Fault detection of broken rotor bar in LS-PMSM using random forests

Juan C. Quiroz[a], Norman Mariun[b,c], Mohammad Rezazadeh Mehrjou[b,c,*], Mahdi Izadi[d,e], Norhisam Misron[c], Mohd Amran Mohd Radzi[b,c]

[a] Sunway University, Bandar Sunway, Selangor, Malaysia
[b] Centre of Advanced Power and Energy Research (CAPER), Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[c] Department of Electrical and Electronic Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[d] Centre for Electromagnetic and Lightning Protection Research (CELP), Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[e] Department of Electrical Engineering, Firoozkooh Branch, Islamic Azad University, Firoozkooh, Iran

ARTICLE INFO

ABSTRACT

This paper proposes a new approach to diagnose broken rotor bar failure in a line start-permanent magnet synchronous motor (LS-PMSM) using random forests. The transient current signal during the motor startup was acquired from a healthy motor and a faulty motor with a broken rotor bar fault. We extracted 13 statistical time domain features from the startup transient current signal, and used these features to train and test a random forest to determine whether the motor was operating under normal or faulty conditions. For feature selection, we used the feature importances from the random forest to reduce the number of features to two features. The results showed that the random forest classifies the motor condition as healthy or faulty with an accuracy of 98.8% using all features and with an accuracy of 98.4% by using only the mean-index and impulsion features. The performance of the random forest was compared with a decision tree, Naïve Bayes classifier, logistic regression, linear ridge, and a support vector machine, with the random forest consistently having a higher accuracy than the other algorithms. The proposed approach can be used in industry for online monitoring and fault diagnostic of LS-PMSM motors and the results can be helpful for the establishment of preventive maintenance plans in factories.

1. Introduction

Electrical motors convert electricity to mechanical energy. They account for two thirds of the total electricity use in industrial sites [1]. As a consequence, electrical machine manufacturers continuously strive to reduce the amount of energy used by motors. The standard IEC/EN 60034-30:2008 proposes IE4 as the highest efficiency for motors [2]. A LS-PMSM consists of a stator and a hybrid rotor. The rotor is comprised of an electricity conducting squirrel-cage and pairs of permanent magnet poles. The efficiency of LS-PMSMs stems from the combination of elements from permanent magnet synchronous motors and induction motors. The LS-PMSM provides (1) high efficiency, similar to permanent magnet synchronous motors, and (2) high starting torque, similar to induction motors [3].

Failures in electrical motors are common and difficult to prevent because motors are generally operated in industrial sites with different types of stress causing failures in various motor parts [4]. This has led to research on methods for early detection of failure in motors, to prevent motor inefficiencies and motor shutdown. In particular, rotor

faults are significant because they exacerbate failures in other parts of the motor [5]. Various sensing techniques have been developed for broken rotor bar detection in electrical motors [5]. For instance, motor current signature analysis (MCSA) is a widely used technique due to its low cost and non-invasive nature [6]. In MCSA, the steady state current of a running motor is collected and recorded. From the recorded signal, features are extracted from the time domain, frequency domain, or time-frequency domain. These features are then used to make a diagnosis of the motor.

Fault analysis in induction motors has been widely applied. MCSA has been used to analyze faults in induction motors, such as rotor faults, bearing faults, eccentricity, misalignment, and stator faults [7–11]. Similar techniques have also been used to analyze vibration [12–16] and acoustic [17] signals of induction motors. The limitation of prior work is that most fault analysis has been applied to induction motors, electrical motors, fans, and gear boxes [7–17]. Yet, fault analysis in LS-PMSMs has been limited to a smaller set of faults, such as rotor faults, static eccentricity faults, and demagnetization [18–21]. Fault analysis in LS-PMSMs also suffers from a number of shortcomings: (1) the use of

mathematical and simulated models to analyze faults, as opposed to using an LS-PMSM machine to collect data for fault detection; (2) the use of steady-state current for fault analysis; and (3) lack of machine learning algorithms for fault detection.

This paper makes three contributions. First, we used an LS-PMSM machine to collect current data while subjecting the motor to different loads. The rotor faults in our LS-PMSM machine were created by physically damaging the rotors of the LS-PMSM. We also analyzed the LS-PMSM starting with an initial load, as opposed to introducing a load after the motor had reached steady state, which is the common practice in prior work [7–11].

Second, we analyzed the transient current from when the motor is started. That is, we started the motor from standstill and waited for the motor to reach steady state, with the current from this transient period used for our analysis. Prior research uses the current from the steady state for fault analysis [7–11]. Finally, our third contribution is that our work is the first to apply machine learning for rotor fault detection in LS-PMSMs. We used random forests for the detection of rotor faults, and assessed the effectiveness of random forests by comparing with a decision tree, a Naïve Bayes classifier, logistic regression, linear ridge, and a support vector machine. To train these machine learning algorithms, we extracted thirteen time domain features from the transient current signal of the LS-PMSM, with the selection of the features based on prior work [22,23]. While machine learning methods have been used for fault detection in induction motors [7–11], to the best of our knowledge this is the first work to present fault analysis in LS-PMSMs by comparing various machine learning algorithms and using features extracted from the transient current signal to train and test these algorithms.

2. Fault detection with machine learning

A random forest is a machine learning algorithm consisting of a number of independent decision trees [24]. A decision tree classifies an instance by testing attributes of the instance at each node of the tree [25]. Each node tests a particular attribute, with the leaves of the tree representing the output labels. Moving down a particular branch of a tree tests particular attributes at each node in order to arrive to an output label. A decision tree is typically built following a greedy approach, with the attribute/feature that results in the best split of the training data being used for the root node, and subsequently the attributes/features that result in the next best splits being used in the children nodes.

In contrast to a decision tree, a random forest uses bagging to build the decision trees in the forest [24]. In bagging, T bootstrap sets are made by sampling with replacement N training examples from the training set, with T indicating the number of trees in the forest. Only 2/3 of each bootstrap set are used to build each tree, with the remaining 1/3, referred to as the out-of-bag data, used to get an estimate of the classification error of each tree. Fig. 1 illustrates the process of building a random forest with T trees.

