

Received 22 November 2023, accepted 5 December 2023, date of publication 29 December 2023, date of current version 26 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3348245

RESEARCH ARTICLE

Bearing Fault Detection and Recognition From Supply Currents With Decision Trees

GIOVANNI BRIGLIA¹, FABIO IMMOVILLI¹, (Senior Member, IEEE), MARCO COCCONCELLI¹, AND MARCO LIPPI¹

Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, 42122 Reggio Emilia, Italy

Corresponding author: Marco Lippi (marco.lippi@unimore.it)

ABSTRACT This paper considers the tasks of detecting and recognizing bearing faults in electric motors from the signals collected from supply currents, using machine learning techniques. In particular, following recent trends in AI, the main point of interest was focused towards interpretable solutions that provide explanations on the decisions taken by the classifiers. For this reason, decision trees were chosen, since they represent a classic machine learning approach which inductively learns tree structures from a collection of observations. Paths along the learnt trees can be easily interpreted as plain classification rules. An extensive experimental comparison shows the strong generalization capabilities of such a classifier. In particular, the present work reports results obtained in a highly challenging scenario, usually overlooked in the literature, where the system is tested on configurations of radial and torsional loads that have not been observed during training. The approach achieves over 90% of accuracy even on this cross-load generalization setting.

INDEX TERMS Fault detection, fault recognition, decision trees, explainable artificial intelligence.

NOMENCLATURE

BALL CHARACTERISTICS

β	Ball contact angle.
<i>BPFI</i>	Ball Pass Frequency Inner or inner race failing frequency.
<i>BPFO</i>	Ball Pass Frequency Outer or outer race failing frequency.
<i>BSF</i>	Ball Spin Frequency or rolling element failing frequency.
D_b	Ball diameter.
D_c	Ball bearing pitch diameter.
F_{be}	Ball bearing characteristic fault frequency, electrical domain.
F_{car}	Ball bearing characteristic fault frequency, mechanical domain (FTF, BPFO, BPFI, BSF).
n	Number of rolling elements.

MACHINE LEARNING

DT	Decision Tree.
KNN	K-Nearest Neighbors.
LR	Logistic Regression.

MOTOR

f	Electric machine supply frequency.
F_r	Shaft rotation frequency.
<i>FTF</i>	Fundamental Train Frequency or cage failing frequency.
i_u, i_v, i_w	Three-phase supply currents.
D1,D2,D3	Level of the bearing damage.
MCSA	Motor Current Signal Analysis.
MUT	Motor under test.
R1,R2,R3	Level of the bearing radial load.
T1,T2,T3	Level of the torque load.

PERFORMANCE METRICS

A	Accuracy.
F1	Harmonic mean between precision and recall.
FN	False Negative.
FP	False Positive.
P	Precision.
R	Recall.
TN	True Negative.
TP	True Positive.

I. INTRODUCTION

Electrical and mechanical fault diagnosis in rotating machinery has been the subject of extensive research in the last

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh¹.

decades with the aim to optimize maintenance and cost savings. Concerning electrical machines, induction motors operating at mains frequency are still widely adopted in industry, mainly because of their low price, ruggedness and reliability. Many works in the scientific literature concern the problem of general condition monitoring of induction machines [1] and fault occurrence within the machine components [2]. Focusing on the mechanical faults, bearing faults are one of the most common failure mode in electrical machines. Bearing faults that are not timely detected may result in reduced performance, degraded efficiency, overheating and malfunctions, up to catastrophic failure of the driven machinery [3].

Fault diagnosis methods based on the analysis of the vibration signals have proven their effectiveness in many scenarios [4]. A common practice is to rely on statistical indicators (e.g., the root mean square value of the vibration velocity [5] or the kurtosis [6]) as a representative of the health status of the electro-mechanical system being monitored. Vibration analysis can be employed as an on-line fault detection tool, but it is usually limited to routine inspections, since the diagnosis equipment is expensive and invasive, requiring dedicated transducers to be installed on the monitored machinery.

Among the non-invasive monitoring methods, motor current signature analysis (MCSA) relies on the monitoring of electrical quantities that are already acquired in the main drive application, e.g., for power metering/energy monitoring, or over-current protection, or to implement the control of an electric drive. Thus, MCSA does not require the installation of additional dedicated transducers. By using signals from the electrical domain, a non-invasive method to diagnose a fault in the system via on-line monitoring of the electrical supply quantities can be obtained [7], [8]. In fact, under various circumstances mechanical signals cannot be directly acquired in field applications: e.g., in remote locations installations such as in-well pumps, when facing harsh environments, or simply because the machinery is difficult to access. Under such conditions, electric signal measurements would be preferable as they are readily available and more immune to external disturbances. Fault detection at the early stage via non-invasive fault diagnosis is preferred, to allow for scheduled maintenance, minimizing system downtime. Suitable signal processing techniques are required to efficiently extract and isolate the fault signatures from raw signal, since fault signatures at incipient stage feature a very small amplitude that is usually buried in noise and which can lead to false positive detection [9].

Recently, also machine learning approaches have been successfully applied to bearing fault detection and classification [10], [11]. Some of the approaches exploit feature extraction and classic machine learning techniques [12], [13], [14], whereas others are based on more recent deep learning architectures [11]. Among the former, few works deal with the current signals, for example by employing a classifier ensemble in combination with discrete wavelet

transform [13] or by using support vector machines with motor stator current spectral features, also in combination with vibration signals [15]; decision trees have instead been used with the stationary wavelet packet transform on the vibration signals in [12] but not on current signals. Among the latter, convolutional neural networks have been widely applied, for example in a transfer learning setting [16] or, again, in combination with discrete wavelet transform [17]. The main limitations of deep learning approaches are their lack of interpretability, as they basically act as “black box” models, and the need for very large data collections for training.

The approach presented in this paper aims to fill this gap in the literature: that of building an automatic, interpretable system for bearing fault detection and recognition across different scenarios, using current signals only. The main contributions of the present work and the novelty with respect to the related literature can be summarized as follows: (i) the use of motor current signals only, rather than the much more investigated domain of vibration signals; (ii) an extremely challenging experimental evaluation scenario, much underrated in the literature, where the generalization capabilities of the fault detection and recognition system are assessed under different radial and torsional load conditions; (iii) the capability of providing explanations for the classification, by using an inherently interpretable machine learning approach: namely, decision trees. Using systems that are interpretable “by design” instead of looking for a-posteriori explanations of black-box models is an argument that has recently found a large consensus in the AI community [18]. Although decision trees have been largely used for fault detection and recognition tasks, also in the context of condition monitoring, they have never been tested on current signals, nor in the challenging cross-load evaluation scenario that are the subject of the present work, and which is of utmost interest for industrial applications.

