Advances in Engineering Software 123 (2018) 13–24

Contents lists available at ScienceDirect

Advances in Engineering Software

journal homepage: www.elsevier.com/locate/advengsoft

Optimizing presetting attributes by softcomputing techniques to improve tapered roller bearings working conditions

Roberto Fernandez Martinez[a,*], Ruben Lostado Lorza[b], Ana A. Santos Delgado[c], Nelson O. Piedra Pullaguari[d]

[a] Department of Electrical Engineering, College of Engineering of Bilbao, University of the Basque Country UPV/EHU, Alameda Urquijo s/n, 48013 Bilbao, Spain
[b] Department of Mechanical Engineering, University of La Rioja, San Jose de Calasanz 31, 26004, Logroño, Spain
[c] Department of Business Studies, Universidad Tecnica Particular de Loja, San Cayetano Alto, Calle París, Loja, Ecuador
[d] Department of Computer Science, Universidad Tecnica Particular de Loja, San Cayetano Alto, Calle París, Loja, Ecuador

ARTICLE INFO

Keywords:
Data mining classification techniques
Finite elements method
Design of experiments
Model parameter tuning and evaluation
Double-row tapered roller bearing

ABSTRACT

Double-row Tapered Roller Bearings are mechanical devices that have been designed to support a combination of loads that are fixed on an optimal presetting to ensure correct working conditions. The emergence of high contact stresses, fatigue spalling and pitting on the bearing railway makes it important to have a tool that enables knowing in advance whether certain presetting loads will lead to excellent working conditions or the opposite. This work proposes a methodology to classify the working condition on the basis of the values of presenting loads on four classes. To achieve this goal, a three-dimensional Finite Element (FE) model was generated. Later, a design of experiments was designed to provide the greatest amount of information by reducing the computational cost of the simulations based on FE models. Then, one of the four classes of working conditions was assigned to each of the experiments. Later, a statistical analysis and machine learning techniques were used to create classification models. Feature transformation and reduction, algorithm parameter tuning and validation methods were used to achieve robust classification models. The best results were obtained based on flexible discriminant analysis. As it provided acceptable accuracy, both the methodology and final model were validated.

1. Introduction

Double-row Tapered Roller Bearings (TRBs) are designed to support the preload (P), axial load (F_a), radial load (F_r) and torque (T) in static and dynamic conditions. TRBs are formed by a pair of inner raceways with a gap of width that are separated by a gap of constant width (δ) between them, an outer raceway and a set of tapered rollers. The gap that separates the inner raceway into two equal parts enables the TRB to be disassembled for maintenance, as well as to be able to apply P. F_a, F_r and T are applied on the outer raceway. The tapered rollers form two columns between the inner and outer raceways. This combination of loads produces contact stresses (S) and local deformations (α) on the bearing raceways, as well as its own rotation and a variation of δ, that are very difficult to predict and validate experimentally. In addition, if the combinations of loads on the TRBs differ from the manufacturer's recommended values, the bearing may malfunction. For example, the combination of a greatly reduced P and a very high F_r and T, could create an S in excess of 1000 MPa and significant local deformations (α_{max}) on the raceway. This could cause harmful defects like pitting and fatigue spalling [1]. In addition to these undesirable tribological problems, a reduced P may cause result in greater stresses in some contact areas of the outer raceway. This could cause the rollers to disengage from the outer raceways [2]. Thus, determining the optimal combination of loads on mechanical devices of this type is normally undertaken by analytical techniques or, alternatively numerical methods like the Finite Element Method (FEM). Although using the FEM provides obvious advantages, it does have some disadvantages. They include high computational expense, especially when the Finite Element (FE) model involves problems in mechanical contact, large displacements and material nonlinearities. A major drawbacks in using FEM for problems of mechanical contact is the mesh size and convergence of the model. Then, if the contact surface between the tapered rollers and raceways is reduced and the mesh size between the two contact bodies is large, the calculation of the contact stresses will be inaccurate [3]. Following this approach, some researchers [4,5] analyzed the influence of the mesh size of the contact bodies on the contact stresses between tapered rollers and raceways when several P and F_r were applied. It is well know that using FEM in isolation (i.e., without

