

Supervised Classification Of induction Motors faults

Radouane El bouisfi¹, Lamiaa El Menzhi¹, Xavier Chimentin²

¹ National School of Applied Sciences, Abdelmalek Essaadi University, Tangier Morocco

² Institut de Thermique, Mécanique, et Matériaux (ITheMM EA7548) - UFR Sciences Reims France.

Abstract. Currently, the environmental challenges have been considered as a strategic issue for most industrial companies around the world, threatening their sustainability and profit; This leads to taking the environmental dimensions seriously and preserving natural resources well, since they are a key criterion for sustainable development. In this context, this work calls for innovative solution and new technologies to support the development and integration of environmental considerations through the implementation of an automated fault detection and diagnosis system in induction machines in order to minimize downtime, increase machine utilization rate, get an idea of remaining machine life based on artificial intelligence (AI) and the analysis of collected data. Using the Pattern Recognition methods, this system aims to support decision making in terms of defect classification, through the following process: the collection of relevant data about the stator currents of two induction machines, powered by a converter, one healthy and the other defective, through the CompactRIO device, then the analysis of the data, using programs developed under LabVIEW software, and the extraction of the indicators to form a database. Based on analysis results, several intelligent methods by classification algorithms can organize the acquired data in order to automate the diagnostic process. Ultimately, the set-up of an alert system to prevent rather than cure. The outcomes showed that the integration of predictive maintenance could help achieve an energy cost recovery equal to 10% of the total costs of an electric motor system. Hence, the premature detection of faults helps to minimize energy expenditure and achieve overall cost savings, which implies energy optimization.

* Corresponding author: elbouisfi.radouane@gmail.com

1 Introduction

Asynchronous machines are being used more and more because of their robustness [1]. They represent up to 90% of electric motors in production processes [2]. However, like any other machine, the induction machine is not immune to malfunction. It can be affected by electrical or mechanical faults in the stator and/or rotor. As part of the optimization and energy monitoring [3], it is necessary to implement a system for the detection and automated diagnosis of induction machine faults, in order to minimize downtime, increase the rate of use of the machine, and to get an idea about the remaining life of the machine.

Indeed, the comparison established between two identical machines, one healthy and the other defective, makes it possible to demonstrate the need for the premature detection of faults so as to minimize energy expenditure and achieve overall cost savings.

Overloading, abrasion, unbalanced loads, or electrical stress can harm any of the components of an induction motor. Bearing, stator, rotor, and other defects are the four categories of motor faults [4].

Since Broken rotor bar defects are feasible in the laboratory, this work is expected on the diagnosis of this failure.

The monitoring of broken bar default gets the scientist's attention; so many methods applied to detect defects such as support vector machine [5], Bayesian network [6] and clustering [7]. Numerous literature reviews are available on monitoring methods [8,9]. From all these used methods, supervised classification analysis is one of the most remarkable approaches. The light gradient boosting machine Classifier (LGBM Classifier) is one of them, one of the effective implementations of Gradient Boosting Decision Tree (GBDT), a popular machine learning algorithm.

Within the framework of rotor bars monitoring, the LightGBM classifier integrated dynamic classification processes for real-time monitoring.

The extracted features play an essential role in the classification, for that many methods used to eliminate unwanted and unimportant features, such as the relief method that is used to select features for the classification of biomedical data. After selecting features and reducing them by eliminating the uncorrelated ones, the importance of dimension reduction comes before starting the classification. In the literature, many methods have applied for dimension reduction, principal component analysis (PCA) [8] and kernel principal component analysis (KPCA) [9], to detect the defect in rolling bearings. A recently developed nonlinear dimensionality reduction technique shows its efficiency in the detection of a fault in rotary based on t-distributed stochastic neighbour embedding (t-SNE) [10].

In this research, a motor current signal is used to classify broken rotor bars faults for five conditions: a healthy rotor bar, one BRB, two BRB, three BRB and four BRB. To reduce the data dimension, the set of statistical features was extracted from the raw time-

series signals and **Feature importance scores** was applied to determine the most significant features for the classification model. Finally, three widely used classification models were trained using the discriminative feature subsets selected by **Feature importance scores**. The conclusive results and comparison indicate that the selected features converged quickly during the training state and therefore provided good performance in the classification models.