Other works in the literature have applied decision trees to bearing fault detection and recognition from supply current signals, although using different feature sets and without a cross-generalization with different radial and torsional loads. In [19], electric damages are mainly considered, and empirical mode decomposition is exploited, using the module of the current vector, while in the present work the instantaneous phase is considered, in order to generalize to different loads. A combination of decision trees and neural networks is instead used in [20] to recognize different induction motor fault conditions: different loads are considered in the experimental evaluation, but no cross-load generalization analysis is performed.

The main contributions of the paper are the following:

- The use of an interpretable machine learning approach, namely decision trees, to perform bearing fault detection and recognition with the intent of providing explanations of the outcome of the condition monitoring system.

- The exploitation of current signals only, without resorting to the (more typically investigated) analysis of vibrations signals.
- The analysis of a particularly challenging scenario of cross-load generalization, where the machine learning system is trained on certain torsional and radial load configurations, and tested on others.
- The code and data used for the experiments were released, to allow reproducibility of the results.

The paper is organized as follows: Section II reviews the preliminaries, including the mechanisms of fault signature generation, the techniques for condition monitoring and an overview of supervised machine learning. Section III introduces the proposed method, comprising current signal preprocessing, decision trees, and post-processing filters. Section IV presents the experimental setup used to gather data under different working conditions and bearing damage, together with the training and validation methods. The section is completed by experimental results discussion, followed by Conclusion and final remarks.

II. PRELIMINARIES

A. BEARING CONDITION MONITORING

Rolling bearings are among the most used components in mechanical engineering since they realize a revolute joint between two bodies. Moreover they are a fundamental part of rotating electric motors since they support the rotating shaft, limiting the friction between rotor and stator by means of rolling elements.

Figure 1 shows an example of a radial bearing. It consists of two concentric rings with inner and outer races machined on them, separated by spherical or cylindrical rolling elements. Rolling elements are uniformly distributed along the circumference by means of a cage, preventing unwanted contacts.

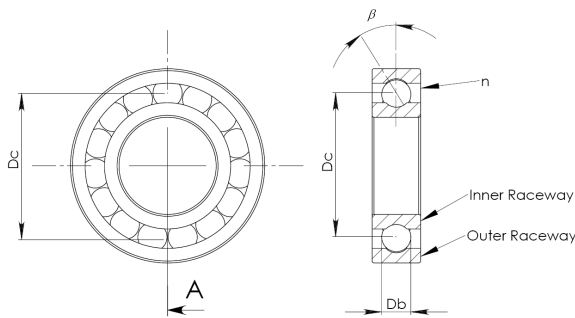


FIGURE 1. Simplified drawing of bearing structure showing the characteristic dimensions.

The kinematics of bearing characteristic frequency components can be computed knowing the bearing’s physical dimensions and the rotating speed of the inner and outer rings [6]. When a fault occurs, the rolling elements produce periodic impacts resulting in vibrations with a characteristic frequency depending on the location of the fault (e.g. on the outer race), the size of the bearing and the working conditions

(i.d. the rotating frequency of the shaft). The following equations of the characteristic frequencies (F_{car}) are obtained by considering the outer ring fixed to the frame and a specific faulted element, namely the cage (FTF), the outer race (BPFO), the inner race (BPFI) and the rolling element (BSF):

$$FTF = \frac{1}{2} F_r \left(1 - \frac{D_b \cos \beta}{D_c} \right) \quad (1)$$

$$BPFO = \frac{n}{2} F_r \left(1 - \frac{D_b \cos \beta}{D_c} \right) \quad (2)$$

$$BPFI = \frac{n}{2} F_r \left(1 + \frac{D_b \cos \beta}{D_c} \right) \quad (3)$$

$$BSF = \frac{D_c}{D_b} F_r \left[1 - \left(\frac{D_b \cos \beta}{D_c} \right)^2 \right] \quad (4)$$

where D_b stands for the ball diameter, D_c for the pitch diameter, n for the number of rolling elements, β for the ball contact angle, according to Fig.1, while F_r is the rotating frequency of the motor shaft. The aim of the condition monitoring of a ball bearing is the early detection of one of these characteristics frequencies (generically designated as F_{car}), e.g. in the vibration signal [6].

Motor Current Signal Analysis (MCSA) is a well-established research field [21], aiming to monitor working condition and to early detect motor malfunctions without disturbing the production. Regarding the bearing diagnostics, the link between mechanical fault components and motor current spectral components is modeled according to two main different effects [22]. Regardless, the mechanical bearing fault components introduce modulation in the machine’s supply currents at frequencies F_{be} :

$$F_{be} = |f \pm k F_{car} \pm j F_r| \quad (5)$$

where k is an integer and $j = \left[0, 1, 0.5 \left(1 - \frac{D_b \cos \beta}{D_c} \right) \right]$ corresponding to a fault on the outer, inner and rolling element respectively. The last frequency contribution is due to the eccentricity effect and often it is negligible with respect to the torque effect [23].

When dealing with realistic not catastrophic faults, the fault signature, being only a small fraction of the supply current, is usually buried in noise or completely swamped by the supply current, so that retrieving bearing fault signature components by means of MCSA is usually a difficult task. According to literature, the torque ripple associated with a realistic severity fault results in characteristic harmonic components on the supply currents whose amplitude 3-4 orders of magnitude smaller than the nominal current of the machine: a value near the accuracy class of common industrial current transducers.

B. MACHINE LEARNING

In this work, bearing fault detection and recognition is considered as a supervised learning tasks. In supervised machine learning, the goal is to learn a function f that

associates a target variable $y \in Y$ to a given set of observable features $x \in X$. The function is inductively learnt from a collection of samples (named data set) that consist in a set of N pairs in the form (x_i, y_i) , each associating a target y_i to a given set of features x_i . For the task of classification, the set Y contains the possible categories (i.e., fault or healthy bearing in case of fault detection). The set X contains all those variables that can be observed and are functional to the prediction of the target variable $y \in Y$.

