Application of Machine Learning to Evaluate Unbalance Severity in Rotating Machines

Dionísio H. C. de S. S. Martins[1,2], Douglas O. Hemerly[1], Matheus Marins[3], Amaro A. Lima[1(✉)], Fabrício L. Silva[2], Thiago de M. Prego[1], Felipe M. L. Ribeiro[3], Sergio L. Netto[3], and Eduardo A. B. da Silva[3]

[1] CEFET-RJ Post Graduate Program in Electrical Engineering (PPEEL), Rio de Janeiro, Brazil
dionisiohmartins@gmail.com, doh4rj@hotmail.com, {amaro.lima,thiago.prego}@cefet-rj.br

[2] CEFET-RJ Department of Mechanical Engineering, Rio de Janeiro, Brazil
fabricio.silva@cefet-rj.br

[3] Program of Electrical Engineering (PEE/COPPE), Federal University of Rio Janeiro (UFRJ), Rio de Janeiro, Brazil
matheusaraujo@poli.ufrj.br, {felipe.ribeiro,sergioln,eduardo}@smt.ufrj.br

Abstract. This paper proposes two modifications in a classification method for unbalancing fault severity analysis in rotating machines based on the unbalancing mass force. The unbalancing severity was categorized into three severity levels, namely High (H), Medium (M) and Low (L). The feature vectors used information from discrete-time Fourier transform (DFT), kurtosis and entropy from the vibration signals. Similarity based Model (SBM) and Kernel discriminant analysis (KDA) techniques were applied in order to evaluate the feature discrimination and reduce the input feature space. All these techniques were tested in a random forest classifier. Test results indicate that non-linear transformations to the feature space combined to random forest can further improve the classification of unbalancing severity defect, by reducing the feature space dimension from 31 to 6.

Keywords: Machine learning · Unbalancing · Condition monitoring

1 Introduction

Rotating machines play a very important role in industrial processes, acting from the production of petrochemical elements to the energy general of an industrial plant. The oil and gas industry is a very demanding market for rotating machines with an amount of two thousand new compressors and five thousand new pumps per year [1]. Considering the financial importance of this industry, the mitigation of failures improving the reliability of the process is a crucial and continuous target.

K. L. Cavalca and H. I. Weber (Eds.): IFToMM 2018, MMS 61, pp. 144–160, 2019.
https://doi.org/10.1007/978-3-319-99268-6_11

The predictive maintenance procedures came up with the development of automation in industrial processes allowing an efficient production monitoring by means of sensors spread all over the plant. These sensors were capable of acquiring several different data from the machinery leading to areas of fault analysis and prediction, which consequently caused production cost reduction, efficiency increase, and undesired halts and spare parts changes avoidance.

One very important cause of defects in machine is the excessive vibration, which could be related to machinery life reduction, production shortage and increase in maintenance costs. Unbalance, misalignment, loose part as many other problems could generate machine vibration.

Vibration due to unbalance is one of the most classical and common problems in rotating machines already addressed in [3] using a centrifugal pump as case study. Another example is shown in [2], where unbalanced rotor lead to cracks in the rotor shaft. A substantial literature can be found about problem of unbalance in rotating machines.

1.1 Previous Researches

Kernel-based algorithms have been used in conjunction with learning functions, reaching good results in the areas of text data, speech data, biological data and complex data structures. These algorithms can be defined on any input space. In machine learning area, there are still problems to be studied using kernel algorithms. The first problem is choosing the type of kernel function for a classification problem [4]. This is bound to the problem of select a representation of the input. This selection may entail big differences in the results. The second problem is when the kernel is fixed and the issue is to choose the best value for the regularization parameter. This can be solved using the regularization path algorithms or employing cross validation methods [5].

The research made by [6] studied a remote system for online condition monitoring and fault diagnosis of gas turbine on offshore oil well drilling platforms based on a kernelized information entropy model. It used entropy for measuring the uniformity of exhaust temperatures that reflects the condition of gas turbine. It also used features in kernel spaces to monitor the condition of the equipment using decision trees as the classification strategy.

In the research done by [7] the experimental bench alignment/balance vibration trainer (ABVT) of Spectra Quest was used to produce the following failure scenarios: misalignment, unbalance, fail in the outer track, fail in the rolling element and fail in the cage. The total failure scenarios produced and analyzed were 1951, resulting in the machinery fault database (MaFaulDa) [8].

The work done by [9] aimed to provide an overview of the similarity-based modeling (SBM) and use this technique to monitor the condition of rotating machines through vibration signals. This research examined the accuracy of this method to detect and diagnose faults in electric motors with variable speed operating conditions. The application of this technique enables the detection of incipient faults in rotating machines.

SBM has been used in many cases of industrial application, like anomaly detection in power plants, modeling airplanes flight paths and fault diagnosis in electric motors. In the work of [10], an automatic fault detector and classifier was proposed using SBM to classify the following defects in rotating machines: bearing defects, misalignment and unbalance. The use of similarity model may be used in two ways: as a standalone classifier or as an auxiliary model to produce features for another classifier. In this article the use of SBM algorithm increased the number of hits in the classification in the MaFaulDa database. The results of this article showed that the SBM model indeed increased the accuracy of the random forest classifier [10].

