FAULT DIAGNOSIS OF ROTARY MACHINES USING MACHINE LEARNING

Iago M. Brandao, Cesar da Costa
Department of Electrical Engineering, IFSP Federal Institute of São Paulo, São Paulo, Brazil
e-mail: i_modesto_br@yahoo.com.br; ccosta@ifsp.edu.br

***Abstract* – Classification of faults in rotary machines using machine learning is gaining attention in the field of science and engineering. In rotating machinery, misalignment is a common fault. This type of fault has been extensively studied in the literature using the vibration signals produced by rotary machines. This study proposes an approach based on machine learning techniques to diagnose misalignments in rotary machines under various conditions. A personalized diagnostic fault approach is proposed to detect misalignment faults. The approach includes three steps. First, the data acquisition model is developed to obtain signals (fault samples). The rotor vibration signals in stationary rotation conditions were obtained by two inductive proximity sensors with analog output, and the data were collected by a data acquisition device. Then, to generate the faulty training samples, each data acquisition signal is transformed to the frequency domain using Fast Fourier Transform (FFT). Finally, using the samples obtained through the feature selection techniques, machine learning algorithms Random Forest, Naïve Bayes and SVM were evaluated, resulting in classifications with different efficiencies. The results show that the SVM algorithm outperforms the Naïve Bayes and Random Forest algorithms when the same number of features is used. The proposed personalized diagnostic approach was applied to detect faults of rotating electrical machines misalignment with success.**

Keywords **– Fault Classification, Fault Diagnosis, Machine Learning, Vibration Analysis.**

I. INTRODUCTION

Rotor balancing is required on all types of rotating machinery, including motors, to ensure smooth machine operation. Rotor balancing involves the entire rotor structure, which is made up of a multitude of parts, including the shaft, rotor laminations, end heads, rotor bars, end connectors, retaining rings (where required), and fans. The design and manufacture of these components must be controlled for achieving stable precision balance. Rotor misalignment generates reaction force in the coupling, which is often a major cause of vibration in machinery [1]- [3].

The motivation of the work focuses on the oil industry's proposal to identify early misalignment failures in its fluid transport machines (solve realistic problems). Currently, the methods used such as vibration analysis, thermal imaging and oil particle analysis have a slow response and require a specialized technician to diagnose the machine. On the other hand, the proposed diagnosis using machine learning uses classification algorithms to accurately identify data set to specific categories of machine failures. An identification algorithm separates specific attributes, in the form of frequency in the data set sample and, from a training set, identifies in the sample, some conclusions about how these failures should be labeled or defined (healthy or not healthy). The machine is healthy or has one of the faults of misalignment taught and trained.

The higher the speed of the rotating electrical machine, the greater the impact of a small misalignment. The early a misalignment is detected, the better for the machine's condition, energy consumption and installation maintenance costs. This type of incipient failure identification can contribute to the planning of maintenance activities and optimize maintenance costs in reference to the production cost, depending on the type of operation [4], [5]. The type of misalignment failure has a great influence on the spectrum. The spectrum depends not only on the parallel or angular misalignment, but also on the coupling, the excitation force, and the rotation speed [6]. To reduce the costs caused by misalignment faults, there is a need for continuous monitoring of the integrity of the rotating machine. Using spectrum for fault analysis requires understanding of multidisciplinary research fields [6].

Often, by using these techniques, misalignment faults can´t be detected without a technical specialist. Recently, machine learning methods have been widely used in fault diagnosis of rotary electrical machines. Misalignment, as the important fault in rotating machinery, exhibits vibration behaviors under different work conditions, and directly affects the reliability and stability of the entire machinery equipment. Several studies are being carried out to apply machine learning to the diagnosis of misalignment faults. Rapid advances have led to the application of machine learning [7], [8], [9] for spectral-based machine diagnostics. Thus, to detect the misalignment faults in rotary machines, many researchers have proposed different intelligent diagnosis methods based on machine learning or artificial intelligence models.

Da Costa et al present a study on a method of diagnosing failures in rotary machines using machine learning techniques [10]. In this study, a support vector machine - SVM algorithm was proposed for fault diagnosis of the rotational misalignment in the rotor. Zhao et al presented the monitoring of rotating machines using machine learning [11]. They showed machine learning techniques and described an application example. The aim of the article was to provide an approach on vibration analysis for failure

Manuscript received 03/21/2022; first revision 06/14/2022; accepted for publication 09/08/2022, by recommendation of Editor Telles Brunelli Lazzarin. http://dx.doi.org/10.18618/REP.2022.3.0013

monitoring and to determine the best method according to the condition presented and the interpretation of the vibration response. Feng et al carried out a study in 2016 about a feature a DNN-based method for diagnosing rotating machine failures [12]. The effectiveness of the proposed method is verified using five datasets of bearings and planetary gearboxes. These data sets contain massive samples involving different health conditions of the rotating machine. Through the results of the diagnosis of these data sets, it is shown that the proposed method can extract failure characteristics from the frequency spectra in an adaptive way for various diagnostic problems and health conditions of the machine. Kankar et al have proposed a study for detection of bearing faults by classifying them using two machine learning methods, namely ANNs and SVMs [13]. Features are extracted from time-domain vibration signals using statistical techniques. Huerta-Rosales et al propose a methodology based on statistical time features (STFs) and support vector machines (SVM) to diagnose a transformer under several SCTs conditions [14].