The rest of this paper is organized as follows: Section 2 presents materials and methods, including information about the experimental testbed and dataset, feature extraction and selection methods, classification techniques, and finally, the proposed method used in this study. Section 3 presents our results validating the proposed method and comparisons with existing methods. Finally, conclusions are provided in Section 4.

2 Materials and Methods

2.1 Broken rotor bar fault

Place The broken rotor bar (BRB) fault is one of the most serious failures and needs to be detected on account of its destructiveness. According to the statistics by the Institute of Electrical and Electronics Engineers (IEEE), approximately 9% of IMs faults are due to the BRB, and 8% are caused by BRB faults according to the statistics from the Electric Power Research Institute (EPRI) [11].

The experimental test data presented in this article shows the performance of the Fourier Transformed (FFT) used for the diagnosis of broken rotor bar fault of the induction machine in no-load operation, based on the analysis of spectral stator current and rotational speed. The results prove the effectiveness of this technique for detecting this type of fault by monitoring the amplitudes of the current harmonics of the induction machine and the sidebands around the fundamental frequency.

Indeed, in healthy Induction Machine, the rotor generates symmetrical currents that cause a forward rotating magnetic field at a synchronous speed. If the rotor bar is cracked, a resultant background rotating field can occur [12]. Consequently, the inverse frequency component of the rotor currents at $-sf_0$ (where s denotes the slip and f_0 denotes the supply frequency) appears. The inverse sequence components act on the stator side and lead to the frequency components of $(1 - 2ks) f_0$ (where $k = 1, 2, 3, \dots$ is any positive integer), and the produced additional frequency information mainly leads to the presence of BRB fault frequency components at $(fr \pm 2ksf_0)$ in radial and axial vibration signals [13]. As a result, a torque ripple and a speed oscillation are generated at the frequency of $2ksf_0$, which acts as a frequency component modulated on the rotation frequency fr in the vibration spectrum that are given by:

$$f_{BRB} = f_r \pm 2k.s.f_0 \quad (1)$$

where f_{BRB} is the sideband frequency related to the BRB fault. The slip s is defined by the mechanical speed (n_m) of the Induction Machine and the synchronous speed n_s as expressed [14]:

$$s = \frac{n_s - n_m}{n_s} \quad (2)$$

Further, the research shows $2sf_0$ is the main modulation component in the BRB fault vibration signals, which helps in identifying BRB faults easily and effectively [15]. In the following analysis, $2sf_0$ is explored and used as the most important fault feature for BRB faults detection.

2.2 Experimental Tested and Dataset Acquisition

The experimental tests are carried out within the Maintenance department of Daher Aerospace Tanger 2 [16]. Since the fault is achievable, two identical three-phase induction machines with a squirrel cage were used, one for the healthy state and the other for a broken bar, for the acquisition of practical signals in real time. The configuration of the two benches is represented in figure 1. The two machines are supplied by an ATV11 Speed variator [17], operating without load. Their characteristics are as follows:

Power	1.5 Kw
Nominal voltage	230/400 V
Rated current	5.10/3 A
Cos (ρ)	0.83
Rated speed	1430 tr/min
Number of pole pairs	2
Number of stator slots	48
Number of rotor bars	28

Table 1: characteristics of the induction machine

For data acquisition, the NI-9246 current input module and the NI-9234 vibration input module [18] with a piezoelectric sensor type 4384 BRUEL & KjaER [19], which are connected directly to the NI CompactROI 9040 [20], which provides access direct hardware to the I / O circuits of each module, are used to record the current of the three phases and the axial and radial vibrations respectively.

The broken rotor bar fault is achieved by drilling a small hole 5.1mm in diameter and 19mm deep at the rotor bar as shown in Figure 1. It is important to note that this procedure does not allow the complete removal of the aluminum material in the notch, which corresponds to a partial breakage of the bar.



Fig. 1. Experimental setup that collects healthy and faulty induction machine data in stator current

The sets of Broken rotor bar used in this dataset can be split into five groups based on their health state: healthy RB, one, two, three and four BRB faults.

The modes of operation that were used to validate the diagnostic procedure are listed in Table 1. In the decision space, each of them is represented by a class.