* Corresponding author.
E-mail addresses: roberto.fernandezm@ehu.es (R. Fernandez Martinez), ruben.lostado@unirioja.es (R. Lostado Lorza), aasantos@utpl.edu.ec (A.A. Santos Delgado), nopiedra@utpl.edu.ec (N.O. Piedra Pullaguari).

https://doi.org/10.1016/j.advengsoft.2018.05.005
Received 8 February 2018; Received in revised form 3 May 2018; Accepted 13 May 2018

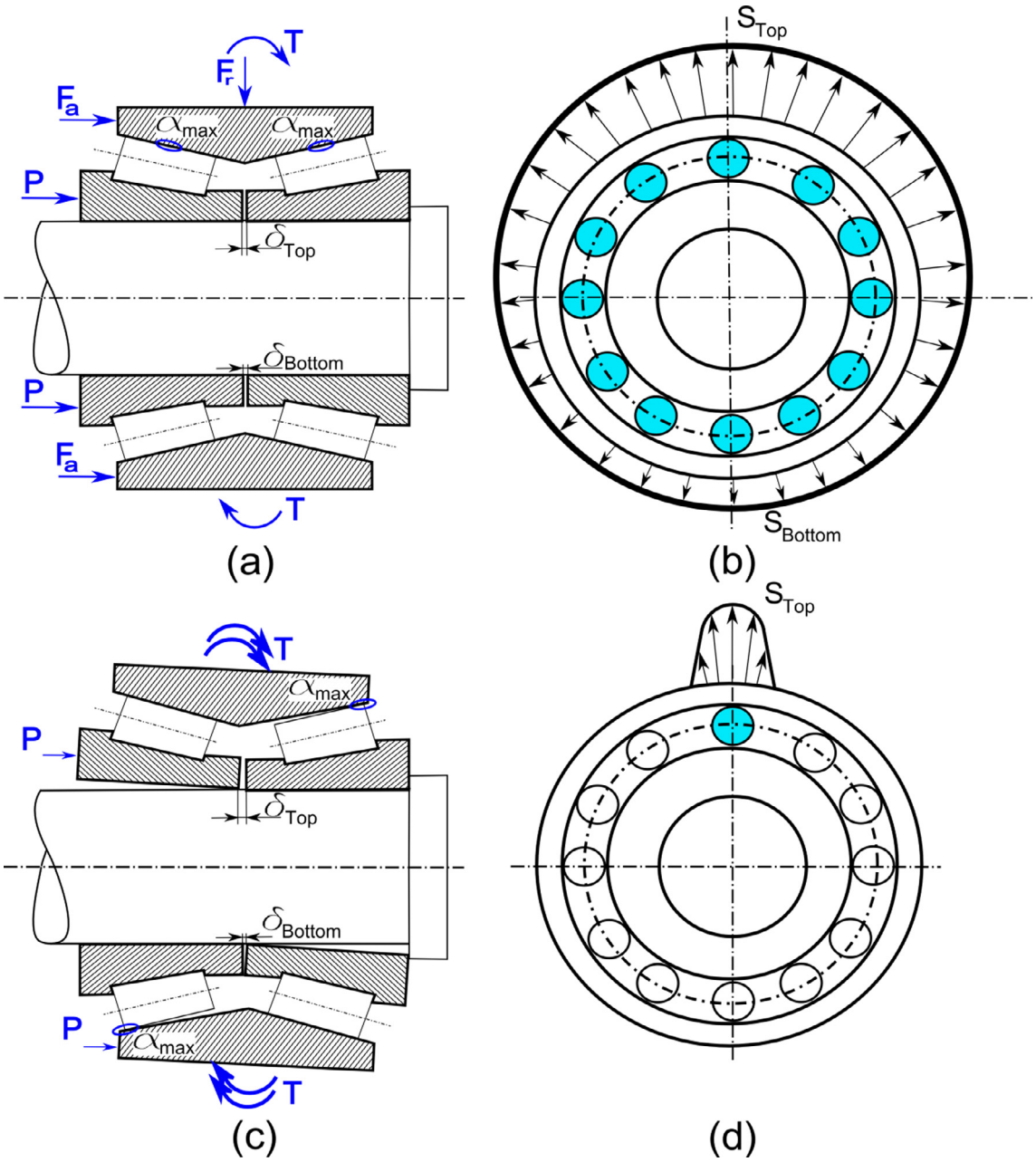

Fig. 1. Main components and the loads on the studied TRB when the external loads on it have standard values: (a) Values of gap δ_{top} and δ_{bottom} that are similar. (b) Contact stress distribution on the outer raceway when there is no roller takeoff. The external loads on the TRB have values that are correct: (c) Different values of gap δ_{top} and δ_{bottom}. (d) Outer raceway contact stress distribution when there is roller takeoff [1].

