore, using the combination obtained for 3 Hz with the other frequencies provided better results than the specific combinations obtained for each frequency. So, according to these results, and the fact that the 3 Hz combination retained fewer principal components (139), i.e., greater dimensionality reduction was achieved, the above combinations were chosen as the optimal

18

| Freq. | Method | Voltages | | | Currents | | | | Rotat. | Accelerometers | | | | Sum. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | C | A | B | C | N | Speed | Ref. | X | Y | Z | PCs |
| 3 Hz | K1 rule | 26 | 28 | 30 | 14 | 15 | 14 | 1 | 1 | 1 | 1 | 4 | 1 | 136 |
| | Scree test | 17 | 18 | 18 | 18 | 18 | 17 | 5 | 2 | 11 | 43 | 189 | 105 | 461 |
| | Hyperopt | 4 | 7 | 8 | 33 | 5 | 19 | 6 | 35 | 1 | 17 | 2 | 2 | 139 |
| 30 Hz | K1 rule | 18 | 18 | 18 | 14 | 14 | 14 | 1 | 1 | 1 | 1 | 9 | 1 | 110 |
| | Scree test | 15 | 18 | 18 | 16 | 17 | 17 | 3 | 1 | 9 | 83 | 171 | 151 | 519 |
| | Hyperopt | 32 | 8 | 4 | 24 | 4 | 22 | 35 | 49 | 3 | 4 | 4 | 22 | 211 |
| Direct supply | K1 rule | 17 | 19 | 18 | 12 | 15 | 15 | 1 | 1 | 1 | 1 | 31 | 1 | 132 |
| | Scree test | 16 | 16 | 16 | 18 | 19 | 19 | 1 | 2 | 12 | 68 | 210 | 176 | 573 |
| | Hyperopt | 6 | 25 | 46 | 2 | 10 | 8 | 2 | 10 | 16 | 38 | 33 | 17 | 213 |

Table 1: Number of principal components selected by the different methods and frequencies. The last column shows the sum of the number of principal components that each method retained, i.e., the number of decision tree inputs.

settings for all the subsequent predictive models.

## 5. Multi-fault early diagnosis method

Since typical problems of this nature are usually solved using time or frequency-domain-based methods, one of their main disadvantages is that expert knowledge on each failure type is required. However, this knowledge is not necessary when artificial intelligence-based approaches are used. These techniques, once trained, are capable of extracting the patterns that are most strongly associated with each fault.

One of the main characteristics of sensor data is the dependent nature of the observations that are highly correlated, because the data consist of a set of adjacent points in time. A common approach for dealing with data of this nature is to use statistical methods such as time series analysis [58]. Most statistical methods require the data to be stationary, which is not the case in this study, because data from the IM start-up (transient data) were used.

As stated earlier, in our data set each instance consists of 540 000 attributes. It is well-known that most Machine Learning algorithms suffer from difficulties when analyzing data sets with a large number of attributes, commonly known as the *curse of dimensionality* [59].
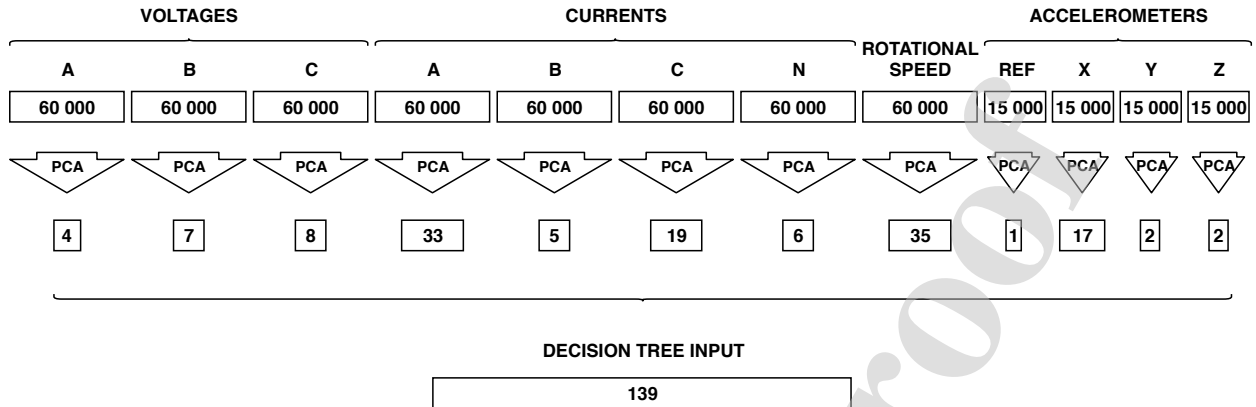
19

Figure 4: Graphical representation of the dimensionality reduction approach by using PCA. The numbers of input measures correspond to the start-up of the IM (i.e., first 5 seconds of operation).

For this reason, PCA was used in this proposal to reduce the number of attributes. As previously explained, PCA is able to reduce the dimension of attributes, by retaining the combination of the most important features in which the maximum variance is retained as possible. Furthermore, when applying PCA to correlated data, one of the properties of the transformed data is that the correlation between attributes is removed [60]. Although the robustness of decision trees leaves them unaffected by the collinearity of attributes, the elimination of the correlation can be beneficial for other Machine Learning methods.

Briefly, in the early multi-fault diagnosis method proposed in this paper, the information provided by the sensors was firstly reduced by means of PCA. Since each sensor measured a different physical magnitude, PCA was independently applied to each one. Then, all the retained principal components were joined as input to the multi-label classification method (a decision tree was used, although it could have been any other multi-label classifier). The output of the method is the predicted label set, where each label is a possible fault.

The graphical illustration of the process, with the number of principal components retained for each sensor measurement, is represented in Figure 4. A dramatic reduction in dimensionality can be observed, where 540 000 input attributes are converted into 139.

Although PCA has been used before for pre-processing single signals [61, 62, 63, 64], the novelty of the method presented here is the way in which multiple signals are combined after

their dimensionality is reduced using PCA. The attributes that are passed to the classifier, a decision tree in our case, are the concatenation of the results obtained from PCA for each of the signals, that include voltages, currents and vibrations (see Figure 4).

The other novelty is that the decision tree is designed to solve multi-label problems, so it is capable of predicting potential faults, even when several are happening at the same time. The evaluation of the model is therefore done using well established measures in the field of multi-label classification. All other previous studies have used conventional single-label prediction methods, even when the fault-detection problem could occur simultaneously.

Despite the simplicity of the approach, the method obtained is robust under varying frequencies and load conditions, is able to detect errors swiftly in the first few seconds, even when there are simultaneous failures and the motor is still in the transition state (when other methods cannot be applied).

## 6. Results and discussion

In this section a thorough experimental program will be reported for assessing the performance of the proposed method. First, in Subsection 6.1, the training of the multi-label classifier trees and their performance tests at each isolated frequency (3 Hz, 30 Hz, and direct supply) will be detailed; second, in Subsection 6.2, the test results of the time interval reductions used for training the models will be described; third, in Subsection 6.3, the test results of the model using all the frequencies (frequency insensitivity) at the same time will be presented; in Subsection 6.4, the test results of another fault type -a Bearing Defect (BD)- will be discussed; and, finally, in Subsection 6.5, an analysis of the multi-fault data set with Fast Fourier Transform will be described.