How function f is defined, and also learnt, clearly depends on the adopted machine learning algorithm and on the underlying hypotheses that are enabled by background knowledge of the problem. For example, in case it is known that a linear dependency between dependent (y) and independent (x) variables holds, function f can be defined as a linear combination of the input features. In the proposed approach, decision trees (DTs) will be taken into consideration: DTs are a kind of machine learning system capable to capture also non-linear dependencies between input and output variables, with the additional characteristic of providing interpretable classification rules. More details will be given in Section III-B.

III. PROPOSED APPROACH

The approach used in the present work for the tasks of bearing fault detection and recognition is hereby summarized. The general setting consists of an induction electric motor containing rolling bearings to support the shaft. The signals collected from the motor consist in the supply currents and in the vibrations related to an uniaxial accelerometer and a triaxial accelerometer. The subject of study of the present work are the supply currents, whereas the study on vibrations will be described in the experiments as a matter of comparison only but they will not be reported in this paper. The aim is to detect bearing failures from the acquired signals, using interpretable machine learning techniques.

In order to describe the overall approach, in the following subsections will be illustrated (i) the pre-processing steps needed for the input signals; (ii) the employed machine learning algorithm; (iii) a post-processing filtering stage that can be exploited to further improve performance.

A. DATA PRE-PROCESSING

Usually statistical scalar indicators are used to provide a collection of parameters representative of the health status of the electro-mechanical system being monitored. Phase current signals are acquired and processed in order to isolate and enhance fault signature data, so that the influence of other operating conditions is minimized. An outline of the proposed signal processing technique used in the present work is summarized hereafter.

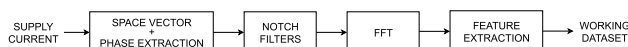


FIGURE 2. Data pre-processing flowchart of the current signal.

Statistical scalar fault indicators are preferred and will be used to assess the response of the classification system signal. Current signal pre-processing steps are visually summarized in Figure 2.

1) CURRENT SPACE VECTOR CALCULATION

The three-phase input current signal is acquired and any residual DC offset due to transducer drift is eliminated. In order to condense the information, the analysis is carried out using the space vector (SV) representation:

$$\vec{i}_s = K \left[i_u(t) + i_v(t)\alpha + i_w(t)\alpha^2 \right] \quad (6)$$

where \vec{i}_s is the resulting current phasor, i_u, i_v, i_w are the three supply currents, $\alpha = e^{j(\frac{2\pi}{3})}$ and $K = \sqrt{\frac{3}{2}}$. Since the vector's magnitude depends on the electric motor load (i.e. the load torque applied to the shaft), the present work focuses on the instantaneous phase angle, in order to condense the information on a single scalar parameter independent from the load condition. Specifically the difference in phase between the SV of the measured currents and a reference 50 Hz phasor is calculated, to obtain the instantaneous phase angle modulation of the SV.

2) NOTCH FILTER

As a second step, the resulting signal is filtered by a series of notch filters in order to remove the harmonics of 50 Hz fundamental mains supply frequency. This has been done with a series of second-order notch filters with central band on 50 Hz and its multiples, up to 500 Hz. Figure 3 demonstrates the effect of the notch filter using the signal of the healthy bearing case as an example. This is an important step in order to thoroughly clean the signal before the extraction of the features, that could be swamped in the omnipresent 50 Hz mains frequency noise. Figure 4 shows the spectra resulting from various types of bearing fault under the same load conditions. Specifically by comparing the Healthy case against the faulty ones, it can be seen that the spectra are different, especially in the low frequency part of the harmonic content.

3) FAST FOURIER TRANSFORM

The third step consists in the computation of the Fast Fourier Transform (FFT) to study the spectrum associated with the previously cleaned signal. The chosen sample rate is 25.6 kHz, which is consistent with the acquisitions made on the test bench. Frequency domain analysis is carried out on the instantaneous phase angle modulation of the SV of the supply currents, in order to assess the presence or increase of defect harmonics modulation in the spectrum of the instantaneous phase.

4) FEATURE EXTRACTION

The last stage of data pre-processing is related to the choice of the frequency slots from which features are to be extracted. As a design choice, 50 Hz wide overlapping frequency slots

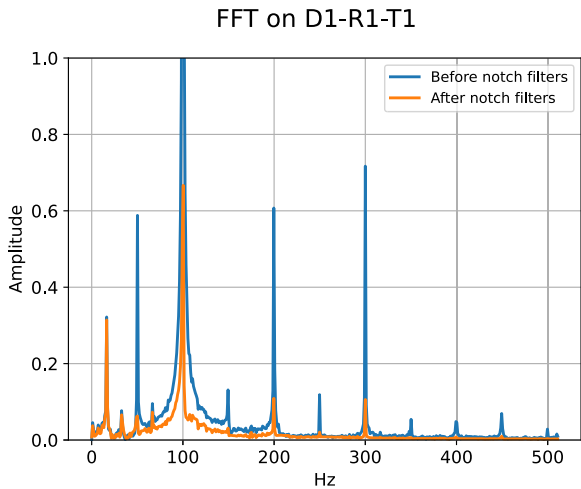


FIGURE 3. FFT of the signal relative to the healthy bearing with and without the notch filters applied, case studied: D1-R1-T1.

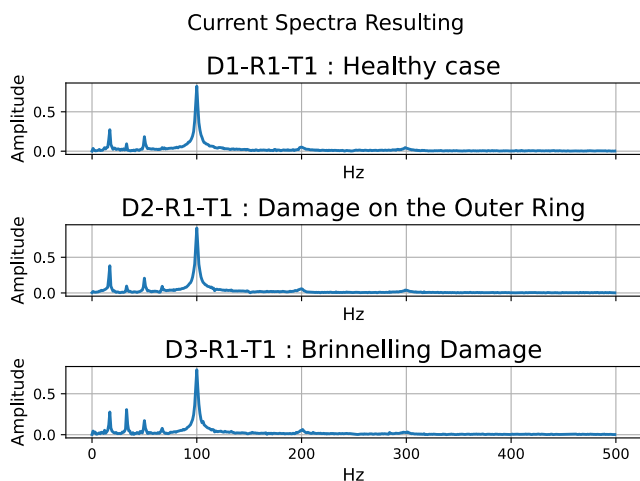


FIGURE 4. Sample spectra resulting from the analysis of the supply current: comparison in case of different fault conditions.