In the work of [11] it was proposed a system to evaluate the fault severity in DC electric motors. After diagnosing the defect in the motor, the degree of failure might be evaluated with the objective of rightly plan the future action to restore the equipment and reduce the unscheduled halt and maintenance costs. The failure scenarios were composed by bearing faults, unbalance and misalignment of the MaFaulDa database that were quantified in 3 levels of severity, namely L (Low), M (Medium) and H (High) without the use of the mechanical background in the quantification process.

Martins et al. [12] proposes a classification method for specifically unbalancing fault severity rotating machines based on the force created by the mass of unbalancing, using only the signals of the unbalance scenario of MaFaulDa database. The study evaluated the capacity to quantify the severity of the unbalancing fault of a rotating machine considering the angular velocity, differing to previous works, which the objective was the severity quantification, not considering the angular velocity, which may increase classification errors. The unbalancing severity was broken down into 3 discrete levels which were L (Low), M (Medium) and H (High). The random forest algorithm was used as a classifier for the unbalancing severity analysis problem.

2 Discriminant Analysis

2.1 Linear Discriminant Analysis

The Linear discriminant analysis aims to find a linear transformation by the maximization of inter-class distance and minimization of intra-class distance. This method explicitly attempts to model the difference between the classes of data. Linear discriminant function is a linear combination of characteristics which is typified by producing maximum separation between two populations.

The LDA is a method that uses information from the categories associated with each pattern to linearly extract the most discriminating characteristics. When used as a classification method, it is a supervised method that use the train data to find the boundary of the classes [13]. In this work, LDA is used only as a vector space transformation and its results are used by another classification algorithm.

The scatter matrices inter-class $\mathbf{S}_b$ and intra-class $\mathbf{S}_w$ is defined by:

$$\mathbf{S}_b = \sum_{i=1}^{G} N_i \left(\overline{\mathbf{x}}^{(i)} - \overline{\mathbf{x}}\right) \left(\overline{\mathbf{x}}^{(i)} - \overline{\mathbf{x}}\right)^T, \text{ and} \tag{1}$$

$$\mathbf{S}_w = \sum_{i=1}^{G} \sum_{j=1}^{N_i} N_i \left(\mathbf{x}_j^{(i)} - \overline{\mathbf{x}}^{(i)}\right) \left(\mathbf{x}_j^{(i)} - \overline{\mathbf{x}}^{(i)}\right)^T, \tag{2}$$

where G is the total number of class or groups, N_i is the number of patterns of classes i, $\mathbf{x}_j^{(i)}$ is the M-dimensional pattern j from class i, the vector $\overline{\mathbf{x}}^{(i)}$ is the unbiased mean vector of class i. The overall mean vector is given by:

$$\overline{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{g} \sum_{j=1}^{N_i} \mathbf{x}_j^{(i)}, \tag{3}$$

where N is the total number of samples.

The principal objective is to establish a projection matrix $\mathbf{W}_{\text{lda}}$ that optimize the ratio of the determinant of the between-class scatter matrix to the determinant of the intra-class scatter matrix known as Fisher criterion obtained from the following equation:

$$\mathbf{W}_{\text{lda}} = \arg\max_{\mathbf{W}} \frac{|\mathbf{W}^T \mathbf{S}_b \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_w \mathbf{W}|}. \tag{4}$$

The Fisher criterion described in Eq. (4) is maximized when the projection matrix $\mathbf{W}_{\text{lda}}$ is composed of the eigenvectors of ${\mathbf{S}_w}^{-1}\mathbf{S}_b$ with at most $(G-1)$ nonzero corresponding eigenvalues [14].

The LDA cannot be used to separate high dimensional problems with accuracy. The other problem is that $\mathbf{S}_w$ is usually singular or mathematically unstable.

2.2 The Kernel Discriminant Analysis

The kernel trick is a way of mapping observations from a general set into an inner product space, without having to compute the mapping explicitly in machine learning algorithms. It can modify the problem from a lower dimension to a higher dimension. Transforming the problem from a lower to a higher dimension makes the approximation function more flexible to the data, reducing the risk of empirical error.

The non-linear mapping of the original space to a larger dimensional feature space commonly leads to a dispersion matrix within poorly placed class. The KDA is useful to figure out the instability issues and the singularity of linear Fisher methods.

To modify a LDA algorithm to a kernel version is necessary to change the $\mathbf{S}_w$ to $\mathbf{S}_w^K$. This can be done by applying the eigenfunction to the scatter matrices $\mathbf{S}_b$ and $\mathbf{S}_w$ in the feature space as shown in [15], or using a intuitive mathematical result as proposed in [16].