In a typical decision tree, the greedy approach to building the tree can result in cases where the weaker features are not used at all. A random forest addresses this by choosing the best split in each node from a random subset of all the available features [24]. The random feature subset used for determining best node splits allows the weaker features to be represented in the random forest. Trees are grown to maximum length and without pruning to get low bias. Low correlation between the trees in the random forest is achieved by randomization as a result of the bootstrap samples and the random selection of features at each split [26]. Random forests have performed well in applications of fault diagnosis in rotating machinery [10,24].

Once the random forest has been built, an instance (x) is classified by passing the instance to each decision tree in the random forest (Fig. 2). Each decision tree classifies the instance by following a particular branch of the tree depending on the outcome from each node. The output of the random forest is then decided by taking the majority
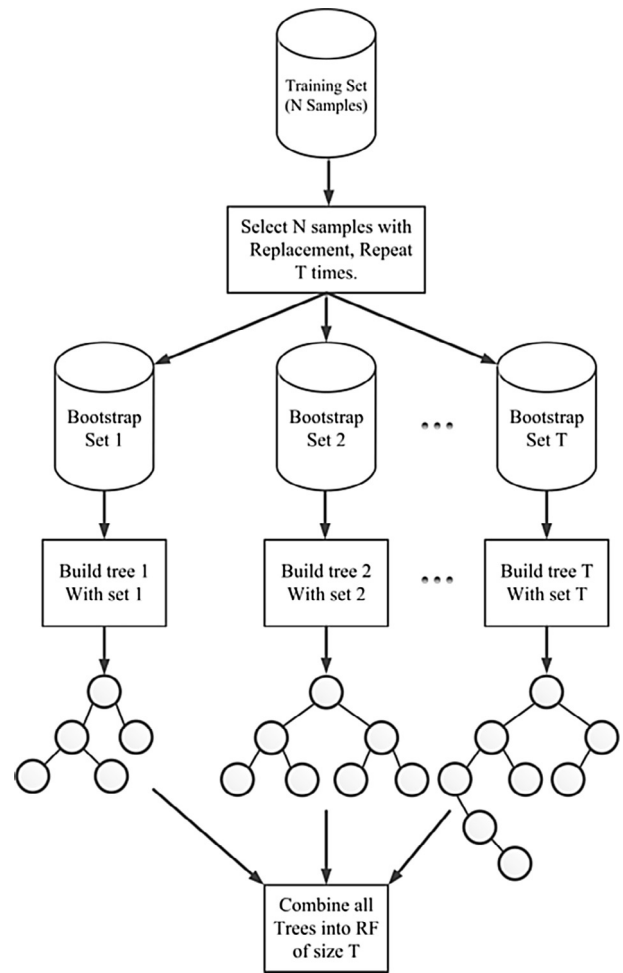

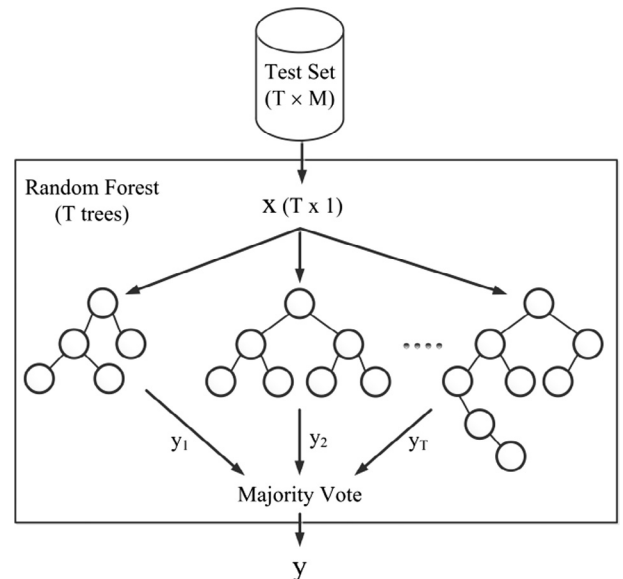**Fig. 1.** Building a random forest with bagging.

Fig. 2. Classifying an instance from the test set by passing the instance to each tree in the forest, and combining the outputs from all the trees using majority vote.

of the outputs from each tree. That is, the output of each tree is considered a vote, with the majority vote determining the output of the random forest.

Random forests provide a way to perform feature selection by using

the importance of each feature derived while building the decision trees [24]. Every non-leaf node in a decision tree is a decision node which tests a single attribute, and based on the decision splits the data. Averaging the impurity decrease for each feature over all the trees in the random forest results in a score, this yields the importance for each feature. We used feature importances to reduce the number of features from 13 features to two features, as further described below.

Random forests have been used for fault monitoring and diagnosis of induction motors and gear boxes, but not for LS-PMSMs. Niu et al. collected data from 21 sensors to detect vibration, current, voltage, and flux signals from an induction motor [7,8]. They extracted features from the time and frequency domains, and compared the performance of support vector machines, linear discriminant analysis, k nearest neighbors, random forests, and the adaptive resonance theory-Kohonen neural network, with random forests performing the best. In [9,10], random forests was used for fault detection, with the number of trees and the number of features selected at each node split optimized by a genetic algorithm. Karabadji et al. used the Weka machine learning library to compare random forests with various types of decision trees on the vibration signals of an industrial fan connected to an electrical motor [12,13]. They also used a genetic algorithm to optimize the type of tree to use and the choice of training and validation sets. In [17], Pandya et al. used acoustic signals for rolling element fault detection. They extracted features using Empirical Mode Decomposition, with modified k nearest neighbors outperforming random forests. Seera et al. proposed a hybrid model combining a Fuzzy Min-Max neural network (FMM) and a random forest [11]. They compared their hybrid model with FMM, a CART decision tree, and a hybrid FMM-CART ensemble. Cerrada et al. used random forests for multi-class fault diagnosis in spur gearboxes, with a genetic algorithm used to select the best subset of features out of 359 features in order to maximize diagnosis accuracy [14].

Our work is the first to use the transient start up current for fault detection. This is in contrast to prior work which has used the steady state current for fault analysis [7–11]. In addition, all applications of random forests to fault detection have been limited to induction motors and they have not been applied to LS-PMSMs. The type of electrical motor has a significant influence on fault detection of the motor due to the differences in structure [18]. Thus, the fault monitoring from induction motors cannot be generalized to LS-PMSMs due to the differences in these two types of motors.