This paper is an extended version of the contribution presented in [10],[15], [16]. It includes a broader range of intelligent fault diagnosis algorithms with additional experimental tests on rotor misalignment. It can be observed in the cited papers an evolution in the methodology in search of the improvement for faults hits.

In this article, the methodology developed can automatically diagnose, in real time, the severity levels in a rotor, measuring the vibration signals through inductive proximity sensors (replacing high-cost accelerometers). Presents recent machine learning techniques that explore the study and construction of algorithms that can learn from their mistakes and make predictions about the dataset. These algorithms operate by building a model from sample inputs to make predictions or decisions guided by the dataset, rather than simply following inflexible, static programmed instructions.

II. OVERVIEW MACHINE LEARNING ALGORITHMS

A. Supervised Learning

According to IBM [17], supervised learning (SL) is a subdivision of machine learning and artificial intelligence. It is defined using labeled data sets to train algorithms that classify data or accurately predict results. As the input data is fed into the model, it adjusts its weights through a reinforcement learning process, which ensures that the model has been properly adjusted. Supervised learning uses a training set to teach models to produce the desired output. This set of training data includes correct inputs and outputs, which allows the model to learn over time. The algorithm measures its accuracy through the loss function, adjusting until the error is sufficiently minimized [18], [19]. Supervised learning can be divided into two types of data mining problems: (i) classification and (ii) regression. Classification uses an algorithm to accurately assign test data to specific categories. It identifies specific entities in the data set and draws some conclusions about how those entities should be labeled or defined. The main classification algorithms are Naïve Bayes, linear classifiers, Support Vector Machines (SVM), decision trees, nearest k-neighbor and random forest. Regression is used to understand the relationship between dependent and independent variables. It is usually used to make projections, such as sales revenue for a particular business. As an example of popular regression algorithms: linear regression, logistic regression, and polynomial regression [14], [17], [18]. Several algorithms are used in supervised machine learning processes. The following are some commonly used learning methods, such as linear regression, SVM, random Forest, among others [19], [20], [21].

B. Naïve Bayes Classification

Naive Bayes is a classification approach that adopts the principle of class conditional independence from the Bayes Theorem. This means that the presence of one feature does not impact the presence of another in the probability of a given outcome, and each predictor has an equal effect on that result. There are three types of Naïve Bayes classifiers: Multinomial Naïve Bayes, Bernoulli Naïve Bayes, and Gaussian Naïve Bayes. This technique is primarily used in text classification, spam identification, and recommendation systems. Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features [22], [23].

The Naïve Bayes classifier is a supervised learning algorithm based on applying Bayes' theorem (1).

$$P(A \setminus B) = \frac{P(B \setminus A) \times P(A)}{P(B)}. \quad (1)$$

Where:

A and B are events.

$P(B \setminus A)$ is probability of B happening given that A has occurred.

$P(A)$ is probability of A to occur.

$P(B)$ is probability of B to occur.

$P(B) \neq 0$.

According to this theorem, it can calculate the probability of event A, conditioned on the event B, by first calculating the probability of the event B conditioned by event A, multiply by the probability of event A and, normalized by the probability of the event B. According to [23], the classifier is called naïve, because it assumes that the attributes of the dataset are independent, which means that each attribute x_i influences the class y, but does not influence other attributes x_1 to x_n.

C. Support Vector Machine

A support vector machine (SVM) is a supervised learning model used for data classification or regression. It is usually applied for classification problems, finding the best hyperplane that maximizes the distance between two classes of data points is maximum. This hyperplane is known as the

decision boundary, separating the classes of data points on both sides of the plane. Given a set of points of 2 types in N dimensional locations, SVM generates a dimensional ($N \geq 0$) hyperplane to separate these points into 2 groups, often configured by what is called SVM kernel. Suppose you have some points of 2 types on a paper that are linearly separable. SVM will find a straight line that separates these points into 2 types and is located as far as possible from all these points [14], [10]. The hyperplane is designed as (2).

$$g(x) = \omega^T x + \omega_0 . \tag{2}$$

Where:

ω is the vector of weights

x is the input vector

ω_0 is the bias.

