ISA Transactions xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatrans

Practice article

Hybrid algorithmic approach oriented to incipient rotor fault diagnosis on induction motors

Ignacio Martin-Diaz[a,b], Daniel Morinigo-Sotelo[a], Oscar Duque-Perez[a], Roque A. Osornio-Rios[c], Rene J. Romero-Troncoso[c,*]

[a] Elect. Eng. Dept. Escuela de Ingenierias Industriales, Sede Paseo del Cauce, University of Valladolid, Paseo del Cauce, 59, 47011, Valladolid, Spain
[b] HSPdigital-CATelematica, DICIS, Universidad de Guanajuato, Carr. Salamanca-Valle km 3.5+1.8, Palo Blanco, 36885, Salamanca, Guanajuato, Mexico
[c] Universidad Autónoma de Querétaro, HSPdigital CA-Mecatronica, Facultad de Ingenieria, Río Moctezuma 249, San Juan del Rio, 76806, Querétaro, Mexico

ARTICLE INFO

Keywords:
Induction motor
Fault diagnosis
Artificial intelligence
Simulated annealing algorithm
Oblique random forests

ABSTRACT

This paper investigates the current monitoring for effective fault diagnosis in induction motor (IM) by using random forest (RF) algorithms. A rotor bar breakage of IM does not derive in a catastrophic fault but its timely detection can avoid catastrophic consequences in the stator or prevent malfunctioning of those applications in which this sort of fault is the primary concern. Current-based fault signatures depend enormously on the IM power source and in the load connected to the motor. Hence, homogeneous sets of current signals were acquired through multiple experiments at particular loading torques and IM feedings from an experimental test bench in which incipient rotor severities were considered. Understanding the importance of each fault signature in relation to its diagnosis performance is an interesting matter. To this end, we propose a hybrid approach based on Simulated Annealing algorithm to conduct a global search over the computed feature set for feature selection purposes, which reduce the computational requirements of the diagnosis tool. Then, a novel Oblique RF classifier is used to build multivariate trees, which explicitly learn optimal split directions at internal nodes through penalized Ridge regression. This algorithm has been compared with other state-of-the-art classifiers through careful evaluation of performance measures not encountered in this field.

1. Introduction

Induction motors (IMs) predominate in industry due to their low manufacturing costs, power-weight ratio and robustness [1,2]. Copper rotor failures can develop into catastrophic breaking due to bar bending or material projection. Early diagnosis of these faults is a primary concern and will guarantee the operation of motors, safeguarding their integrity [3,4].

Rotor fault detection can be achieved monitoring different signals [2]: vibration [5,7], sound [7], acoustic emission [2], temperature [8], air gap magnetic flux [9], instantaneous power [10,11], supply voltage [3] and stator current [1,12]. However, motor current signature analysis (MCSA) has been preferred [13,14] because it is non-invasive, low-cost, and easy to measure. Besides, the motor shutdown is not required.

Many works have proposed fault identification techniques based on the analysis of the stator current. Authors in Ref. [15] improved the Park's Vector (PV) approach for detecting a broken rotor bar (BRB). They tracked the higher harmonic index after the application of elliptic and notch filters on the PV components. In Ref. [16], the authors analysed the zero sequence current (ZSC) with MUSIC to increase the BRB detection reliability. Authors in Ref. [17] tested the ZSC spectrum for detecting rotor asymmetries. The results are promising, but the motor must have a delta connection or the neutral-connected. Although it requires three current sensors, it shows advantages that complement other methods. Samanta et al. [18] used an extended-Kalman based signal conditioner to remove the fundamental component of the stator current signal. The authors claim that their system is fast, accurate and can be implemented online. Bessam et al. [19] presented a NN approach where the Hilbert Transform is used for the diagnosis of BRBs at low load. Authors in Ref. [20] proposed an analytical equation that relates

Abbreviations: IM, induction motor; CBM, condition based monitoring; MCSA, motor current signature analysis; BRB, broken rotor bar; ANN, artificial neural network; DT, decision tree; PCA, principal component analysis; RF, random forests; SA, simulated annealing; ORF, oblique RF; KNN, k-Nearest Neighbours; FFT, fast Fourier transform; PM, performance metric; AUC, area under an ROC curve; ROC, receiver operating characteristic; OOB, out of bag; OVA, one versus all; TPR, true positive rate; FPR, false positive rate

* Corresponding author.
E-mail address: troncoso@hspdigital.org (R.J. Romero-Troncoso).

https://doi.org/10.1016/j.isatra.2018.07.033
Received 6 November 2017; Received in revised form 3 April 2018; Accepted 24 July 2018

the number of BRB to the IM stator vibrations. Abd-el-Malek et al. [21] presented a technique based on the statistical analysis of the stator current envelope for detecting the exact fault location. A simulation under two operating modes validates the method, which otherwise lacks experimental verification. Interesting improvements of actual techniques were recently published, but the diagnosis of incipient faults based on statistical learning is not yet studied. New features, feature selection methods, classifier types, tuning and classi- cation methods continue emerging to enhance systems performance and diagnostic accuracy [12,22]. Thus, as suggested in Ref. [22] improvement of existing diagnostic methods and discovery of new fault indicators is a necessity. Sideband harmonics around the fundamental component of the stator current are the most used fault features [23]. Gyftakis et al. [17,24] propose thresholds for the fault detectability in line-fed IM. However, diagnosis of inverter-fed IM remains to be a topic of interest in the fault diagnosis community. Bruzzese [25] proposes new patterns for IM fed by non-sinusoidal power supplies that demonstrated to detect one or more BRBs.

Recent developments in machine learning algorithms permit to address the automatic diagnosis of IM faults [26]. The standard methodology consists of the following steps: (1) a collection of suited measurements related to the fault of interest; (2) extraction of fault signatures or features; (3) application of a feature reduction or feature selection technique; (4) construction of a classification model. In Ref. [27], the diagnosis is based on the information provided by the discrete wavelet transform. In Ref. [6], the authors used genetic algorithms to select the most significant features and optimize the artificial neural network (ANN) parameters. In Ref. [28], the diagnosis is performed by a multilayer perceptron ANN with statistical parameters as the inputs, whose dimensionality was reduced by Principal Component Analysis (PCA). Authors in Ref. [1] presented an intelligent system based on a combination of stationary wavelet packet transform and multiclass wavelet support vector machines. Researchers in Ref. [29] used an adaptive neuro-fuzzy inference system in combination with decision trees (DT), which permit to build explanatory rules to justify the predictions [30]. Although DT have a low bias, they usually suffer from high variance, which can be solved by combining the predictions of several randomized trees into a single model known as Random Forest (RF). The studies mentioned above suppose an important contribution, and our purpose is to improve some steps of the diagnosis methodology. Our work considers incipient rotor severities in IM fed from the line and inverter. The latter is an important point because this type of feeding introduces noise in the fault signatures, complicating the detection and the diagnosis.

In this paper, we collect and use all fault indicators presented in the recent literature, and for the fault classification, we propose a hybrid approach using the Simulated Annealing (SA) algorithm and the RF classifier. We use a promising version of the RF, known as Oblique RF (ORF). This version uses an oblique node splitting criteria (multivariate decision trees) via Ridge regression that allows improving the classification performance in those cases where the features are correlated. In machine learning, the problem of over fitting is always present when the data available is small. For this reason, we evaluate the effectiveness of the proposed approach with up to 7 experiments, where additional metrics to those used in the literature (Sensitivity, Specificity, ROC) have been considered.

The major contributions of this article are threefold:

- We use all the fault indicators proposed in the recent literature. No previous works have used the indicators proposed in Ref. [25] for the diagnosis of incipient BRB. These indicators were successful at detecting 1, 2, 3 and four broken bars for non-sinusoidal IM supplies.
- We propose the use of the SA algorithm to identify the features with greater discriminant capacity in different experiments. No study has evaluated the discrimination ability of fault features and if that capacity is kept when the IM supply is changed. Unlike genetic algorithms, the optimum design procedures based on the SA are less time consuming, and the optimum solutions obtained may avoid local ones, being feasible both mathematically and practically.
- We use the RF and ORF to classify incipient rotor faults. Their random structure increases their generalization ability. ORF, through its oblique splitting, improves the classification performance in some experiments.

The parts that constitute the present work are organized as follows. Section 2 introduces the theory that justifies the fault signatures suitability. Then, Section 3 describes the algorithms that compose the proposed hybrid approach for diagnosing rotor faults. Experimental results are presented in Section 4 for a line-fed and inverter-fed IM under two different load levels. In this section, a comprehensive comparison between RF, ORF, classification and regression trees (CART) and KNN is shown. The purposed approach is compared with several state-of-the-art techniques found in the literature. Finally, Section 5 concludes this study.

2. Related work

This section presents the theory behind the fault signatures extraction to justify their suitability. In this work, the first set of features has two types of data calculated from the measured stator currents: signatures computed from the raw stator current in the time domain, and signatures obtained by spectral analysis of the same signals.