(e.g., 100-150, 125-175, etc.) are considered, also including a low-frequency window in the the 10-25 Hz range. The spectrum up to 375 Hz was taken into consideration, as the typical failure frequencies of rolling bearings are mainly in this range for the operating speeds reached during the experiments.

For each frequency slot the following features have been extracted from the signal’s spectrum: maximum, minimum, standard deviation, average, kurtosis, skewness, median absolute deviation, score at the 35th percentile, entropy, 35th percentile rank, coefficient of variation, unbiased estimator of the variance of the k -statistic and variance. Columns with zero standard deviation (namely, the 35th percentile from range 100-150 Hz up to 325-375 Hz) were discarded, obtaining a total of 185 features.

B. DECISION TREES

The aim of this work is not just to achieve high performance for the fault detection (or recognition) task,

but also to obtain interpretable predictions. Therefore, an inherently interpretable machine learning model was chosen, specifically Decision Trees (DTs) [24]. A DT is a classic machine learning system, where a tree is inductively learnt from a collection of examples. In the tree, each node is an attribute (or feature) and edges are values (or ranges of values) associated to that attribute. Finally, each leaf is associated to a class. Therefore, each path from the root of the tree down to a leaf is an interpretable classification rule that basically explains the reasons of the classification.

Figure 5 shows an example of a portion of one of the trees learnt in the experiments. If the right-most path of the tree is considered, the classification rule indicates that if the kurtosis of the frequency signal in range [75,125] Hz is larger than 11.037 and the maximum of the signal in range [10,25] Hz is larger than 0.137, then a brinnelling fault is present (orange leaf). If the maximum of the signal in range [10,25] Hz is instead lower than 0.137, the system predicts a fault in the outer ring (green leaf). The left part of the tree continues with more nodes and layers, which are not shown in the chart. The logic behind the classification performed by a DT is thus easy to understand for a human, as it is very similar to the “manual” techniques used in the diagnostics of failures.

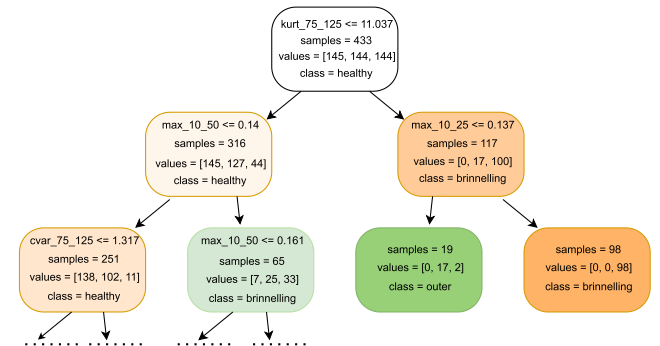


FIGURE 5. An example of a portion of decision tree. Each internal node corresponds to a variable, whereas edges departing from each node correspond to values (or value ranges) of such variable. Leaf nodes correspond to classes. The orange and green leaves indicate a predominance of the brinnelling and outer circle fault classes, respectively.

C. POST-PROCESSING FILTERS

The last component of the proposed approach consists in a post-processing filter that can be applied to refine the predictions made by the decision tree. In fact, while predictions can be made (and thus performance evaluation) sample by sample, in a real-world scenario it is much more appropriate to evaluate faults in a sort of sliding window setting: namely, the occurrence of a fault is confirmed only if the classifier outputs the faulty class in at least B samples out of the last n . Two different types of post-processing filters were implemented.

- 1) *Neighbor-based post-processing*, which consists in modifying the sample-level predictions based on B previous/next predictions in the time series. For example, if $B = 1$ only the previous and next samples are considered (see Figure 6, left). The result is still a set of predictions at the level of single samples.
- 2) *Sliding window post-processing*, which consists in counting the number of damage predictions made in n consecutive samples for a given load configuration, and predicting the fault class only in case such a number is above a fixed threshold. Thus, in this setting, a single prediction is computed for each test load configuration, and not for each sample (see Figure 6, right). This procedure is suitable for many industrial applications.

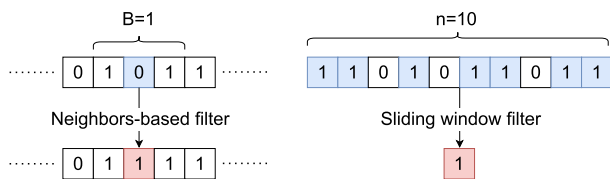


FIGURE 6. Left: neighbors-based filter with $B = 1$. Right: sliding window filtering. In the first case, predictions remain at sample level, whereas in the second case final predictions are made at the sequence level.

IV. EXPERIMENTAL VALIDATION

This section describes the collected data set and the experimental evaluation conducted on such data. All the experiments were run in Python using the `scikit-learn` package for machine learning, and the `Numpy` and `Scipy` for the computation of the features. All the experiments were carried out on a i7-10510U processor PC with 16GB of RAM. To allow reproducibility and to encourage research on the same application scenario, we also made our code and data available via Zenodo: <https://doi.org/10.5281/zenodo.10143055>.

A. DATA SET CONSTRUCTION

The data set used to train and validate the proposed fault detection method originates from a test set obtained by Design Of Experiment (DOE). The data set was introduced in [25] and already used in the literature [26].¹ The test bench comprises the motor under test (MUT) and the experimental setup allows to vary both the torque applied to the shaft as well as the radial load, thus allowing to re-create a variety of operating conditions. The MUT employed in the tests is a 6 poles induction machine and is directly connected to the 50 Hz three-phase mains grid. The nameplate data of the MUT is summarized in Table 1. A second induction machine, fed by a vector control inverter, is employed as the dynamometer/brake, allowing to vary the load torque on the MUT.

¹A similar data set was used in [26], where Long Short-Term Memory networks were used to perform fault detection, yet achieving lower performance with respect to the results described in this paper.

TABLE 1. Nameplate data of the motor under test.

Specification	Value
Nominal Power	1100 W
Number of poles	6
Power Factor	0.76
Nominal Torque	11.5 Nm
Stator Resistance	5.65 Ω

An additional test fixture comprising a crosshead and a pneumatic cylinder provides radial load on the MUT shaft.