### 6.1. Full transient state results

Having determined the optimal number of principal components for the proposed method, multi-label decision trees were trained and tested across each operational range.

In Figure 5 are shown the confusion matrices of each multi-label decision tree for each of the three frequencies. The predicted fault condition is represented on the $X$-axis while the
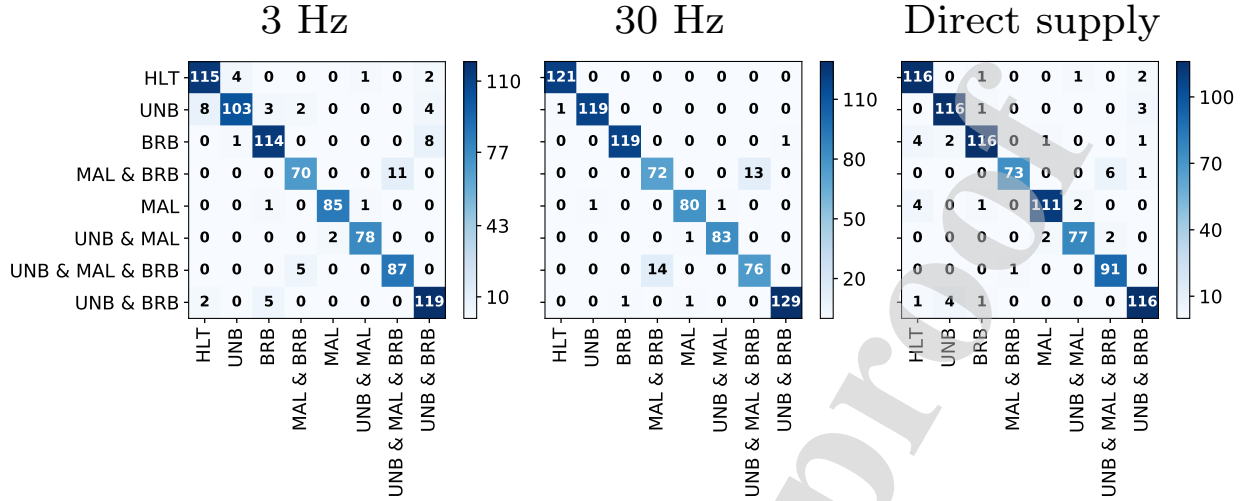
21

Figure 5: Confusion matrices of the three predictive models (one for each frequency). The $X$-axis corresponds to the predicted fault condition, while the $Y$-axis represents the actual fault condition.

actual fault condition is on the $Y$-axis.

The model efficacy is indicated by the high numbers present on the matricesdiagonals. It should be noted that some combinations of failures (such as MAL & BRB) were under-represented and are therefore in a lighter color. According to tables 3- 5, the frequency-specific percentage accuracies are: 93.6% for 3 Hz, 95.5% for 30 Hz, and 95.8% for direct supply.

Although accurate fault detection can be seen at all frequencies, the accuracy at 3 Hz is particularly noteworthy due to the accepted difficulty of detecting faults at this frequency. Good BRB detection rates are also noteworthy, with a 95% percentage accuracy from Figure 5, a type of failure that is difficult to detect [65] using other techniques.

### 6.2. Reducing the time interval

The method still captured the faults within the first five seconds of the IM start-up, despite the non-stationarity of the initial transient dynamics [53]. Therefore we also investigated the response capability in the first 0.5 seconds.

Experiments were executed with different time intervals (from 0.01 seconds to 5). Table 2 depicts the number of sensor measurements at each time interval. As might be expected,

| Time int. (s) | Voltages & currents | Rotational speed | Accelerometers | Total |
|---|---|---|---|---|
| 0.01 | 840 | 120 | 120 | 1 080 |
| 0.02 | 1 680 | 240 | 240 | 2 160 |
| 0.05 | 4 200 | 600 | 600 | 5 400 |
| 0.10 | 8 400 | 1 200 | 1 200 | 10 800 |
| 0.20 | 16 800 | 2 400 | 2 400 | 21 600 |
| 0.50 | 42 000 | 6 000 | 6 000 | 54 000 |
| 0.75 | 63 000 | 9 000 | 9 000 | 81 000 |
| 1.00 | 84 000 | 12 000 | 12 000 | 108 000 |
| 2.00 | 168 000 | 24 000 | 24 000 | 216 000 |
| 5.00 | 420 000 | 60 000 | 60 000 | 540 000 |

Table 2: Total number of sensors measurements within each time interval.

the increase in the number of features was proportional to the size of the time interval.

A classifier was trained and tested with each time interval. Tables 3, 4, and 5 group these results for each of the frequencies under analysis.

As expected, the best results were obtained with the longest time interval (5 seconds), i.e., with the data of the whole experiment. Average ranks were computed and the Hochberg post-hoc procedure [66] was applied, to determine the minimum time interval to use. Figure 6 provides a graphical representation of the data in Table 5. The results highlighted within the square are for the time intervals that were statistically equivalent (using Hochberg procedure) to the best one (i.e., 5 seconds). According to all multi-label metrics, the results of using the signals corresponding to the first 0.5, 0.75, 1, and 2 seconds, were statistically equivalent to the results of using the signals for the first 5 seconds. Only the results for time intervals shorter than 0.5 seconds were significantly worse.

In this regard, in order to quantify the performance reduction from the information constriction of short time intervals, the multi-label approach was investigated on the first 0.5 seconds. Figure 7 shows the confusion matrices of the models trained for the three

23

| Time int. (s) | Macro $F_1 \uparrow$ | Micro $F_1 \uparrow$ | Accuracy $\uparrow$ | $\mathcal{R}$ Loss $\downarrow$ | $\mathcal{H}$ Loss $\downarrow$ | One Error $\downarrow$ |
|---|---|---|---|---|---|---|
| 0.01 | 0.875 | 0.874 | 0.749 | 0.137 | 0.121 | 0.251 |
| 0.02 | 0.895 | 0.892 | 0.781 | 0.123 | 0.103 | 0.219 |
| 0.05 | 0.917 | 0.917 | 0.835 | 0.109 | 0.078 | 0.165 |
| 0.10 | 0.931 | 0.931 | 0.869 | 0.096 | 0.065 | 0.131 |
| 0.20 | 0.949 | 0.948 | 0.895 | 0.075 | 0.049 | 0.105 |
| 0.50 | 0.954 | 0.954 | 0.912 | 0.066 | 0.043 | 0.089 |
| 0.75 | 0.955 | 0.955 | 0.917 | 0.065 | 0.043 | 0.083 |
| 1.00 | 0.954 | 0.954 | 0.918 | 0.063 | 0.044 | 0.082 |
| 2.00 | 0.970 | 0.970 | 0.946 | 0.043 | 0.029 | 0.054 |
| 5.00 | 0.975 | 0.974 | 0.936 | 0.038 | 0.025 | 0.064 |

Table 3: Model results for 3 Hz by the time interval in use.