Operating mode	Types of fault	Load rate
Healthy machine	No defect	0%, 15% et 25%
A broken bar	Rotor fault (Broken Bar)	0%, 15% et 25%
Two broken bars		0%, 15% et 25%
Three broken bars		0%, 15% et 25%
Four broken bars		0%, 15% et 25%

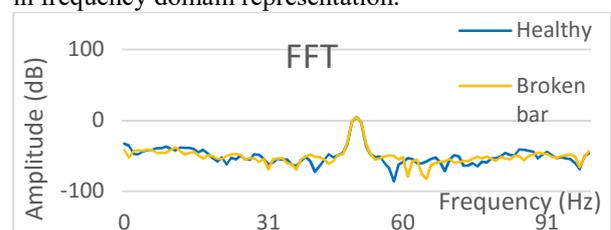
Table 2: Operating mode

All these tests were carried out in order to validate the implementation of our diagnostic process. Some of these measurements will be used to create the training set as well as test sets. These sets will be used to understand pattern recognition algorithms by determining the sensitivity of the calculated parameters and defining the best shape vector.

The data collected on the machine powered by a variable speed drive will be used to demonstrate the effectiveness of the decision algorithms. They also demonstrate the advantage of time and frequency signatures in order to obtain a good separation of the operating modes.

The dataset is a 9527 x 91 matrix including three current signals from various phases and a label added in the last column to separate the rotor bar health states.

The plot of the Fast Fourier Transform of The current signals is provided in Figure 2. It can be seen that there are very subtle differences among the signals in frequency domain representation.



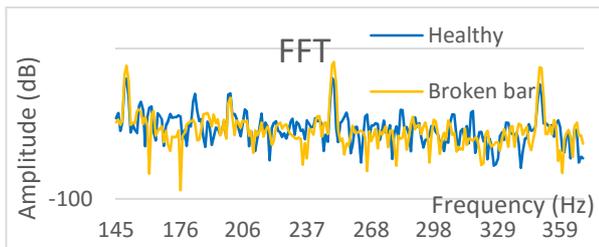


Fig. 2. The spectrum of current for healthy and defected motor

Figure 2 shows the fundamental harmonic of the stator current spectrum of the asynchronous machine operating at no-load. In the one hand, we distinguish a significant increase in the amplitudes of the components whose frequencies are: $f_0(1-2.s)$, $f_0(1+2.s)$ et $f_0(1+2.k.s)$ already present in the spectrum of the healthy machine and due to the natural dynamic eccentricity, with fs is supply frequency, s is slip and $k=1,2,3,\dots$. On the other hand, considering that the rotor is defective, the frequency lines relating to this type of defect are clearly evident in the spectrum. Moreover, the component at the frequency $f_0(3-4.s)$ does not appear in the healthy spectrum but only in the event of a fault.

2.3 Feature extraction

Feature extraction reduces the size of the original raw data by converting it into smaller, more manageable groups for further processing. Most of the time, the raw dataset contains a large number of variables, thus requiring a large amount of compute resources to process. The main goal of this method is to select and convert variables to characteristics, efficiently compacting the data that needs to be processed and accurately describing the original raw data set. Recently, online diagnostics for condition monitoring has gained attention because of its ability to detect emerging faults. However, a small amount of data is not sufficient for diagnosis, so signals measured directly are not suitable for online use. For effective fault diagnosis, a large sample of data is required. Thus, to facilitate the calculation, feature extraction becomes a crucial step that reserves critical information for final decision making.

Twenty-four classical statistical feature parameters for condition assessment were computed from time and frequency domain data: Skewness, kurtosis, Mean, Maximum, Variance, RMS, Crest factor, Standard deviation....

Statistical Feature	Equation	Statistical Feature	Equation
Mean:	$\mu = \frac{1}{N} \sum_{i=1}^N x_i$	Skewness:	$x_{skewness} = \frac{1}{N} \frac{\sum_{i=1}^N (x_i - \mu)^3}{\sigma^3}$
Median:	$median = \left(\frac{N+1}{2}\right)^{th} \text{value}$	Kurtosis:	$x_{kurtosis} = \frac{1}{N} \frac{\sum_{i=1}^N (x_i - \mu)^4}{\sigma^4}$
Standard Deviation:	$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N-1}}$	Energy:	$E = \sum_{i=1}^N x_i^2$
Variance:	$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N-1}$ (13)	RMS:	$x_{rms} = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N}}$
Sum:	$Sum = \sum_{i=1}^N x_i$ (15)	Crest factor:	$C_f = \frac{x_{max}}{x_{min}}$

Tab3. Extracted statistical features from time domains for the feature matrix (x is the current signal).