The MUT is fitted with SKF 6205 deep groove ball bearing at the shaft drive end. Table 2 summarizes the characteristic dimensions supplied by the manufacturer, and the expected fault frequencies when the MUT operates at rated load. A total of three test bearings are employed: one healthy and two faulty with laboratory replicated defects: one with a single defect on the outer race and another with a simulated brinnelling fault, characterized by simultaneous presence of multiple characteristic fault signatures.

TABLE 2. Specifications of the 6205 ball bearing used in the experiments and expected fault frequencies.

Parameter	Value	SI unit
Inner diameter	25	[mm]
Outer diameter	52	[mm]
Number of spheres	9	
Basic static load rating	7800	[N]
f	50	Hz
F_r	15.9	Hz
$BPFI$	86.2	Hz
$BPFO$	57.1	Hz
FTF	6.3	Hz
BSF	75.1	Hz

The original data set contains the signals acquired in 27 different configurations, resulting from 3 different values for radial load (pressure of 0, 3 and 6 bars), 3 different values for load torque (0%, 50% and 100% rated torque) and 3 different bearing statuses (healthy, fault on the outer race, brinnelling). For each of these configurations, the data set provides 256,000 samples, obtained from an acquisition of the supply current signals for 10 seconds, with a sampling frequency of 25.6 kS/sec. As a further pre-processing step, from these 10 seconds of acquisition, 20 examples were extracted adopting a sliding window technique: 1 second of signal, with 0.5 seconds of overlapping. Having 27 different configurations of load and bearing status, the final data set employed in the experimental evaluation thus contains a total of 540 examples.

B. EXPERIMENTAL DATA SETS ANALYSIS

Given this data set, two different tasks are considered: (i) anomaly detection, where no distinction is made between the brinnelling fault and the damage to the outer ring, combining the two into a single class, and thus addressing a binary classification task (healthy vs. faulty bearing); (ii) fault

TABLE 3. Results (A=accuracy, P=precision, R=recall and F1) obtained with decision trees on the binary fault detection task (anomaly detection) considering signal pre-processing only, or the refinement with neighbor-based post-processing filter (using $B = 9$ bits). The last column shows the data size used in the training (Tr), validation (Vl) and testing (Ts) step for all tests.

Test Load	Pre-processing				Neighbor-based filter (B = 9)				Train/Test data size		
	A	P	R	F1	A	P	R	F1	Tr	Vl	Ts
R1	85	92	85	88.3	99	98	100	99.2	288	72	180
R2	76	80	84	82.1	99	99	100	99.5			
R3	74	78	86	81.7	87	84	100	91.3			
T1	87	84	100	91.3	84	81	100	89.6			
T2	77	85	80	82.4	92	97	90	93.5			
T3	80	87	82	84.6	79	99	69	81.4			
Average	84.3	89.5	88.34	88.03	90.37	96.94	93.38	91.32			
R1-T1	100	100	100	100	100	100	100	100	384	96	60
R1-T2	90	95	90	92.3	97	95	100	97.6			
R1-T3	93	97	92	94.9	98	98	100	98.8			
R2-T1	80	77	100	87	68	68	100	80.8			
R2-T2	92	93	95	93.8	97	95	100	97.6			
R2-T3	80	89	80	84.2	95	97	95	96.2			
R3-T1	90	90	95	92.7	90	90	95	92.7			
R3-T2	80	97	72	82.9	98	98	100	98.8			
R3-T3	80	77	100	87	68	68	100	80.8			
Average	84.3	89.5	88.34	88.03	90.37	96.94	93.38	91.32			

recognition, where the three classes are considered separately, thus addressing a multi-class classification task.

In both cases, an extremely challenging scenario is considered: namely, to assess whether the DT is able to detect (or recognize) faults across different load conditions, hence showing *generalization* capabilities. More precisely, the data set is split into a training set and a test set, according to the different radial and torsional loads: the learnt model is basically tested on load configurations that have never been seen during training. Based on the author's experience, this scenario has not yet been considered in the literature. For example, the system is trained only on radial load configurations R1 and R2 (none or half load) and tested on the full load configuration R3. It is worth noticing that this is a much more difficult task with respect to a classic random training/test split: in the latter case, in fact, the task often becomes trivial, since the training set usually contains examples that are very similar to those in the test set – an assumption which seldom holds in real-world applications.

As customary in any machine learning application, the hyper-parameters of the chosen algorithm have to be tuned to maximize performance. To do this, 20% of the examples intended for the training phase were used as a validation set. To find the best hyper-parameters for the DT the `hyperopt` library was used, in particular the `fmin` function. The following hyper-parameters were the subject of tuning: (i) maximum depth of the tree, (ii) minimum number of samples to perform further splits in the tree, (iii) minimum weighted fraction of total sum of input sample weights for a leaf node and (iv) maximum number of features to be considered when looking for the best split.

C. RESULTS

Standard classification metrics were used to measure the performance of the proposed system. In the case of binary

classification, the positive class was defined as the faulty class (with no distinction between fault categories), whereas the negative class corresponds to a healthy bearing. For each data example, it can thus be defined as a true positive (TP) the correct detection of a faulty status; true negative (TN) the correct detection of a healthy status; false positive (FP) as the wrong prediction of a fault in case of a healthy bearing; false negative (FN) as the missing detection of a fault. Accuracy is then defined as the total number of correct predictions out of the total number of test samples:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Recall, instead, corresponds to the ability of the classifier to find all the positive samples (thus taking into account false negatives), whereas precision is the percentage of positive predictions that are indeed correct:

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

As a synthesis between precision and recall, the $F1$ score is also typically used, as the harmonic mean between the two.

In the case of fault recognition, which is a multi-class classification task in the considered scenario, only accuracy is reported, since precision, recall and $F1$ become per-class metrics.

Results on binary classification (fault detection) are reported in Table 3. The test load column refers to the operating conditions used in testing step. For example, "R1" test load means that the test has been done on R1 level of radial load and all levels of torque loads (i.e. R1-[T1, T2, T3]), while "R1-T1" test load means that the test has been done on R1 level of radial load and T1 level of torque load

only (i.e. R1-T1). On the left, the performance achieved when using signal pre-processing only is shown, whereas on the right the performance gain obtained with the neighbor-based post-processing filter is shown (see Section III-C), using $B = 9$ bits. In both cases, these results have to be interpreted as per-sample: that is, the proposed classifier outputs a prediction for any given sample in the test set (i.e., every half a second, using a one-second input window). It is worth noting that the DT shows strong generalization capabilities, achieving almost 85% of accuracy on average, which is enhanced to over 91% when applying the neighbors-based filter.