| Time int. (s) | Macro $F_1 \uparrow$ | Micro $F_1 \uparrow$ | Accuracy $\uparrow$ | $\mathcal{R}$ Loss $\downarrow$ | $\mathcal{H}$ Loss $\downarrow$ | One Error $\downarrow$ |
|---|---|---|---|---|---|---|
| 0.01 | 0.843 | 0.848 | 0.722 | 0.200 | 0.144 | 0.278 |
| 0.02 | 0.880 | 0.884 | 0.785 | 0.171 | 0.111 | 0.216 |
| 0.05 | 0.873 | 0.878 | 0.775 | 0.180 | 0.116 | 0.225 |
| 0.10 | 0.914 | 0.917 | 0.857 | 0.116 | 0.079 | 0.143 |
| 0.20 | 0.911 | 0.914 | 0.851 | 0.119 | 0.082 | 0.149 |
| 0.50 | 0.975 | 0.975 | 0.946 | 0.035 | 0.024 | 0.055 |
| 0.75 | 0.976 | 0.975 | 0.942 | 0.033 | 0.024 | 0.058 |
| 1.00 | 0.977 | 0.977 | 0.944 | 0.030 | 0.022 | 0.056 |
| 2.00 | 0.982 | 0.980 | 0.952 | 0.021 | 0.019 | 0.048 |
| 5.00 | 0.984 | 0.983 | 0.955 | 0.026 | 0.016 | 0.045 |

Table 4: Model results for 30 Hz by the time interval in use.

| Time int. (s) | Macro $F_1 \uparrow$ | Micro $F_1 \uparrow$ | Accuracy $\uparrow$ | $\mathcal{R}$ Loss $\downarrow$ | $\mathcal{H}$ Loss $\downarrow$ | One Error $\downarrow$ |
|---|---|---|---|---|---|---|
| 0.01 | 0.853 | 0.855 | 0.698 | 0.174 | 0.137 | 0.302 |
| 0.02 | 0.888 | 0.890 | 0.762 | 0.158 | 0.103 | 0.238 |
| 0.05 | 0.909 | 0.911 | 0.823 | 0.111 | 0.083 | 0.177 |
| 0.10 | 0.923 | 0.924 | 0.836 | 0.103 | 0.071 | 0.164 |
| 0.20 | 0.932 | 0.932 | 0.856 | 0.088 | 0.063 | 0.145 |
| 0.50 | 0.955 | 0.954 | 0.902 | 0.062 | 0.043 | 0.098 |
| 0.75 | 0.960 | 0.960 | 0.920 | 0.054 | 0.037 | 0.080 |
| 1.00 | 0.963 | 0.964 | 0.927 | 0.042 | 0.034 | 0.073 |
| 2.00 | 0.964 | 0.965 | 0.929 | 0.046 | 0.033 | 0.071 |
| 5.00 | 0.982 | 0.982 | 0.958 | 0.027 | 0.017 | 0.042 |

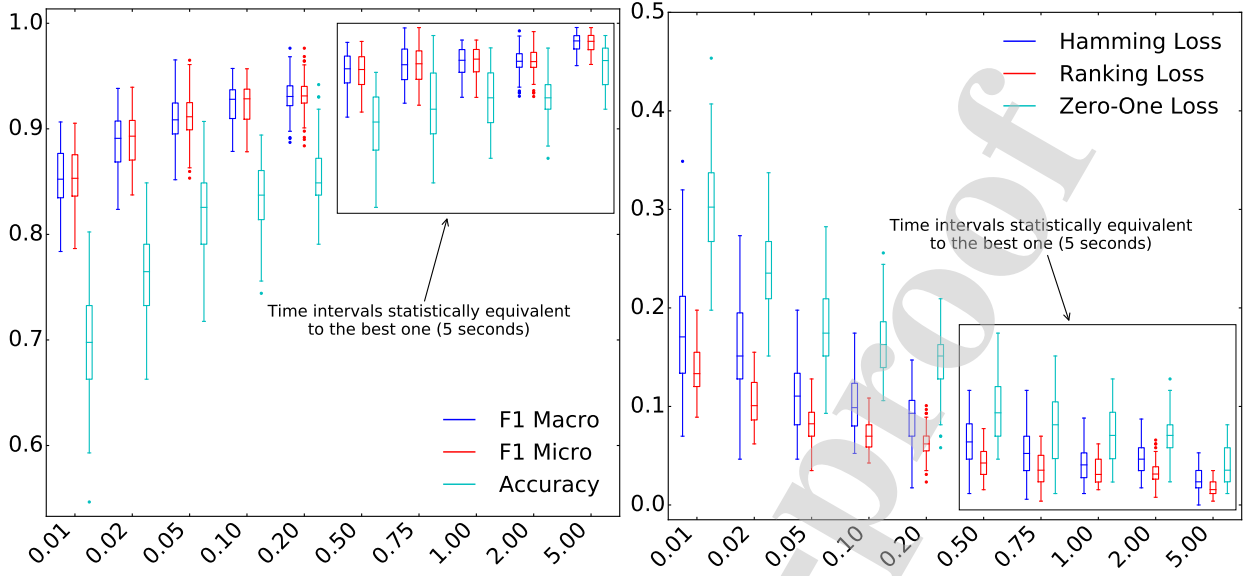Table 5: Model results for direct supply by the time interval in use.

24

Figure 6: Multi-label evaluation metrics of a classifier trained with signals from the first 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 0.75, 1, 2, and 5 seconds. The $X$-axis depicts the time interval in seconds, while the $Y$-axis refers to the metric value in range $[0 - 1]$. The plotted results show the direct supply of the frequency. The results statistically equivalent to the use of the signals of the first 5 seconds are highlighted within a square.

frequencies. The predicted fault condition is represented on the $X$-axis while the actual fault condition is on the $Y$-axis.

Despite the constricted transient information within just the first 0.5 seconds, the multi-label decision tree still provides fault detection accuracies over 90%: 91.1% at 3 Hz, 94.5% at 30 Hz and 90.2% for direct supply.

### 6.3. Insensitivity to IM operating frequency

A model that is insensitive to the operating frequency of an IM might seem useless, because the frequency is commonly known. However, training a separate model for each operating frequency might not be feasible in some scenarios. Our method can also be trained and deal with data acquired from IMs operating at different frequencies and the resulting model will still be able to classify fault conditions accurately. This is because the transformation of sensor measurements by means of PCA yields features that highlight the fault conditions, so multi-label decision trees can learn to classify them, regardless of the
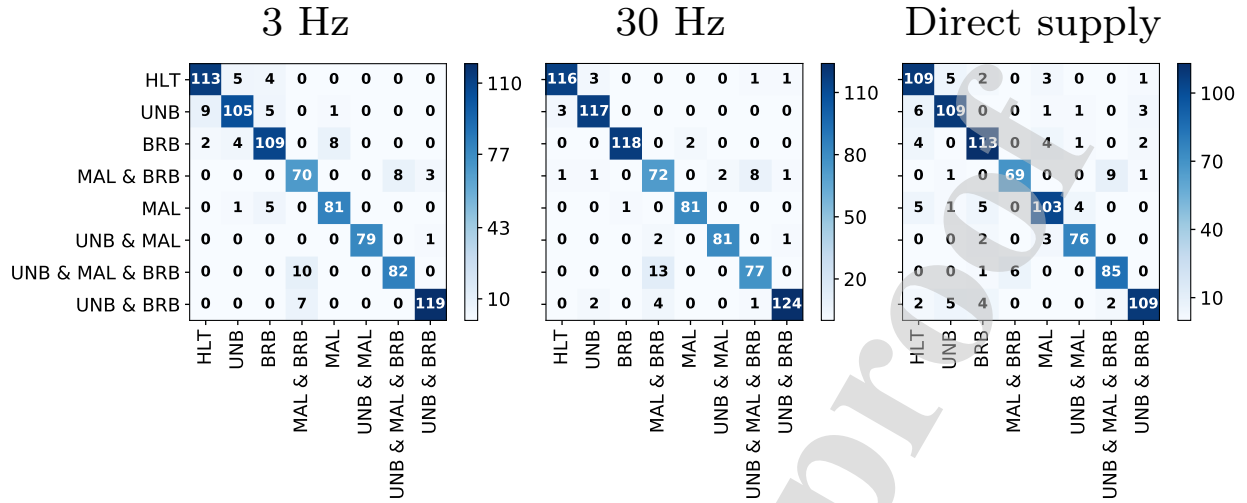
25

Figure 7: Confusion matrices of the three predictive models trained with 0.5 seconds (one for each frequency). The $X$-axis corresponds to the predicted fault condition, while the $Y$-axis represents the actual fault condition.

operating frequency.