2.1. Time-domain fault signatures

A statistical analysis of the raw data (stationary period of the stator current) permitted to calculate fifteen fault-signatures in the time-domain. These signatures are described in Table 1.

2.2. Spectral fault signatures

As it is well known, a BRB fault occurs after the development of small cracks at the junction between the bar and the end ring [31]. Consequently, the resultant signatures for cracked and BRBs on the current spectrum are the effects due to rotor circuit asymmetries [32]

Table 1
Statistical features extracted from the stator current in time domain.

Time-domain features		
Feature	Symbol	Expression
First Moment	m_1	$\frac{1}{N}\sum x[n]$
Second Moment	m_2	$\frac{1}{N}\sum (x[n]-\bar{x}[n])^2$
Third Moment	m_3	$\frac{1}{N}\sum (x[n]-\bar{x}[n])^3$
4th Moment	m_4	$\frac{1}{N}\sum (x[n]-\bar{x}[n])^4$
Second Cumulant	c_2	$m_2 - m_1^2$
Third Cumulant	c_3	$m_3 - 3m_1m_2 + 2m_1^3$
4th Cumulant	c_4	$m_4 + 3m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4$
Skewness	*Skew*	$\frac{m_3}{(\sqrt{m_2})^3}$
Kurtosis	*Kurt*	$\frac{m_4}{(\sqrt{m_2})^4}$
Absolute mean	AM	$\lvert\bar{x}\rvert$
Peak value	PV	$\frac{1}{2}(max(x[n]) - min(x[n]))$
Squared root value	SRV	$\left(\frac{1}{N}\sum\sqrt{\lvert x\rvert}\right)^2$
Crest factor	CF	PV/RMS
Shape factor	SF	$RMS/\lvert\bar{x}\rvert$
RMS value	RMS	$\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}[x[n]-\bar{x}]^2}$

producing multiple frequency components around the fundamental frequency given by (1).

$$f_{BRB} = (1 \pm 2ks)f_1, \tag{1}$$

where s is the slip, k is any integer, and f_1 is the fundamental frequency. Hence, Fast Fourier Transform (FFT) is applied to each stator current signal under stationary conditions, and six fault signatures are calculated.

The remaining four signatures considered in this paper for cage monitoring in the frequency domain were proposed in Ref. [25], and they are defined next, (2)–(5):

$$\Gamma_5 = \frac{I_{(7-2s)\omega}}{I_{5\omega}} \tag{2}$$

$$\Gamma_7 = \frac{I_{(5+2s)\omega}}{I_{7\omega}} \tag{3}$$

$$\Gamma_{11} = \frac{I_{(13-2s)\omega}}{I_{11\omega}} \tag{4}$$

$$\Gamma_{13} = \frac{I_{(11+2s)\omega}}{I_{13\omega}} \tag{5}$$

These indicators are suitable because they are independent of load-torque oscillations, and they were useful in diagnosing IM fed from non-sinusoidal voltage sources [25]. We will use the fault features mentioned above to identify incipient fault severities, previous to a full BRB. No previous studies have used this data in the diagnosis of BRB.

3. Methodology

This work proposes a hybrid approach for diagnosing rotor faults in IM that consists of a feature selection process and a classification stage. The goal of the first step is to find the subset of features with the most fault discriminatory capability using a Simulated Annealing algorithm. This group of fault patterns is the input to the RF-based fault classifier in the second step of the process. An Oblique RF (ORF) ensemble is built from multivariate trees to learn optimal directions at internal nodes through regularization.

3.1. Feature selection via SA

The main objective of the feature selection stage is to seek the combination of features that provide an optimal prediction of the target variable with a reasonable computational effort. The problem of feature selection is inherently a complex optimization problem belonging to the class of NP-hard problems. Therefore, this selection can be posed as an optimization problem where a “goodness of fit measure” or a performance measure (*PM*) is maximized or minimized. In the fault diagnosis context, metrics such as Cohen's Kappa, Accuracy, and Area Under a ROC curve (AUC) have been suggested [33], which means that features that maximize one of them are labelled as candidates to be used in the classification stage. Feature selection procedures also help to reduce computational time and take care of the curse of dimensionality, a problem that arises in the design of supervised classifications.

Depending on the desired complexity, feature selection can be carried out using *wrapper* or *filter* methods [34]. The latter perform feature selection based on the properties of the dataset itself. The *wrapper* methods generate feature subsets by searching and evaluating using a classification algorithm. Due to the computational time and the feature space complexity, the exhaustive enumeration is not an option to search for the best feature subset. Many methods are available to reduce the search space [34,35], as the branch and bound technique, the sequential forward and backward search and the min-max approach. Soft computing techniques, such as ANNs or evolutionary algorithms including genetic algorithms [36], can also be used.

Another alternative is the use of a class of optimizers called direct methods, which do not use derivatives to find the settings with optimal compressive strength and evaluate the prediction equation many more times than derivative based optimizers. One of these procedures is the SA algorithm [37]. In this paper, this technique has been used to perform the feature selection stage. The SA algorithm, named as an analogy to the process of metal cooling [37], is a local search meta-heuristic that can be applied to discrete optimization problems. This method avoids local optima easily by allowing hill-climbing moves towards worse objective function values in search for a global optimum. By accepting these worse objective function values, the algorithm examines other areas of the search space and therefore is less greedy.

Algorithm 1 describes the feature selection process using the SA technique. An initial subset of features is selected and used to estimate the performance of the model (denoted here as PM_1). *PM* is the performance metric to maximize, and T is a value called temperature that changes over iterations. The current feature subset is slightly changed, and another model is created with an estimated performance measure value of PM_2. If the new model improves the previous one (i.e., $PM_2 > PM_1$), the new feature set is accepted. However, if this model is worse than the previous one, it may still be accepted based on a probability $P_a{}^i$, where i is the iteration number. This probability is configured to decrease over time, so the acceptance of a suboptimal configuration becomes unlikely as i increases. This probability of accepting a worse features subset permits the exploration of the search space and the avoidance of local optima. The process continues for some pre-specified number of iterations, and the best feature subset of all the iterations is used.

Algorithm 1
Simulated Annealing (SA) pseudocode for feature selection.

Input: Training data: $\mathbf{D} = \{(x_1,y_1),(x_j,y_j), \ldots,(x_n,y_n)\}$
Output: Best configuration PM^* found by the algorithm
Start with an initial random generation of features
for iterations $i = l$ to T ***do***
  Alter randomly the best current feature set
  Data Standardization
  Tune the model using this current feature set
  Estimate the performance of the tuned model (PM_i)
  if $PM_i > PM_{best}$ ***then***
    Update the best current feature set
    Set $PM_{best} = PM_i$
  else
    Compute the probability of accepting the current feature set:
      $P_i^a = exp\left(\frac{PM_{best} - PM_i}{T}\right)$
    Generation of a random number R in the range [0,1]
    if $P_i^a \leq R$ ***then***
      Update the best current feature set
      Set $PM_{best} = PM_i$
    else
      Retain the best current feature set
    end if
  end if
end for
Choose those features related to the greatest PM_i throughout all iterations.
Output the best con guration PM^*.

3.2. Random forests

Random Forest was introduced by Breiman in Ref. [38] and has been used effectively as an algorithm for classifier induction. Most of

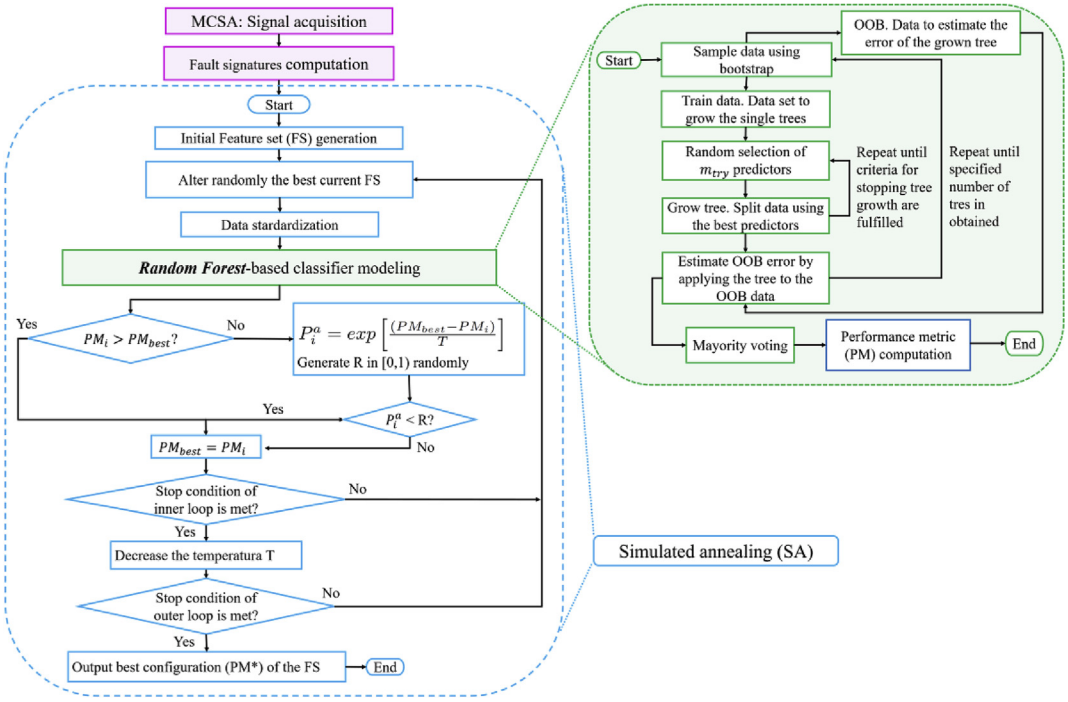

Fig. 1. Flowchart of the hybrid approach.