3. Line start permanent magnet synchronous motors

An important obstacle for ordinary PMSMs is that they need an inverter for starting, which is not economical for many single speed applications. The squirrel-cage equipped permanent magnet motor-the LS-PMSM-provides a high efficiency motor with high starting capability, but without the need of a drive system [3,27]. LS-PMSMs can now reach super premium efficiency levels [28]. The structure of LS-PMSM comprises (1) a single or three-phase stator similar to an induction motor and (2) a hybrid rotor containing conductor bars and permanent magnets. The squirrel-cage bars in an electrical machine produce adequate high starting torque when the motor is run from standstill. Similar to asynchronous motors, squirrel-cage bars in LS-PMSM develop the startup performance during motor run up by enabling the rotor to have direct-on-line movement. When the load on the motor is unbalanced or the rotation speed is fluctuated, the squirrel-cage bars lessen the counter-rotating fields of the air gap, which otherwise would lead to significant losses [29].

Fig. 3 depicts the cross-section of one pole of a three-phase, four pole LS-PMSM. The rotor bar produces the starting torque as a result of the induction current in the bars establishing an electromagnetic field, which interacts with the rotatory field and subsequently pulls the motor toward the synchronism. Despite the induced current, the copper loss in the bars is negligible because of the synchronous operation.
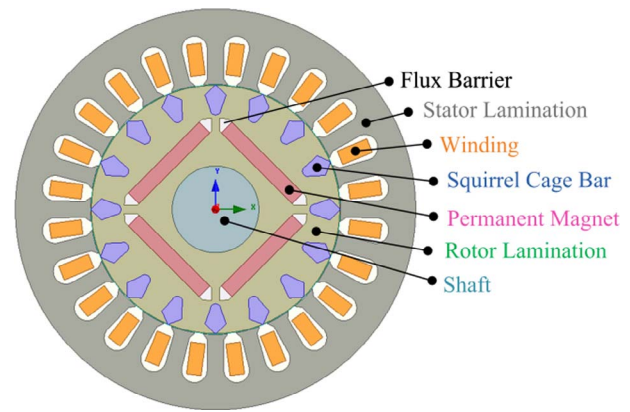

Fig. 3. The structure of three phase LS-PMSM (Four Pole).

Broken rotor bar is a major fault in squirrel cage motors. The presence of any damage in the rotor bar brings about secondary failures in the motor, leading to serious malfunctions of the motor. Once a rotor bar breaks, the condition of the neighboring bars also deteriorates over time due to the increased stress. Moreover, any defect in the bar influences the flux distribution. Thus, our work focuses on the detection of broken rotor bar faults.

Our work is most closely related to the work in [29,30]. In [30], Merjou et al. present broken rotor bar fault detection in LS-PMSMs, but their results are obtained from modeling with a simulation software using FEM at healthy condition and under fault. Merjou et al. do use transient current signals for their analysis, but these transient current signals are part of the simulation and not obtained from an actual LS-PMSM machine. Their analysis is limited to analyzing the current spectrum in the time domain by extracting statistical features. In [31], the Hilbert transform was used to extract the envelope of the current signal. From the envelope, time domain features were extracted, such as mean-index, RMS, skewness, kurtosis, among others. We improve on the work by Merjou et al. by conducting rotor bar fault detection on an LS-PMSM machine, whereas in [30,31] a software simulation is used for the fault analysis. We also take the additional step of including a machine learning algorithm trained with features from a healthy motor and a motor having a broken rotor bar. No machine learning method is presented in [30,31].

The remaining prior fault detection in LS-PMSMs includes eccentricity faults and demagnetization. In [18], Karami et al. analyze the eccentricity fault during the steady state operation of an LS-PMSM motor. The analysis is done using a software simulation of a three-phase LS-PMSM. Their work also uses analysis on the frequency domain using power spectral density (PSD) analysis. In [27], the irreversible demagnetization of an LS-PMSM is analyzed using transient analysis and the two-dimensional finite-element method. The analysis and results are based on a model, and not a physical LS-PMSM. In [21], the two-dimensional time-stepping finite element method is used to analyze the transient performance of a LS-PMSM simulated model. In [32], Lu et al. examined the behavior of an LS-PMSM during demagnetization condition along with the causes leading to the demagnetization. While a LS-PMSM machine is used to gather parameters and to understand the performance of the machine, the actual analysis is done on a mathematical model of the machine. Lu et al. were also one of the first to introduce machine learning, specifically a neural network, to detect demagnetization in LS-PMSM [32]. In [33–35], demagnetization in LS-PMSMs is also analyzed by using a simulation model.

In contrast to prior work, we used an LS-PMSM machine to collect current data for fault analysis, whereas prior work has solely used software simulation of LS-PMSMs. We also used transient current signal for our fault detection, as opposed to using the steady-state current. Lastly, prior fault analysis has studied demagnetization, whereas we
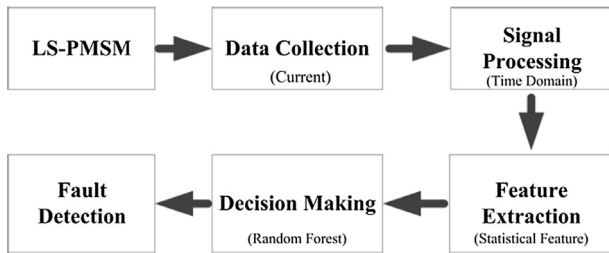
Fig. 4. The flowchart diagram of fault diagnosis system.

focus on rotor faults.

4. Fault detection algorithm

Fig. 4 presents a flowchart of the fault detection method applied in this research. First, the test workbench was set with the LS-PMSM having two different conditions: (1) healthy and (2) faulty—one broken rotor bar. The stator current signal was collected during the motor startup using a Hall-Effect current sensor. Next, the acquired current signals were preprocessed by truncating the data to the first 40 periods of the current signal. The data was also cleaned by ensuring that the start and stop points of the data fell in zero. In the next stage, 13 time domain features were extracted (discussed below) from the current signal. Finally, the features were used to train and test the random forest.

4.1. Setup configuration

Fig. 5 shows the schematic of the test facilities used to examine the broken rotor bar fault. The system consisted of a three phase power supply, an LS-PMSM, a torque and speed sensor, mechanical couplings, a magnetic powder brake acting as a load, a DC source, a Hall-Effect current sensor, an oscilloscope, and a computer.