Multiple operation frequency-insensitive models were trained with all of the data-set examples. Table 6 shows the results obtained for each time interval size. As expected, and consistent with Section 6.2, the best fault detection accuracy of 91.6% occurs for the longest time interval. Notably, the detection accuracy only dropped to 87.6% for time intervals of 0.5 seconds. Operation frequency-insensitive models therefore also appear appropriate for transient start-up time fault detection.

## 6.4. Testing another type of fault: bearing defect

Bearing defects (BDs) are common IM faults widely studied in the literature [8, 16, 67, 68, 69]. As a defective bearing is worn down, it will produce a complete mechanical breakdown. This makes almost impossible to experimentally reproduce this fault together with others on a test bench. In other words, although this kind of fault (as any other) can theoretically be predicted in combination with other faults, in practice it is almost impossible to obtain a data set with BDs and other faults occurring at the same time.

In this study, properly functioning examples and BD examples were employed, to prove
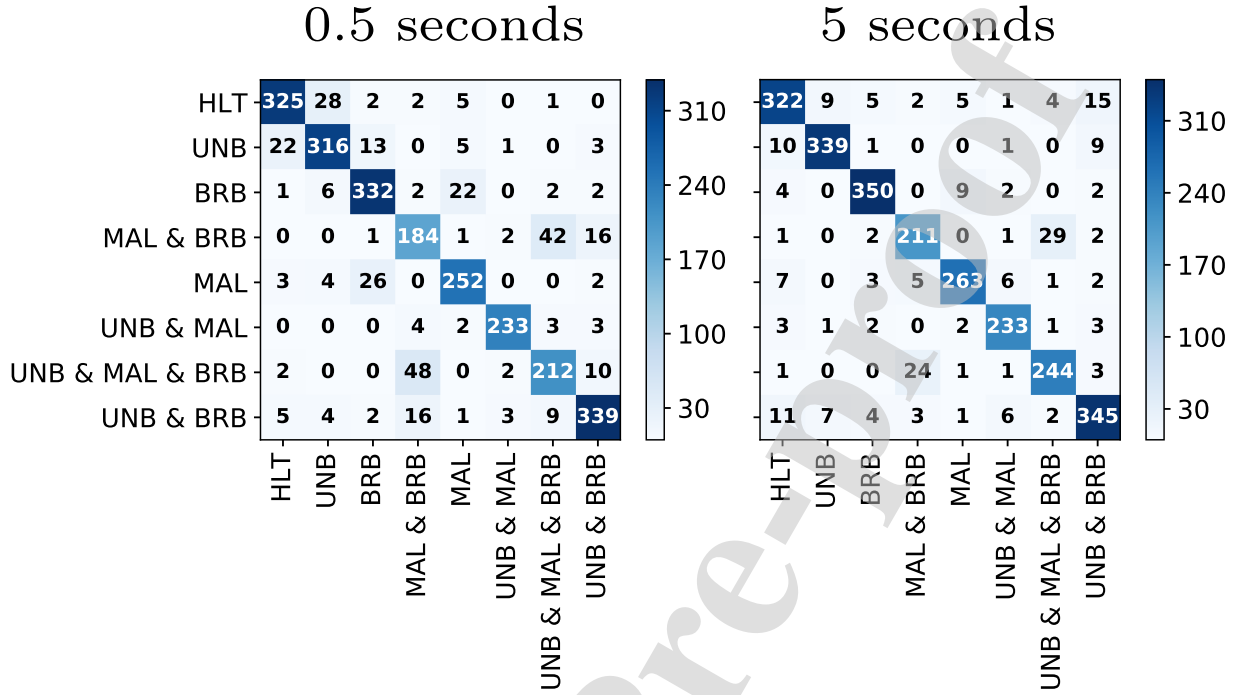
26

## 0.5 seconds



## 5 seconds

Figure 8: Confusion matrices of the general predictive models trained with all examples of all operating frequencies. The matrix on the left corresponds to first 0.5 seconds of IM operation, while the one on the right corresponds to the first 5 seconds (full transient state). The $X$-axis corresponds to the predicted fault condition, while the $Y$-axis represents the actual fault condition.

| Time int. (s) | Macro $F_1$ ↑ | Micro $F_1$ ↑ | Accuracy ↑ | $\mathcal{R}$ Loss ↓ | $\mathcal{H}$ Loss ↓ | One Error ↓ |
|---|---|---|---|---|---|---|
| 0.01 | 0.866 | 0.868 | 0.749 | 0.183 | 0.123 | 0.251 |
| 0.02 | 0.875 | 0.876 | 0.761 | 0.174 | 0.116 | 0.239 |
| 0.05 | 0.895 | 0.896 | 0.800 | 0.143 | 0.098 | 0.200 |
| 0.10 | 0.902 | 0.903 | 0.810 | 0.128 | 0.091 | 0.190 |
| 0.20 | 0.911 | 0.911 | 0.830 | 0.122 | 0.084 | 0.170 |
| 0.50 | 0.937 | 0.938 | 0.876 | 0.085 | 0.059 | 0.124 |
| 0.75 | 0.945 | 0.946 | 0.891 | 0.081 | 0.051 | 0.109 |
| 1.00 | 0.932 | 0.932 | 0.865 | 0.088 | 0.064 | 0.135 |
| 2.00 | 0.958 | 0.958 | 0.908 | 0.056 | 0.040 | 0.092 |
| 5.00 | 0.960 | 0.959 | 0.916 | 0.050 | 0.039 | 0.084 |

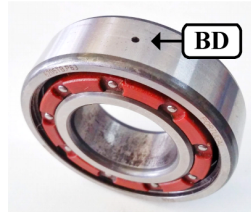Table 6: Model results for all frequencies depending on the time interval in use.

27

Figure 9: Induced Bearing Defect (BD) fault for the experimentation.

| Time int. (s) | 3 Hz | 30 Hz | Direct supply | All frequencies |
|:---:|:---:|:---:|:---:|:---:|
| 0.5 | 95.9% | 97.0% | 98.7% | 96.3% |
| 5 | 95.1% | 100.0% | 97.0% | 96.0% |

Table 7: Classification results in terms of accuracy for Bearing Defect (BD) fault condition.

that a model trained using our approach could also learn to predict these types of faults to high levels of accuracy.