the applications in literature were built by univariate decision trees such as CART or C4.5. RF structure increases the expected diversity among the individual classifiers since tree-based classifiers are unstable and small changes in the training set results in considerable changes in the classifier predictions. The general structure of RF is shown in Fig. 1. This is a user-friendly algorithm since it only has two tuning parameters: the number of trees in the forest (n_{tree}) and the number of randomly sampled variables as candidates at each split (m_{try}). RF uses a bagging procedure along with a random feature selection. The main steps to construct a random forest classifier are the following:

1. Given a training set $T = \{(x_1, y_1), ..., (x_N, y_N)\}$ with N samples $x_i = (x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(p)})$ and being their corresponding class labels $y_i \in \{C_1, ..., C_C\}$ for $i = 1,2, ..., N$; draw n_{tree} samples from such training set via bootstraping.
2. From each bootstrap sample build a univariate classification tree t algorithm as it is explained next. Choose $m_{try} \leq p$ features randomly at each node from the original set of features. Then, use the *Gini Index* to compute the optimal univariate orthogonal split for the current step, only with this subset of features. Fix the value of m_{try} during the forest construction and grow the trees to the leaf nodes without performing pruning.
3. Predict unseen data from the test set by majority voting from the n_{tree} trees as shown in Fig. 1.

Note that, the random feature selection of the second step contributes to the reduction of the feature space dimensionality and improves RF accuracy over bagging alone significantly [38]. The essential feature of this algorithm is that the optimal thresholds for the random selection of single features in every split correspond to mutually orthogonal one-dimensional hyperplanes [39].

3.3. Oblique Random Forests

Random Forest algorithms consist of a group of univariate decision trees. Other flavours are based on multivariate decision trees, like oblique decision trees, which are also useful for classification problems. Whereas CART and other standard univariate classification trees split the data using just a single feature at the same time, oblique decision trees use the information of multiple variables instead. For instance, in rotation forest [40], the oblique split directions are computed from the principal components of feature subsets of the training data. This algorithm was reported as a good alternative for improving classification tasks. The use of supervised approaches to defining optima split directions is a distinguishing feature of ORF [39]. This technique exploits the benefits of splitting the feature space at a node through a wide range of linear discriminative node models, all of them employing different optimization objectives.

In this work, the Ridge regression method was chosen, due to the properties of regularized node models, to improve the classification results regarding those built with the original version of RF.

The ensemble generation process in ORF is similar to the previously explained for RF. The only difference lies in the construction of the oblique decision trees.

3.3.1. Oblique decision trees

Consider an $(n \times p)$-dimensional data matrix $\mathbf{X}$, defined by the number of samples n and by the number of features p. The variable Y is generally referred to as the response variable and belongs to class $y = 1$, indicating the fault state, or to class $y = -1$, indicating healthy state. ORF uses all the selected m_{try} variables to learn the optimal split direction using a supervised model. Ridge regression aims to improve determination of the regression coefficients and, as expressed in (6), puts further constraints on the β_j's parameters in the linear model to control the variance among highly correlated features by imposing a penalty on the coefficients.

$$PRSS(\lambda) = \underset{\boldsymbol{\beta}}{argmin} + \lambda \tag{6}$$

where λ controls the shrinkage of the regression coefficients and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ is the regression parameter. The parameter β_j, $j = 1, \ldots, p$, represents the effect size of covariate j on the response. By convention, $\mathbf{X}$ is assumed to be standardized and y is assumed to be centered.

In this case, the expression has a penalty term on the β's instead of just minimizing the residual sum of squares. This penalty term λ, which is a pre-chosen constant, multiplies the squared norm of the β. This

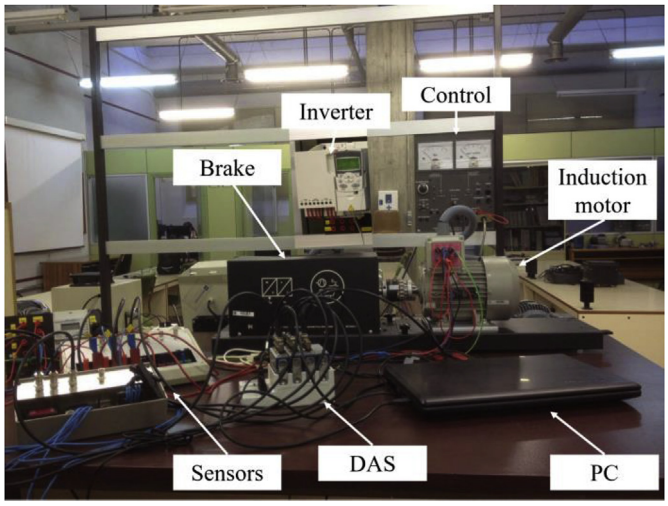

Fig. 2. Experimental setup.

means that if the β's take on large values, the optimization function is penalized. Smaller β's, or β's that are close to zero, would be preferred to drive the penalty term small. Regarding this node choice, the model is optimized for

$$\hat{\beta}_{ridge}(\lambda) = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}_P)^{-1}\mathbf{X}^T\mathbf{y} \tag{7}$$

The regularization parameter λ permits to adapt the classifier to an optimal split direction β_{ridge} in between two extremes [39]. When $\lambda = 0$, β_{ridge} may indicate maximal correlation with class or response values, similar to Linear Discriminant Analysis. $\lambda > > 1$ indicates highest data variation, similar to PCA. At each node, the model parameter λ can be adapted to the out-of-bag (OOB) samples available at the node in question [39].

Consider a sample $\mathbf{x} = [x_1 \ldots x_{mtry}]^T$ in an m_{try}-dimensional feature space, the oblique decision tree r at a particular node m results in a non-orthogonal hyperplane at each decision node as follows:

$$r_m(\mathbf{x})\text{: } \beta_m^T\mathbf{x} > \boldsymbol{c_m} \tag{8}$$

with the β_m coefficients representing the projection for the split for a threshold c_m. All samples $\mathbf{X}$ are projected into β_m and the optimal split-point c_m is identified using the *Gini Index* [39]. The threshold c_m maximizing the decrease in the *Gini Index* is chosen and samples are separated accordingly. The best split between both subsets is found iterating recursively until both classes are separated [39].

3.3.2. ORF tuning

As mentioned above, an RF algorithm is formed by a large combined ensemble of trees. Remember that the tuning of the two hyper-parameters is required to control the generation of DT in the ensemble. One parameter is the subspace dimensionality m_{try}, which determines the number of features sampled randomly at each node and the randomization of the classifier construction. The variability of the trees in the forest is characterized by m_{try}, which needs to be chosen to ensure a low correlation between individual trees. The other parameter is the ensemble size from which the final label is assigned to the leaf node. The ensemble decision can be pooled regarding different strategies [38]. In this work, a normalized number of votes described by a probability $p \in [0,1]$ is chosen due to its robustness against over-fitting [39]. Trees are built as illustrated in Algorithm 2.

Algorithm 2
Pseudocode for Oblique Random Forest (ORF).

Input: Training data: $\mathbf{D} = \{(x_1,y_1),(x_j,y_j), \ldots,(x_n,y_n)\}$
Output: Classification result.
1: **for** $b = 1$ to B do
2: Draw a bootstrap sample Z_b of size N (with replacement) from the training data and grow a tree from this bootstrap sample
3: Build a RF tree T to the bootstrap data through recursive repetition of the following steps for each leaf node of the tree, until node size $n_{\min}$ is reached.
1. Select m features at random from the s selected features.
2. Find the best split direction β in the subspace spanned by these features according to the predefine node model. Choose the best s feature.
3. Tune the hyperparameters λ on the out-of-bag samples available at that node.
4. Split the node into two daughter nodes through the fitted values used to identify the threshold c.
5. Grow the tree to full depth without pruning.
4: **end for**
5: Output the ensemble of trees: $\{T_b\}_1^B$
6 Predict a new point x:
$\hat{C}_{RF}^B(x) = majority\ vote\left\{\hat{C}_1^B\right\}$ being $\hat{C_b}(x)$ the class prediction of the b-th RF tree.