2.4 Features selection

Feature selection is the process of detecting and eliminating irrelevant, less useful, and redundant features, and of discovering the most appropriate inputs for a classification model. Feature importance scores play an important role in a predictive modelling project, including providing insight into the data, insight into the model, and the basis for dimensionality reduction and feature selection that can improve the efficiency and effectiveness of a predictive model on the problem.

Feature importance refers to a class of techniques for assigning scores to input features to a predictive model that indicates the relative importance of each feature when making a prediction.

The relative scores can highlight which features may be most relevant to the target, and the converse, which features are the least relevant. This may be interpreted by a domain expert and could be used as the basis for gathering more or different data.

Although, most importance scores are calculated by a predictive model that has been fit on the dataset. Inspecting the importance score provides insight into that specific model and which features are the most important and least important to the model when making a prediction. This is a type of model interpretation that can be performed for those models that support it.

Finally, feature importance can be used to improve a predictive model. This can be achieved by using the importance scores to select those features to delete (lowest scores) or those features to keep (highest scores). This is a type of feature selection and can simplify the problem that is being modelled, speed up the modelling process (deleting features is called dimensionality reduction), and in some cases, improve the performance of the model.

2.5 Support Vector Machines

After feature selection using Features importance score, three different classifier algorithms are applied to classify the fault signature. To begin with, Support Vector Machine (SVM) was chosen. SVM is a supervised machine learning algorithm that can be used for both classification and regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features

you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot).

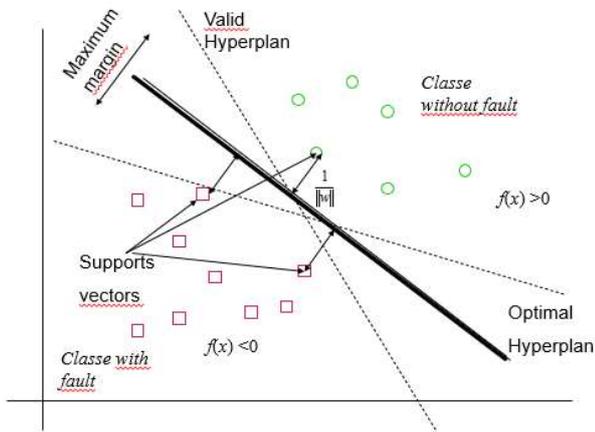


Fig. 3. Principle of SVMs

Support Vectors are simply the coordinates of individual observation. The SVM classifier is a frontier that best segregates the two classes (hyper-plane/ line).

2.6 Random Forest Classifier

For both regression and classification, random decision trees are one type of collective learning. During the training phase, this approach constructs a large number of decision trees, then defines the class that is the mode of the classification classes and predicts the mean value for regression of each individual tree. The benefit of RF over a decision tree is that it can avoid the problem of overfitting. In RF, the significant parameters are the number of trees, maximum depth, number of features in each split, and number of sample leaves. In most situations, the large number of trees aids in the computation of correct findings by providing steady results. In general, the number of characteristics determines the variety of a tree in a forest, with fewer features resulting in uncorrelated trees and more features resulting in correlated trees. As a result, while a high number of trees might enhance speed, they also add to the size and computational cost of the system. The above-mentioned parameters are evaluated over a wide range of values in this study.

2.7 LightGBM Classifier

Light Gradient Boosting Machine is a widely-used machine learning algorithm, due to its efficiency, accuracy, and interpretability. LightGBM achieves state-of-the-art performances in many machine learning tasks, such as multi-class classification [21], click prediction [22], and learning to rank [23].

Most decision tree learning algorithms develop trees by level (depth), like the following image:

Light GBM grows tree vertically while other algorithm grows trees horizontally meaning that Light GBM grows tree leaf-wise while other algorithms grow level-wise. It will choose the leaf with max delta loss to grow. When growing the same leaf, Leaf-wise algorithm can reduce more loss than a level-wise algorithm.

