Article

Condition Monitoring of Bearing Faults Using the Stator Current and Shrinkage Methods

Oscar Duque-Perez[1,*], Carlos Del Pozo-Gallego[1], Daniel Morinigo-Sotelo[1], Wagner Fontes Godoy [2]

[1] Department of Electrical Engineering. University of Valladolid. Valladolid, Spain;
[2] Department of Electrical Engineering. Universidade Tecnologica Federal do Parana. Cornelio Procopio, Brazil;
* Correspondence: oscar.duque@eii.uva.es

Abstract: Condition monitoring of bearings is an open issue. The use of the stator current to monitor induction motors has been validated as a very advantageous and practical way to detect several types of faults. Nevertheless, for bearing faults the use of vibrations or sound generally offers better results in the accuracy of the detection although with some disadvantages related to the sensors used for monitoring. To improve the performance of bearing monitoring, it is proposed to take advantage of more information available in the current spectra, beyond the usually employed, incorporating the amplitude of a significant number of sidebands around the first eleven harmonics, growing exponentially the number of fault signatures. This is especially interesting for inverter-fed motors. But, in turn, this leads to the problem of overfitting when applying a classifier to perform the fault diagnosis. To overcome this problem, and still exploit all the useful information available in the spectra, it is proposed to use shrinkage methods, which have been lately proposed in machine learning to solve the overfitting issue when the problem has much more variables than examples to classify. A case study with a motor is shown to prove the validity of the proposal.

Keywords: condition monitoring, bearings, machine learning, current spectra.

1. Introduction

Induction motors are a fundamental part of many production processes due to their inherent robustness, low cost and reliability, among other advantages. However, they are not fault-free, with bearings being the component that accounts for the greatest percentage of total failures [1].

The signals that are most frequently used for bearing fault detection are vibration and acoustic noise [2]. However, the use of the stator current to monitor the motor provides some practical advantages related to the simplicity and non-invasive characteristics of the sensors. These advantages are especially relevant in industrial facilities where some motors can run simultaneously [3,4]. The use of current has proven its effectiveness in detecting faults such as broken bars and eccentricity [2], but in the case of faulty bearings, it faces technical difficulties that hinder its successful implementation. Mainly, the low energy of the vibrations associated to the fault, which makes it difficult to distinguish in the current spectrum the frequency components related to the fault that may be buried in the noise [1,5,6]. Besides, for inverter-fed motors the noise is higher and there are other harmonics present in the spectrum which complicates even more the detection of the faulty related components [7]. Consistently, in [8] denoising techniques are applied to highlight the faulty components in the current spectrum. Other advanced spectral techniques have also been proposed such as wavelets [9,10], Short-Time Fourier Transform [11], Gabor spectrogram [11] Hilbert-Huang Transform [12], Empirical Mode Decomposition [13] MUSIC [13,14], space vector angular fluctuation method [15]. These techniques have the drawback of a high computational cost.

The first stage related to the detection of the fault provides with some fault signatures that will feed the second stage of the process, the diagnosis. A wide variety of algorithms has been proposed to diagnose faulty bearings such as Artificial Neural Networks [16,17], Support Vector Machines [12], [18,19], K-nearest neighbors [20,21], supervised fuzzy-neighborhood density-based clustering [22], random forest [23], bagging, boosting and stacking methods [24], Common Vector Approach [25], Decision Trees [26,27]. Maximum margin classification [28], Bayes classifier [29], Euclidean Distance Minimization [30] and Bayesian inference [31]. The aforementioned algorithms were mostly applied making use of the known fault signatures related to bearing faults, limited to just a few signatures (usually just the sideband around the main harmonic, as it will be shown in Section 2). However, for challenging cases this is a restricted use of the information available in the spectra. In the case of inverter-fed motors, it can be used the information related to sidebands around the harmonics introduced by the supply [3]. In [32] it is shown how the effects of different types of bearing faults are spread over the spectrum, being for some cases more notorious for the 3rd and 7th harmonics and, in other cases, for high frequency odd harmonics; besides, these effects are different depending on the operating characteristics of the motor. In [33] the cases of excessive and defective lubrication are analyzed showing how these situations produce changes in the amplitudes of different sidebands around different harmonics, showing also a variation depending on the load.

Consequently, in this paper it is proposed to take advantage of the information available over the spectrum, considering not only the main frequency but also different odd and even harmonics (up to the 11th one) and including many more sidebands than is usual in literature. This way, instead of using just a few signatures to feed the algorithm as is usual in literature, almost one thousand signatures are used in this proposal.

The drawback of feeding the classifier with a high number of signatures is a clear risk of overfitting. Overfittings arises when the model has learned the data too well, leading to a small error in the training set (used to build the model) but poor prediction ability. Besides, overfitting is intensified by the presence of noise in the data [34] and, precisely, dealing with motors fed from inverters, a significant presence of noise is to be expected. A solution to minimize the overfitting problem is to apply shrinkage techniques. These techniques perform shrinking the values of the coefficients in the trained model. There are several versions of shrinkage techniques depending on the degree of shrinkage of each coefficient. If some of them can be set to zero, then the method is known as Lasso (Least Absolute Shrinkage and Selection Operator) and consequently the number of signatures in the classifier is reduced, obtaining simpler models. If the value of the coefficients is reduced, but all the signatures are included in the model, the method is known as Ridge Regression. The technique known as Elastic Nets is a way of combining both Lasso and Ridge Regression.