4. Experimental results

After introducing the theoretical concepts of the proposed approach, a condition monitoring task of an IM using real data from experimental tests was conducted. The proposed automatic diagnosis of rotor faults is tested with a dataset of experimental currents belonging to different rotor states. Fig. 2 shows the experimental test bench. The star-connected IM is fed from the line and an ABB inverter, model ACS355, which is programmed with a V/f linear control. The motor load is a magnetic powder brake that permits to set two load levels: medium (LL1; *slip* ≈ 0.03) and high (LL2; *slip* ≈ 0.05). The stator current is measured by a Hall effect current transducer by LEM. The operating frequency is 50 Hz for both power supplies with motor functioning at steady state. The data is acquired with an NI 9215 module from National Instruments and an NI cDAQ-9174 base platform with a sampling frequency of 50 kHz and a sampling time of 10 s.

The different bar damage degrees are simulated by drilling a hole in one of the rotor bars as shown in Fig. 3. Five rotor condition states are

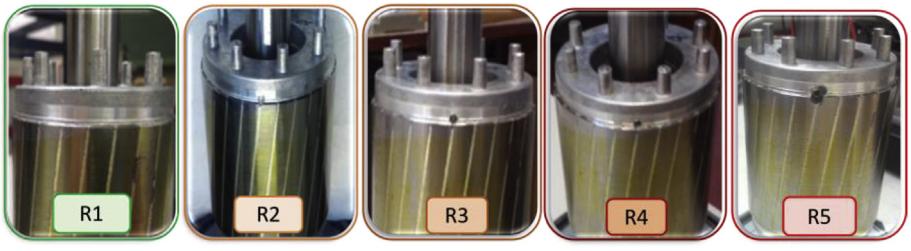

Fig. 3. Different rotor conditions.

considered. The rotor is healthy (condition R1) in the first tests. Next, a 4.2 mm depth hole is drilled in one of the rotor bars, and this is the incipient faulty rotor state (R2). The third rotor condition (R3) is achieved with a 9.4 mm depth hole. The fourth rotor condition (R4) is more severe and is created by drilling completely the same rotor bar. Finally, the fifth rotor condition (R5) is produced by widening the hole diameter to 3.5 mm. 300 signals were collected at those indicated operating settings.

All these signals in the time domain were processed with the MATLAB® software to extract the fault signatures introduced in Section 2. The result of this feature extraction process was a matrix with 300 rows (observations), and 21 columns (features) that is the input to the hybrid approach. The pre- and post-data processing were performed under the free software environment for statistical computing known as R [41].

4.1. Application of the hybrid approach for rotor fault diagnosis

The hybridization of the two heterogeneous algorithmic techniques involves three steps. In the first stage, the SA is applied for selecting the features. The initial data is divided into two sets: the training set contains two-thirds of that data, and the testing set the remaining one-third. Algorithm 1 is run using 5-repeated ten-fold cross-validation and within a *wrapper* setting where a proper performance metric (PM) is optimized, i.e., the overall accuracy and the unweighted Kappa statistic in this case. The second stage aims to tune RF and ORF classifiers. The data is formed by the features selected in the first stage and according to the experiments described in Table 2. In the final third stage, the classifiers are evaluated with the testing set, and several performance metrics are computed to check which class is better classified. In addition, these results are compared with the CART and KNN classifiers since they are widely used techniques in the literature. The implementation used in this work is based on the following R packages: “caret” [33], “randomForest” [42] and “obliqueRF” [43], which are available at the CRAN repository.

To find the best model a grid-search through the algorithm parameters has been carried out. For the RF-based classifiers, the parameter grid has been established as follows: $m_{try} \in [[\sqrt{p}/2], p]$, p being the number of condition features and $\sqrt{p}/2$ the nearest integer function, and the number of trees in the ensemble $n_{trees} \in [1500]$. $n_{trees} = 500$ and $m_{try} = 3$ were the best parameter settings for each model and experiment within the exhaustive search, for both RF and ORF. For the KNN algorithm, the number of neighbours for tuning purposes was 1, 3, 5, 7, and 9. The Minkowski distance was used. In the next section, the feature selection results by SA algorithm are shown. Then, the classifier performance evaluation considering a one versus all (OVA) approach is developed. Seven experiments have been designed to analyse the hybrid approach behaviour under different conditions (see Table 2).

4.2. Results of the feature selection through SA

Fig. 4 and 5 illustrate the optimization process. Fig. 4 shows the values of the performance metrics considered during the iterative procedure. They increase monotonically as the prediction ability improves. Fig. 5 presents the error evolution, including the OOB-error in the training process, as the number of trees increases. In Fig. 4, the best mean value of the performance metrics is achieved at the 121 iteration: Accuracy has a mean value of 0.9780 and Kappa a value of 0.9725. Table 2 shows the fault features selected and ranked (from left to right) by the SA algorithm. Experiments 1, 2, and 3 correspond to the motor fed from the line. Experiment 1, when the motor load is lower, shows the highest number of features. The number of selected features is lower as the motor load level increases (experiment 2). However, the number of features is reduced when the two load levels are considered together (experiment 3). Experiments 4, 5, and 6 correspond to the inverter-fed motor. The SA algorithm selects the same number of features for the two load levels (experiments 4 and 5), but only the 50% of them match. Unlike experiment 3, the features selected in experiment 6 (that also considers both load levels together) have increased. These results indicate that a more homogeneous data, regarding the motor load level, is necessary when the motor is fed by the inverter. This contrasts with the line-fed motor case. Having data corresponding to several load levels is not as critical as in the case of the inverter-fed IM, where a higher number of features is required. To summarize, the most selected features in the experiments considered are: Γ_5, Third Cumulant (c3), Kurtosis (appearing in 6 experiments), LSH, RMS (appearing in 5 experiments), Second Cumulant (c2), Shape Factor (SF), and Skewness (appearing in 4 experiments). Most of them are time-domain features.

4.3. Ridge Regression tuning in Oblique Random Forest

In Fig. 6, we can visualize the coefficients of the Ridge regression for ORF classifier corresponding to the experiment 7 without SA feature selection. Each curve corresponds to a variable or feature. It shows the path of its coefficient against the logarithmic values of λ. The candidate grid values of λ are those in the range [10e-5, 10e5] in steps of 0.1.

On the other hand, in Fig. 7 we can see the confidence intervals that represent error estimates for the misclassification error (red dots). The lowest point in the curve indicates the optimal lambda (λ = 5.0119e-5) i.e. the logarithmic value of λ that best minimized the error in the OOB samples and controls the overall strength of the penalty. The axis above both Figs. 6 and 7 indicates the number of non-zero coefficients at the current λ.

4.4. Classification results for an one-versus-all (OVA) scenario

Table 3 presents a comparison of performance results between the Random Forest (RF) and the Oblique-Random Forest (ORF) methods, with and without a feature selection with the SA algorithm. Accuracy, Specificity, and Sensitivity [44] are the performance metrics chosen to analyse the classifier effectiveness. It is noticeable that the classification performance is sensitive to the motor load level in the case of line-fed IM. However, RF achieves better results compared to ORF when the IM is fed with the ABB inverter. The classification results are better without the feature selection with the SA algorithm in experiments 1 and 2. The ORF achieves the best results in experiment 7 that consider the data of all other six experiments together (two power supplies and two motor load levels). On the other hand, KNN achieves similar results regarding

Table 2
Description of the features selected by the SA-ORF algorithm according to the feeding and load conditions analysed.

# Experiment	Type of feeding	Load level	Number of features	Condition features
1	Line	LL1	14	LSH, RHS, Γ_5, Γ_7, Γ_{11}, m_4, c_1, c_3, c_4, *Skew*, *Kurt*, AM, PV, SF
2	Line	LL2	10	Γ_{11}, c_1, c_3, c_4, *Skew*, *Kurt*, SRV. CF. SF, RMS
3	Line	Both levels	7	Γ_5, c_2, c_4, *Kurt*, PV,SF, RMS
4	ABB inverter	LL1	8	LSH, Γ_5, m_1, c_2, c_3, c_4, *Skew*, RMS
5	ABB inverter	LL2	8	LSH, Γ_5, Γ_{13}, c_2, c_3, *Kurt*, AM, SRV
6	ABB inverter	Both levels	11	LSH, Γ_5, Γ_7, Γ_{13}, m_1, c_2, c_3, *Skew*, *Kurt*, AM, RMS
7	Both feedings	Both levels	10	LSH, RHS, Γ_5, m_4, c_1, c_3, *Kurt*, AM, SF,RMS

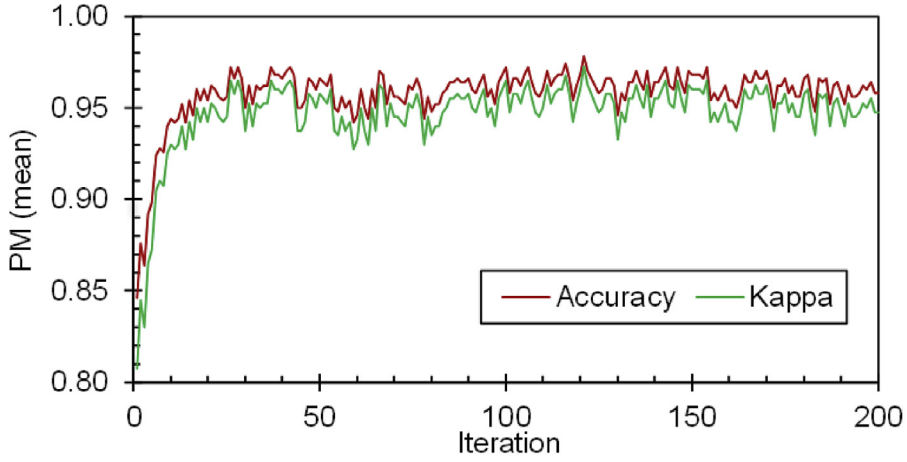

Fig. 4. Estimated Accuracy and Kappa versus SA iterations.