KDA is capable of projecting the sample data onto matrix $\mathbf{B}$, in order to separate the samples as discriminating as possible. Being $\mathbf{K}_c$ the centralized $\mathbf{K}$, thus ${\mathbf{S}_t}^K = \mathbf{K}_c\mathbf{K}_c$ is the total scatter matrix, $\mathbf{S}_b^K = N^{-1}\mathbf{K}_c\mathbf{Y}\mathbf{Y}^T\mathbf{K}_c$ is the between-class scatter matrix and $\mathbf{S}_w^K = \mathbf{S}_t^K - \mathbf{S}_b^K$ is the within-class scatter matrix. Assuming that the dataset consists of $\{(\mathbf{x}_n, y_n)\}_{n=1}^N$, where $\mathbf{x}_n$ is the input vector and y_n is its associated class given by $y_n \in \{1, \ldots, G\}$. The indicator matrix $\mathbf{Y}$ is built with elements $Y(n,i)$ corresponding to row n and column (class) i defined as [17,18]:

$$Y(n,i) = \begin{cases} \sqrt{\dfrac{N}{N_i}} - \sqrt{\dfrac{N_i}{N}} & y_n = i. \\ -\sqrt{\dfrac{N_i}{N}} & \text{otherwise.} \end{cases} \tag{5}$$

This algorithm has the capability of minimizing the trace of $\mathbf{B}^T\mathbf{S}_w^K\mathbf{B}$ that corresponds to the intra-class distance and maximizing the trace of $\mathbf{B}^T\mathbf{S}_b^K\mathbf{B}$ that corresponds to the inter-class distance. In this way, KDA is capable of maximizing the separability of the samples in the dimensionality reduced space as described mathematically in the following equation:

$$\mathbf{B}_{\text{kda}} = \arg\max_{\mathbf{B}} \left\{ \text{Tr}\left(\left(\mathbf{B}^T {\mathbf{S}_W}^K \mathbf{B} \right)^{-1} \mathbf{B}^T {\mathbf{S}_b}^K \mathbf{B} \right) \right\}. \tag{6}$$

3 Similarity Based Model

The Similarity-based modelling, or simply SBM, is a nonparametric empirical modelling method that uses pattern recognition from historical data to generate estimates of the current values of each feature (dimension) in a set of feature vectors [19]. The SBM can be used as a classifier itself, or as an auxiliary model to generate features to train a further classifier. This is the case of the present work where the SBM provides the feature vectors to the random forest classifier. The key aspect within the use of the SBM is the possibility of reduction the feature vectors quantity still keeping the classification accuracy and performance by selecting the most important vectors in terms of its capacity to represent a given class [10]. As presented in [10], in the SBM methodology a system state at a time n is given by a vector:

$$\mathbf{x}_n = [x_n(1), x_n(2), \ldots, x_n(M)]^T, \tag{7}$$

where M is the number of different features.

After gathering L instances of $\mathbf{x}_n$, a $L \times M$ matrix $\mathbf{D} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L]^T$ is created storing historical data of $\mathbf{x}_n$ representing a state space or a class.

The selection of the most representative feature vectors is evaluated by a linear combination with all states contained in $\mathbf{D}$, that is:

$$\hat{\mathbf{x}}_n = \mathbf{D}^T \mathbf{w}_n, \tag{8}$$

where

$$\mathbf{w}_n = \left(\mathbf{D} \otimes \mathbf{D}^T\right)^{-1} (\mathbf{D} \otimes \mathbf{x}_n) \tag{9}$$

$$= \mathbf{G}^{-1} \mathbf{a}_n, \tag{10}$$

where operator $\otimes$ represents a similarity function.

These similarity functions $s(\mathbf{x}_i, \mathbf{x}_j)$ can quantify the similarity between two vectors $\mathbf{x}_i$ and $\mathbf{x}_j$, for $0 \leq s(\mathbf{x}_i, \mathbf{x}_j) \leq 1$. The closer $s(\mathbf{x}_i, \mathbf{x}_j)$ is to 0, the less $\mathbf{x}_i$ is similar to $\mathbf{x}_j$. Therefore in Eq. (10), $\mathbf{a}_n$ gives us the similarity between the current state and the class representative state in $\mathbf{D}$, and $\mathbf{G}$ provides weights to transform matrix $\mathbf{a}_n$ for each state in $\mathbf{D}$. A special case is when $\mathbf{G} = \mathbf{I}$ which is called auto-associative kernel regression (AAKR). In this case if $\mathbf{x}_n$ belongs to $\mathbf{D}$, $\hat{\mathbf{x}}_n = \mathbf{x}_n$ only for one state in $\mathbf{D}$.

One example of a similarity function is the inverse Euclidean similarity given by [20]:

$$s(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{1 + \gamma ||\mathbf{x}_i - \mathbf{x}_j||_p}. \tag{11}$$

By doing so, a similarity threshold must be evaluated and defined, for example, $s(\mathbf{x}_i, \mathbf{x}_j) \geq 0.7$, which means a reduced feature vectors space will remain with the most representative vectors for that given class and then this new state space is further provided as a training set to the random forest classifier.

4 Fault Emulator System

The experimental system used to simulate fault scenarios was based on the equipment known as alignment/balance vibration trainer (ABVT) as shown in Fig. 1. This equipment is commonly used for the study of usual fault problems in rotary machines in a controlled environment. This type of system is very useful, since the study of real machine failures in full operation is very costly, especially for the industrial scenario.