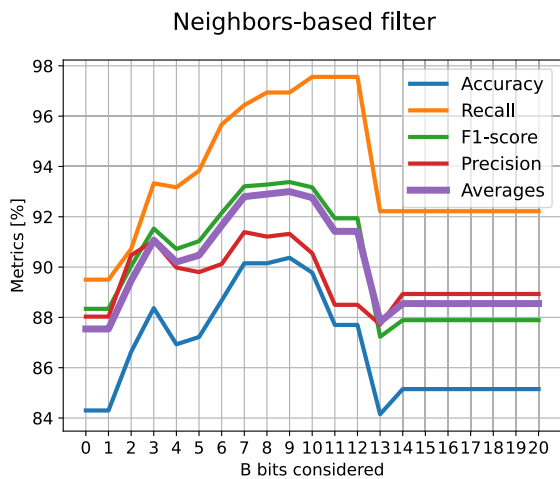


FIGURE 7. Performance of the neighbor-based filter as a function of the considered number of B bits in the post-processing phase.

Figure 7 shows the average of the performance metrics as a function of the considered number B of bits in the neighbor-based post-processing filter. This analysis motivates the choice of $B = 9$ for the best case, despite each B value investigated resulting in better performance compared to the initial point (where $B=0$ signifies a non-contributing filter).

As a further evaluation, the application of the sliding window post-processing filter is considered. In this case, for each load configuration the 20 predictions made on the available test samples are considered, and a fault for the overall sequence is confirmed in case the number of predictions of the positive class exceeds a chosen threshold between 1 and 20; as shown in Figure 8, it is evident that damage is consistently detected across all threshold levels. However, when no damage is present, achieving significant results requires the use of higher threshold levels. It was also considered a setting where the positive predictions have to be consecutive. The results presented in Figure 9 illustrate the performance of this filter. They demonstrate that when a threshold level of 12 is applied, the anomaly detection is highly accurate. Similarly, for lower threshold values, the filter consistently detects damages; whereas, as the threshold level increases, the accuracy of the anomaly detection gradually degrades. Conversely, for cases without

TABLE 4. Accuracy for anomaly detection, 2 classes (pre-processing only).

Test Load	DT	KNN	LR
R1	85	78	71
R2	76	62	32
R3	74	73	65
T1	87	67	66
T2	77	69	47
T3	80	72	67
R1-T1	100	87	70
R1-T2	90	78	78
R1-T3	93	87	78
R2-T1	80	72	75
R2-T2	92	75	67
R2-T3	80	85	52
R3-T1	90	85	67
R3-T2	80	67	62
R3-T3	80	85	67
Average	84.3	76.1	64.3

TABLE 5. Accuracy for fault recognition, 3 classes (pre-processing only).

Test Load	DT	KNN	LR
R1	67	74	57
R2	47	46	34
R3	68	63	46
T1	67	46	34
T2	57	58	54
T3	67	62	41
R1-T1	77	80	63
R1-T2	88	75	77
R1-T3	92	80	80
R2-T1	63	45	43
R2-T2	65	47	45
R2-T3	53	55	35
R3-T1	87	72	62
R3-T2	65	65	52
R3-T3	95	85	43
Average	70.5	63.5	51.1

TABLE 6. Accuracy for anomaly detection and fault recognition, varying the range of frequencies analyzed (pre-processing only).

Test Load	Anomaly Detection (Healthy/Faulty)			Fault recognition (Healthy/Outer fault/Brinnelling)		
	0-375	0-125	0-75	0-375	0-125	0-75
R1	85	81	72	67	67	56
R2	76	73	71	47	44	52
R3	74	73	71	68	65	47
T1	87	82	72	67	57	56
T2	77	76	69	57	61	47
T3	80	80	72	67	66	47
R1-T1	100	98	80	77	67	67
R1-T2	90	90	80	88	88	53
R1-T3	93	97	88	92	87	53
R2-T1	80	88	90	63	70	68
R2-T2	92	92	72	65	75	43
R2-T3	80	78	73	53	53	60
R3-T1	90	95	87	87	92	65
R3-T2	80	73	72	65	65	47
R3-T3	80	90	67	95	97	40
Average	84.3	84.4	75.7	70.5	70.3	53.4

damages, the opposite trend is observed. In both cases, a single prediction is obtained for every considered load configuration.

In order to assess the validity of the performance achieved by the DT classifier with respect to other predictors, the

TABLE 7. Examples of interpretable classification rules learned via decision trees.

1)	IF $\max_{[25,75]} > 0.069$ AND $\text{skew}_{[75,125]} > 3.081$ THEN BRINNELLING
2)	IF $\text{kurt}_{[75,125]} > 11.037$ AND $\max_{[10,25]} > 0.137$ THEN BRINNELLING
3)	IF $\text{kurt}_{[75,125]} > 11.037$ AND $\max_{[10,25]} \leq 0.137$ THEN OUTER RING FAULT
4)	IF $\max_{[25,75]} \leq 0.069$ AND $\text{entropy}_{[75,125]} \leq 3.393$ AND $\text{skew}_{[75,125]} > 3.01$ AND $\max_{[25,75]} \leq 0.053$ THEN OUTER RING FAULT
5)	IF $\text{kurt}_{[50,100]} \leq 11.9$ AND $\max_{[10,25]} \leq 0.231$ AND $\text{kurt}_{[75,125]} \leq 9.516$ AND $\max_{[10,50]} \leq 0.125$ AND $\text{std}_{[10,50]} \leq 0.132$ THEN HEALTHY

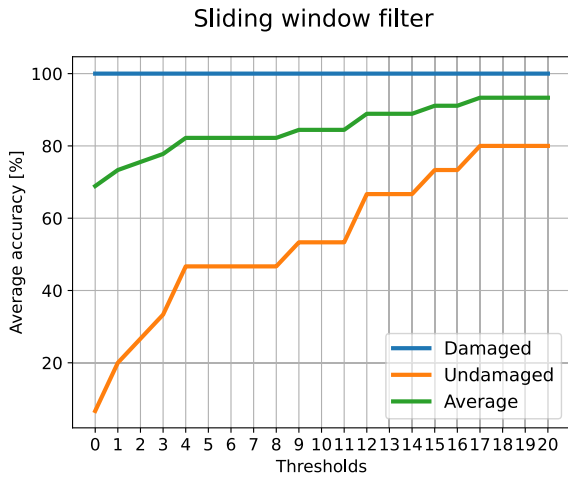


FIGURE 8. Performance metrics of the sliding window post-processing filter, as a function of the positive class threshold.