BD examples were artificially provoked by drilling a through-hole on the outer race with a 1.191 mm tungsten drill bit, as shown in Figure 9. A total of 724 examples were used for testing the classification of a fault due to a BD. Of these, 363 corresponded to a healthy condition, and the remaining 361 to the BD fault condition. Isolating each frequency, 242 (122 HLT, 120 BD) were obtained at 3 Hz, 241 (121 HLT, 120 BD) were obtained at 30 Hz, and 241 (120 HLT, 121 BD) were obtained at direct supply. The type and number of measurements are the same as previously described in Subsection 4.1. The results, in terms of accuracy, are reported in Table 7.

A graphical representation of the classification results, using confusion matrices, is shown in Figure 10, where the top row refers to the models using 0.5 seconds of data, and the bottom row refers to the models using the full transient state (5 seconds of data).

*6.5. Weakness of FFT when simultaneous faults occur*

On the other hand, as mentioned in Section 2, several works have been successfully proposed for detecting the occurrence of faults in induction motors. Indeed, in most of these proposals, the identification of faults is performed through classical frequency domain
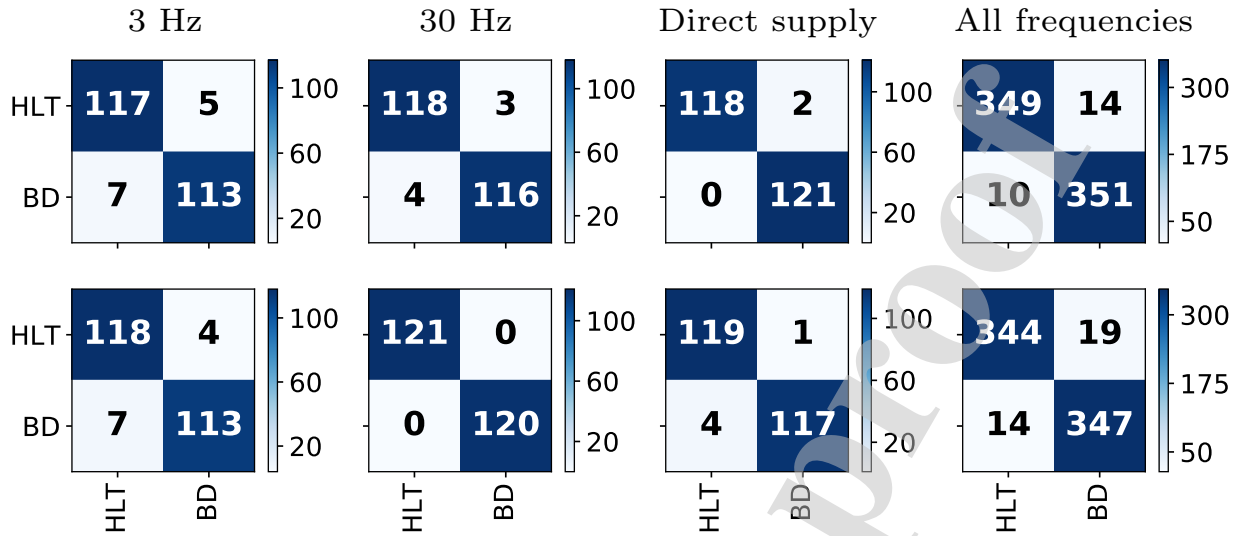
Figure 10: Confusion matrices of the four predictive models trained (one for each frequency and one operation frequency-insensitive). The $X$-axis corresponds to the predicted fault condition, while the $Y$-axis represents the actual fault condition. The matrices of the top row refers to the first 0.5 seconds of the IM operation, while the bottom row refers to the first 5 seconds (full transient state).

analysis (i.e., Fourier Transform applied to acquired vibration signals or stator current signatures). In this regard, and although the occurrence of faulty conditions in electric machines may be detected by means of calculating specific fault-related features (i.e., fault-related frequency features following classical motor current signature analysis—MCSA), the identification of simultaneous or multiple faulty conditions represent a major issue. Under this statement and considering the types of failures evaluated in this study, the unbalance (UNB) and misalignment (MAL) conditions are special scenarios that may produce similar effects on classical physical magnitudes that are acquired during the continuous monitoring of the rotatory electrical machine, such as vibrations or stator current signals.

Thereby, by analyzing the theoretical effects produced by the UNB and MAL conditions, both scenarios produce effects that are hard to distinguish. In the case of UNB condition due to a centrifugal force that generates a high vibration amplitude equal to $1 \times$ RPM ($1\times$ rotational speed). On the other hand, the MAL condition causes high radial and/or axial vibrations that usually produces their dominant frequency components at $2 \times$ RPM, but

29

also at $1 \times$ RPM [70]. Consequently, both faulty conditions tend to generate similar fault-related frequency components over an estimated stator current spectrum. Indeed, those components have normally been estimated with the characteristic frequency component for the UNB condition as follows:

$$f_{ecc} = f_s \left[ 1 + -k \left( \frac{1-s}{p} \right) \right] = f_s + -kf_r \tag{7}$$

where, $f_s$ is the electrical supply frequency, $f_r$ is the rotational frequency, $s$ is the per-unit slip, $p$ is the number of pole pairs, and $k = 1, 2, 3 \ldots$. If the number of pole pairs is equal to one, these equations are reduced to its simplified form, as presented.

Likewise, the characteristic fault-related frequencies associated with the MAL condition are estimated as follows [71]:

$$f_{sb} = f_s + -kf_r \tag{8}$$

Therefore, when performing the theoretical analysis of the effects that these faults may produce in relation to the corresponding stator current signal, it is concluded that both faulty conditions may appear masked or overlapping each other. This makes it difficult to identify and differentiate failures during condition assessment and, in addition, reduces the reliability of the condition monitoring strategy. In this regard, in order to experimentally demonstrate that the occurrence (isolated or simultaneous) of these faulty conditions, UNB and MAL, produces similar effects and that their fault-related frequency components appear overlapped, the stator current signals of the considered experimental test bench are analyzed.

Accordingly, Figure 11 shows the frequency spectra obtained by means of applying the FFT technique to the acquired stator current signature when the supply frequency, to feed the IM, was set at 30 Hz and direct supply (60 Hz), respectively. For both supply frequencies, the stator current spectra is shown around the closest to $f_s + f_r$, respectively.

Moreover, from these obtained spectra, it should be highlighted that the $f_s + f_r$ components of the UNB, MAL and UNB+MAL conditions appear at the same place with slightly different amplitudes. On the other side, despite the slightly differences between the amplitudes of each frequency component, these are not big enough as to identify the machine condition because the amplitude of these frequency components will change depending on
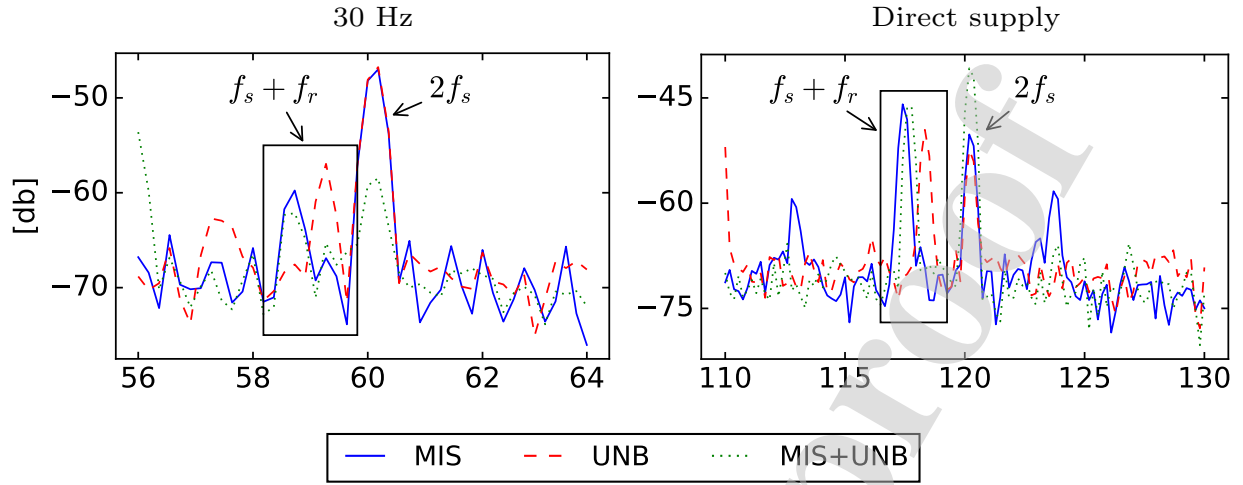
30