To monitor correctly the state of the bearings and be really useful for maintenance purposes it is it is essential to be able to distinguish between different states of deterioration and detect incipient faults before they develop into critical ones. With this purpose, it has been simulated in the laboratory a progressive deterioration of an induction motor bearing via the contamination of the lubrication introducing particles of silicon carbide in the bearing grease. This process tries to emulate conditions usually present in the industry that produces bearing wear related to the use, to inadequate lubrication or to the proper contamination of the grease in open ball bearings.

In this paper, it is proposed the use of a large number of fault signatures obtained from the current spectra to monitor bearing failures. A case study is presented where five states of deterioration of the bearing are considered giving rise to a problem of multiclassification. The improved in the performance of different classifiers when using such fault signatures is shown by comparing with the performance with the usual signatures used in these studies. Then, to deal with the problem of overfitting, shrinkage techniques are applied, comparing the performance of Lasso, Ridge Regression and Elastic Nets, proving their validity to diagnose bearing failures.

2. Fault Signatures

When a bearing defect appears, a radial motion between rotor and stator will occur modifying the airgap of the motor thus changing the airgap field. These modifications in the airgap can be

interpreted as a combination of bidirectional rotating eccentricities [35], which implies that the defect affects the stator current and, therefore, it is possible to monitor it in the current spectra. The radial motion generates harmonics in the stator current at frequencies given by (1)

$$f_{fault} = | f_1 \pm n f_v | \quad (1)$$

where f_1 is the main supply frequency, n is an entire number and f_v is the vibration characteristic frequency. f_v depends on the type of bearing fault (outer or inner race, balls and train defect) with expressions that are function of the geometry and composition of the bearing [7].

The fault frequencies given by (1) are the result of taking into account the deviations in the main component of the airgap field. When the motor is fed by a power converter, the harmonics level is increased and so is the noise level, hampering the detection of the fault signatures. Nevertheless, the presence of these harmonics can be used to increase the available information also considering the deviations produced in the fields as a consequence of these harmonics. Even when the motor is directly fed from the line, since the supply is hardly ever perfectly sinusoidal, the number of fault signatures can be increased too considering the harmonic introduced by the supply. Therefore, (1) can be generalized by (2) where k is the order of the current harmonic.

$$f_{fault} = | k f_1 \pm n f_v | \quad (2)$$

Considering (2), the number of fault signatures can be increased, resulting in a smaller or larger number of variables depending on the value of k and n. In [7,32] the first sideband around the 5th and 7th harmonics is employed. In this paper, it is proposed to use a larger number of signatures, considering more harmonics, and other sidebands in addition to the first one as well. In the case study of Section 4, the first eleven current harmonics are used to feed the classifier, considering the first eleven sidebands around those harmonics. As each sideband is composed of two values, there are 242 signatures for each characteristic bearing fault frequency, resulting in 968 signatures. Table 1 summarizes the information regarding the proposed fault signatures and the comparison with the traditional approach.

Table 1. Bearing fault signatures considered in the traditional and in the proposed approach.

	Traditional approach	Proposed approach
Bearing fault frequencies	BPFO, BPFI, FTF; BSF	BPFO, BPFI, FTF; BSF
Harmonics considered	1	1, 2, …, 11
Sidebands around harmonic	1	1, 2, …, 11
Number of fault signatures	8	968

3. Diagnosis

The next step after selecting the candidate fault signatures is to choose and train the classifier. Many classification algorithms are available, with a wide variety of them already proposed to perform diagnosis tasks in induction motors. With the purpose of analyzing the improvement in the performance of the classifier when using the fault signatures presented in the previous section, the Matlab 2019a Classification learner App has been used. In this app there are available different types of classifiers: Decision trees, discriminant analysis, logistic regression classifiers, Naïve Bayes classifiers, support vector machines, nearest neighbor classifiers and ensemble classifiers. In each group, there are several classifiers available. Using these classifiers, it has been proved (as shown in the results section) the huge increase in performance of all the classifiers when using the 968 fault signatures instead of the usual 8 signatures.

However, when using such high number of signatures, and with a reduced number of tests, the risk of overfitting is certain. Shrinkage techniques allow to make use of all the predictors but shrinking the coefficients towards zero, hence, reducing variance [36]. If applied in linear models (which has the advantage in terms of interpretability of the model), it performs as follows: let x_i be the m predictors (or fault signatures in the context of condition monitoring) and y_i the response for

the n cases of the problem. A linear model tries to estimate the m+1 coefficients (b_0,.., b_m). Using least squares fitting approach, b_i are selected to minimize (a).

$$\sum_{i=1}^{n}\left(y_i - b_0 - \sum_{j=1}^{m} b_j x_{ij}\right)^2 \quad (3)$$

To perform the shrinkage, a second term is added to (a), $\lambda \sum_{j=1}^{m} b_j^2$, which acts as shrinkage penalty. Its influence depends on the value of λ, which is a tuning parameter that increases or decreases the penalty. For higher values of λ, the penalty grows and the coefficient estimated will tend to zero, which implies that the estimation is somehow penalized, sacrificing some of the performance on the training set but with the aim of improving its predictive capacity with future observations. The penalty applies to all the coefficients but the intercept, b_0, since this term is just an estimation of the mean when the predictors are zero [36].