soft computing or Data Mining (DM) techniques) for a design to optimize the operating conditions or predict bearing failures involves a computational cost that is unacceptable. For example, to predict the distribution of contact stresses on a hub that was mounted on a double-row TRB, Lostado et al. [6] used various types of regression models on the basis of DM techniques. The distribution was determined as a function of *P, F_r* and the friction coefficient. The regression models obtained could be used as viable alternatives in the design phase of this type of mechanical device. Also, Lostado et al. [7] used a combination of FEM and DM to determine the maximum load capacity of double-row TRBs in order to optimizing the operating conditions of these devices. In this case, the work focused on finding a combination of input loads (*P, F_a, F_r* and *T*) in which the contact stress ratio of both rollers rows was close to 25%, and F_r was the maximum. Other researchers have used classification techniques based on machine learning methods in order to determine in advance the failure and malfunction of bearings. Most of these works that are based on classification techniques for predicting bearing failure have been based solely on experimental data. For example, for a fault diagnosis of roller bearings, Sugumaran et al. [8] used a decision tree and a kernel-based neighborhood score, multi-class, Support Vector Machine (SVM). In this case, the conditions of the roller bearings that were studied had good bearing, bearings with inner race faults, bearings with outer race faults, and bearings with inner and outer race faults. Other authors including Gryllias and Antoniadis [9] used SVM approach for an automated diagnosis of rolling bearing fault detection. The effectiveness of this method produces promising results and has the potential use in fault diagnosis of rolling bearings. More recently, Kankar et al. [10] used artificial neural networks (ANN) and SVM to detect and diagnose mechanical faults in ball bearings. The vibration response that was obtained and analyzed for the various

defects were used to train and test the ANN and SVM. The results indicated that the algorithms mentioned above can be used for automated diagnoses of bearing faults. Jiang et al. [11] proposed a novel method based on empirical wavelet transform and ambiguity correlation classification for the fault diagnosis in rolling bearings. The work demonstrated experimentally that the proposed method can effectively diagnose three different operating conditions of rolling element bearings with higher detection rates than provided by SVM and back-propagation ANN.

The current paper searches, using classification techniques based on an approach that combines FEM and DM, the optimal operating conditions in TRB. The DM techniques that were used to classify the categorical response that defines the bearing conditions can be grouped according into supervised dataset and unsupervised dataset [12]. In the supervised classification, the class label for each data instance in the dataset is known and new instances are predicted by a constructed model. However, no instance is labeled in the unsupervised classification, and instances are grouped by a specific set of similarity measures. Both supervised and unsupervised techniques were used. Based on the poor results with the unsupervised techniques and because every instance was labeled, more attention was given to supervised techniques in this work. The studied unsupervised classifying algorithm was model-based clustering that was based on finite normal mixture modeling. However, the number of studied supervised classifying algorithms was higher. In this work, they were studied divided into three groups – linear classification methods, nonlinear classification methods and classification trees and rule-based methods.

The main objective in this work was to construct accurate models with which to automatically classify with higher accuracy the presetting operating conditions of a double-row TRB from several pre-assigned combination of P, F_a, F_r and T loads that define this operation (e.g., bad, poor, good or excellent). Use of classification models will provide additional information about the operating conditions and the possibilities of poor conditions, such as high contact stresses, high local deformation, pitting and fatigue spalling, etc. In order to achieve this, the proposed machine learning methods were applied to the dataset that was obtained from the FE model of the double-row TRB under several loads that define the operating conditions following the Design of Experiments (DoE). Then, several models were created and trained. Meanwhile, feature reduction and transformation, and tuning of parameters were carried out. Later, the models that were selected were tested to determine their degree of generalization on the basis of to several robustness criteria [13–15].