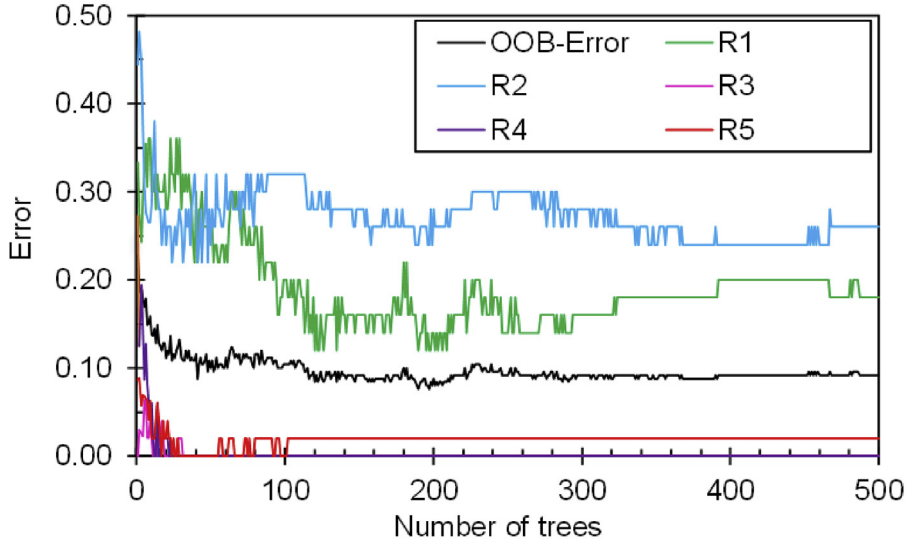

Fig. 5. Error evolution during the optimization process for feature selection using SA in a case of the IM fed by the inverter.

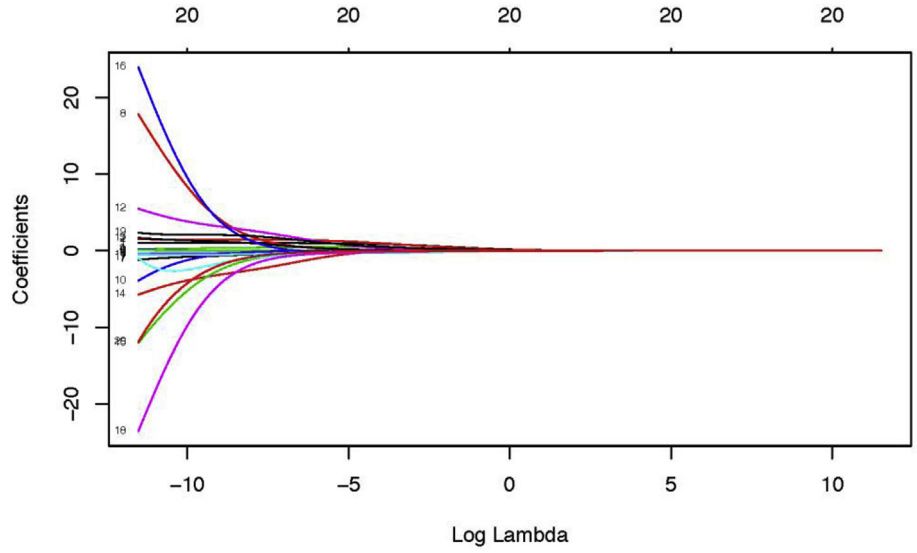

Fig. 6. Coefficients of the Ridge regression for ORF classifier corresponding to the experiment 7 without SA feature selection.

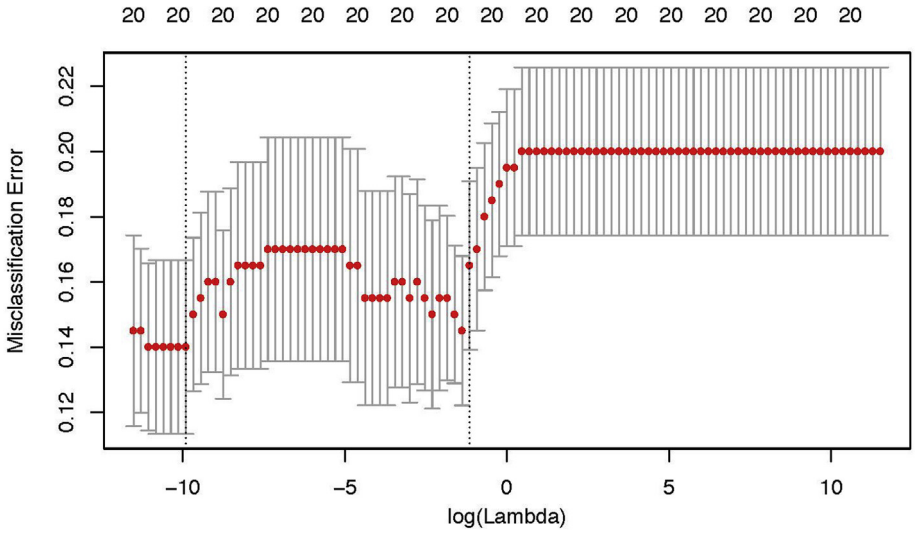

Fig. 7. Confidence intervals estimation for the misclassification error. The vertical dotted line shows the log value of lambda that best minimized the error in the OOB samples. The numbers across the top are the number of non-zero coefficient estimates.

Table 3
Performance results for RF and ORF with and without feature selection (FS) with SA. The comparative study handles several conditions of load and type of IM feeding.

# Experiment	Without Feature Selection						With SA Feature Selection					
	RF			ORF			RF			ORF		
	Acc.	Specif.	Sensit.	Acc.	Specif.	Sensit.	Acc.	Specif.	Sensit.	Acc.	Specif.	Sensit.
1	**1.00**	**1.00**	**1.00**	0.88	0.40	1.00	0.92	0.60	1.00	0.80	0.00	1.00
2	0.96	1.00	0.95	**1.00**	**1.00**	**1.00**	0.96	0.80	1.00	0.96	1.00	0.95
3	0.92	0.60	1.00	0.92	1.00	0.60	0.88	0.70	0.92	0.84	0.30	0.97
4	0.80	0.40	0.90	**0.88**	**0.40**	**1.00**	0.80	0.20	0.95	0.84	0.40	0.95
5	0.84	0.20	1.00	0.84	0.20	1.00	**0.88**	**0.60**	**0.95**	0.84	0.60	0.90
6	0.92	0.80	0.95	0.86	0.40	0.97	**0.98**	**0.90**	**1.00**	0.94	0.70	1.00
7	0.88	0.70	0.92	0.86	0.45	0.96	0.90	0.65	0.96	0.92	0.60	1.00

Table 4
Performance results for CART and k-NN with and without feature selection with SA. The comparative study handles several conditions of load and type of IM feeding.

# Experiment	Without Feature Selection						With SA Feature Selection					
	CART			KNN			CART			KNN		
	Acc.	Specif.	Sensit.	Acc.	Specif.	Sensit.	Acc.	Specif.	Sensit.	Acc.	Specif.	Sensit.
1	0.82	0.92	0.40	0.88	1.00	0.85	0.92	0.71	1.00	**1.00**	**1.00**	**1.00**
2	0.92	0.90	1.00	**1.00**	**1.00**	**1.00**	0.84	0.71	0.89	0.88	0.90	0.80
3	0.90	0.60	0.97	0.96	0.95	1.00	0.96	1.00	0.95	0.84	0.90	0.60
4	0.68	0.20	0.80	0.88	1.00	0.40	0.88	0.40	1.00	0.88	0.40	1.00
5	0.80	1.00	0.75	0.80	0.40	0.90	0.80	1.00	0.75	0.68	0.00	0.85
6	0.84	0.90	0.82	0.82	0.20	0.97	0.76	0.60	0.80	0.82	0.92	0.40
7	0.92	0.65	0.99	0.84	0.90	0.60	0.90	0.75	0.94	0.85	0.89	0.75

Table 5
One-sided Binomial test with a significance level of 0.05 applied for RF ORF, CART and KNN with SA feature selection.