This way of applying the penalty so performing the shrinkage on the estimated coefficients is known as Ridge Regression. It has the disadvantage that the shrinkage is applied to all the coefficients but none of them are set to zero, so all the predictors are included in the solution, which for problems with a large number of predictors (as in the problem dealt in this paper) leads to lose the interpretability of the model. A way of tackling this problem is to change the penalty term into $\lambda \sum_{j=1}^{m} |b_j|$, or in statistical terms, to change an l_2 penalty for an l_1 one [36]. The use of an l_1 norm has the inconvenience of turning the function to minimize into non-differentiable, although there are available methods to proceed with the minimization, such as proximal gradient ones [37]. This way of considering the penalty gives rise to the method known as Lasso. As opposed to Ridge Regression, with Lasso, some variables are canceled, so performing as variable selection, depending the number of the variables to be selected on the value of λ (as λ grows, less variables are selected).

Lasso was first applied to linear regressions and lately is receiving much attention being proposed to regularize a wide variety of statistical models [38]. In accordance to Occam's razor principle, simpler models are preferable, as long as they predict well the training data, since they are more likely to generalize well to unseen data [39]. With this principle in mind, Logistic Regression has been chosen as base model in which to apply the shrinkage technique. Logistic regression is adapted to classification problems since has a discrete outcome. It is based on the logistic function given by (4) which is suitable to be used in classification since its outcome can be inferred as a probability since runs between 0 and 1, and its elongated S-shape offers the advantage that the same additional input influences less the outcome for values near zero or one [40,41]. For binary classification, a threshold value of 0.5 is defined to assign the outcome to one class or the other, which in condition monitoring would be healthy or faulty. When the aim is to distinguish among different states of failure there are several classes into which the outcome can be classified. This multiclassification is performed via the one-versus-all approach as represented in the flow chart in Figure 1. This way, several binary classifiers are trained (as many as classes), where each classifier confronts one class against the rest. Finally, the outcome is assigned to the class where the probability is highest.

$$f(x) = (1 + e^{-x})^{-1} \quad (4)$$

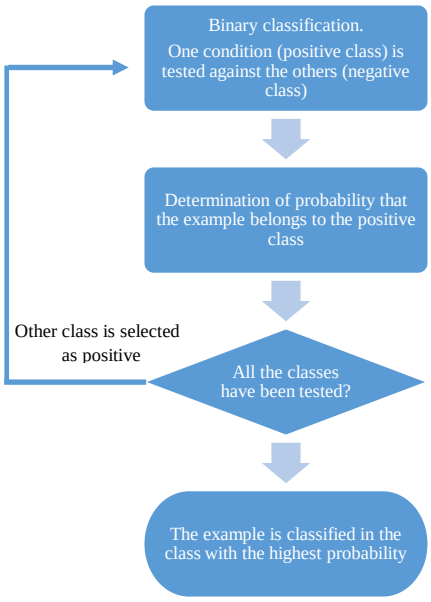


Figure 1. Methodology for multiclass classification.

4. Results

4.1 Test Bench

The tested induction motor is a two pole pair squirrel cage motor, star connected, with a rated power of 0.75 kW at 400 V and a rated current of 1.9 A at a rated speed of 1395 RPM. The tests were performed at two levels load, low (almost no load) and high (rated speed) using a magnetic powder brake. The data were collect using a DAC PCI-6250 M model (16 analogue inputs - 16-bit 1 MS/s) of National Instruments and Hall effect sensors of LEM. The sampling frequency was 25 kHz with a sampling time of 10 s (steady state).

For different supply conditions were considered (Table 2). The first one (S1) represents the motor directly fed from a 400 V utility supply. Supply S2 is the motor fed by an inverter (ABB) at 50 Hz and with a switching frequency of 4 kHz. For S3, the operating frequency was changed to 25 Hz, and for S4 the switching frequency was established at 5 kHz).