2. Materials and methods

2.1. Combination of loads and operating conditions for double-row tapered roller bearing

Fig. 1(a) shows how the combination of loads (P, F_a, F_r and T) are applied on each of the TRB parts. P loads are applied on the inner raceway, whereas F_a, F_r and T loads are applied on the outer raceway. The inner raceway has two equal parts, which are separated by δ to enable the TRB to be disassembled, as well as to be able to apply the P. It is understood that how the loads that affect the TRB are combined, the bearing may malfunction. For example, if the combined loads on the TRB are equal in value to what the manufacturer recommended, the δ on the top (δ_{top}) should be the same as that on the bottom (δ_{bottom}) (Fig. 1(a)). Also, the outer raceway's top and bottom zones have contact stresses with a value other than zero (Fig. 1(b)). In this situation, all rollers maintain mechanical contact with the inner and outer raceways at all times. Also, there will not be high values of contact stresses and the localized deformations (α). This will avoid the appearance of fatigue spalling and pitting. In contrast, if the values of the loads on the bearing were not correct (i.e., high values of F_r and T, highly combined with low values of P), the gap on the top (δ_{top}) and the bottom (δ_{bottom}) they would have been different. Consequently, there would be faulty operation of the TRB (Fig. 1(c)). Further, there a value of normal contact stresses of nearly zero for only a small area of the outer raceway's upper zone (Fig. 1(d)). In this situation, the contact stresses on the top zone of the outer raceway will be very high (even higher than 1000 MPa), whereas the outer raceway's bottom zone has a contact stress value of zero. This would lead to the appearance of pitting and fatigue spalling and, therefore, to a malfunction of the TRB. In order to quantify the take-off effect of the contact rollers with both raceways, the S ratios for each of the two columns of rollers of the TRB are defined by the below equation.

$$Contact\ Ratio\ (S) = \frac{S_{top}}{S_{bottom}} \cdot 100 \tag{1}$$

where S_{top} is the top contact stress and S_{bottom} is the bottom contact stress obtained for each of the two columns of rollers. Detachment of the TRB rollers may be prevented if the contact ratio for each of the two columns of rollers (S_1 and S_2) exceeds 20%.

Also, in order to quantify the bearing raceways rotation effect and thus determine the malfunction of the bearing, the gap difference (Δδ) is defined by the below equation.

$$Gap\ Difference\ (\Delta\delta) = \delta_{top} - \delta_{bottom} \tag{2}$$

where δ_{top} and δ_{bottom} are the gaps between the inner raceways that correspond to the TRB's upper part and lower part respectively. An optimum condition for the correct functioning of the TRB would be one whose Δδ has been reduced as much as possible.

However, to minimize normal localized deformation of the outer raceway in order to have no pitting or fatigue spalling, Eq. (3) is defined on the basis of Eschmann et al. [16].

$$\alpha_{max} = 0.0001 \cdot d \tag{3}$$

where d is the average diameter of the tapered roller with a value equal to 11.3 mm. In this case, and as a general rule, the value of α_{max} should not exceed the outer raceway at any point by more than 0.00113.

Once the requirements in which the TRB can function properly have been determined, the following ranges of δ, α, S_1 and S_2 have been considered in this paper for classification of the working condition into four categories (bad, poor, good and excellent). With these requirements the number of instances labeled according to each class were: for the class “bad”, 15 instances for training and 4 instances for testing, for the class “poor”, 14 instances for training and 16 instances for testing, for the class “good”, 27 instances for training and 6 instances for testing, and finally, for the class “excellent” 25 instances for training

Table 1
Ranges of Δδ, α_{max}, S_1 and S_2 considered for classification of working conditions.

	Excellent		Good		Poor		Bad	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
α_{max}	6.05e−4	7.25e−4	7.25e−4	8.65e−4	8.65e−4	9.3e−4	9.3e−4	–
Δδ [mm]	1.5e−2	2.5e−2	2.0e−2	3.5e−2	3.0e−2	4.0e−2	4.0e−2	–
S_1 [%]	24	43	17	35	15	27	–	–
S_2 [%]	27	50	20	50	10	31	–	–

and 6 instances for testing. Table 1 summarizes the range proposed in the current paper to classify the mentioned working condition.