3 Results and Discussion

In order to evaluate the performance of the classifier models, we use feature importance to select the discriminant indicators of each class, which allows to remove the unimportant indicators using the importance of the derived characteristic. The results are shown in the graph below:

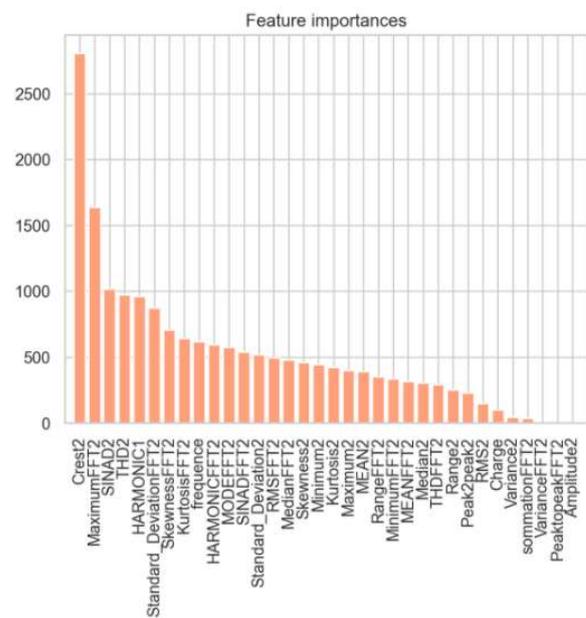


Fig. 4. Graph of feature importance

The final dataset is a 9527 x 16 matrix for each current phase including three labels added, one to separate the rotor bar health states in addition to frequency and load.

The classifier’s performance was assessed using the confusion matrix presented in Figure 5. The evaluation parameters shown in Equations provided in Table 4, were calculated from this confusion matrix.

$\text{Precision} = \frac{TP}{TP + FP} \quad (18)$	$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (19)$
$\text{Specificity} = \frac{TN}{FP + TN} \quad (20)$	$F1 = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (21)$
$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (22)$	

Table 4. Performance evaluation parameters

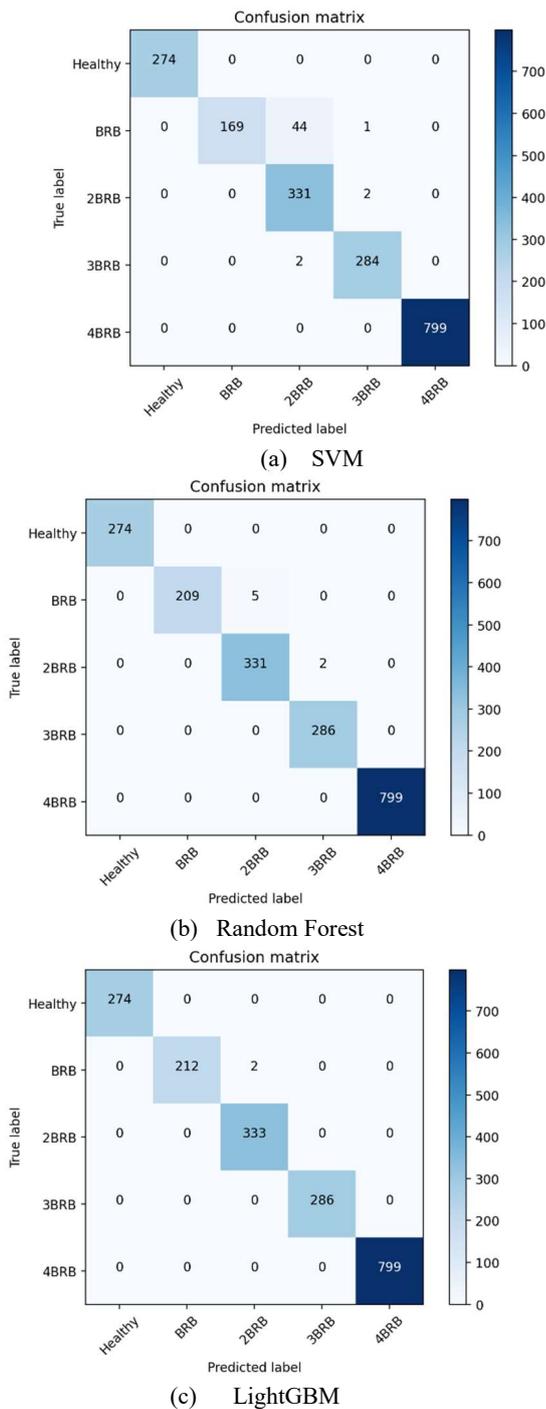


Fig. 5. Confusion matrix for (a) SVM, (b) Random Forest, and (c) LightGBM Classifiers.