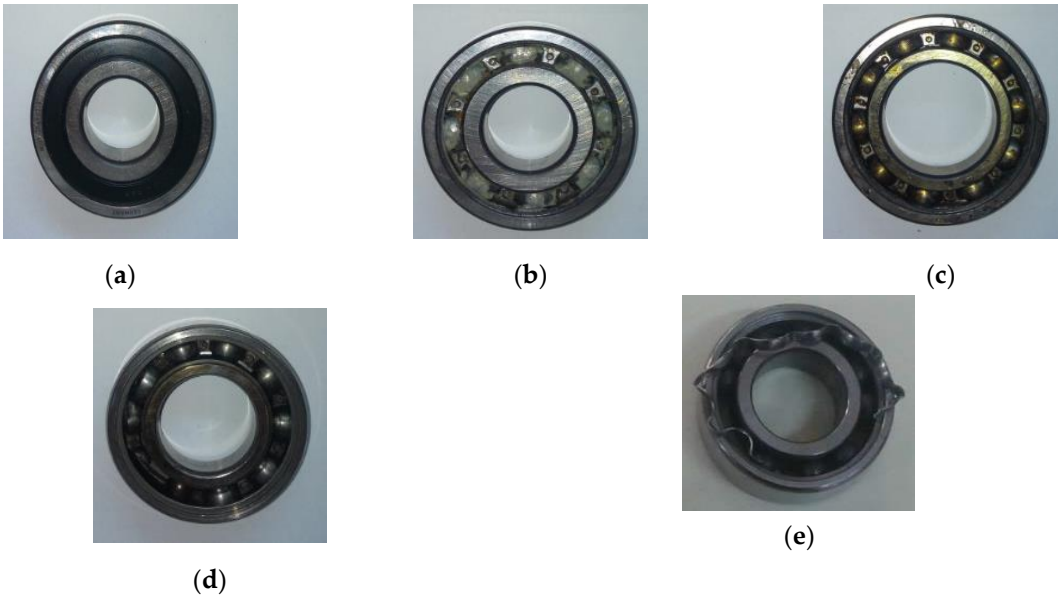
Table 2. Supplies tested.

Supply Identification	Power source	Operating frequency	Switching frequency
S1	utility	50 Hz	-
S2	Power converter	50 Hz	4 kHz
S3	Power converter	25 Hz	4 kHz
S4	Power converter	50 Hz	5 kHz

.

To initiate the tests, a new SKF Explorer 6004 bearing was used, performing the corresponding tests to represent the healthy condition. Then, to provoke the progressive wear of the bearing, the lubricant grease was contaminated using silicon carbide, a ceramic material with high resistance to erosion, corrosion and high thermal cycling. This process was established to simulate industrial environment conditions that lead to the degradation of the bearing such as inadequate lubrication, overloads or lubricant contamination (especially relevant to open bearings). During this process, five condition states were defined according to the degradation of the bearing, as summarized in Table 3. After assembling the new bearing, 20 tests per supply were run corresponding to the healthy state (C1). Then, the bearing was first contaminated and the motor run unloaded for 12 hours to lead the

bearing to the "incipient fault" condition (C2). In this condition, 15 tests were performed for each supply. The process of running the motor unloaded and contaminating the grease was repeated giving way to "intermediate fault" condition (C3), with 15 tests per supply, "developed fault" with 10 tests (C4) and "complete breakdown" (C5) with 10 tests for each supply. Figure 2 presents pictures of the bearing in each of the conditions, showing the evolution of the fault along the tests.



(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 2. Evolution of the bearing along the tests: (**a**) healthy state, (**b**) incipient fault, (**c**) intermediate fault (**d**) developed fault (**e**) complete breakdown.

Table 3. Bearing conditions tested.

Condition	Evolution of the fault	Number of tests per supply
C1	healthy state	20
C2	incipient fault	15
C3	intermediate fault	15
C4	developed fault	10
C5	complete breakdown	10

4.2. Classification with 968 fault signatures

In order to show the improvement in the classification when using the whole of the fault signatures as proposed in Section 2, next the results obtained using the Matlab 2019a Classification learner App are presented. 5-fold cross validation was used. Table 4 summarizes the results obtained with the App with the accuracy for the classification in each of the five bearing conditions at low and high load. All the algorithms included in the App have been tested showing the one that has the best performance for each tested case (depending on the load and the supply) and its accuracy. The same procedure has been applied feeding the algorithms with eight inputs, following the traditional procedure of considering just the first sideband around the vibration characteristics frequencies, according to (2).

Table 4. Comparison between the traditional (8 fault signatures) and proposed approach (968 fault signatures) with the algorithms included in the Matlab 2019a Classification learner App.

Supply Identification	Load	Best accuracy 968 fault signatures	Best algorithm 968 fault signatures	Best accuracy 8 fault signatures	Best algorithm 8 fault signatures
S1	Low	95.7 %	KNN[1]	40 %	SVM[2]
	High	92.9 %	SVM[2]	40 %	SVM[2]
S2	Low	88.6 %	SVM[2]	32.9 %	BT[5]
	High	98.6 %	SVM[2]	44.3 %	BT[5]
S3	Low	81.4 %	BT[5]	41.4 %	SVM[2]
	High	97.1 %	LD[4]	60 %	KNN[1]
S4	Low	85.7 %	LD[4]	41.4 %	SVM[2]
	High	98.6 %	BT[5]	42.9 %	SVM[2]

[1] Fine KNN, [2]Quadratic Support Vector Machines, [3]Gaussian Naive Bayes, [4]Linear Discriminant, [5]Bagged Tress

According to the results shown in Table 4, it is very clear that the use of the fault signatures related to a bigger number of sidebands around more harmonics outperforms the use of just eight fault signatures. The huge improvement in the performance is observable in the all the cases, for the different supplies, operating frequencies, switching frequencies and loads. The results are in general better for high load since the energy associated to the harmonics is higher. It is also remarkable that there is a variety of selected algorithms, being Support Vector Machines the most repeated although, in some cases, Gaussian Naive Bayes, Linear Discriminant, Fine KNN and Bagged Trees perform better. This discrepancy adds difficulty to the selection of a classifier valid for all the operating conditions. An algorithm that performs well for all the cases would be desirable. Besides, as it was stated earlier, the use of a big number of signatures (much bigger than the number of tests) may lead to overfitting, losing the trained algorithms the ability to generalize when classifying new observations. To take into account this situation, shrinkage is applied as explained in Section 3.