# Experiment	RF		ORF		CART		KNN	
	p-Value	Hypothesis	p-Value	Hypothesis	p-Value	Hypothesis	p-Value	Hypothesis
1	0.0982	Accepted	0.6167	Accepted	0.0038	Rejected	0.0038	Rejected
2	0.0274	Rejected	0.0274	Rejected	0.0982	Accepted	0.0000	Rejected
3	0.1034	Accepted	0.3073	Accepted	0.0013	Rejected	0.0038	Rejected
4	0.6167	Accepted	0.4207	Accepted	0.2340	Accepted	0.2340	Accepted
5	0.2340	Accepted	0.4207	Accepted	0.6167	Accepted	0.9532	Accepted
6	0.0002	Rejected	0.0056	Rejected	0.4437	Accepted	0.81396	Accepted
7	0.0057	Rejected	0.0008	Rejected	0.0008	Rejected	0.1285	Accepted

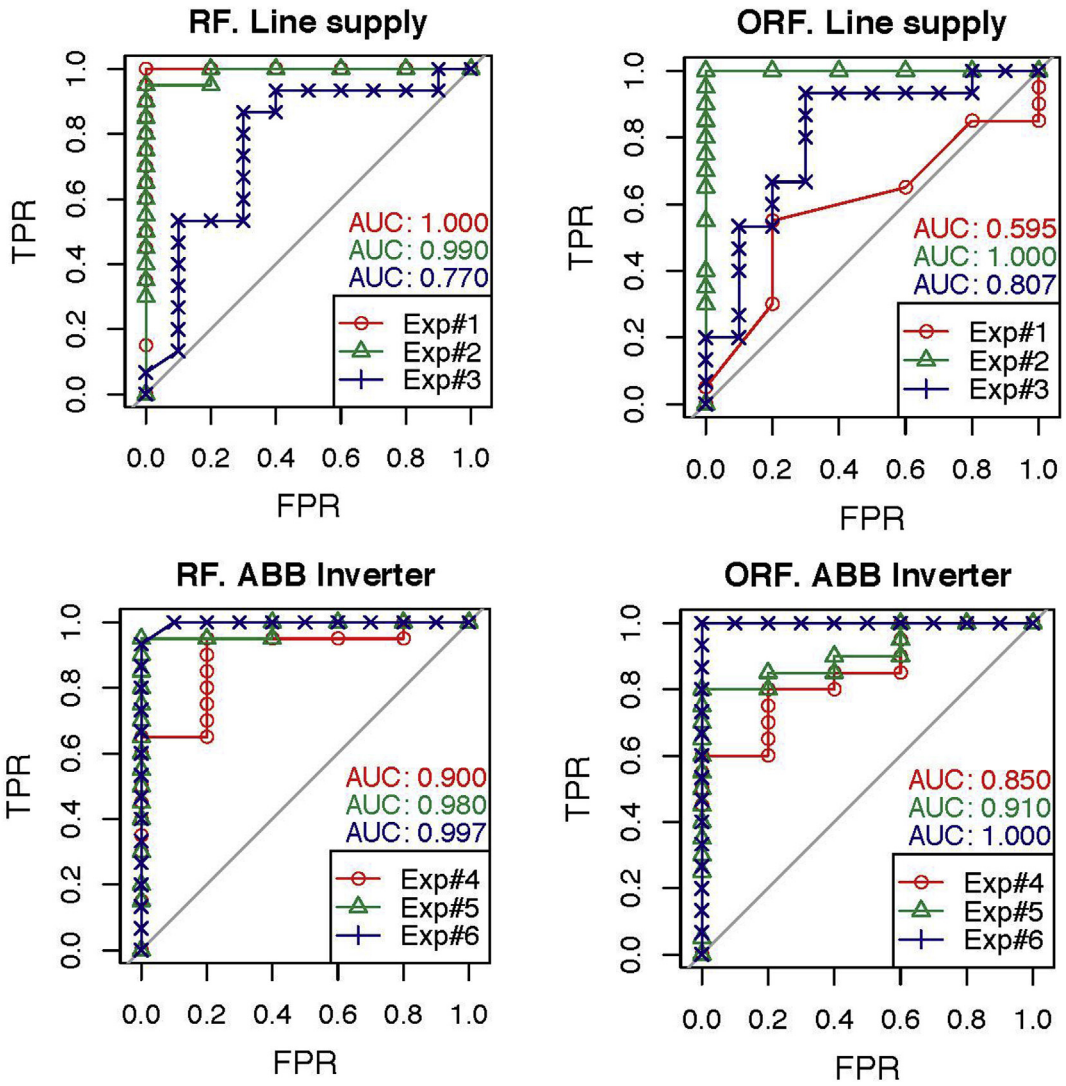

Fig. 8. ROC Curves grouped by classifier and by the feeding for RF and ORF. They are organized by rows according to the feeding: First row, line-fed IM; Second row, inverter-fed IM.

RF-based classifiers when the IM is fed directly from the line (see Table 4). Nevertheless, KNN decreases its performance when the IM is fed with the ABB inverter. For the case of CART with SA feature selection and IM line-fed (experiments from 1 to 3) the results are close to the previous ones. Nonetheless, CART without SA feature selection for this same case and for the rest of the cases worsen its results as it can be seen in Table 4.

Nonetheless, the data test set has the same number of samples for each class. This implies that when applying an OVA-type classification, i.e., R1 against the rest, the resultant binary problem become unbalanced. In this case, accuracy is not a representative metric, and it is advisable to perform a hypothesis test to know if the model's accuracy is better than the majority-class proportion in the data. The results of a binomial test, with a significance level of 0.05, are shown in Table 5. This test checks if accuracy is better than the no-information rate, i.e., the percentage of the data largest class. The interest of this one-sided test is to prove if the classification rate is or not better than the probability of selection by chance. In the experiments 1, 3, 4, and 5 (for RF and ORF), experiments 2, 4, 5, and 6 (for CART) and experiments 2, 4, 5, 6 and 7 (for KNN) the p-value is more significant than the 0.05 significance level (see Table 5), so the null hypothesis cannot be rejected. This means that there is no evidence to prove that the accuracy values are statistically significant for the performance of the classification.

Fortunately, there are other performance indicators, as the ROC curves, which can be used for the analysis of imbalanced scenarios [44]. ROC curves were designed as a general method that, given a collection of continuous data points, determine an effective threshold such that values above the threshold correspond to a specific class. The ROC curve is created by evaluating the class probabilities for the model across a continuum of thresholds. For each candidate threshold, the resulting true positive rate (TPR), or Sensitivity, is represented versus the false positive rate (FPR). An optimal classifier that separates completely the two classes would lay on the upper left-hand border of the TPR vs FPR plane. On the other hand, an ineffective classifier would result in a ROC curve close to the 45-degree line in the TPR vs FPR plane, giving an AUC of 0.5.

The first row of Figs. 8 and 9 shows a comparison of different ROC curves for the line-fed induction motor cases (experiments 1, 2, and 3). These curves agree with the results presented in Tables 3 and 4 Evident differences arise depending on the motor load level. As it can be seen in Fig. 8, the ORF classifier applied to the experiment 1 shows the poorest performance. It is not much better than random guessing. However, the RF classifier displays an optimal behaviour for the same experiment 1. In contrast, the ORF classifier outperforms the RF classifier in the experiments 2 and 3. The second row of Figs. 8 and 9 shows the