The experiments with three different algorithms are performed by applying the Python Scikit-learn library, which offers various classification approaches that are suitable for predicting simultaneously multiple outputs.

From the values we get from the confusion matrix, it can be said that the LightGBM preforms better (99.89%) than that of Random Forest and SVM to predict the true class. Tables 5-6-7 lists the performance parameters.

States	Precision	Recall	F1-score	Support
Healthy	1.00	1.00	1.00	274
1 BRB	1.00	0.99	1.00	214
2 BRB	0.99	1.00	1.00	333
3 BRB	1.00	1.00	1.00	286
4 BRB	1.00	1.00	1.00	799

Table 5. The results of three classifiers in terms of six evaluation parameters for LightGBM

States	Precision	Recall	F1-score	Support
Healthy	1.00	1.00	1.00	274
1 BRB	1.00	0.79	0.88	214
2 BRB	0.88	0.99	0.93	333
3 BRB	0.99	0.99	0.99	286
4 BRB	1.00	1.00	1.00	799

Table 6. The results of three classifiers in terms of six evaluation parameters for SVM

States	Precision	Recall	F1-score	Support
Healthy	1.00	1.00	1.00	274
1 BRB	1.00	0.98	0.99	214
2 BRB	0.99	0.99	0.99	333
3 BRB	0.99	1.00	1.00	286
4 BRB	1.00	1.00	1.00	799

Table 7. The results of three classifiers in terms of six evaluation parameters for Random Forest

	SVM	Random Forest	LightGBM
Accuracy	0.974291	0.996327	0.998950

Table 8. Accuracy of different used algorithms of classification

It can be seen from Table 4 that the classification performance has been the most enhanced by using LightGBM Classifier. The accuracy may be further improved by training a model with a larger training set and further tuning the algorithms. Its excellent pattern recognition capabilities can be effectively utilized for the fault classification of electrical machines in the presence of disturbing noises.

Conclusion

In this work, a motor-current data-driven strategy is provided for fault detection of broken rotor bars and assessing the severity of defects in noisy conditions, based on a mix of statistical features, feature importance score, and machine learning algorithms. To look at the most effective features, 30 statistical features are extracted using the motor current signal as the original input for each phase. 16 features are selected using feature importance, which are used to train three different classifiers for fault diagnosis. A series of simulations are run to determine the optimum accuracy utilizing classification models for evaluating the performance of different classifiers. The major objective is to decrease duplicate data and the original data set's dimension, resulting in lower model power and computing complexity. Our study provides a new

framework and a new strategy that have been tested on an experimental setup. From the comparative surveys, we have found that the Light Gradient Boosted Machine Classifier is the best performer in terms of classification. The proposed system was tested using only the current signals from five different motor states. One condition is the healthy state, and the other conditions include one to four broken bars. The overall accuracy of fault classification is satisfactory, showing that the proposed method is promising for real time application. Additionally, to authenticate the efficacy of this model, a comparison with some recent work on the same dataset and several different datasets was presented, and the comparison indicates that the proposed method, employing time and frequency features has comparable accuracy while requiring less computational complexity than other methods. This indicates that current signal collected from a sensor less system is as effective as the conventional sensor-based system. In our future research work, we integrate the short circuit fault between coils in order to evaluate the classifier in case of multi-class. In addition, we analyze the database to refine it and make it more discriminating and optimal by using new methods of selecting indicators.