4.3. Classification with 968 fault signatures applying shrinkage

Previous section has shown that the classification improves hugely when more information available in the spectra is considered. In this section, shrinkage methods are applied with the double purpose of selecting an algorithm with good performance independently of the operating conditions and of avoiding the problem of overfitting (prone to appear due to the high number of fault signatures, much higher than the number of tests). As explained in Section 3, two different types of shrinkage methods are considered: Ridge regression where all the inputs are considered in the classification, and Lasso that performs variable selection (considering a higher or lower number of inputs in the classification depending on the value of the penalty parameter). A third method is included in the comparison, Elastic nets, which can be considered as a mixture between Lasso and Ridge regression.

To build the algorithms and measure their performance, the data sets for each case study are divided into two different sets, the training set consisting of 70% of the cases and the test set with the other 30% of the data. Table 5 shows the performance, measured in terms of accuracy, for the three shrinkage methods, for each supply and load. It can be observed that the results are very good for all the cases, although with some differences in the performance among the cases analyzed, as it also happened for the algorithms considered in the previous section. The three shrinkage methods perform well, although in general, Lasso obtains the best accuracy, therefore, if a single method were to be selected, Lasso would be the candidate. In this selection it has also been taken into account that, since Lasso eliminates variables from the classifier, the model obtained gains in interpretability and computational cost. For that reason, next, a deeper analysis on the performance of Lasso is presented, taking especially into consideration the influence of the penalty parameter, since the number of variables selected (and, consequently, the characteristics of the model) depend on this parameter.

Table 5. Comparison of Accuracy (%) obtained using different methods of shrinkage with 968 fault signatures.

Supply Identification	Load	Lasso	Elastic Nets	Ridge Regression
S1	Low	100	100	100
	High	90.48	90.48	85.71
S2	Low	95.24	90.48	80.95
	High	95.24	100	100
S3	Low	80.95	76.19	76.19
	High	90.48	95.24	95.24
S4	Low	90.48	80.95	85.71
	High	100	95.24	85.71

As it was stated in Section 3, the main way in which Lasso avoids overfitting is by feature selection, which is controlled by adjusting the regularization parameter λ. The bigger λ, more parameters bj in (3) will be zero, that is, the corresponding predictors will not be considered when designing the classifier. Therefore, if a high value of λ is chosen, it is much less likely to result in overfitting, besides, the computational cost is highly reduced. The drawback is that, if less predictors are considered, the performance of the classifier will be reduced. Therefore, a trade-off must be reached to select the best value for the regularization parameter to obtain a good classifier performance with less computational cost. The selection of the value of the regularization parameter has been performed considering the train set. No additional validation set has been considered since the number of tests per case study is low and this would have led to training, validating and test sets with very few data in each one. Figure 3 shows the evolution of the accuracy depending on the regularization parameter and Table 6 shows the selected value for each supply and load condition. The value chosen for the classifier is the highest one that achieves the best accuracy for that supply and load, since for smaller values of λ the computational cost would be higher. If a unique value were to be chosen for all the supplies, 0.05 could be selected when operating at high load and 0.02 at low load. If the common value of 0.02 were to be chosen, the performance of the algorithm in this case would decreased around 5%, although the computational cost would decreased.

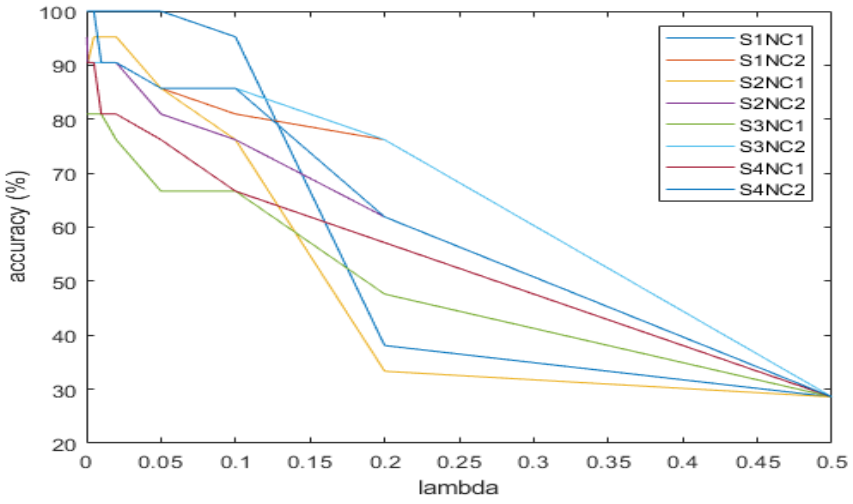


Figure 3. Performance of Lasso classifier depending on the value of the regularization parameter fir different supplies (S1-S6) and loads (HL: high load, LL: low load).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2019 doi:10.20944/preprints201907.0309.v1

Peer-reviewed version available at Energies 2019, 12, 3392; doi:10.3390/en12173392

Table 6. Regularization parameter selected for different supplies and level of load.