Experimental tests were performed on a two-pole-pair LS-PMSM to evaluate the accuracy of fault detection achieved through the integration of time domain features and a random forest algorithm. Experimental data was directly obtained from the test motors by using the test bench shown in Fig. 6(a). The magnetic powder brake was coupled to the LS-PMSM to produce four different levels of starting loads: 0 N m, 0.5 N m, 1.0 N m, and 1.5 N m. The brake was controlled with current from the DC power supply. The broken rotor bar fault was made by drilling into the rotor bars, with an example shown in Fig. 6(b). Table 1 lists the parameters of the LS-PMSM used in this research.

In fault monitoring of motors, stator current signal is commonly used due to its noninvasive nature, making it easy to measure and analyze in industry. In LS-PMSM, the squirrel-cage bars carry appreciable current during the startup stage, while little or no current when the machine operates in steady-state [36]. Thomson et al. proposed monitoring and analyzing the stator current during motor startup [37]. Transient current is an accurate method for detection of the failure in
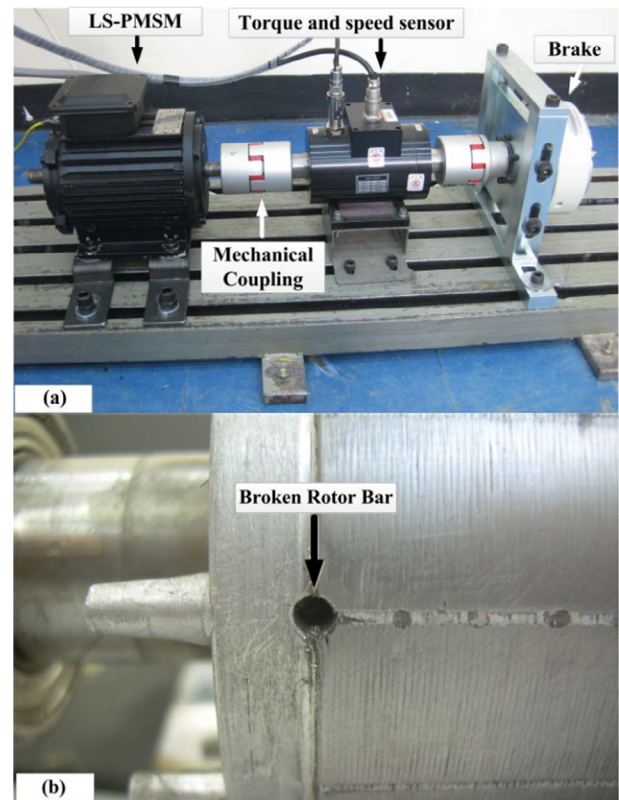

Fig. 5. Schematic of the LS-PMSM test facility.

Fig. 6. (a) Experimental test bench and (b) Rotor with one bar breakage.

Table 1
Motor specifications.

Parameters	Value
Rated output power (HP)	1
Rated voltage (V)	415
Rated frequency (Hz)	50
Rated torque (N m)	4.8
Rated speed (rpm)	1500
Rated current (A)	1.3
Starting torque	2.3
Number of poles	4
Connection	Y
Number of stator slots	24
Number of rotor slots	16

an induction machine [37,38]. Broken bars generate extra components in the stator current that depend on the rotor speed, and these can be measured during the motor start up. However, these extra components fade away or coincide with other components, like saturation-induced components that do not contain information related to the fault, when the motor reaches steady state [39]. One of the challenges of using the current during startup, and one of the main reasons why it has not been commonly used for fault detection, is the complexity of extracting fault characteristics as the startup current is extremely unsteady and its period is quite short [40]. Recent analysis of transient current signature in induction motors [28,29] has shown that transient stator current is independent of load conditions that make it suitable for fault detection. In [30,31], the effect of load for fault detection in induction motors during the transient state was not considered. However, in LS-PMSMs the starting torque decreases whenever there is a broken rotor bars [32]. In this paper, we also consider the effect of load on the starting time.

The current signal data was acquired with the Pico Scope 4424 oscilloscope and its accompanying software. For each test, the current
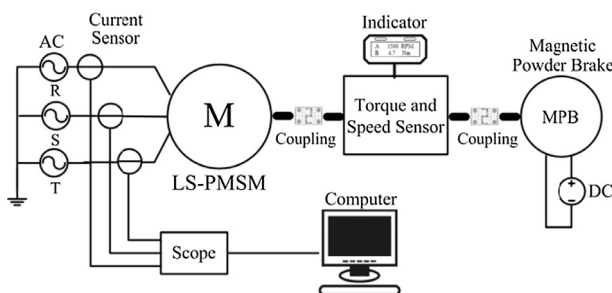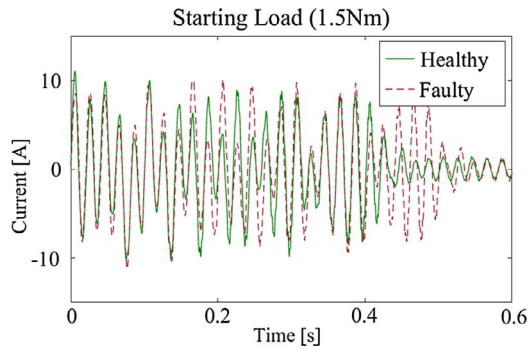
Fig. 7. Experimental current signal in high-load condition for a healthy and a faulty motor.

signal in one phase was recorded. The stator current signal is suitable for the acquisition system, which employs a Hall sensor (LEM-LTS25-NP) for this signal condition. A sampling rate of 5 kHz and a resolution of 12b were used for recording the signals. The number of 40 periods of startup current signal was selected for feature calculation, with approximately 4K sampled data used. We used the period of signal instead of sampled data because the fundamental frequency is not fixed in the normal electrical supply. To analyze the data, both the transient state and a portion of the steady state of the current signal were considered.

Forty tests were performed for each condition. A total of 320 tests were done based on healthy and faulty motor with four different levels of starting load. The torque was measured by a Dacell-TRB-10Ktorque meter, with the speed measured by a MP-981. Fig. 7 shows the current signal extracted from healthy and faulty motors running under high load conditions.