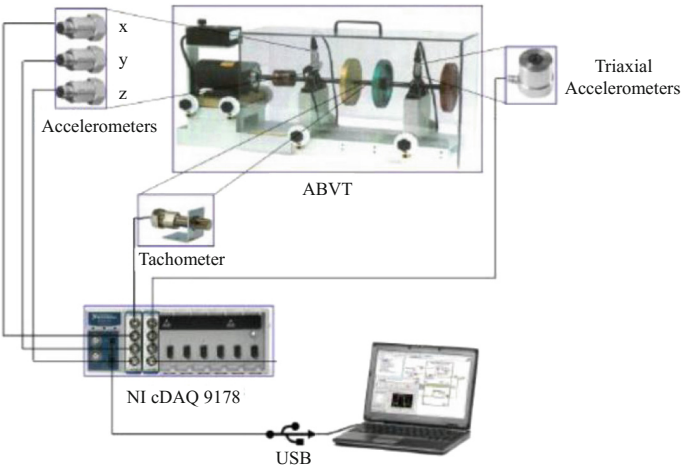

Fig. 1. Alignment balance vibration trainer.

4.1 Equations of Motion

In order to develop a dynamic model of the ABVT system, a rotor-shaft-bearing system shall be considered. For that, the equations of motion for individual disc and shaft are derived in a frame attached to the base of the system using Lagrange equation. Considering the kinetic energy of a rigid disc, the respective kinetic energy, strain energy due to bending and Rayleigh dissipation function due to viscous material damping of the flexible shaft, the kinetic energy of the unbalance mass associated with the disc and the force applied on the rotor from the bearings due to the movement of the shaft relative to them, the equations of motion for the considered system are obtained [21]:

$$\mathbf{M_r}\ddot{\mathbf{q}} + \mathbf{D_c}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{f}, \tag{12}$$

where $\mathbf{M_r}$ is the assembled inertia matrix, $\mathbf{D_c}$ is the assembled matrix coefficient to global velocity vector including gyroscopic, Coriolis, damping effects representation, and $\mathbf{K}$ is the assembled matrix coefficient to global displacement vector including bending stiffness, circulatory, parametric stiffness matrices due to base motion and the bearing stiffness. $\mathbf{f}$ is a load vector containing the effects of mass unbalance, and $\mathbf{q}$ is the displacement vector for disc and shaft nodes.

4.2 Critical Speed

A critical speed of the system is the angular speed which matches one of its natural frequencies. However, finding the natural frequencies of a stationary rotor is not enough to determine the critical speed, once the natural frequency of the rotor depends on the rotor angular speed. Therefore, it is important to compute the natural frequency by considering the effect of the rotor.

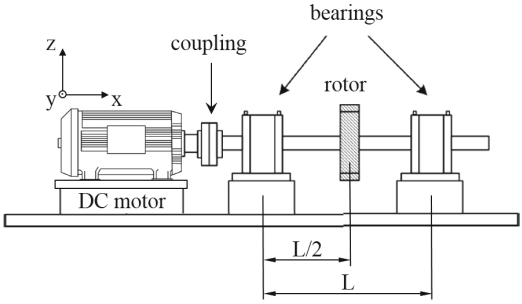

Fig. 2. Rotor positioning scheme.

The main characteristics of the ABVT are: it has a direct current motor with a power of 183.8 W, its range of speed varies in the range of [700; 3780] RPM, the length of its shaft is 520 mm and the diameter of it is 16 mm. About the rotor, its thickness is 16 mm, its diameter is 152.4 mm, and as presented in the Fig. 2, it is mounted in between two bearings with a thickness of 12 mm and distance of bearing spacing (L) of 390 mm. Neglecting the dissipative effects and the free vibration of the shaft and rotor, the Eq. (13) became:

$$\mathbf{M_s}\ddot{\mathbf{q}} + \mathbf{K_s}\mathbf{q} = \mathbf{0}. \tag{13}$$

Assuming a modal solution for the Eq. (13),

$$\mathbf{q} = \mathbf{a}e^{i\omega t}, \tag{14}$$

yields the eigenvalue problem

$$\left[-\omega^2\mathbf{M_s} + \mathbf{K_s}\right]\mathbf{a} = \mathbf{0}, \tag{15}$$

from which the approximate natural frequencies, eigenvectors and approximate eigenfunctions can be obtained considering ω as the angular velocity and $\mathbf{a}$ as a vector of the same size of $\mathbf{q}$ with scalar elements. For the presented characteristics and considering the developed Eqs. (13, 14 and 15), the natural frequencies are obtained (Table 1).

Table 1. Natural frequencies.

Modes	Natural frequency (Hz)
1	64.52
2	65.8
3	248.07
4	474.12
5	1053.22
6	1063.94

5 Features Extraction

Feature extraction is useful for the recognition of failures in rotating machine bearings because it reduces the complexity of the learning problem for the classification algorithm [22].

Feature extraction has the great advantages to drastically reducing the size of the information vector compared to the number of samples in the time series. An excessively large vector hampers the task of learning the classifier affecting system performance. In addition, this stage of the process allows the use of a vector with greater discrimination capacity of the faults classes than that if obtained using in the original time series [11].

The vibration signals are acquired by piezoelectric accelerometers in the axial, radial and tangential directions of the motor. These are processed with the purpose of extracting the characteristics of the signals that are able to differentiate the failures types.