The equations for the support vectors (SVM) of each class are given as (3, 4).

$$\omega^T x + \omega_0 \geq 1, \ \forall x \in \omega_1 \tag{3}$$

$$\omega^T x + \omega_0 \leq -1, \ \forall x \in \omega_2 . \tag{4}$$

Where:

ω_1 and ω_2 correspond to each class.

To discover the optimal hyperplane (training), the quadratic problem minimizing presented in (3, 4) must be solved by (5, 6).

$$J(\omega) \equiv \frac{1}{2}\|\omega\|^2 . \tag{5}$$

Where:

$J(\omega)$ is the optimal hyperplane (training). Subject to (5).

$$y_i(\omega^T x_i + \omega_0) \geq 1 . \tag{6}$$

Where:

$y_i = \pm 1$ is a class indicator for each training data x_i, also known as support vector.

The final decision function can be obtained by (7).

$$\omega = \sum_{i=1}^{N} y_i \alpha_i x_i . \tag{7}$$

Where:

α_i are the Lagrange multipliers.

When non-linear and non-separable patterns appear, the resulting linear classifier is obtained by (8).

$$g(x) = \sum_{i=1}^{N} y_i \alpha_i K(x_i, x) + \omega_0 . \tag{8}$$

Where:

x_i is assigned to ω_1 if $g(x) > 1$ or ω_2 if $g(x) < 0$.

$K(x_i, x)$ is a kernel function.

The most used kernels in SVM are linear, polynomial, RBF - Radial Basis Function, and sigmoidal. The Radial Basis Function kernel, also known as a Gaussian, it maps inputs to a feature space with a higher dimension. In this study, the used kernel function is the radial basis function (RBF) and is represented by (9).

$$K(x_i, x) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2} \right). \tag{9}$$

Where:

σ is called the kernel scale and this parameter affects distributing complexity of data in the feature space.

To control the SVM generalization capability, a misclassification parameter C is also defined by (10).

$$K(x_i, x_j) = K(x_i, x_j) + \frac{1}{C}\delta_{ij} . \tag{10}$$

Where:

C is penalized parameter and appropriate value of this parameter increase the classification performance.

δ_{ij} is Kronecker symbol.

The selection of SVM parameters C and σ has an important influence on the classification accuracy. Both parameters are chosen by the user [14].

D. Random Forest

Random Forest is a machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in Machine Learning. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. The "forest" references a collection of uncorrelated decision trees, which are then merged to reduce variance and create more accurate data predictions [10].

E. Measures for the Evaluation of Classification Algorithms

There are different methods of evaluating models for supervised learning and its classification algorithms.

1) Accuracy: This measure can be defined as the ratio of the number of correctly classified examples according to the total number of classified examples (11).

$$A = \frac{number\ of\ correctly\ classified\ examples}{total\ number\ of\ cases} . \tag{11}$$

2) Kappa coefficient: The Kappa Coefficient is a statistical method for evaluating the level of agreement or reproducibility between two data sets. Kappa coefficient is a measure of agreement for two methods of classification algorithms which seek to measure the agreement between observed and expected proportions. The Kappa coefficient is defined by (12).

$$K = \frac{(P_o - P_e)}{(1 - P_e)} . \tag{12}$$

Where:

P_o is the relative acceptance rate.

P_e is the hypothetical acceptance rate.

Kappa magnitude can be interpreted as indicated in Table 1. When the agreement is complete between the two datasets $K = 1$.

TABLE I
Kappa Magnitude

Value K	Level Agreement
< 0.20	Not Agreement
0.21 - 0.40	Minimum Agreement
0.41 – 0.60	Reasonable Agreement
0.61 – 0.80	Substantial Agreement
0.81 – 1.00	Perfect Agreement

3. Confusion Matrix

It is often important in practical problem solving distinguish certain types of errors. In cases where it is necessary to distinguish more types of errors result of the classification is shown in the form of two-dimensional matrix, where each row of the matrix corresponds to one class and record number of examples where it is forecasted class, and each column of the matrix is also marked by a class. On the diagonal of the matrix is the number of correct classified examples, while other elements of the matrix indicate the number of examples that were incorrectly classified as some of the other classes. It can be concluded that the use of a confusion matrix allows better analysis of different types of errors. Figure 4 illustrates the confusion matrix used in this work.

III. DATA PRE- PROCESSING

Data pre-processing can significantly benefit the performance of supervised learning models and can be a tool to reduce the complexity of algorithms and models used in machine learning. Data cleansing, normalization, transformation, feature extraction and feature elimination are the main data pre-processing methods. In this section, we will discuss two more suitable preprocessing methods for this work: (i) feature extraction and (ii) feature selection methods. It's important to notice that these pre-processing techniques can be often interchangeable, the problem to be solved that will direct what is the best order to be used.