Figure 11: Conventional condition monitoring assessment based on the stator current spectrum at the closest region to $f_s + f_r$ when the supply frequency is set to 30 Hz (left) and direct supply (right). The $X$-axis depicts the frequency in Hz, while the $Y$-axis refers to the amplitude in db.

the severity of the individual faults. Hence, the number of false negatives will be higher and the performance of the fault identification will be affected. Summing up, through this experimental validation, it has been shown that classical condition monitoring strategies based on frequency domain analysis by means of the FFT have critical disadvantages when dealing with the identification of multiple fault sources that appear simultaneously, such as the UNB and MAL conditions in rotating machines like IM.

## 7. Conclusions and future lines

The problem of motor fault diagnosis has normally been addressed by using signal processing based on time domain, frequency domain, and time-frequency domain analyses. Nevertheless, the large data volumes collected by sensors for monitoring the motor operations are a source of information for the study of new multi-fault identification techniques, for faster and more efficient fault identification.

In this work, a new technique based on multiple sensor information for early diagnosis of single, combined, and simultaneous faulty conditions in IMs has been proposed, implemented and validated. In this regard, there exist several aspects that must be highlighted;

31

first, the proposed model is able to perform with high-performance during the analysis of the transient state because the PCA allows to retain a reduce number of significant features for each available physical measurement, that are vibrations, stator currents, voltages and rotational speed. Additionally, through the application of the PCA, the most representative information related to each evaluated condition is extracted without the need to use expert knowledge about each specific type failure. Secondly, despite the simplicity of the proposed approach, the consideration of PCA and multi label decision trees facilitates the identification and evaluation of multiple simultaneous failure conditions because the decision trees are specially adapted to this type of problems. Finally, the obtained results shown the effectiveness, robustness and capability for the extremely early detection with high-performance results even within the first 0.5 seconds of operation, during transient and steady state regimes. In this sense, it should be highlighted that the early fault diagnosis at low frequencies, such as 3 Hz, is a very challenging task and, the proposed method has demonstrated that achieves good results at different frequencies: 3 Hz, 30 Hz, and direct supply; and even, when variable load conditions are considered.

Therefore, those results are of special interest, because classical signal processing is not useful for light loads, and faults such as a broken rotor bar are very difficult to detect under these conditions [65]. Moreover, when the load at which the IM is working is known in advance, even better results could be achieved by training several models, one for each of the load conditions, and by selecting the right model as a function of the motor operating load. Although the best classification ratios were obtained from the models for each specific operating frequency, it has been demonstrated that the proposed method can also be trained using measures from variable operation frequency conditions without too much accuracy shrinkage. It may therefore be concluded that the proposed method is insensitive to the operational frequency.

With the aim of demonstrating that the proposed method is suitable for motors with other types of malfunctions, and additional test to diagnose a bearing-related fault has been conducted. Thus, it has demonstrated the capability of the proposed method to learn to diagnose BD fault with high accuracy rates. It may also be concluded that the method is

32

capable of efficiently solving binary classification problems, although when our PCA and decision trees strategy was used with both multi-label and binary problems, the accuracy of the resultant models was greater for the multi-label problems. Therefore, it is the first time that a model applied to multi-fault diagnosis in induction motors has been proposed and evaluated using the specific performance metrics for multi-label problems: accuracy, F1, Hamming loss, one-error, and rank loss.

As a future line of research, one promising option could be the use of supervised projections methods as alternatives to PCA, such as Linear Discriminant Analysis (LDA), which makes use of the labels to calculate the directions of data projection, and for which Wang et al. [72] have already presented an adaptation to the multi-label case.

Straightforward fault identification normally works by taking into account whether the fault is present or not. However, it is usually more useful to know when the fault is starting to occur. As a future line of research, our interest will be directed at the detection of fault severity or different magnitudes of malfunctioning.

## Acknowledgements

## References

[1] Y. Liu, A. M. Bazzi, A review and comparison of fault detection and diagnosis methods for squirrel-cage induction motors: State of the art, ISA Transactions 70 (2017) 400 – 409. `doi:10.1016/j.isatra.2017.06.001`.

[2] R. Liu, B. Yang, E. Zio, X. Chen, Artificial intelligence for fault diagnosis of rotating machinery: A review, Mechanical Systems and Signal Processing 108 (2018) 33 – 47. `doi:10.1016/j.ymssp.2018.02.016`.

[3] N. Bayar, S. Darmoul, S. Hajri-Gabouj, H. Pierreval, Fault detection, diagnosis and recovery using artificial immune systems: A review, Engineering Applications of Artificial Intelligence 46 (2015) 43 – 57. `doi:10.1016/j.engappai.2015.08.006`.

[4] Z. Gao, C. Cecati, S. X. Ding, A survey of fault diagnosis and fault-tolerant techniques – part I: Fault diagnosis with model-based and signal-based approaches, IEEE Transactions on Industrial Electronics 62 (6) (2015) 3757–3767. `doi:10.1109/TIE.2015.2417501`.

[5] I. Martin-Diaz, D. Morinigo-Sotelo, O. Duque-Perez, R. A. Osornio-Rios, R. J. Romero-Troncoso, Hybrid algorithmic approach oriented to incipient rotor fault diagnosis on induction motors, ISA Transactions 80 (2018) 427 – 438. `doi:10.1016/j.isatra.2018.07.033`.

[6] Y. Lei, F. Jia, J. Lin, S. Xing, S. X. Ding, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, IEEE Transactions on Industrial Electronics 63 (5) (2016) 3137–3147. `doi:10.1109/TIE.2016.2519325`.

[7] J. J. Saucedo-Dorantes, M. Delgado-Priet, R. A. Osornio-Rios, R. de Jesus Romero-Troncoso, Multifault diagnosis method applied to an electric machine based on high-dimensional feature reduction, IEEE Transactions on Industry Applications 53 (2017) 3086 – 3097. `doi:10.1109/TIA.2016.2637307`.

[8] K. H. Hui, M. H. Lim, M. S. Leong, S. M. Al-Obaidi, Dempster-shafer evidence theory for multi-bearing faults diagnosis, Engineering Applications of Artificial Intelligence 57 (2017) 160 – 170. `doi:10.1016/j.engappai.2016.10.017`.

[9] G. Tsoumakas, I. Katakis, Multi-label classification: An overview, International Journal of Data Warehousing and Mining 3 (3) (2007) 1–10,12–13. `doi:10.4018/jdwm.2007070101`.