Supply	Load	Regularization parameter
S1	High	0.02
	Low	0.05
S2	High	0.0003
	Low	0.02
S3	High	0.01
	Low	0.01
S4	High	0.005
	Low	0.005

So far, accuracy has been used to measure the performance of the classifier. Obviously, from an algorithmic point of view, it is important to classify all the states correctly, and therefore to achieve the highest possible accuracy. However, from a condition monitoring point of view, some misclassifications are more relevant than others, being especially relevant to predict the first and fifth states correctly, that is, the healthy and completely faulty conditions. With this purpose, Tables 7-10 show the confusion matrices resulting from applying Lasso classifier for the four supplies and two load conditions. It can be observed how for the healthy state, for 48 instances (there are six true healthy states in each of the eight cases) only one case (S3, low load) is misclassified. And even if this case can be considered as a false negative, this instance is classified as an incipient fault, not as a more developed one. In the same way, for the 24 complete faulty cases (three for each of the eight cases) only one is misclassified (again, S3 at low load), being predicted as an intermediate fault. Finally, it is relevant to point out that for all the 168 cases to classify, 13 are not correctly classify but just four of them are classify more than one class away from the true class.

Table 7. Confusion matrices for supply S1 applying lasso classifier.

		Low Load					High Load				
		Predicted class					Predicted class				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
True class	C1	6	0	0	0	0	6	0	0	0	0
	C2	0	5	0	0	0	0	3	0	0	1
	C3	0	0	4	0	0	0	1	4	0	0
	C4	0	0	0	3	0	0	0	0	3	0
	C5	0	0	0	0	3	0	0	0	0	3

Table 8. Confusion matrices for supply S2 applying lasso classifier.

		Low Load					High Load				
		Predicted class					Predicted class				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
True class	C1	6	0	0	0	0	6	0	0	0	0
	C2	1	3	0	0	0	0	3	0	0	1
	C3	0	0	5	0	0	0	1	4	0	0
	C4	0	0	0	3	0	0	0	0	3	0
	C5	0	0	0	0	3	0	0	0	0	3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 July 2019 doi:10.20944/preprints201907.0309.v1

Peer-reviewed version available at Energies 2019, 12, 3392; doi:10.3390/en12173392

Table 9. Confusion matrices for supply S3 applying lasso classifier.

		Low Load					High Load				
		Predicted class					Predicted class				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
True class	C1	5	1	0	0	0	6	0	0	0	0
	C2	0	4	0	0	0	0	4	0	0	0
	C3	0	0	3	1	1	1	1	3	0	0
	C4	0	0	0	3	0	0	0	0	3	0
	C5	0	0	0	1	2	0	0	0	0	3

Table 10. Confusion matrices for supply S4 applying lasso classifier.

		Low Load					High Load				
		Predicted class					Predicted class				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
rue class	C1	6	0	0	0	0	6	0	0	0	0
	C2	0	4	0	0	0	0	4	0	0	0
	C3	0	0	4	0	1	0	0	5	0	0
	C4	0	0	1	2	0	0	0	0	3	0
	C5	0	0	0	0	3	0	0	0	0	3

5. Discussion

A procedure for the diagnosis of induction motor bearings has been presented. The main purpose of the proposal is to maintain the good performance of existing methods that use vibrations or sound as inputs but using the stator current. So far, the monitoring of the current has not achieved as good performance as the use of the other variables mentioned, but since has some clear advantages related to the necessary sensors it is advisable to have a procedure that allows to use the current. To achieve this goal it has been proposed to take advantage of more information that can be extracted from the spectra beyond what is commonly used, but with no extra computational cost.

It has been shown that the use of much more information greatly improved the performance of the diagnosis, which has been proved by means of 24 classifiers (available in the Matlab Classification learner app). However, it must be taken into account that detection and diagnosis are interlinked. There is no use in expecting a good diagnosis performance if the fault signatures obtained during the detection process are of a bad quality. Conversely, although there were high informative fault signatures, if the diagnosis stage is bad designed, the whole process will suffer. Besides, the chosen algorithm must be in accordance with the available variables. Therefore, it has been selected a type of classifier that can perform well with the particular conditions of the problem, where there are much more fault signatures that cases to classify. Shrinkage methods have been chose since they allow to perform in those condition avoiding the problem of overfitting.

Three shrinkage methods have been compared, Lasso, Ridge regression and Elastic nets and all of them have proved to achieve a very good performance in the cases analyzed. Although all three meet the expectations, Lasso has been chosen to analyze its results in greater depth since this method selects variables, providing simpler and more interpretable models. For the analysis of the performance of Lasso, the confusion matrices for eight different scenarios have been provided and analyzed. Although from an algorithmic point of view, it is important to classify all the states correctly, from a maintenance perspective, it is especially relevant the presence of false positives or false negatives concerning the healthy and complete fault conditions. That is, some misclassifications

are more relevant than others are. For example, wrong predictions between conditions corresponding to intermediate and incipient faults are not likely to have important repercussions but, on the contrary, a misclassification between states healthy and complete fault will surely have further implications. It has been shown that the predictions obtained with the proposed method matches the expectations form a condition monitoring perspective.