4.2. Signal processing and feature calculation

After preprocessing, those features related to broken rotor bar was extracted from the raw signal (time domain signal). Statistical features like average, variance, skewness, and kurtosis were used as a quick test for changes in the pattern of signals. These parameters are commonly used for statistical analysis of the current and vibration signals in the time domain [41]. If variation in the condition of the motor causes a change in the current signal, monitoring this signal can provide detection information. The statistical features may be dimensional or non-dimensional. Dimensional parameters include mean-index, Root Mean Square (RMS), Root-sum-of-squares level (RSS), Peak–Peak value and Energy. Non-dimensional parameters include pulse index, waveform index (Shape Factor), impulsion index, peak index (Crest factor), tolerance index (Margin factor), skewness index, and kurtosis index [42]. Han et al. compared the features obtained from time domain analysis of the steady state current signal and indicated the ability and efficiency of these features for detection of different faults [43]. The definitions of the features used in current work are the following:

Dimensional parameters:
Mean-index

$$X_{Mean} = \frac{1}{N} \sum_{i=1}^{N} X_i \tag{1}$$

RMS

$$X_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} X_i^2} \tag{2}$$

RSSQ

$$X_{RSS} = \sqrt{\sum_{i=1}^{N} |X_i^2|} \tag{3}$$

Peaktopeak

$$X_{PP} = max(X) - min(X) \tag{4}$$

Energy

$$X_{Energy} = \sum_{i=1}^{N} X_i^2 \tag{5}$$

Non-dimensional parameters:
Shape Factor

$$X_{Wi} = X_{RMS} \bigg/ \frac{1}{N} \sum_{i=1}^{N} |X_i| \tag{6}$$

Impulsion

$$X_{Ii} = max|X| \bigg/ \frac{1}{N} \sum_{i=1}^{N} |X_i| \tag{7}$$

Crest factor

$$X_{Pi} = max|X|/X_{RMS} \tag{8}$$

Margin factor

$$X_{Ti} = max|X| \bigg/ \left(\frac{1}{N} \sum_{i=1}^{N} |X_i|^{1/2} \right)^2 \tag{9}$$

Peak-to-average power ratio

$$X_{PA} = (max(X))^2/(X_{RMS})^2 \tag{10}$$

Variance

$$X_{Va} = \frac{1}{N-1} \sum_{i=1}^{N} |X_i - X_{Mean}|^2 \tag{11}$$

Skewness

$$X_{Si} = \frac{1}{N} \sum_{i=1}^{N} \frac{(X - X_{Mean})^3}{\sigma^3} \tag{12}$$

Kurtosis

$$X_{Ki} = \frac{1}{N} \sum_{i=1}^{N} \frac{(X - X_{Mean})^4}{\sigma^4} \tag{13}$$

For the equations presented, X is a signal, N is number of sampled data points of the signal, and σ is the standard deviation calculated as follows:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X - X_{Mean})^2} \tag{14}$$

4.3. Experimental setup for classification

We used the Python Scikit-learn library for the random forest implementation [44], to conduct the test and training, and for the comparison with the other machine learning algorithms. In this implementation, the random forest combines the decision trees by averaging their probabilistic prediction, instead of having each decision tree vote for a single class [44]. We used the Gini impurity to measure the quality of a split, and selected among $\sqrt{p}$ maximum random features for each node split, where p is the total number of features. Random forests were tested with 10, 100, 200, 500, and 1000 trees. We further compared the performance of the random forest to a single CART decision tree [25], Gaussian Naïve Bayes classifier [45], logistic regression [46], linear ridge classifier [47], and a support vector machine (SVM) with a radial basis function (RBF) kernel [48]. For all experiments, we measured out of sample performance using the Area under the ROC Curve (AUC) from 5-fold cross-validation. After training the random forest, we used the feature importances for feature selection [14,49]. The results show how the accuracy of the random forest changed as a factor of the number of features used, starting with the 13

Table 2
The accuracy results of random forest for broken rotor bar detection based on different number of features. Means and standard deviations (STD) over folds are reported.

	Features		All features	Mean-index, Impulsion, Shape Factor	Mean-index, Impulsion
Accuracy	10 Trees	mean	99.5	99.3	99.3
		STD	0.007	0.008	0.012
	100 Trees	mean	99.6	99.6	99.6
		STD	0.007	0.006	0.006
	200 Trees	mean	99.8	99.6	99.6
		STD	0.003	0.006	0.006
	500 Trees	mean	99.8	99.6	99.6
		STD	0.003	0.006	0.006
	1000 Trees	mean	99.7	99.6	99.6
		STD	0.005	0.006	0.006

maximum features and subsequently removing the less important features.

5. Results and discussion

Table 2 presents the results of a random forest algorithm trained with different feature sets and number of trees in the forest. First, we used all 13 features to train and test the random forests. The performance for a random forest with 100 trees was 99.6%. However, the performance remains consistent independent of the number of decision trees. Next, we reduced the number of features by using the feature importances of the random forest. Fig. 8 shows a plot of the importance of each of the 13 features. The three most informative features were mean-index, impulse factor, and shape factor. These three features were used to train and test another set of random forests. The performance of a random forest with 100 trees and using these three features was also 99.6%, the same as when using all 13 features. In addition, the performance of the random forest when using three features is independent of the number of trees in the forest. Lastly, using only the two top features, mean-index and impulse factor, gives results comparable to using only three features.