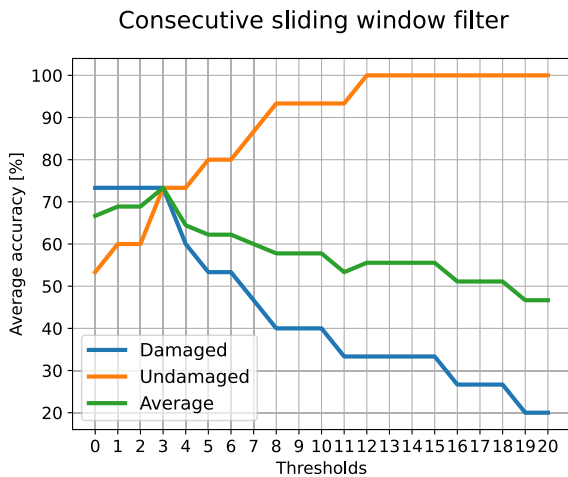


FIGURE 9. Performance metrics of the sliding window post-processing filter, as a function of the positive class threshold. The additional constraint of consecutive predictions for the positive class is considered.

proposed approach is compared against Logistic Regression (LR) and K-Nearest Neighbors (KNN). The former was chosen as a representative of the class of linear models, to assess whether the dependency between features and classes can be expressed in the form of a linear function. The latter was chosen as a further test of generalization capabilities: in

fact, KNN is a representative of distance-based approaches, which typically work well only when test examples have a high similarity with respect to training examples – a setting that only partially matches the considered scenario where load configuration are different between training and test sets. Tables 4 and 5 report results in terms of accuracy. The DT stands out as the best choice when compared to its competitors. It is evident that linear models are ill-equipped to capture the non-linear dependencies between dependent and independent variables. As a result, LR shows, on average a 20% lower accuracy when compared to the DT. Conversely, when compared to the KNN classifier, the discrepancy is slightly smaller, with an accuracy difference of more than 10%.

In order to assess the contribution of the features extracted from the whole frequency range, Table 6 summarizes the accuracy obtained with features extracted from reduced spectra of frequencies, both in the 2-class and in the 3-class settings. The results show that the most important features are those extracted in the interval [0-125] Hz, whereas considering the larger interval [0-375] gives only a slightly improvement for all problems. Considering only the [0-75] Hz interval produces, instead, a significant degradation in performance.

As explained throughout the paper, the proposed fault detection is implemented using a DT in order to have interpretable solutions. With the aim of identifying the most frequent patterns that represent classification rules, the tree structures has been analyzed to find those tree portions that appear across several scenarios with different radial and torsional load. Table 7 shows some examples of rules that have been extracted with this procedure. It can be easily observed that rules involving the faulty classes, especially brinnelling, are typically much shorter, meaning that the observation of a smaller number of features is often sufficient to perform classification. For example, the first rule takes into account two features in order to recognize a brinnelling fault: if the maximum within the [25,75] Hz interval and the skewness of the [75,125] Hz are above some threshold (identified by the learning process) then a brinnelling occurs. The second and third rules are interesting because they distinguish brinnelling from the outer ring fault: in both cases the kurtosis within the [75,125] Hz interval is above 11, whereas the maximum within the [10,25] Hz interval is used

to discriminate between the two fault categories. In general, it can be observed from the rules extracted from the DT how the low frequencies (below 125 Hz) are more informative to recognize and discriminate the faults (especially the maximum value, the skewness and the kurtosis), which is not surprising given the expected fault frequencies reported in Table 2. This observation also corroborates the results reported in Table 6.

As a final test for the effectiveness of features, for each combination of radial and torsional load, a DT was constructed using only the six most important features:² it achieved above 77.3% in average accuracy over the load settings for the binary task, and over 63.1% for the 3-classes. This result suggests that this reduced set of features is very informative for the detection of a fault (accuracy is not far from 84.3% obtained with the whole set of features) but not entirely sufficient for fault recognition, given the decrease in performance from 70.5%, obtained with all the features.

V. CONCLUSION

This paper described a proposed machine learning approach based on decision trees to perform bearing anomaly detection and fault recognition in electric machines, from current signals spectra only, using an inherently interpretative model, namely decision trees. The proposed approach has been tested in a highly challenging scenario, where the generalization capabilities of the classifier were assessed across different radial and torsional load conditions using realistic faults reproduced in laboratory. This setting is under-considered in the literature. A broad experimental evaluation, carried out under different load conditions, confirms the suitability of the methodology. In the binary detection problem, the proposed approach achieves an average precision and recall of 89.5% and 88.34%, respectively, in this challenging cross-load setting. With a post-processing, the performance further improves to 96.94% and 93.38%. In the recognition problem (3 classes), the average accuracy score attained is approximately 70%. Overall, the proposed method is shown to be able not only to detect and recognize faults with a very high accuracy, but also to provide explanations of the performed classifications that can be easily interpreted by humans.