A. Feature Extraction

The feature extraction aims to transform the original features into a new set constructed by the combinations of the original set (dimension reduction). It is usually required when we have few variables and we need to combine them to generate a larger dimensional space and possibly with greater predictive power, or it can be used when, through the knowledge domain, it is known that the combination of some variables translates identifiable behaviors in the data. An advanced feature extraction technique is Fast Fourier Transform (FFT). The Fourier transform (FT) defines that a periodic time-domain waveform can be represented by a weighted sum of sines and cosines. The same waveform can then be represented in the frequency domain as an amplitude-phase pair for each frequency component. The frequency spectrum of a signal is basically the frequency components (spectral components) of that signal. Spectral analysis refers to the representation of current signals in the frequency domain. The result is basically the frequency components (spectral components) of that signal in the frequency domain.

B. Feature Selection

Feature selection is a method of identifying and finding the most irrelevant and redundant information possible, allowing machine learning models to be trained and executed faster by using only the data that really is most discriminating in the prediction process. In this work, some pre-processing techniques are used, such as: (i) coefficient of statistical variation; (ii) Boruta algorithm; (iii) Recursive Feature Elimination (RFE) algorithm.

IV. EXPERIMENTAL PROCEDURE

A several data set was collected from the smart experimental setup. The data acquisition parameters for these tests were defined by a sampling frequency of 1KHz and 2000 samples. Figure 1 presents the experimental setup composed by: (1) 3-phase induction motor 220/380V, 0.25 cv, 4 poles, 1710 rpm; (2) variable speed controller with voltage vector control; (3) experimental rotor kit consisting of one disc measuring 160 mm in diameter and 16 mm in thickness, into which holes were punched to introduce misalignment; (4) rectified shaft 16 mm in diameter and 255 mm in length; (5) rolling bearings; (6) ball bearings with a reference frequency of 38,000 rpm and dynamic load of 351 kgf; (7) helical coupling, allowing an angular misalignment of 5° at a reference frequency of 25,000 rpm and rated torque of 2.3 Nm; (8) two proximity sensors eddy current (model IWRM, manufacturer Baumer); (9) data acquisition device.

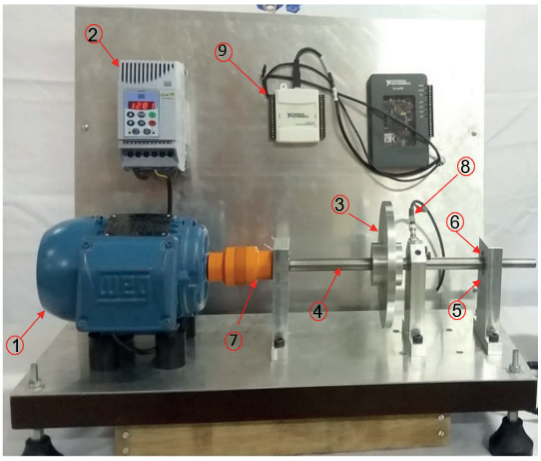

Fig. 1. Experimental setup.

The disc from the rotor was used to study misalignment. To perform misalignment fault simulation, a trial mass of predetermined weight was introduced. In the first

experiment, a mass of 0.006 kg was attached to the rotor shaft disc. The misalignment thus created produced a mechanical vibration in the machine's structure. The rotor vibration signals in stationary condition of rotation were transduced by two inductive proximity sensors, resolution of 5 μm and sensing distance from 0 to 2 mm installed near the displacements of the rotating shaft in the directions of the *x*- and *y*-axis with a sampling rate of 1000 data points per second. The rotational speed was 1800 rpm. The data sets were generated under balanced (test mass not inserted), and eight different unbalanced (test mass inserted in different positions). Vibration signals were collected during 500 acquisition cycles, under a rate of 2KHz.

A. Proposed Machine Learning Algorithms

The block diagram of the algorithm of the personalized diagnosis method for the diagnosis of incipient failures in rotary electrical machines based in machine learning techniques is shown in Figure 2. The different stages of the proposed system are discussed here: (i) Data acquisition; (ii) Feature extraction (Fast Fourier Transformer – FFT); (iii) Feature selection (Coefficient variation, Boruta algorithm, Recursive feature elimination); (iv) Training set (Naïve Bayes, Support vector machine, Random Forest), and (v) Test results (Accuracy, Kappa Coefficient, and Confusion Matrix).