Author Contributions: Conceptualization, Oscar Duque-Perez, Daniel Morinigo-Sotelo and Wagner Fontes Godoy; Data curation, Wagner Fontes Godoy; Methodology, Oscar Duque-Perez; Software, Carlos Del Pozo-Gallego; Validation, Daniel Morinigo-Sotelo; Writing – original draft, Oscar Duque-Perez; Writing – review & editing, Daniel Morinigo-Sotelo.

References

1. Immovilli, F.; Bellini, A.; Rubini, R.; Tassoni, C. Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison. *IEEE Trans. Ind. Appl.* **2010,** *46-4*, 1350–1359.
2. Henao H. et al. Trends in fault diagnosis for electrical machines. *IEEE Ind. Electron. Mag.*, **2014**, *8-2*, 31–42.
3. Akin, B.; Toliyat, H.A.; Orguner, U.; Rayner, M. PWM Inverter Harmonics Contributions to the Inverter-Fed Induction Machine Bearing Fault Diagnosis. in IEEE Applied Power Electronics Conference, APEC 2007 - Twenty Second Annual, pp. 1393-399.
4. Bellini, A.; Filippetti, F.; Tassoni, C.; Capolino, G.-A.. Advances in diagnostic techniques for induction machines. *IEEE Trans. Ind. Electron.*, **2008**, *55-12*, 4109–4126.
5. Zarei, J. ; Poshtan, J. An advanced Park's vectors approach for bearing fault detection. *Tribology International* **2009** *42*, 213– 219.
6. Blodt, M. ; Granjon, P.; Raison, B.; Rostaing, G. Models for bearing damage detection in induction motors using stator current monitoring. *IEEE Trans. Ind. Electron.*, **2008**, *55-4*, 1813–1822.
7. Duque, O.; Pérez, M.; Moríñigo, D. Detection of bearing faults in cage induction motors fed by frequency converter using spectral analysis of line current in IEEE International Conference on Electric Machines and Drives, 2005 pp. 17-22.
8. Zhou, W.; Habetler, T.G.; Harley, R.G. Bearing Fault Detection Via Stator Current Noise Cancellation and Statistical Control. *IEEE Trans. Ind. Electron.*, **2008**, *55-12*, 4260–4269.
9. Singh, S.; Kumar, A.; Kumar N. Motor Current Signature Analysis for Bearing Fault Detection in Mechanical Systems. *Procedia Materials Science*, **2014**, *6*, pp.171–177.
10. Deekshit Kompella, K.C.; Venu Gopala Rao, M.; Srinivasa Rao, R. DWT based bearing fault detection in induction motor using noise cancellation. *Journal of Electrical Systems and Information Technology*, **2016**, *3-3*, 411–427.
11. Rosero, J.; Cusido, J.; Garcia Espinosa, A.; Ortega J. A.; Romeral, L. Broken Bearings Fault Detection for a Permanent Magnet Synchronous Motor under non-constant working conditions by means of a Joint Time Frequency Analysis in IEEE International Symposium on Industrial Electronics, 2007. ISIE 2007, pp. 3415-3419.
12. Kablaa, A.; Mokrani, K. Bearing fault diagnosis using Hilbert-Huang transform (HHT) and support vector machine (SVM). *Mechanics & Industry*, **2016**, *17-3*.
13. Camarena-Martinez, D.; Osornio-Rios, R.; Romero-Troncoso, R.J.; Garcia-Perez, A. Fused Empirical Mode Decomposition and MUSIC Algorithms for Detecting Multiple Combined Faults in Induction Motors. *Journal of Applied Research and Technology*, **2015**, *13-1*, 160-167.
14. Boudinar, A.H. ; Benouzza, N.; Bendiabdellah, A.; Khodja, A. Induction Motor Bearing Fault Analysis Using a Root-MUSIC Method. *IEEE Trans. Ind. Appl.*, **2016**, *52-5*, 3851–3860.
15. Arkan, M.; Çaliş, H.; Tağluk, M.E. Bearing and misalignment fault detection in induction motors by using the space vector angular fluctuation signal. *Electrical Engineering*, **2005**, *87-4*, 197–206.
16. Godoy, W.F.; da Silva, I.N.; Goedtel, A.;. Palácios, R.H.C.; Gongora, W.S. Neural approach for bearing fault classification in induction motors by using motor current and voltage, in IEEE International Joint Conference on Neural Networks (IJCNN), 2014, pp. 2087 – 2092.
17. Ballal, M.S. ; Khan, Z.J.; Suryawanshi, H.M.; Sonolikar, R.L. Adaptive Neural Fuzzy Inference System for the Detection of Inter-Turn Insulation and Bearing Wear Faults in Induction Motor. *IEEE Trans. Ind. Electron.*, **2007**, *54-1*, 250–258.