5.1 Fast Fourier Transform

The Fast Fourier Transform (FFT) is an algorithm capable of calculating the Discrete Fourier Transform (DFT) efficiently, decreasing the amount of arithmetic operations to be performed. k is the frequency index relative to Ω_0, where $\Omega_0 = 2\pi/N$ is the fundamental frequency. The discrete frequency Ω of the DFT spectrum is obtained by $k\Omega_0$. The Eq. (16) calculates the DFT.

$$X[k] = \sum_{-\infty}^{\infty} x\,[n]\, e^{-jk\Omega_0 n}, \tag{16}$$

where $X[k]$ represents the signal in the discrete frequency domain, $x[n]$ is the signal in the discrete time domain, n is the number of samples and N is the number of samples of $x[n]$.

5.2 Kurtosis

Kurtosis is defined as the fourth-order central moment. It is used to recognize impulsive vibration signals, because it is robust when used to classify noisy signals [23]. Equation (17) represents the calculation of the kurtosis $K(u)$ of a random variable u.

$$K(u) = \frac{\mathrm{E}[u-\mu]^4}{\sigma^4}, \tag{17}$$

where $\mathrm{E}[\cdot]$ is the expected value operator, μ is the mean $\mathrm{E}[u]$ and σ is the standard deviation $\sqrt{\mathrm{E}[(u-\mu)^2]}$.

5.3 Entropy

The entropy can be understood as a measure capable of evaluating the degree of unpredictability of a random variable. Equation (18) shows the calculation of the entropy $H(u)$ for a discrete random variable.

$$H(u) = -\sum_{i=0}^{C-1} p(u_i) \cdot \log(p(u_i)), \tag{18}$$

where $p(u_i)$ is the probability density of the discrete random variable u at u_i, u_i is the i-th value of u and $0 < C < \infty$ is the number of values of u.

5.4 Feature Vector

For each signal that comes from the accelerometers positioned in the axial, radial and tangential directions of the closest and farther bearings from the motor, using the FFT, three characteristics were obtained that correspond to the value of the spectrum amplitude of the frequencies R_f (motor rotation frequency), $2R_f$ and $3R_f$ [22], totalizing 18 amplitude values.

For each accelerometer signal, kurtosis and entropy were also calculated, yielding 12 additional characteristics, which were kurtosis of the vibration signals in the axial, radial and tangential direction of the next bearing; kurtosis of the vibration signals in the axial, radial and tangential direction of the bearing away; entropy of the vibration signals in the axial, radial and tangential direction of the next bearing; entropy of the vibration signals in the axial, radial and tangential direction of the bearing away.

The rotation frequency of the motor was also obtained through the FFT of the tachometer signal, totalizing 31 features in the feature vector such as proposed in [22].

6 Classification

The random forest is a powerful machine learning algorithm which can be used for either regression and classification that creates an ensemble of decision trees during the training phase. This technique has been proved as an excellent tool to defect detection and prediction [11,22]. At the present work the random forest is used as a classifier. This algorithm makes use of the bootstrap aggregation or, bagging [24], statistical technique which in turn comes from the bootstrap idea. While the bootstrap creates several subsamples from the population, the bagging algorithm makes several subsamples from the same sample which is the training set containing the feature vectors with its respective output classes.

Each feature vector creates or "grows" one decision tree, which means the algorithm can create hundreds or thousands of decision trees. That is the reason of the name "forest". As the quantity of subsets are increased, the possibility

of overfitting the model decreases. When a new vector comes during the test phase each tree will points to a class and a voting procedure is than executed to determine final random forest result which will be the most voted class among all the trees.

The decision trees are known as naive classifiers, it means small variations in the training set can cause significant changes in the classifier balance and thus the bagging procedure comes to mitigate this problem [12], and aims to reduce variance [24]. This procedure however is not completely efficient, because every feature in the vector may cause same influence when a new split is made during the tree creation phase. This can result in quite similar or correlated trees. To avoid this problem the random forest algorithm leaves one feature out randomly to ensure high uncorrelated trees. As an example using a vector with a set of feature dimensions $\{f_0; \ldots; f_{L-1}\}$, one is removed, for instance, f_i, making that dimension null, ϕ_i, thus the resulting dimensionality reduced vector is represented by $\{f_0; ...; \phi_i; ...; f_{L-1}\}$, as shown in Fig. 3.