[10] M.-L. Zhang, Z.-H. Zhou, A review on multi-label learning algorithms, IEEE Transactions on Knowledge and Data Engineering 26 (8) (2014) 1819–1837. `doi:10.1109/TKDE.2013.39`.

[11] S. A. S. A. Kazzaz, G. Singh, Experimental investigations on induction machine condition monitoring and fault diagnosis using digital signal processing techniques, Electric Power Systems Research 65 (3) (2003) 197 – 221. `doi:10.1016/S0378-7796(02)00227-4`.

[12] A. Garcia-Perez, R. J. Romero-Troncoso, E. Cabal-Yepez, R. A. Osornio-Rios, J. A. Lucio-Martinez, Application of high-resolution spectral analysis for identifying faults in induction motors by means of sound, Journal of Vibration and Control 18 (11) (2012) 1585–1594.

[13] P. Naderi, Modified magnetic-equivalent-circuit approach for various faults studying in saturable double-cage-induction machines, IET Electric Power Applications 11 (2017) 1224–1234(10). `doi:10.1049/iet-epa.2016.0782`.

[14] P. Gangsar, R. Tiwari, Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms, Mechanical Systems and Signal Processing 94 (2017) 464–481.

[15] A. Choudhary, D. Goyal, S. L. Shimi, A. Akula, Condition monitoring and fault diagnosis of induction motors: A review, Archives of Computational Methods in Engineering`doi:10.1007/s11831-018-9286-z`.

[16] T. W. Rauber, F. de Assis Boldt, F. M. Varejão, Heterogeneous feature models and feature selection applied to bearing fault diagnosis, IEEE Transactions on Industrial Electronics 62 (1) (2014) 637–646.

[17] A. J. Marques Cardoso, S. M. A. Cruz, D. S. B. Fonseca, Inter-turn stator winding fault diagnosis in three-phase induction motors, by park's vector approach, IEEE Transactions on Energy Conversion 14 (3) (1999) 595–598. `doi:10.1109/60.790920`.

[18] F. Duan, R. Živanović, Condition monitoring of an induction motor stator windings via global optimization based on the hyperbolic cross points, IEEE Transactions on Industrial Electronics 62 (3) (2015) 1826–1834. `doi:10.1109/TIE.2014.2341563`.

[19] H. Mahmoud, A. Abou-Elyazied Abdallh, N. Bianchi, S. M. El-Hakim, A. Shaltout, L. Dupré, An inverse approach for interturn fault detection in asynchronous machines using magnetic pendulous oscillation technique, IEEE Transactions on Industry Applications 52 (1) (2016) 226–233. `doi:10.1109/TIA.2015.2478882`.

[20] G. Betta, C. Liguori, A. Paolillo, A. Pietrosanto, A dsp-based fft-analyzer for the fault diagnosis of rotating machine based on vibration analysis, IEEE Transactions on Instrumentation and Measurement 51 (6) (2002) 1316–1322. `doi:10.1109/TIM.2002.807987`.

[21] V. Thomas, K. Vasudevan, V. Kumar, Online cage rotor fault detection using air-gap torque spectra, IEEE Transactions on Energy Conversion 18 (2) (2003) 265–270. `doi:10.1109/TEC.2003.811718`.

[22] J. R. Stack, T. G. Habetler, R. G. Harley, Bearing fault detection via autoregressive stator current modeling, IEEE Transactions on Industry Applications 40 (3) (2004) 740–747. `doi:10.1109/TIA.2004.827797`.

[23] R. Yan, R. X. Gao, X. Chen, Wavelets for fault diagnosis of rotary machines: A review with applications, Signal Processing 96 (2014) 1 – 15, time-frequency methods for condition based maintenance and modal analysis. `doi:10.1016/j.sigpro.2013.04.015`.

[24] E. H. E. Bouchikhi, V. Choqueuse, M. E. H. Benbouzid, Current frequency spectral subtraction and its contribution to induction machines' bearings condition monitoring, IEEE Transactions on Energy Conversion 28 (1) (2013) 135–144. `doi:10.1109/TEC.2012.2227746`.

[25] M.-Y. Chow, S. Yee, P. Mangum, A neural network approach to real-time condition monitoring of induction motors, IEEE Transactions on Industrial Electronics 38 (6) (1991) 448–453. `doi:10.1109/`

41.107100.

[26] V. N. Ghate, S. V. Dudul, Cascade neural-network-based fault classifier for three-phase induction motor, IEEE Transactions on Industrial Electronics 58 (5) (2011) 1555–1563. `doi:10.1109/TIE.2010.2053337`.

[27] H. Keskes, A. Braham, Recursive undecimated wavelet packet transform and dag svm for induction motor diagnosis, IEEE Transactions on Industrial Informatics 11 (5) (2015) 1059–1066. `doi:10.1109/TII.2015.2462315`.

[28] J. D. Martínez-Morales, E. R. Palacios-Hernández, D. Campos-Delgado, Multiple-fault diagnosis in induction motors through support vector machine classification at variable operating conditions, Electrical Engineering 100 (1) (2018) 59–73. `doi:10.1007/s00202-016-0487-x`.

[29] P. Konar, P. Chattopadhyay, Multi-class fault diagnosis of induction motor using hilbert and wavelet transform, Applied Soft Computing 30 (2015) 341 – 352. `doi:10.1016/j.asoc.2014.11.062`.

[30] C. Shen, D. Wang, F. Kong, P. W. Tse, Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier, Measurement 46 (2013) 1551 – 1564. `doi:10.1016/j.measurement.2012.12.011`.

[31] R. Romero-Troncoso, A. Garcia-Perez, D. Morinigo-Sotelo, O. Duque-Perez, R. Osornio-Rios, M. Ibarra-Manzano, Rotor unbalance and broken rotor bar detection in inverter-fed induction motors at start-up and steady-state regimes by high-resolution spectral analysis, Electric Power Systems Research 133 (2016) 142 – 148. `doi:10.1016/j.epsr.2015.12.009`.

[32] I. T. Jolliffe, Principal Component Analysis, Springer Series in Statistics, Springer New York, New York, NY, 1986. `doi:10.1007/978-1-4757-1904-8`.

[33] I. T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374 (2065) (2016) 20150202.

[34] E. Roman-Rangel, S. Marchand-Maillet, Inductive t-SNE via deep learning to visualize multi-label images, Engineering Applications of Artificial Intelligence 81 (2019) 336 – 345. `doi:10.1016/j.engappai.2019.01.015`.

[35] W. R. Zwick, W. F. Velicer, Comparison of five rules for determining the number of components to retain., Psychological Bulletin 99 (3) (1986) 432–442. `doi:10.1037/0033-2909.99.3.432`.

[36] H. F. Kaiser, The application of electronic computers to factor analysis, Educational and Psychological Measurement 20 (1) (1960) 141–151. `doi:10.1177/001316446002000116`.

[37] R. B. Cattell, The scree test for the number of factors, Multivariate Behavioral Research 1 (2) (1966) 245–276. `doi:10.1207/s15327906mbr0102_10`.

[38] J. Ruscio, B. Roche, Determining the number of factors to retain in an exploratory factor analysis

using comparison data of known factorial structure., Psychological Assessment 24 (2) (2012) 282–292. doi:10.1037/a0025697.