18. Zarei, J.; Arefi, M.M.; Hassani, H. Bearing fault detection based on interval type-2 fuzzy logic systems for support vector machines in 6th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO), 2015, pp. 1 – 6.
19. Salem, S.B.; Bacha, K.; Chaari, A. Support vector machine based decision for mechanical fault condition monitoring in induction motor using an advanced Hilbert-Park transform. *ISA Transactions*, **2012**, *51-5*, 566-572.
20. Bouguerne, A. ; Lebaroud, A.; Medoued A.; Boukadoum, A. Classification of induction machine faults by K-nearest neighbor in 7th International Conference on Electrical and Electronics Engineering (ELECO), 2011, pp. I-363-I-366
21. Tian, J.; Morillo, C.; Azarian M.H.; Pecht, M. Motor Bearing Fault Detection Using Spectral Kurtosis-Based Feature Extraction Coupled With K-Nearest Neighbor Distance Analysis. *IEEE Trans. Ind. Electron.*, **2016**, *63-3*, 1793–1803.
22. Farajzadeh-Zanjani, M.; Razavi-Far, R.; Saif, M. Diagnosis of Bearing Defects in Induction Motors by Fuzzy-Neighborhood Density-Based Clustering in IEEE 14th International Conference on Machine Learning and Applications, 2015, pp. 935-940.
23. Peng H.-W.; Chiang, P.-J. Control of mechatronics systems: Ball bearing fault diagnosis using machine learning techniques in 8th Asian Control Conference (ASCC), 2011, pp. 175-180.
24. Santos, S.P.; Costa, J.A.F. A Comparison between Hybrid and Non-hybrid Classifiers in Diagnosis of Induction Motor Faults in 11th IEEE International Conference on Computational Science and Engineering, CSE '08, 2008, pp. 301-306.
25. Ergin, S.; Tezel, S.; Gulmezoglu, M.B. DWT-based fault diagnosis in induction motors by using CVA in International Symposium on Innovations in Intelligent Systems and Applications (INISTA), 2011, pp. 129–132.
26. Cunha Palácios, R.H.; da Silva, I.N.; Goedtel, A.; Godoy, W.F. A comprehensive evaluation of intelligent classifiers for fault identification in three-phase induction motors. *Electric Power Systems Research*, **2015**, 127, 249–258.
27. Sugumaran, V.; Ramachandran, K.I. Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. *Mechanical Systems and Signal Processing*, **2007**, *21-5*, 2237-2247.
28. Zeng, M.; Yang, Y.; Zheng, J.; Cheng, J. Maximum margin classification based on flexible convex hulls for fault diagnosis of roller bearings. *Mechanical Systems and Signal Processing*, **2016**, *66-67*, 533–545.
29. Mbo'o, C.P.; Hameyer, K. Fault Diagnosis of Bearing Damage by Means of the Linear Discriminant Analysis of Stator Current Features From the Frequency Selection. *IEEE Trans. Ind. App.*, **2016**, *52-5*, 3861-3868.
30. Montechiesi, L.; Cocconcelli, M.; Rubini, R. Artificial immune system via Euclidean Distance Minimization for anomaly detection in bearings. *Mechanical Systems and Signal Processing*, **2016**, *76*, 380-393.
31. Wang, D. An extension of the infograms to novel Bayesian inference for bearing fault feature identification. *Mechanical Systems and Signal Processing*, **2016**, 80, 19–30.
32. Frosini, L.; Bassi, E.; Fazzi, A.; Gazzaniga, C. Use of the stator current for condition monitoring of bearings in induction motors in Proceedings of the 2008 International Conference on Electrical Machines, 2008, pp. 1-6.
33. Morinigo-Sotelo, D.; Duque-Perez, O; Perez-Alonso, M. Assessment of the lubrication condition of a bearing using spectral analysis of the stator current in International Symposium on Power Electronics Electrical Drives Automation and Motion (SPEEDAM), 2010, pp. 1160 - 1165.
34. Domingos, P. *The Master Algorithm*. Basic Books, New York, 2015.
35. Benbouzid, M. E. H.; Kliman, G. B. What stator current processing based technique to use for induction motor rotor faults diagnosis? *IEEE Trans. Energy Convers.*, **2003**, *18-2*, 238–244.
36. James, G.; Witten, D.; Tibshirani, T.H.R. *An introduction to statistical learning* Springer 2013.
37. Amaral Santos, P.L.; Imangaliyev, S.; Schutte, K.; Levin, E. Feature selection via co-regularized sparse-group Lasso. in *Machine learning, optimization and big data*, P.M Pardalos, P. Conca, G. Giuffrida and G. Nicosi, Eds. Springer, 2016, pp. 118-131.
38. Tibshirani, D R. Regression shrinkage and selection via the lasso: a retrospective. *J. R. Statist. Soc. B*, **2011**, *73-3*,273–282.
39. Zhang, X. Regularization. In *Encyclopedia of Machine Learning*. 2010 Edition. | Editors: Claude Sammut, Geoffrey I. Webb. Springer, Boston, MA, 2010

443 40. Kleinbaum, G.; Klein, M. *Logistic Regression*. 3rd ed.; Spinger Science+Business Media New York, 2010.
444 41. Pampel, F.C. *Logistic Regression. A Primer*. Sage Publications. 2000.