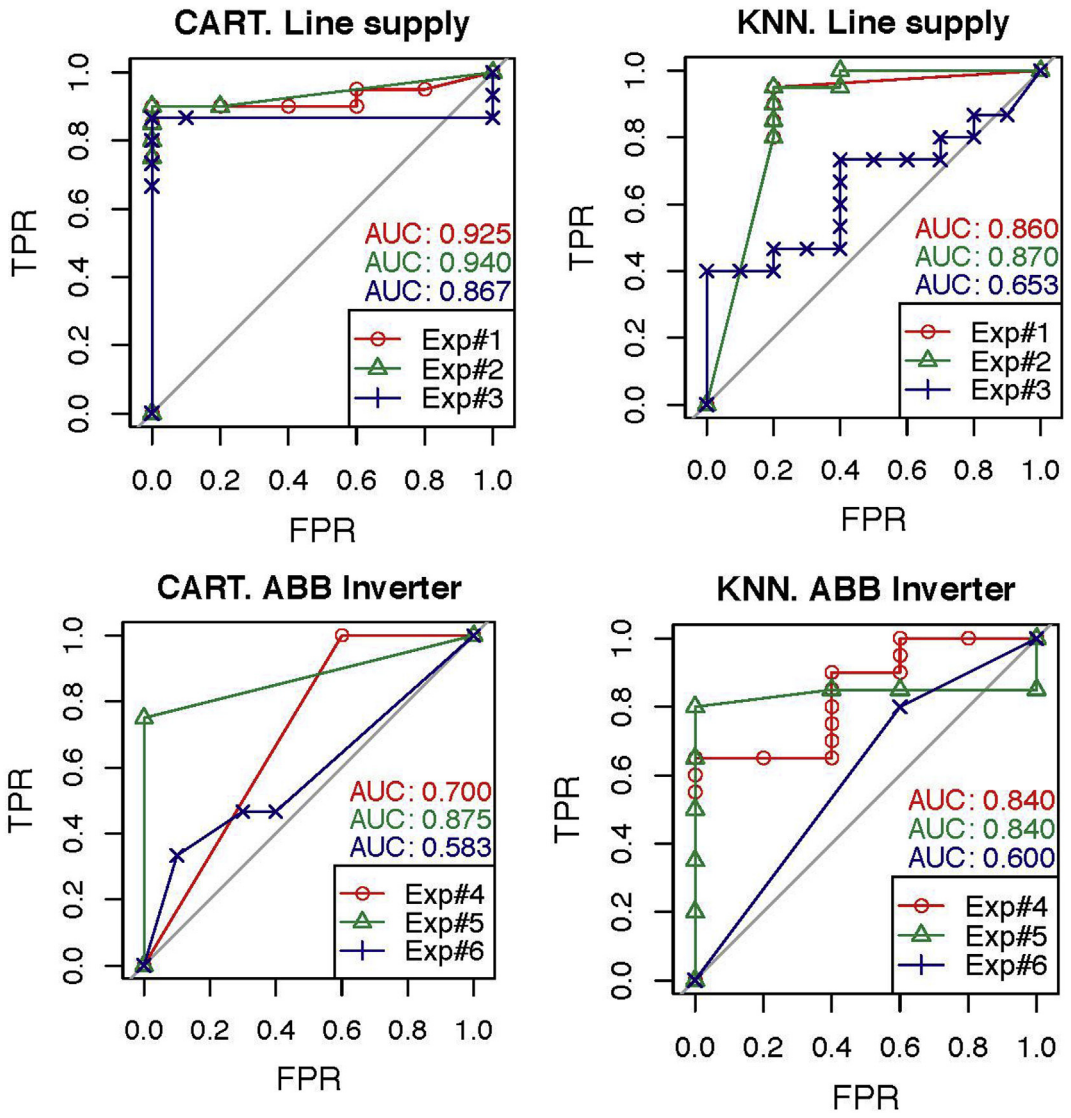

Fig. 9. ROC Curves grouped by classifier and by the feeding for CART and KNN. They are organized by rows according to the feeding: First row, line-fed IM; Second row, inverter-fed IM.

comparison of different ROC curves for the inverter-fed motor cases (experiments 4, 5, and 6). In Fig. 8, the RF classifier outperforms the ORF classifier in the experiments 4 and 5. In experiment 6, the ORF presents slightly better performance than the RF, being almost optimal for both cases. As can be seen in Fig. 9, CART achieved good results for the experiments related to the IM line feeding (experiments 1, 2, and 3). However, when the IM is fed through the inverter the results of CART worsened. The worst CART results are obtained for experiment 6, where data corresponding to both IM load levels under inverter feeding are mixed. In Fig. 9, the results of KNN are worse with regard ORF and RF except for experiment 4 where the results are comparable to the rest of the classifiers.

Finally, Fig. 10 shows the comparison of different ROC curves for the experiment 7 (data from both loads and both feedings). The ORF classifier outperforms the RF classifier. The KNN classifier decreases its performance in this experiment regarding the previous ones. CART presents intermediate results between ORF and KNN. In view of the results, we can say that the KNN classifier worsens its predictive behaviour when data obtained under different operating conditions of the IM are used, i.e. where load levels and power source types are mixed. CART reduces its performance when the IM is inverter-fed. This fact may be due to the intrinsic noise introduced by the inverter for which CART classifier is very sensitive. RF and ORF are superior for diagnosing rotor bar failures with data obtained under these conditions. RF improves the variance reduction of bagging by reducing the correlation between the trees with the tree-growing process through random selection of the input variables. Nonetheless, the orthogonal decision tree may fail to capture the geometrical structure of the data samples. ORF, through Ridge Regression, instead of using a single-feature based orthogonal classification, employs a penalized multivariate regression to perform partition. On the other hand, KNN is a lazy learning classifier since it simply uses the training data itself to assign directly the test observations according to a distance-based strategy.

5. Conclusions

This paper has presented a classification problem based on a novel hybrid approach composed by an SA algorithm used for feature selection and an ensemble composed of multivariate decision trees, known as ORF. The proposed methodology permits the design of a fault diagnosis system applied to rotor faults using raw data from the stator current signal. A set of fault signatures are computed from these signals, processing them in the time and frequency domains. For the first time, all fault patterns published in the literature have been used. Their

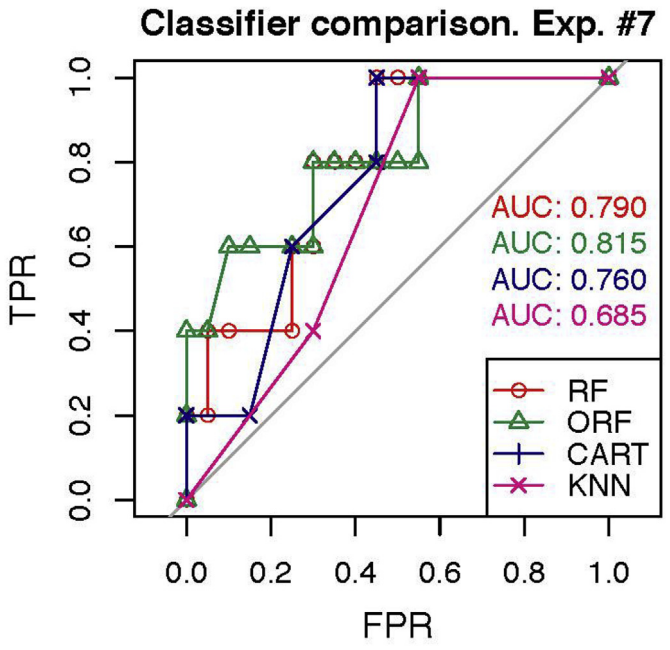

Fig. 10. ROC Curves for the classifiers corresponding to the experiment 7.

discriminant ability for distinguishing different rotor conditions has been studied. These fault patterns are used to determine the rotor health of the induction motor using different classifiers. Seven experiments were designed and permitted to evaluate the influence of the motor power supply and the motor load level on the feature selection and the classification performance. Two different voltage supplies were used, and the motor load had two levels. The number of fault severities considered was five (from a healthy state (R1) to a full broken rotor bar (R5)). The study results are very interesting. Regarding the feature selection, the left-hand side harmonic (LSH) is essential for diagnosis, as it has been corroborated by many works. A novel outcome is that the Γ_5 (another fault pattern in the frequency domain), RMS and the Kurtosis (time-domain features) prove to be relevant for the diagnosis as the results of Table 2 demonstrate.

These last two seem to be robust features since the technique has given the best results in experiments 3 and 6, where data come from mixed IM loads per power source as well as in experiment 7 where load levels and power source types are mixed. The feature selection using the SA discards around a 50% of the original features for the subsequent classification phase. The outcome of the feature selection stage differs depending on the power supply and motor load level. Therefore, this study is necessary for the design of any fault diagnosis system. The classification results are also different for the three methods studied. The ORF was superior when the two load levels experiments contributed to the training data, regardless of the power supply. However, ORF and RF showed a similar performance in other cases. The results of KNN are worse except for experiment 4 (data from LL1 with inverter feeding) where the results are comparable to the rest of the classifiers. CART achieved good results for the experiments related to the line feeding, particularly when hybridized with Simulated Annealing algorithm. However, when the IM is fed through the inverter the results of CART worsened. Another result is the computational requirements lower when the training data is reduced after the feature selection stage supervised by the SA algorithm.

Acknowledgement

This work has been partially supported by the Mexican Council of Science and Technology (CONACyT), under the doctoral scholarship 598078, by a mobility grant from the University of Valladolid, Spain, and by the Spanish Ministerio de Economia y Competitividad (MINECO) and FEDER program in the framework of the Proyectos I + D del Subprograma de Generacion de Conocimiento, Programa Estatal de Fomento de la Investigacion Cientifica y Tecnica de Excelencia (ref: DPI2014-52842-P).