2.2. Modeling and experimental validation of the double-row TRB

The three-dimensional FE model of the bearing that are proposed for study of the different combination of loads reproduces in this case a half of double-row TRB. These conditions of symmetrical geometry with symmetrical loading conditions were taken into consideration in order to reduce the computational cost of the simulations. The three-dimensional FE model that was proposed included tapered rollers, inner and outer raceways and the hub on which to mount the TRB. Although quadratic shape functions can potentially predict the mechanical contact existing between raceways and rollers with greater accuracy, the combination of these quadratic functions together with full integration needs a considerable higher computational cost. Also, this problem increases at the same time the number of FE simulations required increases. In this case, a combination of 3D finite elements with 8 and 6 nodes and the use of linear functions with full integration were considered for assembly of the proposed TRB FE models. The geometrical dimensions of the TRB that was studied in this case were: a bore diameter of 78 mm, an outer diameter of 130 mm, a length of 90 mm, and 25 tapered rollers per each column. Because the roughness of the hub that was studied in this case is relatively high, a coefficient of friction of 0.2 between hub and inner raceway was assumed. However, because there is less friction between the rollers and the inner raceway, as well as between rollers and outer raceway, a coefficient of 0.001 was assumed [17]. The complete TRB FE model was formed by 73,484 elements, 84,009 nodes and 3289 degrees of freedom. In contrast to the node-to-segment classical contact detection algorithm, the segment-to-segment contact detection algorithm is able to reduce the error in calculating contact stresses obtained with increasingly smaller mesh densities [18,19]. In this case, with the goal of improving detection of mechanical contact between the different parts of the TRB, a segment-to-segment contact detection algorithm was selected to develop the FE analysis [5]. The Jacobian factor for the entire FE models was always greater than 0.6 and for this reason, elements with zero volume were not generated. In addition, the aspect ratio, defined as the ratio between the longest edge of an element and its shortest edge, never exceeded a value of 10:1 for any of the elements. The raceways and tapered rollers were modeled with use of linear elastic and isotropic steel. The latter had a Young's modulus (E) of 208 GPa and a Poisson's ratio (ν) of 0.29. The hub was modeled on the assumption that $E = 200$ GPa and $\nu = 0.29$ [17]. Fig. 2(a) shows the different component parts of the proposed TRB. Fig. 2(b) shows the final mesh size considered after the FE model adjustment.

The proposed FE model was adjusted and experimentally validated so that the results obtained from it were as realistic as possible. This process were based on the adjustment of the mesh size of the pairs of contact (raceways and rollers) and on the adjustment of the relative displacements between the raceways. This process was performed using P values of 300, 400, 500, and 600 N and an F_r value of 2000 N, as in the previous work undertaken by Lostado et al. [20]. Eight computers with Intel Xeon processor, CPU 3.4 GHz (two processors with four cores each) and 4.00 GB (RAM) were used for simulating all the proposed FE models.

2.2.1. Adapting the TRB to the rollers and raceway mesh size

Non-linearity in problems of mechanical contact is known to be one of the main difficulties in the use of FEM to obtain realistic contact stresses. With this type of non-linear problem, the proposed FE models have limited contact surface between them and the size of mesh is very large, this method may yield contact stresses that are unrealistic. This is usually solved by making the mesh between the pairs of contacts smaller, i.e., increasing the number of nodes in contact, or increasing the degree of the polynomial of the shape functions of the proposed FE model [3]. This adjustment process for a FE model of a cylindrical roller bearing was performed by Demirhan and Kanber [17]. In this case, the effect of element size on contact stress between the raceway and rollers was studied, as well as the relative displacement of the inner and outer raceways. Also, the contact stresses from the different mesh sizes proposed between the rollers and the raceway were compared to theoretical data [21,22]. The adjustment of the model was completed when the difference between the results obtained from the FE model and those obtained theoretically and also experimentally did not differ significantly. In a way that is similar in the current paper to Demirhan's work, FE models with contact areas contained localized fine mesh were created successively by using increasingly smaller mesh size. This method of creating successive meshes by use of with increasingly smaller mesh sizes was developed until there was little difference in the contact stresses between those that occurred on these the outer raceway FE models and those from the theoretical models. Hertz's theory for cylindrical bodies [22] was the basis of the theoretical models to which the different mesh sizes of the proposed FE model were compared. In this case, the Mean Absolute Percentage Error (MAPE) was used to determine when the mesh size that are proposed for FE models are valid. The MAPE is defined as follows in the below equation.