Given that two features are sufficient to yield high performances and that the performance of the random forest is independent of the
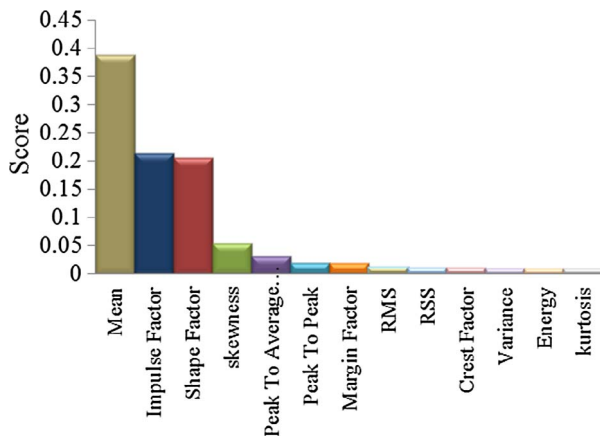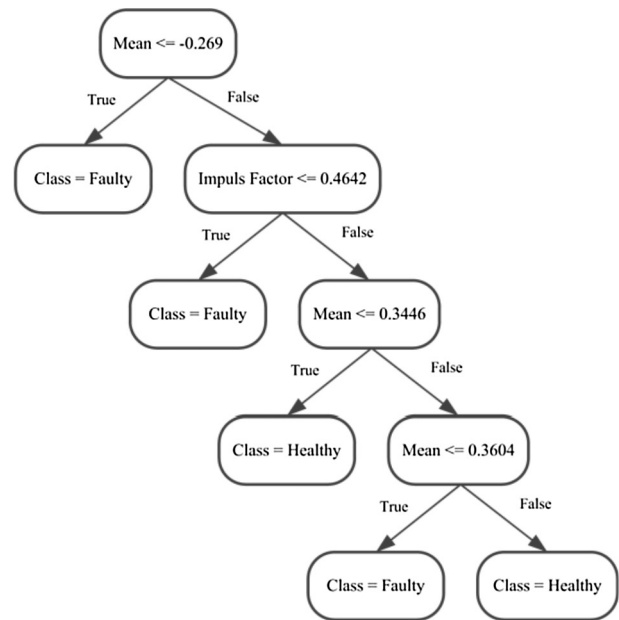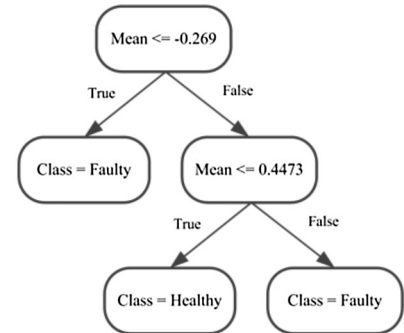

Fig. 8. Feature importance for the 13 features.

(a)

(b)

Fig. 9. (a) CART decision tree built and trained using mean-index and impulse factor, and (b) CART decision tree built and trained using mean-index and impulse factor, but only mean-index is used in the nodes of the decision tree.

number of trees, we hypothesized whether all the trees in the random forest were the same when using only two features. To test this hypothesis, we trained and tested a single CART decision tree using different random seeds, and examined the resulting topologies of the decision trees.

Fig. 9 illustrates two example decision trees generated using different random seeds. The first decision tree (Fig. 9(a)) has depth four and a total of nine nodes. The decision nodes of the first decision tree test different values of the mean-index and the impulse factor to determine whether the motor is healthy or faulty. For example, following one path of the tree indicates that if the mean-index of the startup current sample is less than or equal to -0.269, then the motor is faulty. Another path of the decision tree shows that if the mean is greater than -0.269, the impulse factor is less than or equal to 0.4642, and the mean-index is less than 0.3446, then the motor is healthy. In contrast, the second decision tree (Fig. 9(b)) has depth two and a total of five nodes. In particular, the second decision tree relies solely on the mean-index feature to determine whether the motor is faulty or healthy. This shows that even when using only two features, the random forest will contain diverse trees which make it less likely to over fit to the training set.

To assess the effectiveness of random forest, we tested baseline algorithms on our transient current data set from the LS-PMSM. The AUC

Table 3

AUC comparison for broken rotor bar data, means and standard deviations (STD) over folds are reported.

Classifier	All features		Mean-index, impulsion, shape factor		Mean-index, impulsion	
	mean	STD	mean	STD	mean	STD
Random forest	**99.6**	**0.01**	**99.6**	**0.01**	**99.6**	**0.01**
CART	98.8	0.01	97.5	0.03	96.9	0.04
Naïve Bayes	96.9	0.05	96.3	0.07	95	0.1
Logistic regression	94	0.09	93.8	0.1	93.8	0.1
Linear ridge	93.9	0.1	93.8	0.1	93.8	0.1
SVM	93.8	0.1	94.4	0.1	95.2	0.1

for the random forest (with 100 trees) compared to the other algorithms is reported in Table 3. We tested all the algorithms using all 13 features, using only three features (mean-index, impulsion factor, and shape factor), and using only two features (mean-index and impulsion factor). The random forest outperforms all of the other classifiers. For some of the methods, such as CART and Naïve Bayes, reducing the number of features leads to minimal drop in performance. For the SVM, reducing the number features results in a minimal increase in performance.

The approach presented here can detect whether the LS-PMSM is healthy or faulty by using a random forest to analyze transient current signal data from the LS-PMSM. Current signature analysis has the advantage of being non-invasive. In addition, feature importances derived from building the random forest was be used to reduce the number of features, with the random forest still maintaining high prediction accuracy. The performance of the random forest was independent of the number of trees in the forest. Finally, even though the random forest performed the best, as highlighted bold in the table, all the other methods performed well with accuracies of over 90%.

In this case study, random forest outperformed the other classifiers. A random forest combines the predictions of the decision trees in the forest by averaging their predictions. This combination of the individual estimations made by the trees in the forest lowers the variance of the model, which gives it a predictive advantage over a single, simple classifier [24].

6. Conclusions

In this paper, we presented a new approach for the fault detection of rotor bars in LS-PMSM using random forests. We collected the transient current signal during the startup phase of a healthy LS-PMSM machine and an LS-PMSM machine with rotor faults. During the data collection, the LS-PMSM machines were subjected to different load conditions. Experimental results indicate the validity and reliability of the random forest fault detection method. The approach attains a high correct rate of diagnosis of 98.8%. The accuracy of the random forest was found to be independent of the number of trees. With feature importances, we identified mean-index and impulsion as the two most important features. With these two features alone, the random forest achieved a performance of 98.4%. Comparing the random forest to other machine learning algorithms showed that random forest performed the best. However, all algorithms achieved accuracies over 90%, which gives practitioners the confidence and flexibility to choose various types of algorithms and not be limited to random forests.

Acknowledgments

References

[1] P. Walde, C. Brunner, Energy-efficiency policy opportunities for electric motor-driven systems, IEA Energy Pap. (2011), http://dx.doi.org/10.1787/5kgg52gb9gjd-en.