References

1. T. Jean-Claude, *Diagnostic des machines électrique*. Traité EGEM, série Génie électrique, pages 269–272 (2011).
2. Motor Reliability Working Group IEEE Industry Applications Society. "Report of Large Motor Reliability Survey of Industrial and Commercial Installations, Part I". IEEE Transactions on Industry Applications. Vol. IA-21, issue 4, p.853-864. July 1985.
3. United States Industrial Electric Motor Systems Market Opportunities Assessment. U.S. Dept. of Energy, Washington, DC, USA, 1998.
4. 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. Appl. 2016, 52, 3861–3868.
5. Sri, J.; Senanayaka, L.; Kandukuri, S.T.; Van Khang, H.; Robbersmyr, K.G. *Early Detection and Classification of Bearing Faults Using Support Vector Machine Algorithm*. In Proceedings of the 2017 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD), Nottingham, UK, 20–21 April 2017.
6. Li, Z.; Zhu, J.; Shen, X.; Zhang, C.; Guo, J. *Fault diagnosis of motor bearing based on the Bayesian network*. Procedia Eng. 2011, 16, 18–26.
7. Tian, J.; Azarian, M.H.; Pecht, M. *Rolling Element Bearing Fault Detection Using Density-Based Clustering*. In Proceedings of the 2014 International Conference on Prognostics and Health Management, Cheney, WA, USA, 22–25 June 2014.
8. Xie, Y. *A Fault Diagnosis Approach Using SVM with Data Dimension Reduction by PCA and LDA Method*. In Proceedings of the 2015 Chinese Automation Congress (CAC), Wuhan, China, 27–29 November 2015.
9. Deng, F. Ren, B. *Fault Diagnosis of Rolling Bearing Using the Hermitian Wavelet Analysis, KPCA and SVM*. In Proceedings of the 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), Shanghai, China, 16–18 August 2017.
10. Tu, D. Zheng, J. Jiang, Z. Pan, H. *Multiscale Distribution Entropy and T-Distributed Stochastic Neighbor Embedding-Based Fault Diagnosis of Rolling Bearings*. Entropy 2018, 20, 360.
11. A. Choudhary, D. Goyal, S.L. Shimi, A. kula, *Condition Monitoring and Fault Diagnosis of Induction Motors: A Review*. Arch. Comput. Methods Eng. 2018, 25, 1–18.
12. P.A. Delgado-Arredondo, D. Morinigo-Sotelo, R.A. Osornio-Rios, J.G. Avina-Cervantes, H. Rostro-Gonzalez, R.D.J. Romero-Troncoso. *Methodology for fault detection in induction motors via sound and vibration signals*. Mech. Syst. Signal. Process. 2017, 83, 568–589.
13. Y. Gritli, A.O. Di Tommaso, F. Filippetti, R. Miceli, C. Rossi, A. Chatti. *Investigation of motor current signature and vibration analysis for diagnosing rotor broken bars in double cage induction motors*. In Proceedings of the International Symposium on Power Electronics Power Electronics, Electrical Drives, Automation and Motion, Sorrento, Italy, 20–22 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1360–1365.
14. J. Antoni, G. Xin, N. Hamzaoui. *Fast computation of the spectral correlation*. Mech. Syst. Signal. Process. 2017, 92, 248–277.
15. G. Singh, V. Naikan. *Detection of half broken rotor bar fault in VFD driven induction motor drive using motor square current MUSIC analysis*. Mech. Syst. Signal. Process. 2018, 110, 333–348.
16. DAHER Homepage, <https://www.daher.com/>, last accessed 2021/05/10.
17. Schneider Electric, <https://www.se.com/ww/en/product/ATV11HU29M2U>, last accessed 2021/09/10
18. National Instruments, <https://www.ni.com/en-lb/support/model.ni-9246.html>, last accessed 2021/09/10.
19. BRUEL & KjaER, <https://www.bksv.com/fr/transducers/vibration/accelerometers/4384>, last accessed 2021/05/10
20. National Instruments, <https://www.ni.com/en-lb/support/model.crio-9040.html>, last accessed 2021/09/10.
21. P. Li. *Robust logitboost and adaptive base class (abc) logitboost*. arXiv preprint arXiv:1203.3491, 2012.

22. M. Richardson, E. Dominowska, and R. Ragno. *Predicting clicks: estimating the click-through rate for new ads*. In Proceedings of the 16th international conference on World Wide Web, pages 521–530. ACM, 2007.
23. Christopher JC Burges. *From ranknet to lambdarank to lambdamart: An overview*. *Learning*, 11(23-581):81, 2010.