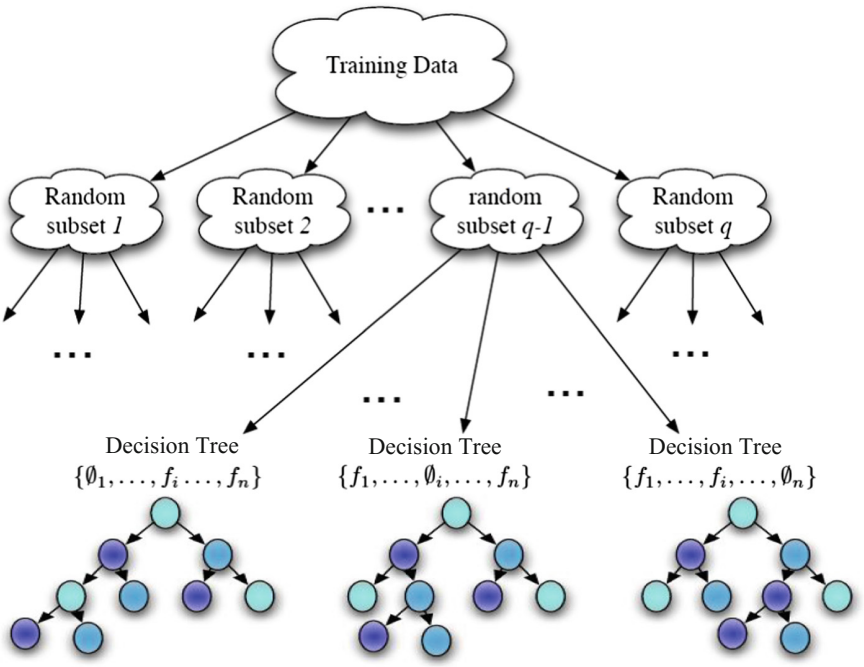

Fig. 3. Random forest algorithm diagram.

In a three classes classification problem, the confusion matrix is a 3×3 matrix as shown in Table 2. The number of elements of class C_i is given by $N_i = m_{1i} + m_{2i} + m_{3i}$. C_i is the class to be recognized by the classifier and for practical purposes in this work, it represents the severity levels low, medium and high (L, M and H), which will be defined in Sect. 7.

Table 2. Confusion matrix of a three classes classification problem.

Output	Target classes		
	C_1	C_2	C_3
C_1	m_{11}	m_{12}	m_{13}
C_2	m_{21}	m_{22}	m_{23}
C_3	m_{31}	m_{32}	m_{33}

A typical tune process for the random forest classification algorithm is choosing the number of trees the maximizes the classification performance. There are several ways of measuring the performance of classification algorithms and the one used in this work is the average intra-class accuracy η, given by

$$\eta = \frac{1}{3}\left(\frac{m_{11}}{N_1} + \frac{m_{22}}{N_2} + \frac{m_{33}}{N_3}\right). \tag{19}$$

7 Training and Test Sets

There are 333 different fault scenarios in the database with different rotating speeds. These scenarios had 7 different levels obtained with 7 different unbalancing masses: 6 g (49 fault scenarios), 10 g (48 scenarios), 15 g (48 scenarios), 20 g (49 scenarios), 25 g (47 scenarios), 30 g (47 scenarios), 35 g (45 scenarios). The resulting unbalancing fault severity effects were separated into 3 discrete levels: H (High), M (Medium) and L (Low), then creating the output classes

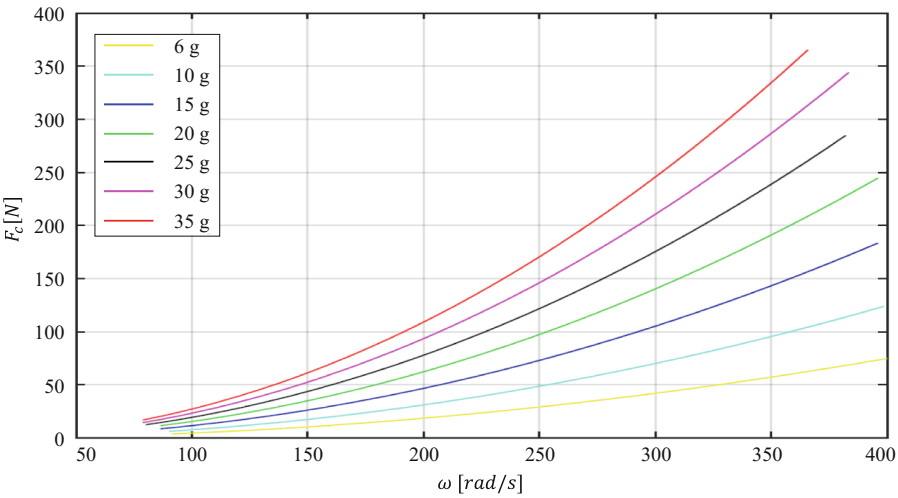

Fig. 4. Unbalancing force × angular velocity for all unbalancing masses separately.

of the classifier following similar reasoning as [25]. This classes separation was based on the magnitude of the unbalancing force represented by equation

$$F_c = mr\omega^2, \tag{20}$$

in which m is the unbalancing mass, r is the mass distance from the spin axis in the center of the inertial disk, and ω is the angular velocity [12].

The $F_c \times m$ curves for each individual mass are shown in Fig. 4. As in experiment 4 from [12], the severity levels are divided as L ($0 \leq F_c < 39.47$ N), M ($39.47 \leq F_c < 118.43$ N) and H ($F_c \geq 118.43$ N), which leads to 31% of the data in L class, 38% in M class and 31% in H class.