[39] L. Breiman, Classification and regression trees, Routledge, 2017.

[40] J. R. Quinlan, C4. 5: programs for machine learning, Elsevier, 2014.

[41] J. Maudes, J. J. Rodríguez, C. García-Osorio, N. García-Pedrajas, Random feature weights for decision tree ensemble construction, Information Fusion 13 (1) (2012) 20–30. doi:10.1016/j.inffus.2010.11.004.

[42] J.-F. Díez-Pastor, C. García-Osorio, J. J. Rodríguez, Tree ensemble construction using a GRASP-based heuristic and annealed randomness, Information Fusion 20 (2014) 189–202. doi:10.1016/j.inffus.2014.01.009.

[43] C. Pardo, J. F. Diez-Pastor, C. García-Osorio, J. J. Rodríguez, Rotation forests for regression, Applied Mathematics and Computation 219 (19) (2013) 9914–9924. doi:10.1016/j.amc.2013.03.139.

[44] Á. Arnaiz-González, J. F. Díez-Pastor, C. García-Osorio, J. J. Rodríguez, Random feature weights for regression trees, Progress in Artificial Intelligence 5 (2) (2016) 91–103.

[45] J. Read, Scalable multi-label classification, Ph.D. thesis, University of Waikato (2010). URL https://researchcommons.waikato.ac.nz/bitstream/handle/10289/4645/thesis.pdf

[46] E. Spyromitros-Xioufis, G. Tsoumakas, W. Groves, I. Vlahavas, Multi-target regression via input space expansion: treating targets as inputs, Machine Learning 104 (1) (2016) 55–98.

[47] W. Waegeman, K. Dembczyński, E. Hüllermeier, Multi-target prediction: a unifying view on problems and methods, Data Mining and Knowledge Discovery 33 (2) (2019) 293–324.

[48] Á. Arnaiz-González, J.-F. Díez-Pastor, J. J. Rodríguez, C. García-Osorio, Study of data transformation techniques for adapting single-label prototype selection algorithms to multi-label learning, Expert Systems with Applications 109 (2018) 114–130. doi:10.1016/j.eswa.2018.05.017.

[49] Á. Arnaiz-González, J.-F. Díez-Pastor, J. J. Rodríguez, C. García-Osorio, Local sets for multi-label instance selection, Applied Soft Computing 68 (2018) 651 – 666. doi:10.1016/j.asoc.2018.04.016.

[50] M. Kordos, Á. Arnaiz-González, C. García-Osorio, Evolutionary prototype selection for multi-output regression, Neurocomputingdoi:10.1016/j.neucom.2019.05.055.

[51] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[52] R. Kohavi, et al., A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Vol. 14, Montreal, Canada, 1995, pp. 1137–1145.

[53] K. Kim, A. G. Parlos, Model-based fault diagnosis of induction motors using non-stationary signal

segmentation, Mechanical Systems and Signal Processing 16 (2-3) (2002) 223–253. `doi:10.1006/mssp.2002.1481`.

[54] G. Madjarov, D. Kocev, D. Gjorgjevikj, S. Džeroski, An extensive experimental comparison of methods for multi-label learning, Pattern recognition 45 (9) (2012) 3084–3104. `doi:10.1016/j.patcog.2012.03.004`.

[55] J. L. Horn, A rationale and test for the number of factors in factor analysis, Psychometrika 30 (2) (1965) 179–185. `doi:10.1007/BF02289447`.

[56] J. S. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, in: Advances in neural information processing systems, 2011, pp. 2546–2554.

[57] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, D. D. Cox, Hyperopt: a python library for model selection and hyperparameter optimization, Computational Science & Discovery 8 (1) (2015) 014008. `doi:10.1088/1749-4699/8/1/014008`.

[58] R. H. Shumway, D. S. Stoffer, Time Series Analysis and Its Applications, Springer Texts in Statistics, Springer International Publishing, Cham, 2017. `doi:10.1007/978-3-319-52452-8`.

[59] R. E. Bellman, Adaptive control processes: a guided tour, Vol. 2045, Princeton university press, 2015.

[60] T. Wang, J. Qi, H. Xu, Y. Wang, L. Liu, D. Gao, Fault diagnosis method based on FFT-RPCA-SVM for Cascaded-Multilevel Inverter, ISA Transactions 60 (2016) 156–163. `doi:10.1016/J.ISATRA.2015.11.018`.

[61] J. A. Cadzow, B. Baseghi, T. Hsu, Singular-value decomposition approach to time series modelling, in: IEE Proceedings F (Communications, Radar and Signal Processing), Vol. 130, IET, 1983, pp. 202–210.

[62] D. Lowe, Feature space embeddings for extracting structure from single channel wake eeg using rbf networks, in: Neural Networks for Signal Processing VIII. Proceedings of the 1998 IEEE Signal Processing Society Workshop (Cat. No. 98TH8378), IEEE, 1998, pp. 428–437.

[63] C. James, D. Lowe, Single channel analysis of electromagnetic brain signals through ica in a dynamical systems framework, in: 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 2, IEEE, 2001, pp. 1974–1977.

[64] W. L. Woon, D. Lowe, Can we learn anything from single-channel unaveraged meg data?, Neural Computing & Applications 13 (4) (2004) 360–368.

[65] O. E. Hassan, M. Amer, A. K. Abdelsalam, B. W. Williams, Induction motor broken rotor bar fault detection techniques based on fault signature analysis–a review, IET Electric Power Applications 12 (7) (2018) 895–907. `doi:10.1049/iet-epa.2018.0054`.

[66] Y. Hochberg, A sharper Bonferroni procedure for multiple tests of significance, Biometrika 75 (4) (1988) 800–802. `doi:10.1093/biomet/75.4.800`.

[67] C. Li, J. V. de Oliveira, M. Cerrada, F. Pacheco, D. Cabrera, V. Sanchez, G. Zurita, Observer-biased

bearing condition monitoring: From fault detection to multi-fault classification, Engineering Applications of Artificial Intelligence 50 (2016) 287 – 301. doi:10.1016/j.engappai.2016.01.038.

[68] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, H. Henao, Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based on Statistical-Time Features and Neural Networks., IEEE Transactions on Industrial Electronics 60 (8) (2012) 3398–3407.

[69] W. Caesarendra, T. Tjahjowidodo, A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing, Machines 5 (4) (2017) 464–481. doi:10.3390/machines5040021.

[70] M. Tsypkin, The origin of the electromagnetic vibration of induction motors operating in modern industry: Practical experience, analysis and diagnostics, TRANSACTIONS ON INDUSTRY APPLICATIONS 53 (2017) 1669 – 1676. doi:10.1109/PCICON.2016.7589239.

[71] G. B. Carlos Verucchi, José Bossio, G. Acosta, A misalignment detection in induction motors with flexiblecoupling by means of estimated torque analysis and mcsa, Mechanical Systems and Signal Processing 80 (2016) 570 – 581. doi:https://doi.org/10.1016/j.ymssp.2016.04.035.

[72] H. Wang, C. Ding, H. Huang, Multi-label linear discriminant analysis, in: K. Daniilidis, P. Maragos, N. Paragios (Eds.), Computer Vision – ECCV 2010, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 126–139. doi:10.1007/978-3-642-15567-3_10.

**\*Conflict of Interest**

There is no conflict of interest with the concerned persons or organizations

There is no conflict of interest with the concerned persons or organizations