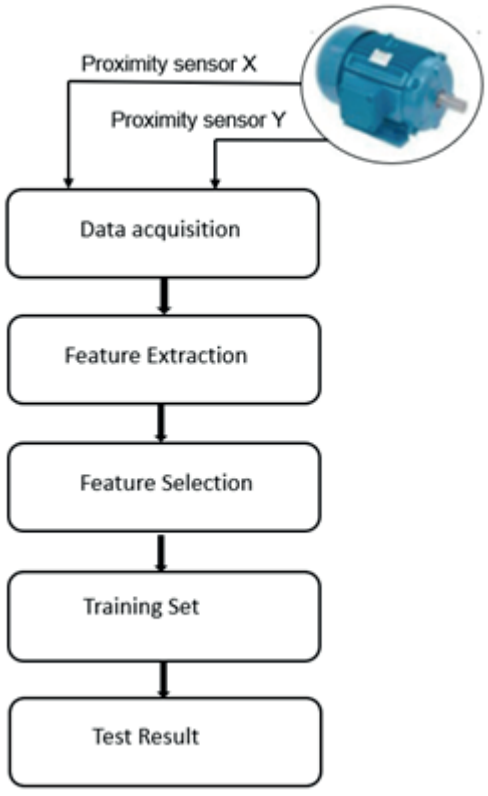

Fig. 2. Block diagram of algorithms implemented.

1. Data Acquisition

Vibration signals were collected in real time by a data acquisition device (DAQ) through two analog channels of 12 bits and 10k samples per second connected to two eddy current proximity sensors. Eleven different types of tests were performed: (i) one not balanced – N/B (trial mass not inserted); (ii) two different balanced: BLC1 and BLC2 (trial mass inserted in different positions) and (iii) eight different unbalanced: DBL1 to DBL4 (trial mass inserted in different positions). The tests performed with specific features are listed in Table II.

TABLE II
Performed Tests

Types	Description
N/B	Not Balanced
BLC1	Balanced 1
BLC2	Balanced 2
DBL1	Unbalanced
DBL1+	Unbalanced
DBL2	Unbalanced
DBL2+	Unbalanced
DBL3	Unbalanced
DBL3+	Unbalanced
DBL4	Unbalanced
DBL4+	Unbalanced

2. Feature Extraction

The performance of a fault identification system depends on the input features. A small feature dataset does not guarantee a complete system description, e.g., different rotor misalignment conditions dataset can lead to misclassification. Many features can increase the discrimination capacity, but do not guarantee the addition of relevant information related to rotor misalignment failure, that is, the information can be redundant. The elaboration of models from datasets with many features requires more computational time, and the processing time grows exponentially. Thus, algorithms must be developed to extract significant information from features, eliminate redundancy and reduce datasets. Once there is a consolidated dataset that represents the vibration signals in the time domain, it is necessary to extract resources from these signals to the frequency domain to reduce the complexity of the classification model developed. The vibration signal was transformed into the frequency domain using the Fast Fourier Transform (FFT). FFT was used in this work to reduce modelling complexity, since the use of temporal data requires more complex algorithms and other machine learning techniques (cross-validation techniques). Once the temporal signal is transformed into a frequency spectrum, a range of frequencies is observed, as well as better-established features. Considering that the FFT is more suitable for stationary signals, feature selection methods are used to remove noise generated by non-stationary signals inherent to the developed mechanical system.

3. Feature Selection

Since there is a data set with a dimension of 2000 variables, being 1000 from the X-axis vibration signal spectrum and 1000 from the Y-axis vibration signal spectrum, and a relatively high volume of records, it is necessary understand if all variables help to identify the health status of the rotor, if they do not add value or interfere

in some way, as well as reducing the necessary computation time. To extract useful information from these high volumes of data, it's necessary to use techniques to reduce noise or redundant data. This is where feature selection plays an important role. Not only does it help train your model faster, but it also reduces its complexity, facilitates interpretation, and improves accuracy. To obtain greater reliability of the variables that best describe a healthy and faulty rotor condition, three distinct variable selection techniques were used in the step of model training. Three feature selection techniques were used: (i) coefficient of variation (CV); (ii) Boruta algorithm and (iii) Recursive Feature Elimination (RFE). These techniques were used in parallel, receiving the entire data set of features and delivering the most relevant features at the end. Boruta's algorithm produced an importance graph, where all the features used can be seen in an agglutinated form and some with a much higher importance than the others. From the feature selection stage, the relationships between the main features, also known as frequencies, were obtained. It should be noted that features in the form "CXX.X" represent the frequency "XX.X Hz". Figure 3 shows the relationship between the frequency C30.0 and C60.0 with one dataset of healthy conditions and four datasets of rotor unbalance fault conditions (DBL1, DBL2, DBL3 and DBL4), as shown in Table 2.

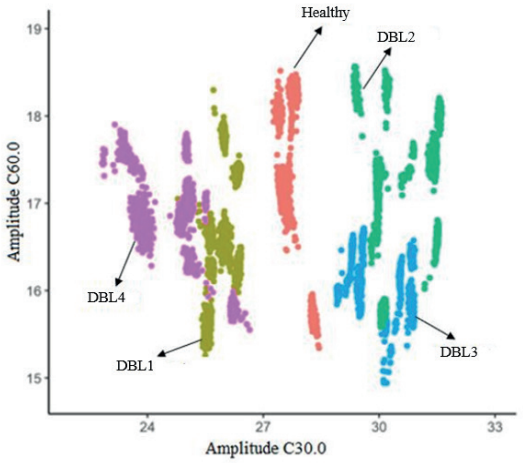

Fig. 3. Dataset of faults obtained by feature selection techniques.