The separation of the data set for this work was done using cross-validation, which is a statistical technique used to estimate the performance of an algorithm based on a prediction model. In the pattern recognition area, there are several cross-validation methods used to asses the efficiency of the classifier independent of the test set chosen. In this work the k-fold method with 10 folds was used. Following this methodology, the original data set is randomly divided into 10 parts (or folds) with approximately equal sizes. 3 of these 10 parts are chosen randomly as the test set and the other 7 parts are used as the training set. This process leads to $K = 120$ possible combinations for test and training sets. Finally, the 120 classification results as combined through simple average in order to obtain a single result

$$\bar{\eta} = \frac{1}{K} \sum_{k=0}^{K-1} \eta_k, \tag{21}$$

where η_k is the average intra-class accuracy of the classification algorithm using the k-th k-fold combination. Examples of test and training fold combinations are shown in Fig. 5.

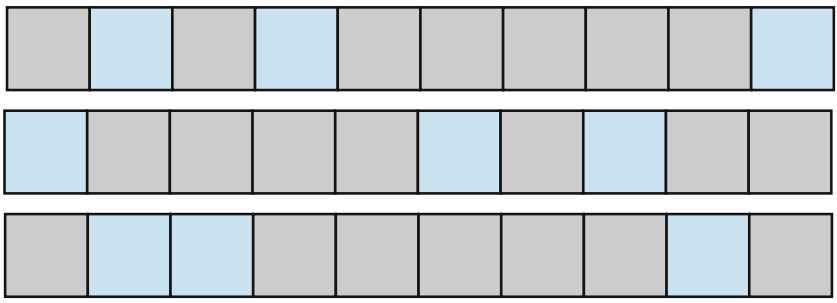

Fig. 5. Examples of combination of the k-fold strategy used: light square for test and dark squares for training folds.

8 Results and Discussion

The work in [12] used the random forest as the classification approach with the feature vector composed by the 31 features described in Sect. 5.4. Table 3 shows the average confusion matrix of the 120 k-fold configurations as in [12], with $\bar{\eta} = 93.73\%$ overall classification accuracy.

Table 3. Confusion matrix of the unbalancing severity analysis as in [12].

Output	Target classes (%)		
	L	M	H
L	$\mathbf{93.32 \pm 4.85}$	4.84 ± 3.54	0 ± 0
M	6.68 ± 4.85	$\mathbf{92.05 \pm 3.40}$	4.18 ± 3.89
H	0 ± 0	3.11 ± 2.62	$\mathbf{95.83 \pm 3.89}$

In order to improve the classification process, two different mapping approaches were applied on the vector space of the unbalancing severity analysis database: SBM and KDA.

8.1 Experiment 1: SBM + Random Forest

The research described in [10] used SBM together with random forest was used to improve the classification of the whole MaFaulDa database. The same method was applied to the unbalancing severity analysis problem of [12].

Table 4. SBM + random forest confusion matrix of the unbalancing severity analysis.

Output	Target classes (%)		
	L	M	H
L	$\mathbf{94.59 \pm 4.35}$	6.44 ± 4.18	0.00 ± 0.00
M	4.23 ± 3.28	$\mathbf{92.13 \pm 3.24}$	3.51 ± 3.31
H	0.00 ± 0.00	2.99 ± 2.58	$\mathbf{95.70 \pm 4.11}$

Table 4 shows the result of SBM combined with random forest in the unbalancing severity analysis problem, reaching $\bar{\eta} = 94.14\%$ overall classification accuracy, which is 0.41% greater than the results achieved by just using random forest with the original feature vector space. This can be explained by the already high accuracy of random forest and the low number of subjects for each class.

8.2 Experiment 2: KDA + Random Forest

Another approach used to map the feature vector space before applying to the random forest was the KDA method. Table 5 shows the result of KDA combined with random forest in the unbalancing severity analysis problem, reaching $\bar{\eta} = 93.31\%$ overall classification accuracy, which is 0.42% less than the results achieved by just using random forest with the original feature vector space. The advantage of using KDA is the dimensionality reduction from 31 dimensions used on the original random forest approach, to 6 dimensions used in the combination of KDA and random forest.

Table 5. KDA + random forest confusion matrix of the unbalancing severity analysis.

Output	Target classes (%)		
	L	M	H
L	$\mathbf{93.03 \pm 4.44}$	7.22 ± 4.26	0.00 ± 0.00
M	6.81 ± 4.25	$\mathbf{89.58 \pm 4.14}$	2.62 ± 3.30
H	0.16 ± 0.77	3.20 ± 2.61	$\mathbf{97.33 \pm 3.35}$

9 Conclusion

This work proposes a modification in a classification method to further improve it's capacity to discriminate the severity of the unbalancing fault of a rotating machine considering the angular velocity, by transforming the feature space using kernel discriminant analysis (KDA).

Experimental results show that by combining KDA with random forest, the discrimination in L, M and H of the original method proposed in [12] is preserved with a reduction in dimensionality from 31 to 6, a reduction of almost 81%. Also a sightly improvement in accuracy was observed using KDA with random forest, which may not impact on the system performance statistically. The problem configuration was guided to a well balanced amount of class elements leading to a less challenging task and did not take the full benefits of the non-linear techniques applied. The usage of mechanical standards that consider the unbalancing fault in rotating machines could provide more significance to the experiment also making the task more challenging.