[2] International Electrotechnical Commission, IEC 60034-30-1: Efficiency Classes of Line Operated AC Motors (IE Code), 1.0, 2014. < https://webstore.iec.ch/publication/136 > (accessed March 28, 2016).

[3] A. Hassanpour Isfahani, S. Vaez-Zadeh, Line start permanent magnet synchronous motors: challenges and opportunities, Energy. 34 (2009) 1755–1763, http://dx.doi.org/10.1016/j.energy.2009.04.022.

[4] A.H. Bonnett, G.C. Soukup, Cause and analysis of stator and rotor failures in three-phase squirrel-cage induction motors, IEEE Trans. Ind. Appl. 28 (1992) 921–937, http://dx.doi.org/10.1109/28.148460.

[5] M.R. Mehrjou, N. Mariun, M. Hamiruce Marhaban, N. Misron, Rotor fault condition monitoring techniques for squirrel-cage induction machine—a review, Mech. Syst. Signal Process. 25 (2011) 2827–2848, http://dx.doi.org/10.1016/j.ymssp.2011.05.007.

[6] D. Hwang, Y. Youn, J. Sun, Y. Kim, Robust diagnosis algorithm for identifying broken rotor bar faults in induction motors, J. Electr. Eng. Technol. 8 (2014) 742–749, http://dx.doi.org/10.5370/JEET.2013.8.?.742.

[7] G. Niu, B.S. Yang, Fault Diagnosis: Performance Evaluation of Multi-sensors Signals and Classifiers for Faults Diagnosis of Induction Motor, 한국소음진동공학회 학술발표논문집 2006 (2006) 1–6.

[8] G. Niu, J.-D. Son, A. Widodo, B.-S. Yang, D.-H. Hwang, D.-S. Kang, A comparison of classifier performance for fault diagnosis of induction motor using multi-type signals, Struct. Heal. Monit. 6 (2007) 215–229, http://dx.doi.org/10.1177/1475921707081110.

[9] X.D.T. Han, B.-S. Yang, S.-J. Lee, in: J. Mathew, J. Kennedy, L. Ma, A. Tan, D. Anderson (Eds.), Application of Random Forest Algorithm in Machine Fault Diagnosis, Springer London, 2006, pp. 779–784. < http://link.springer.com/chapter/10.1007/978-1-84628-814-2_82 > (accessed February 11, 2016).

[10] B.-S. Yang, X. Di, T. Han, Random forests classifier for machine fault diagnosis, J. Mech. Sci. Technol. 22 (2008) 1716–1725, http://dx.doi.org/10.1007/s12206-008-0603-6.

[11] M. Seera, C.P. Lim, S. Nahavandi, C.K. Loo, Condition monitoring of induction motors: a review and an application of an ensemble of hybrid intelligent models, Expert Syst. Appl. 41 (2014) 4891–4903, http://dx.doi.org/10.1016/j.eswa.2014.02.028.

[12] N.E.I. Karabadji, H. Seridi, I. Khelf, L. Laouar, in: A. Abelló, L. Bellatreche, B. Benatallah (Eds.), Decision Tree Selection in an Industrial Machine Fault Diagnostics, Springer Berlin Heidelberg, 2012, pp. 129–140. < http://link.springer.com/chapter/10.1007/978-3-642-33609-6_13 > (accessed March 28, 2016).

[13] N.E.I. Karabadji, I. Khelf, H. Seridi, L. Laouar, in: T. Fakhfakh, W. Bartelmus, F. Chaari, R. Zimroz, M. Haddar (Eds.), Genetic Optimization of Decision Tree Choice for Fault Diagnosis in an Industrial Ventilator, Springer Berlin Heidelberg, 2012, pp. 277–283. < http://link.springer.com/chapter/10.1007/978-3-642-28768-8_29 > (accessed March 28, 2016).

[14] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, C. Li, Fault diagnosis in spur gears based on genetic algorithm and random forest, Mech. Syst. Signal Process. 70–71 (2016) 87–103, http://dx.doi.org/10.1016/j.ymssp.2015.08.030.

[15] R.K. Patel, V.K. Giri, Feature selection and classification of mechanical fault of an induction motor using random forest classifier, Perspect. Sci. (2016), http://dx.doi.org/10.1016/j.pisc.2016.04.068.

[16] H.W. Peng, P.J. Chiang, Control of mechatronics systems: ball bearing fault diagnosis using machine learning techniques, 2011 8th Asian Control Conf. IEEE, Kaohsiung, Taiwan, 2011, pp. 175–180.

[17] D.H. Pandya, S.H. Upadhyay, S.P. Harsha, Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN, Expert Syst. Appl. 40 (2013) 4137–4145, http://dx.doi.org/10.1016/j.eswa.2013.01.033.

[18] M. Karami, N. Mariun, M.R. Mehrjou, M.Z.A.Ab. Kadir, N. Misron, M.A. Mohd Radzi, Static eccentricity fault recognition in three-phase line start permanent magnet synchronous motor using finite element method, Math. Probl. Eng. 2014 (2014), http://dx.doi.org/10.1155/2014/132647 12 pages.

[19] M.R. Mehrjou, N. Mariun, M. Karami, N. Misron, M.A.M. Radzi, Performance analysis of line-start permanent magnet synchronous motor in presence of rotor fault, in: 2014 IEEE Student Conf. Res. Dev., IEEE, 2014, pp. 1–4. http://dx.doi.org/10.1109/SCORED.2014.7072978.

[20] X. Tang, X. Wang, Research of the demagnetization mechanism of line-start permanent magnet synchronous motor under operating condition of sudden reversal, in: 2014 17th Int. Conf. Electr. Mach. Syst., 2014, pp. 1981–1984. http://dx.doi.org/10.1109/ICEMS.2014.7013826.

[21] W. Lu, M. Liu, Y. Luo, Y. Liu, Influencing factors on the demagnetization of line-start permanent magnet synchronous motor during its starting process, in: 2011 Int. Conf. Electr. Mach. Syst., 2011, pp. 1–4. http://dx.doi.org/10.1109/ICEMS.2011.6073886.

[22] J. Wang, S. Liu, R.X. Gao, R. Yan, Current envelope analysis for defect identification and diagnosis in induction motors, J. Manuf. Syst. 31 (2012) 380–387, http://dx.doi.org/10.1016/j.jmsy.2012.06.005.