The second technique is the Recursive Elimination of Variables and seeks to understand which is the minimum set of variables that best describe the response variable, using all possible variables and excluding a fixed number of variables in each iteration. The Coefficient of Variation statistics is the third technique used to select variables and was used with the objective of removing frequencies from the spectrum that had little or no change in the tests performed, representing constants between the health and fault rotor condition. Thus, the categories equivalent to the tests carried out can be better discriminated. Table 3 presents the parameters used to define the feature selection algorithms.

TABLE III
Algorithms Selection Features

Algorithms	Parameters
Boruta	maxRuns = 150
Recursive Feature Elimination	Method='lvq', preProcess='scale', trControl=trainControl (method='repeatedcv', number=10, repeats=3)
Coefficient Variation Statistics	maxCV = 0.05 (or 5%)

4. Training Set

The Naïve Bayes, Random Forest, and Support Vector Machine (SVM) techniques are three of the most popular classification algorithms. To evaluate the performance of used. They were used because they are available to any user and are easy to use. All code sources are RStudio software packages. RStudio is an open-source integrated development environment for the R programming language for graphing and statistical calculations. Table 4 presents the parameters used to define the classification algorithms.

TABLE IV
Algorithms Classification

Algorithms	Parameters
Naïve Bayes	Method = 'nb', fl=2, usekernel=True, adjust=1
Support Vector Machine (SVM)	Kernel='polynomial', degree=2, gamma=3, coef0=0.01, cost=10, nu=0.5, épsilon=0.1
Random Forest	Method='rf', ntree=100, tuneLength=3, allowParallel=True, trControl=trainControl (method='cv', number=5)

Using the samples obtained through the feature selection, the Random Forest, Naïve Bayes and SVM algorithms were evaluated, which allowed obtaining classifications with different efficiency. The obtained results show that the SVM algorithm has a better performance than the Naïve Bayes, and Random Forest algorithms when the same number of features is used. The results of the three machine learning algorithms are presented in Table 5. The SVM algorithm obtained the best learning in the accuracy and kappa metrics.

TABLE V
Accuracy Percentage of 3 Machine Learning Algorithms

Random Forest		Naïve Bayes		Support Vector Machine	
Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
97,20%	96,90%	97,44%	97,18%	98,67%	98,54%
96,50%	96,10%	95,57%	95,11%	99,08%	98,99%

The number of features in each subset is defined by the threshold value obtained in the feature selection. In the field

of machine learning, when it comes to the statistical classification problem, a confusion matrix, also known as an error matrix, is a specific table that allows the visualization of the training performance of an algorithm. Each row of the matrix represents data in a predicted classification, while each column represents data in an actual classification. The SVM algorithm was trained, and the confusion matrix was obtained. Figure 4 shows a confusion matrix. There are 11 class and 801 samples in each class. The main diagonal (738, 712, 773…to 801) gives the correct predictions. That is, the cases where the actual values and the model predictions are the same. Looking at the N/B (Not Balanced) column, of the 801 N/B predicted by the model (sum of column N/B), 738 were N/B, while 55 were BLC1 (Balanced 1) incorrectly predicted to be N/B and 8 were BLC2 (Balanced 2) incorrectly predicted to be N/B.

In the BLC1 (Balanced 1) column of the 801 BLC1 predicted by the model (sum of column BLC1), 712 were BLC1, while 65 were N/B (Not Balanced) incorrectly predicted to be BLC1 and 24 were BLC2 (Balanced 2) incorrectly predicted to be BLC1. In the BLC2 (Balanced 2) column of the 801 BLC2 predicted by the model (sum of column BLC2), 773 were BLC2, while 08 were B/N (Not Balanced) incorrectly predicted to be BLC2 and 20 were BLC1 (Balanced1) incorrectly predicted to be BLC2. On the other columns corresponding to unbalance faults (DBL1 to DBL4+) gave the correct predictions. The actual values and model predictions are the same.