$$MAPE_{y_i} = \frac{1}{m}\sum_{k=1}^{m}\left|\frac{Y_{y_i\,FEM} - Y_{y_i\,TH}}{Y_{y_i\,TH}}\right| \tag{4}$$

where, $Y_{yi\,FEM}$ and $Y_{yi\,TH}$ are the contact stresses of each FE model with its various proposed mesh sizes and those of the theoretical model. k is one of each contact stress obtained for the nodes of to the localized contact zone, whereas m is the total number of nodes that are considered for each mesh size. y_i is the adjustment parameter, which corresponds in this case to the contact area elements' mesh size. Table 2 shows the MAPE obtained in the adjustment of mesh size when the dimensions of the elements in the contact area are 5.1, 1.0 and 0.2 mm for both, raceways and rollers. Also, Table 2 shows the MAPE for preloads P of 300, 400, 500, and 600 N; F_r is 2000 N and the average computational time necessary is studied for each mesh.

In this case, a MAPE that is considered to be valid is lower than 6%, whereas the maximum computational cost is 14,400 min (10 days). Both the MAPE and the computational cost are attainable goals and valid for any problem that is solved by use of FEM. According to these selected criteria, the elements that meet this requirement are those that have a mesh size of 0.2 mm (MAPE and a lower computational cost than the threshold considered for all studied P preloads).

2.2.2. Adjusting the TRB on the basis of relative displacements of the raceways

The relative displacement of the inner and outer raceways was considered in validating the FE model once the mesh size had been adjusted. The relative displacement of the inner and outer raceways is used mainly to verify that there is an appropriate definition of the mesh size of the FE model, the coefficients of friction and the elastic properties (E and ν). Two Red Crown comparator pencils, Fig. 3(a), are used to measure relative displacement experimentally. In the FE model, it is measured between two nodes that belong to the plane of symmetry of the model itself (Fig. 2(a)). As in the mesh size adjustment process, the P values of the loads were 300, 400, 500 and 600 N and the F_r value was 2000 N. F_r was applied to the TRB by a load cell (HBM U3 that had 5KN of capacity). The P was applied by a screw drove a steel sleeve along on the inner raceway. P loads were measured by a strain gauge affixed to a steel sleeve that pushed the inner raceway as the screw turned. Fig. 3(b) shows the differences between the relative displacements obtained from test bench and those from the FE models for various values of P and F_r. The dashed line indicates the experimentally produced relative displacement, whereas the solid line indicates that produced by the FE models. These pairs of curves show that there has been a reduction in the difference between the relative displacements obtained. Also,