REFERENCES

- [1] A. Garcia-Perez, R. D. J. Romero-Troncoso, E. Cabal-Yepez, and R. A. Osornio-Rios, "The application of high-resolution spectral analysis for identifying multiple combined faults in induction motors," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 2002–2010, May 2011.
- [2] A. H. Bonnett, "Root cause AC motor failure analysis with a focus on shaft failures," *IEEE Trans. Ind. Appl.*, vol. 36, no. 5, pp. 1435–1448, Sep. 2000.
- [3] L. Frosini and E. Bassi, "Stator current and motor efficiency as indicators for different types of bearing faults in induction motors," *IEEE Trans. Ind. Electron.*, vol. 57, no. 1, pp. 244–251, Jan. 2010.
- [4] J. R. Stack, T. G. Habetler, and R. G. Harley, "Fault-signature modeling and detection of inner-race bearing faults," *IEEE Trans. Ind. Appl.*, vol. 42, no. 1, pp. 61–68, Jan. 2006.
- [5] *Interstate Council for Standardization, Metrology and Certification*, Standard ISO 10816-1: Eval. Mach. Vibrat, 1995.
- [6] R. B. Randall, *Vibration-based Condition Monitoring*. Hoboken, NJ, USA: Wiley, 2011.
- [7] I. P. Georgakopoulos, E. D. Mitronikas, and A. N. Safacas, "Detection of induction motor faults in inverter drives using inverter input current analysis," *IEEE Trans. Ind. Electron.*, vol. 58, no. 9, pp. 4365–4373, Sep. 2011.
- [8] T. A. Shifat and J. W. Hur, "An effective stator fault diagnosis framework of BLDC motor based on vibration and current signals," *IEEE Access*, vol. 8, pp. 106968–106981, 2020.
- [9] S. Choi, B. Akin, M. M. Rahimian, and H. A. Toliyat, "Performance-oriented electric motors diagnostics in modern energy conversion systems," *IEEE Trans. Ind. Electron.*, vol. 59, no. 2, pp. 1266–1277, Feb. 2012.
- [10] S. Ganesan, P. W. David, P. K. Balachandran, and D. Samithas, "Intelligent starting current-based fault identification of an induction motor operating under various power quality issues," *Energies*, vol. 14, no. 2, p. 304, Jan. 2021.
- [11] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020.
- [12] F. B. Abid, M. Sallem, and A. Braham, "Optimized SWPT and decision tree for incipient bearing fault diagnosis," in *Proc. 19th Int. Conf. Sci. Techn. Autom. Control Comput. Eng. (STA)*, Mar. 2019, pp. 231–236.
- [13] R. Nishat Toma and J.-M. Kim, "Bearing fault classification of induction motors using discrete wavelet transform and ensemble machine learning algorithms," *Appl. Sci.*, vol. 10, no. 15, p. 5251, Jul. 2020.
- [14] K. Yatsugi, S. E. Pandarakone, Y. Mizuno, and H. Nakamura, "Common diagnosis approach to three-class induction motor faults using stator current feature and support vector machine," *IEEE Access*, vol. 11, pp. 24945–24952, 2023.
- [15] P. Gangsar and R. Tiwari, "A support vector machine based fault diagnostics of induction motors for practical situation of multi-sensor limited data case," *Measurement*, vol. 135, pp. 694–711, Mar. 2019.
- [16] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.
- [17] C. E. Sunal, V. Dyo, and V. Velisavljevic, "Review of machine learning based fault detection for centrifugal pump induction motors," *IEEE Access*, vol. 10, pp. 71344–71355, 2022.
- [18] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019.
- [19] J. Mao, F. Chen, B. Jiang, and L. Wang, "Composite fault diagnosis of rotor broken bar and air gap eccentricity based on park vector module and decision tree algorithm," in *Proc. CAA Symp. Fault Detection, Supervision Saf. Tech. Processes (SAFEPROCESS)*, Jul. 2019, pp. 701–706.
- [20] M. Seera, C. P. Lim, D. Ishak, and H. Singh, "Fault detection and diagnosis of induction motors using motor current signature analysis and a hybrid FMM-CART model," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 1, pp. 97–108, Jan. 2012.
- [21] R. R. Kumar, M. Andriollo, G. Cirrincione, M. Cirrincione, and A. Tortella, "A comprehensive review of conventional and intelligence-based approaches for the fault diagnosis and condition monitoring of induction motors," *Energies*, vol. 15, no. 23, p. 8938, Nov. 2022.
- [22] M. Blodt, P. Granjon, B. Raison, and G. Rostaing, "Models for bearing damage detection in induction motors using stator current monitoring," *IEEE Trans. Ind. Electron.*, vol. 55, no. 4, pp. 1813–1822, Apr. 2008.
- [23] F. Immovilli, A. Bellini, R. Rubini, and C. Tassoni, "Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison," *IEEE Trans. Ind. Appl.*, vol. 46, no. 4, pp. 1350–1359, Jul. 2010.
- [24] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

²Feature importance was computed with the `scikit-learn` library.

- [25] F. Immovilli and M. Cocconcelli, "Experimental investigation of shaft radial load effect on bearing fault signatures detection," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 2721–2729, May 2017.
- [26] F. Immovilli, M. Lippi, and M. Cocconcelli, "Automated bearing fault detection via long short-term memory networks," in *Proc. IEEE 12th Int. Symp. Diag. Electr. Mach., Power Electron. Drives (SDEMPED)*, Aug. 2019, pp. 452–458.



GIOVANNI BRIGLIA received the B.S. and M.S. degrees in mechatronic engineering from the University of Modena and Reggio Emilia, Italy, in 2021 and 2023, respectively. His research interests include condition monitoring, reinforcement learning, causal inference, multi-agent systems, and in general artificial intelligence techniques for industrial applications.



FABIO IMMOVILLI (Senior Member, IEEE) received the M.S. and Ph.D. degrees in mechatronic engineering from the University of Modena and Reggio Emilia, Italy, in 2006 and 2011, respectively. Since November 2019, he has been an Associate Professor in electric converters, machines and drives with the Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia. He is the author or coauthor of more than 55 technical articles. His research interests include electric machine condition monitoring, electric power converters, machines for energy conversion from renewable energy sources, and thermoacoustics.



MARCO COCCONCELLI received the M.S. degree in mechanical engineering and the Ph.D. degree in applied mechanics from the University of Bologna, Bologna, Italy, in 2003 and 2007, respectively. In 2007, he joined the University of Modena and Reggio Emilia, Reggio Emilia, Italy, where he is currently an Associate Professor in applied mechanics with the Department of Sciences and Methods for Engineering. His research interests include diagnostics of bearing and gear, condition monitoring of machineries and history of mechanisms, and machine science. He is a member of the Italian Branch of the International Federation for the Promotion of Mechanism and Machine Science (IFTOMM-Italy).



MARCO LIPPI received the Ph.D. degree in computer and automation engineering from the University of Florence, in 2010. He is currently an Associate Professor with the Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, Italy. His work focuses on machine learning and artificial intelligence, with applications to several areas, including condition monitoring, natural language processing, legal informatics, transportation, and medicine. He received the "E. Caianiello" Prize for the Best Italian Ph.D. Thesis in the field of Neural Networks, in 2012.

...