HYPOTHESIZED CLASS

TRUE CLASS	S/F	BLC1	BLC2	DBL1	DBL1+	DBL2	DBL2+	DBL3	DBL3+	DBL4	DBL4+
S/F	738	65	8	0	0	0	0	0	0	0	0
BLC1	55	712	20	0	0	0	0	0	0	0	0
BLC2	8	24	773	0	0	0	0	0	0	0	0
DBL1	0	0	0	801	0	0	0	0	0	0	0
DBL1+	0	0	0	0	801	0	0	0	0	0	0
DBL2	0	0	0	0	0	801	0	0	0	0	0
DBL2+	0	0	0	0	0		801	0	0	0	0
DBL3	0	0	0	0	0	0	0	801	0	0	0
DBL3+	0	0	0	0	0	0	0	0	801	0	0
DBL4	0	0	0	0	0	0	0	0	0	801	0
DBL4+	0	0	0	0	0	0	0	0	0	0	801

Fig. 4. Confusion matrix referring to the SVM model.

Considering that the Accuracy (ACC) is calculated by the number of correctly classified examples divided by the total number of cases (Eq. 11).

1) Class N/B (Not Balanced):

$$Acc = \frac{738}{738+55+8} = 92\%$$

2) Class. BLC1 (Balanced 1):

$$Acc = \frac{712}{712+65+24} = 88\%$$

3) Class BLC2 (Balanced 2):

$$Acc = \frac{773}{773+20+8} = 96\%$$

4) Class DBL1, DBL1+, DBL2, DBL2+, DBL3, DBL3+, DBL4, DBL4+ (Unbalanced):

$$Acc = \frac{801}{801} = 100\%.$$

V. CONCLUSION

In this paper, a methodology based in machine learning for assessing a rotor's health condition under misalignment fault is presented. The methodology developed can diagnose eleven severity levels in a rotor by measuring the vibration signals from two inductive proximity sensors with analog output and collected by a data acquisition device in real time. The feature selection allows obtaining the optimum set of features, first, by selecting the features that present the most relevant information related to the rotor performance and, after, by reducing the dimensional space. The Boruta´s algorithm implementation to select features allowed reducing from an extensive number of features a set of only 2000 features, i.e., 1000 for x-axis, and 1000 for y-axis. In reducing the dimensional space, the Boruta method presents a more satisfactory performance than the Recursive Elimination of Variables and Coefficient of Variation statistical method, simplifying the classification process. The SVM classifier can classify among eleven severities of rotor with an accuracy of 98,67%. The results also demonstrate that the SVM classifier performs better than the Naïve Bayes and Random Forest algorithms under the same experimental setup. Finally, the classification results show that the present personalized fault diagnosis method is effectiveness to identify the faults in misalignment. The proposed personalized fault diagnosis method based on machine learning can solve many problems, such as providing complete fault samples for various intelligent diagnosis, ensuring that the fault signal is pure without noise interference, etc.

REFERENCES

[1] J. Z. Szabo, "Vibration diagnostic test for effect of unbalance", *in 16th International Conference on Intelligent Engineering System*, Lisbon, Portugal, pp. 81-85, July 2012.

[2] T. H. Patel, A. K. Darpe, "Vibration response of misaligned rotor", *J. Sound Vibration*, 325 (2009), pp. 609-628, Aug 2009.

[3] J. M. Bossio, G. R. Bossio, C. H. De Angelo, "Angular misalignment in induction motor with flexible coupling", *in 35th Annual Conference of IEEE Industrial Electronics,* 2009, pp. 1033-1038, Febr 2009.

[4] M. Lal and R. Tiwari, "Multi-fault identification in simple rotor-bearing-coupling systems based on forced response measurements", *Mech. Mach. Theory*, vol. 51, pp. 87–109, May 2012.

[5] J. K. Sinha and K. Elbhbah, "A future possibility of vibration-based condition monitoring of rotating

machines", *Mech. Syst. Signal Process.*, vol. 34, nº 1–2, pp. 231–240, Jan 2012.

[6] M. Monte, F. Verbelen, B. Vervisch, B. "The use of orbitals and full spectra to identify misalignment", Wicks A. (eds) Structural Health Monitoring, in *Conference Proceedings of the Society for Experimental Mechanics Series*. Springer, vol.5. pp. 215-222, Feb 2014.

[7] B. Luo, H. Wang, H. Liu, B. Li, F. Peng, "Early Fault Detection of Machine Tools Based on Deep Learning and Dynamic Identification", *IEEE Transactions on Industrial Electronics*, vol. 66, Issue 1, pp.509- 518, Feb 2019.

[8] G. Toh and J. Park, "Review of Vibration-Based Structural Health Monitoring Using Deep Learning", *Applied Sciences*, vol. 10, nº 5: 1680, March 2020.

[9] Martin-Diaz, D. Morinigo-Sotelo, O. Duque-Perez, T. J. Romero-Troncoso, "An experimental comparative evaluation of machine learning techniques for motor fault diagnosis under various operating conditions", *IEEE Transactions on Industry Applications*, vol. 54, nº 3, pp.2215-2224, Feb 2018.

[10] A. Pinheiro, I. M. Brandão, C. Da Costa, "Vibration Analysis in Turbomachines Using Machine Learning Techniques", *European Journal of Engineering Research and Science*, vol. 4, nº 2, pp. 12-16, Feb 2019.