The categorization of fault severity in Low, Medium and High is a first step in the development of a predictive maintenance system. In order to do so, other types of fault such as misalignment and bearing defects should be studied and a categorization with more divisions or a numerical estimation for severity should be also studied.

References

1. Pareschi, D., Bittanti, S., Pesatori, A.: Safety, Performance and innovation in Oil and Gas Rotating Machines. In: mcT 2012 - Tecnologie per il petrolchimico (2012)
2. Serra, R.C.: Detecção de Trincas em Eixos de Bombas Centrífugas em Operação, através da técnica de análise de Vibrações. Instituto de Pesquisas Energéticas e Nucleares, São Paulo (1995)
3. Taneja, S.: Effect of unbalance on performance of centrifugal pump. Int. J. Sci. Technol. Res. **2**, 56–60 (2013)
4. Bach, F.R., Thibaux, R., Jordan, M.I.: Computing regularization paths for learning multiple kernels. In: Advances in neural information processing systems, pp. 73-80 (2005)
5. Evgeniou, T., Micchelli, A., Pontil, M.: Learning multiple tasks with kernel methods. J. Mach. Learn. Res. 615-637 (2005)
6. Wang, W., Xu, Z., Tang, R., Li, S., Wu, W.: Fault detection and diagnosis for gas turbines based on a kernelized information entropy model. Sci. World J. **617162**, 13 (2014)
7. Pestana, D.V., Zambrano, R.L., Lima, A.A., Prego, T.M., Netto, S.L., and da Silva, E.A.B.: The influence of feature vector on the classification of mechanical faults using neural network. In: Proceedings of the IEEE Seventh Latin American Symposium on Circuits and Systems (2016)
8. MaFaulDa - Machinery Fault Database. http://www02.smt.ufrj.br/~offshore/mfs/. Accessed 20 Feb 2018
9. Wegerich, S.W.: Similarity based modeling of time synchronous averaged vibration signals for machinery health monitoring. In: Aerospace Conference, pp. 3654–3662 (2004)
10. Marins, M.A., Ribeiro, F.M.L., Netto, S.L., da Silva, E.A.B.: Improved similarity data base model for the classification of rotating-machine failures. J. Franklin Inst. 1–18 (2017)
11. Martins, D.H.C.S.S., Prego, T.M., Lima, A.A.: Classificação de Severidade de Falhas em Máquinas Rotativas usando Random Forest e Redes Neurais Artificiais. In: XIII Encontro Nacional de Inteligência Artificial e Computacional, pp. 13–24 (2016)
12. Martins, D.H.C.S.S., Prego, T.M., Lima, A.A., Hemerly, D.O., Silva, F.L.: Predictive Maintenance based on unbalance severity analysis of rotating machines. In: 24th ABCM International Congress of Mechanical Engineering (2017)
13. Thomaz, C.E., Gilson, A.G.: A kernel maximum uncertainty discriminant analysis and its application to face recognition. In: International Conference on Computer Vision Theory and Applications, pp. 341–346 (2009)
14. Fukunaga, K.: Introduction to statistical pattern recognition, 2nd edn. Academic press, San Diego (2013)
15. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Regularized discriminant analysis for the small sample size problem in face recognition. In: Pattern Recognition Letters, pp. 3079–3087 (2003)
16. Yang, J., Zhang, D., Frangi, A.F., Yang, J.Y.: Two-dimensional PCA: a new approach to appearance-based face representation and recognition. In: 26th IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 131–137 (2004)
17. Ye, J.: Least squares linear discriminant analysis. In: Proceedings of the 24th international conference on Machine learning, pp. 1087–1093 (2007)

18. Zhao, Z., Sun, L., Yu, S., Liu, H., Ye, J.: Multiclass probabilistic kernel discriminant analysis. In: International Joint Conference on Artificial Intelligence, pp. 1363–1368 (2009)
19. Wegerich, S.: Similarity based modeling of vibration features for fault detection and identification. Sens. Rev. **25**(2), 114–122 (2005)
20. Marins, M.A.: Classificação de Falhas em Máquinas Rotativas Utilizando Métodos de Similaridade e Random Forest. Universidade Federal do Rio de Janeiro, UFRJ (2016)
21. Das, A.S., Dutt, J.K., Ray, K.: Active vibration control of unbalanced flexible rotor-shaft systems parametrically excited due to base motion. Appl. Math. Model. **34**, 2353–2369 (2010)
22. Martins, D.H.C.S.S., Pestana-Viana, D., Prego, T.M., Lima, A.A., Zambrano-López, R., Netto, S.L.: Diagnóstico de falhas em máquinas rotativas utilizando Random Forest. In: XXXIV Simpósio Brasileiro de Telecomunicações, pp. 916–920 (2016)
23. Saidi, L., Ali, J.B., Fnaeich, F.: The use of spectral kurtosis as a trend parameter for bearing faults diagnosis. In: 15th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), pp. 394–399 (2004)
24. Gareth, J., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning. 1st edn. (corrected 6th printing 2015), pp. 316–321. Springer, New York (2013)
25. Hussein, A., Irina, T.: A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling. J. Sound Vib. **369**, 246–265 (2016)