References

[1] Keskes H, Braham A, Lachiri Z. Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet svm. Elec Power Syst Res Apr. 2013;97:151–7.
[2] Mehrjou MR, Mariun N, Hamiruce Marhaban M, Misron N. Rotor fault condition monitoring techniques for squirrel-cage induction machine - a review. Mech Syst Signal Process Nov. 2011;25:2827–48.
[3] Nemec M, Drobnic K, Nedeljkovic D, Fiser R, Ambrozic V. Detection of broken bars in induction motor through the analysis of supply voltage modulation. IEEE Trans Ind Electron Aug. 2010;57:2879–88.
[4] Nandi S, Toliyat HA, Li X. Condition monitoring and fault diagnosis of electrical motors a review. IEEE Trans Energy Convers Dec. 2005;20:719–29.
[5] Patel RK, Giri V. Feature selection and classification of mechanical fault of an induction motor using random forest classifier. Perspect Sci Sep. 2016;8:334–7.
[6] Nguyen N-T, Lee H-H, Kwon J-M. Optimal feature selection using genetic algorithm for mechanical fault detection of induction motor. J Mech Sci Technol Mar. 2008;22:490–6.
[7] Delgado-Arredondo PA, Morinigo-Sotelo D, Osornio-Rios RA, Avina-Cervantes JG, Rostro-Gonzalez H, de Jesus Romero-Troncoso R. Methodology for fault detection in induction motors via sound and vibration signals. Mech Syst Signal Process Jan. 2017;83:568–89.
[8] Garcia-Ramirez AG, Morales-Hernandez LA, Osornio-Rios RA, Benitez-Rangel JP, Garcia-Perez A, de Jesus Romero-Troncoso R. Fault detection in induction motors and the impact on the kinematic chain through thermographic analysis. Elec Power Syst Res Sep. 2014;114:1–9.
[9] Cabanas MF, Pedrayes F, Rojas CH, Melero MG, Norniella JG, Orcajo GA, Cano JM, Nuno F, Fuentes DR. A new portable, self-powered, and wireless instrument for the early detection of broken rotor bars in induction motors. IEEE Trans Ind Electron Oct. 2011;58:4917–30.
[10] Eltabach M, Antoni J, Najjar M. Quantitative analysis of noninvasive diagnostic procedures for induction motor drives. Mech Syst Signal Process Oct. 2007;21:2838–56.
[11] Drif M, Cardoso AM. Airgap-eccentricity fault diagnosis, in three-phase induction motors, by the complex apparent power signature analysis. IEEE Trans Ind Electron Mar. 2008;55:1404–10.
[12] Martin-Diaz I, Morinigo-Sotelo D, Duque-Perez O, de Jesus Romero-Troncoso R. An experimental comparative evaluation of machine learning techniques for motor fault diagnosis under various operating conditions. IEEE Trans Ind Appl May-June 2018;54:2215–24.
[13] Elbouchikhi E, Choqueuse V, Benbouzid M. Induction machine bearing faults detection based on a multi-dimensional music algorithm and maximum likelihood estimation. ISA Trans Jul. 2016;63:413–24.
[14] Georgakopoulos IP, Mitronikas ED, Safacas AN. Detection of induction motor faults in inverter drives using inverter input current analysis. IEEE Trans Ind Electron Sep. 2011;58:4365–73.
[15] Gyftakis KN, Cardoso AJM, Antonino-Daviu JA. Introducing the filtered park's and filtered extended park's vector approach to detect broken rotor bars in induction motors independently from the rotor slots number. Mech Syst Signal Process Sep. 2017;93:30–50.
[16] Morinigo-Sotelo D, Romero-Troncoso R, Panagiotou PA, Antonino-Daviu J, Gyftakis KN. Reliable detection of rotor bars breakage in induction motors via music and zsc methods. IEEE Trans Ind Appl Mar./Apr. 2018;54:1224–34.
[17] Gyftakis KN, Antonino-Daviu JA, Garcia-Hernandez R, McCulloch MD, Howey DA, Cardoso AJM. Comparative experimental investigation of broken bar fault detectability in induction motors. IEEE Trans Ind Appl Mar./Apr. 2016;52:1452–9.
[18] Samanta AK, Naha A, Routray A, Deb AK. Fast and accurate spectral estimation for online detection of partial broken bar in induction motors. Mech Syst Signal Process Jan. 2018;98:63–77.
[19] Bessam B, Menacer A, Boumehraz M, Cherif H. Detection of broken rotor bar faults in induction motor at low load using neural network. ISA Trans Sep. 2016;64:241–6.
[20] Martinez J, Belahcen A, Muetze A. Analysis of the vibration magnitude of an induction motor with di erent numbers of broken bars. IEEE Trans Ind Appl May/Jun. 2017;53:2711–20.
[21] Abd-el Malek M, Abdelsalam AK, Hassan OE. Induction motor broken rotor bar fault location detection through envelope analysis of start-up current using hilbert transform. Mech Syst Signal Process Sep. 2017;93:332–50.
[22] Liu Y, Bazzi AM. A review and comparison of fault detection and diagnosis methods for squirrel-cage induction motors: state of the art. ISA Trans Sep. 2017;70:400–9.
[23] Deleroi W. Squirrel cage motor with broken bar in the rotor-physical phenomena and their experimental assessment. Proc. of int. Conf. On electrical machines. 1982. p. 767–70.
[24] Martin-Diaz I, Morinigo-Sotelo D, Duque-Perez O, Arredondo-Delgado P, Camarena-Martinez D, Romero-Troncoso R. Analysis of various inverters feeding induction motors with incipient rotor fault using high-resolution spectral analysis. Elec Power Syst Res Nov. 2017;152:18–26.

[25] Bruzzese C. Analysis and application of particular current signatures (symptoms) for cage monitoring in nonsinusoidally fed motors with high rejection to drive load, inertia, and frequency variations. IEEE Trans Ind Electron Dec. 2008;55:4137–55.
[26] Xue X, Zhou J. A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery. ISA Trans Jan. 2017;66:284–95.
[27] Bouzida A, Touhami O, Ibtiouen R, Belouchrani A, Fadel M, Rezzoug A. Fault diagnosis in industrial induction machines through discrete wavelet transform. IEEE Trans Ind Electron Sep. 2011;58:4385–95.
[28] Ghate VN, Dudul SV. Optimal MLP neural network classifier for fault detection of three phase induction motor. Expert Syst Appl Apr. 2010;37:3468–81.
[29] Yang B-S, Oh M-S, Tan ACC, et al. Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference. Expert Syst Appl Mar. 2009;36:1840–9.
[30] Seera M, Lim CP, Loo CK. Motor fault detection and diagnosis using a hybrid FMM-cart model with online learning. J Intell Manuf Dec. 2016;27:1273–85.
[31] Karvelis P, Georgoulas G, Tsoumas IP, Antonino-Daviu JA, Climente-Alarcon V, Stylios CD. A symbolic representation approach for the diagnosis of broken rotor bars in induction motors. IEEE Trans Ind Inf Oct. 2015;11:1028–37.
[32] Bellini A, Filippetti F, Franceschini G, Tassoni C, Kliman G. Quantitative evaluation of induction motor broken bars by means of electrical signature analysis. IEEE Trans Ind Appl Sep./Oct. 2001;37:1248–55.
[33] Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, the R Core Team. Benesty M, Lescarbeau R, Ziem A, Scrucca L, Tang Y, Candan C, Hunt T. Caret: classification and regression training. 2017. R package version 6.0-76.
[34] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res Mar. 2003;3:1157–82.
[35] Rauber TW, de Assis Boldt F, Varejao FM. Heterogeneous feature models and feature selection applied to bearing fault diagnosis. IEEE Trans Ind Electron Jan. 2015;62:637–46.
[36] Cerrada M, Zurita G, Cabrera D, Sanchez R-V, Artes M, Li C. Fault diagnosis in spur gears based on genetic algorithm and random forest. Mech Syst Signal Process Mar. 2016;70:87–103.
[37] Bohachevsky IO, Johnson ME, Stein ML. Generalized simulated annealing for function optimization. Technometrics 1986;28(3):209–17.
[38] Breiman L. Random forests. Mach Learn Jan. 2001;45:5–32.
[39] Menze BH, Kelm BM, Splitthoff DN, Koethe U, Hamprecht FA. On oblique random forests. Proc. Springer ECML PKDD'11, vol. 2. Sep. 2011. p. 453–69. Berlin, Heidelberg.
[40] Rodriguez JJ, Kuncheva LI, Alonso CJ. Rotation forest: a new classifier ensemble method. IEEE Trans Pattern Anal Mach Intell Oct. 2006;28:1619–30.
[41] R Core Team R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
[42] Liaw A, Wiener M. Classification and regression by random forest. R News 2002;2(3):18–22.
[43] Menze B, Splitthoff N. obliqueRF: oblique random forests from recursive linear model splits. 2012. R package version 0.3.
[44] Martin-Diaz I, Morinigo-Sotelo D, Duque-Perez O, Romero-Troncoso RJ. Advances in classifier evaluation: novel insights for an electric data-driven motor diagnosis. IEEE Access Oct. 2016;4:7028–38.