[11] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R. X. Gao, "Deep learning and its applications to machine health monitoring", *Mech. Syst. Signal Process*, vol. 115, pp. 213–237, January 2019.

[12] J. Feng, L. Yaguo, L. Jing, Z. Xin, L. Na, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data", *Mechanical Systems and Signal Processing*, vol. 72, pp. 303-315, May 2016.

[13] P. K. Kankar, S. C. Sharma, S. P. Harsha, "Fault diagnosis of ball bearings using machine learning methods", *Expert Systems with Applications*, vol. 38, nº 3, pp.1876-1886, March 2011.

[14] J. R. Huerta-Rosales, D. Granados-Lieberman, A. Garcia-Perez, D. Camarena-Martinez, J. P. Amezquita-Sanchez. "Short-Circuited Turn Fault Diagnosis in Transformers by using Vibration Signals, Statistical Time Features, and Support Vector Machines on FPGA", *Sensor.* , vol. 21, nº 11: 3598, May 2021.

[15] G. K. Yamamoto, C. Da Costa, J. S. S. Sousa, "A smart experimental setup for vibration measurement and imbalance fault detection in rotating machinery", *Case Studies in Mechanical Systems and Signal Processing*, vol. 4, pp 8-18, Dec 2016.

[16] C. Da Costa, R.S. Gama, C. E. Nascimento, I. M. Brandão, E. C. Medeiros, M. H. Mathias, "Orbit Analysis for Imbalance Fault Detection in Rotating Machinery", *Journal of Electrical and Electronics Engineering (IOSR-JEEE)*, vol.13, Issue 1, pp 43-53, March 2018.

[17] IBM, "What is supervised learning? IBM Cloud Learn Hub, Supervised Learning". (2020) [Online]. Available: https://www.ibm.com/cloud/learn/supervised-learning#toc-what-is-su-d3nKa9tk

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Review - Deep learning", *Nature*, vol. 521**,** pp. 436–444, May 2015.

[19] C. G. Mattera, J. Quevedo, T. Escobet, H. R. Shaker, and M. Jradi, "Fault Detection and Diagnostics in Ventilation Units Using Linear Regression Virtual Sensors", *in 2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, Rabat, Morocco, pp. 1-6, Jan 2019.

[20] Q. Yang, D. Guo, W. Yao, J. Cai, and C. Mei, "Support Vector Machines based Rotor Fault Diagnosis with Improved Particle Swarm optimization", *in Chinese Automation Congress (CAC)*, Hangzhou, China, pp. 4321-4324, Feb 2019.

[21] N. Iman, and T. Ahmad, "Improving Intrusion Detection System by Estimating Parameters of Random Forest in Boruta", *in International Conference on Smart Technology and Applications (ICoSTA),* Surabaya, Indonesia, pp. 1-6, April 2020.

[22] X. Shi, J. Liang, L. Ye, and B. Hu, "A method of fault diagnosis based on PCA and Bayes classification", *in 8th World Congress on Intelligent Control and Automation,* Jinan, pp. 5628-5631, July 2010.

[23] X. Xing, J. Luo, Z. Jia, Y. Li, and Q. Han, "Automated Fault Detection for Web Services using Naïve Bayes Approach", *in IEEE 10th International Conference on Software Engineering and Service Science (ICSESS),* Beijing, China, pp. 336-339, March 2019.

BIOGRAPHIES

Iago M. Brandão, was born in São Paulo, SP, Brazil. He received the B.Sc. degree in Automation and Control Engineering from the IFSP –Federal Institute of São Paulo, SP, Brazilin 2018. He is currently Machine Learning Engineer. His research interest includes Artificial Neural Network, machine learning, induction motors, sensors, diagnostic, electrical machines, and FPGA.

Cesar da Costa, was born in Rio de Janeiro, RJ, Brazil. He received the B.Sc. degree in electronic and electrical engineering from the CEFET-RJ, Federal Center of Technological Education Celso Suckow da Fonseca and Nuno Lisboa University in 1975 and 1980 respectively. He received the M.Sc. degree in mechanical engineering from Taubate University, Taubate, SP, Brazil, and the Ph.D. degree in mechanical engineering from UNESP-Universidad Estadual Paulista Julio de Mesquita Filho, Guaratingueta, SP, Brazil in 2005 e 2011, respectively. He did sandwich doctoral stage, PDEE-CAPES, in the IST-Institute Superior Tecnico, Lisbon, Portugal in 2009. He is currently post-doctoral and professor of automation and control engineering in the IFSP –Federal Institute of Education, SP, Brazil. His research interests include Fuzzy controller, Artificial Neural Network, machine learning, electrical machines, FPGA, and Industry 4.0.