[23] V. Ghorbanian, J. Faiz, A survey on time and frequency characteristics of induction motors with broken rotor bars in line-start and inverter-fed modes, Mech. Syst. Signal Process. 54–55 (2015) 427–456, http://dx.doi.org/10.1016/j.ymssp.2014.08.022.

[24] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, http://dx.doi.org/10.

1023/A:1010933404324.

[25] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and Regression Trees, first ed., Chapman and Hall/CRC, New York, 1984.

[26] A. Verikas, A. Gelzinis, M. Bacauskiene, Mining data with random forests: a survey and results of new tests, Pattern Recognit. 44 (2011) 330–349, http://dx.doi.org/10.1016/j.patcog.2010.08.011.

[27] G.-H. Gyu-Hong, J. Kang, Jin Hur, H. Hyuk Nam, J.-P. Jung-Pyo Hong, G.-T. Gyu-Tak Kim, Analysis of irreversible magnet demagnetization in line-start motors based on the finite-element method, IEEE Trans. Magn. 39 (2003) 1488–1491, http://dx.doi.org/10.1109/TMAG.2003.810330.

[28] H.A. Toliyat, G.B. Kliman, Handbook of Electric Motors, second ed., CRC Press, 2004.

[29] J. Pyrhonen, T. Jokinen, V. Hrabovcova, Design of Rotating Electrical Machines, Wiley, 2009.

[30] M.R. Mehrjou, N. Mariun, M. Karami, M.A. Mohd Radzi, N. Misron, Statistical features analysis of transient current signal for broken bars fault detection in LS-PMSMs, in: Smart Instrumentation, Meas. Appl. (ICSIMA), 2015 IEEE Int. Conf., IEEE, 2015.

[31] M.R. Mehrjou, N. Mariun, M. Karami, N. Misron, M.A. Mohd Radzi, Broken rotor bar detection in LS-PMSMs based on statistical features analysis of start-up current envelope, in: Smart Instrumentation, Meas. Appl. (ICSIMA), 2015 IEEE Int. Conf., IEEE, 2015.

[32] X. Lu, K.L.V. Iyer, K. Mukherjee, N.C. Kar, Study and detection of demagnetization in line start permanent magnet synchronous machines using artificial neural network, 2012 15th Int. Conf. Electr. Mach. Syst. IEEE, Sapporo, 2012, pp. 1–6.

[33] W.f. Lu, H.s. Zhao, S. Liu, Demagnetization conditions comparison for line-start permanent magnet synchronous motors, in: 2014 17th Int. Conf. Electr. Mach. Syst., 2014, pp. 48–52. http://dx.doi.org/10.1109/ICEMS.2014.7013449.

[34] M. Niazazari, M. Mirsalim, S. Mohammadi, Analytical framework for analysis and demagnetization study of a slotted solid-rotor line-start permanent-magnet synchronous motor, in: Power Electron. Drive Syst. Technol. Conf. (PEDSTC), 2014 5th, 2014, pp. 494–499. http://dx.doi.org/10.1109/PEDSTC.2014.6799425.

[35] J.X. Shen, P. Li, M.J. Jin, G. Yang, Investigation and countermeasures for demagnetization in line start permanent magnet synchronous motors, IEEE Trans. Magn. 49 (2013) 4068–4071, http://dx.doi.org/10.1109/TMAG.2013.2244582.

[36] G.B. Kliman, Transient Rotor Fault Detection in Induction and Synchronous Motors, U.S. Patent 5,030,917, 1991.

[37] W. Thomson, N. Deans, R. Leonard, A. Milne, Monitoring strategy for discriminating between different types of rotor defects in induction motors., in: 18th Univ. Power Eng. Conf., Surrey, UK, 1983, pp. 241–246.

[38] S. Elder, J.F. Watson, W.T. Thomson, Fault detection in induction motors as a result of transient analysis, in: Fourth Int. Conf. Electr. Mach. Drives, IET, London, 1989, pp. 182–186.

[39] F. Briz, M.W. Degner, P. Garcia, D. Bragado, Broken rotor bar detection in line-fed induction machines using complex wavelet analysis of startup transients, IEEE Trans. Ind. Appl. 44 (2008) 760–768, http://dx.doi.org/10.1109/TIA.2008.921382.

[40] Z. Zhengping Zhang, Z. Zhen Ren, W. Wenying Huang, A novel detection method of motor broken rotor bars based on wavelet ridge, IEEE Trans. Energy Convers. 18 (2003) 417–423, http://dx.doi.org/10.1109/TEC.2003.815851.

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

[42] M.O. Mustafa, G. Nikolakopoulos, G. Georgoulas, Fault classification of broken rotor bars in induction motors based on envelope current analysis, in: 2015 IEEE 13th Int. Conf. Ind. Informatics, IEEE, 2015, pp. 795–800. http://dx.doi.org/10.1109/INDIN.2015.7281838.

[43] T. Han, B.-S. Yang, W.-H. Choi, J.-S. Kim, Fault diagnosis system of induction motors based on neural network and genetic algorithm using stator current signals, Int. J. Rotat. Mach. 2006 (2006) 1–13, http://dx.doi.org/10.1155/IJRM/2006/61690.

[44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[45] H. Zhang, The optimality of naive Bayes, AA 1 (2004) 3.

[46] H.-F. Yu, F.-L. Huang, C.-J. Lin, Dual coordinate descent methods for logistic regression and maximum entropy models, Mach. Learn. 85 (2010) 41–75, http://dx.doi.org/10.1007/s10994-010-5221-8.

[47] R.M. Rifkin, R.A. Lippert, Notes on Regularized Least Squares, Technical Report MIT-CSAIL-TR-2007-025, Massachusetts Institute of Technology, Cambridge, USA, 2007.

[48] T.-F. Wu, C.-J. Lin, R.C. Weng, Probability estimates for multi-class classification by pairwise coupling, J. Mach. Learn. Res. 5 (2004) 975–1005.

[49] D. Cabrera, F. Sancho, R.-V. Sánchez, G. Zurita, M. Cerrada, C. Li, R.E. Vásquez, Fault diagnosis of spur gearbox based on random forest and wavelet packet decomposition, Front. Mech. Eng. 10 (2015) 277–286, http://dx.doi.org/10.1007/s11465-015-0348-8.