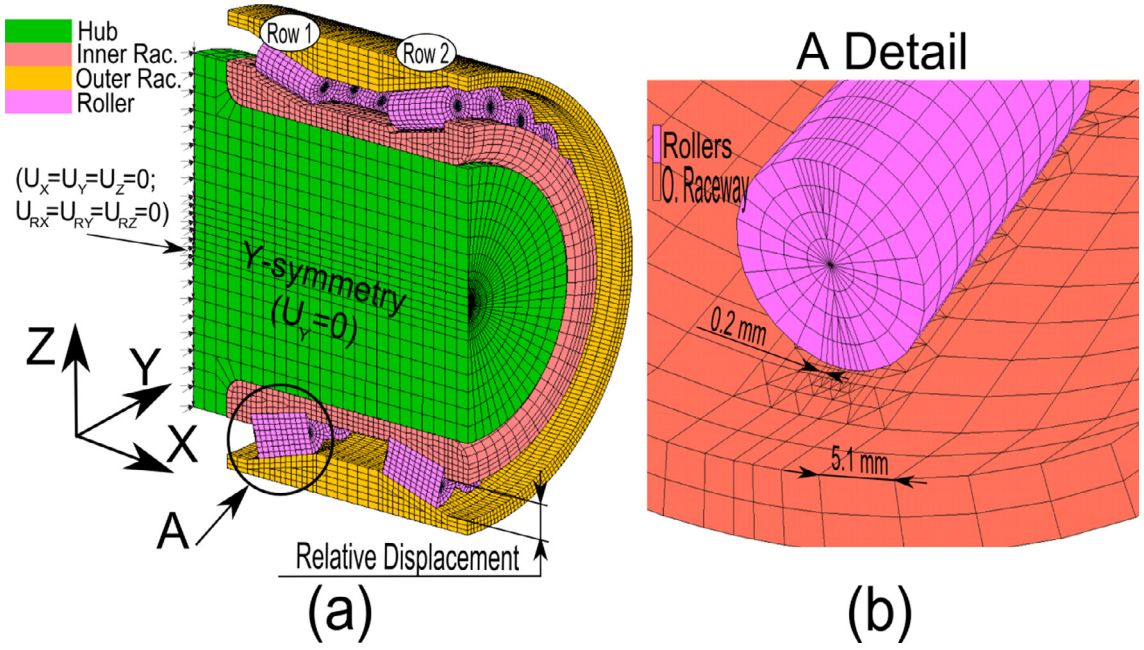

Fig. 2. (a) Main component parts of the TRB. (b) Final mesh size considered after the FE model adjustment.

Fig. 3(b) shows that the greater relative displacement occurred when the lowest P was applied, and the minimum relative displacement was obtained when the highest P was applied to both FE models and experimental data. Fig. 3(b) also shows as the preload rises, the raceway roller contact stiffness rises and the TRB deformation declines. Fig. 3(c) shows in detail the different relative displacement reached values experimentally and through the FE models. This stiffness variation agrees well with the load–deflection relationship that Harris and Kotzalas [1] and Kania [23] reported previously, which indicates that the higher value of preload in a TRB, the greater its rigidity is. From Fig. 3(b), it can be deduced that the relative displacements with the FE models, if the mesh size valid, were in agreement with the experimental data and the FE model's preset parameters ($E_{bearing}$ = 200 GPa, E_{hub} = 208 GPa, ν = 0.29 and coefficients of friction = 0.001 and 0.2). Although the loads used in the optimization process were higher (see Table 4) than those loads used in the adjustment process, the preset parameters as well as the mesh size described in Section 2.2.1 were set for all FE simulations of the optimization process [24]. This consideration was mainly due to the fact that none of the parts of the TRB suffered a permanent plastic deformation since the loads used in the optimization process were within of the TRB work zone according to the manufacturer [25].

2.3. Design of experiments

The number of variables/factors (different loads that are based on an optimal presetting) and levels for each variable that is considered can be too large to cover the entire space of possibilities. Thus, the application of a DoE in this work to reduce the large number of combinations provides a good compromise between the number of samples that is desirable for training models and practical aspects of FEM simulations. In this work, a 3^k full factorial DoE with three levels was conducted. Table 3 provides the values of factors and levels necessary for implementation of the DoE on the basis of the 3^k full factorial.

After establishing the factors and levels that appear as in Table 3, the design matrix and load combinations were generated. In this case, it was necessary to conduct 81 experiments or FE simulations to cover all possibilities and, subsequently, to optimize presetting loads to improve working conditions for a double-row TRB. How many experiments or FE simulations were required, as well as the values of the loads to simulate are indicated in the design matrix (Table 4).

2.4. Multivariate analysis and clustering analysis of the data

Before using a classification method, all dataset variables were normalized between 0 and 1. This was done to balance the weight of the variables and improve the statistical analysis. Then, an exploratory analysis of the variance and correlation was conducted followed by statistical analyses of the variance. The latter consisted of parametric and nonparametric: ANOVA [26], Bartlett [27], Brown–Forsyth [28] and Fligner–Killeen [29]. The analyses sought to determine each feature's significance to the class. Also, these analyses were conducted to determine if the four classes differed significantly and contained any redundant features. In addition, a Principal Component Analysis (PCA) yielded a new dataset that had new features, which subsequently was studied. PCA was used to simplify the problem. The former is a popular method that is used successfully for this propose [30–32]. It produces

Table 2
MAPE and the FE model's computational cost of the FE model while the mesh density is adjusted.

Mesh size (mm)	MAPE (%)				Computational cost (min)
	P = 300 N F_r = 2000 N	P = 400 N F_r = 2000 N	P = 500 N F_r = 2000 N	P = 600 N F_r = 2000 N	
5.1	53.60	53.37	52.86	52.34	6912
1.0	22.34	22.56	22.64	22.68	8496
0.2	4.58	4.68	4.78	4.88	11,952

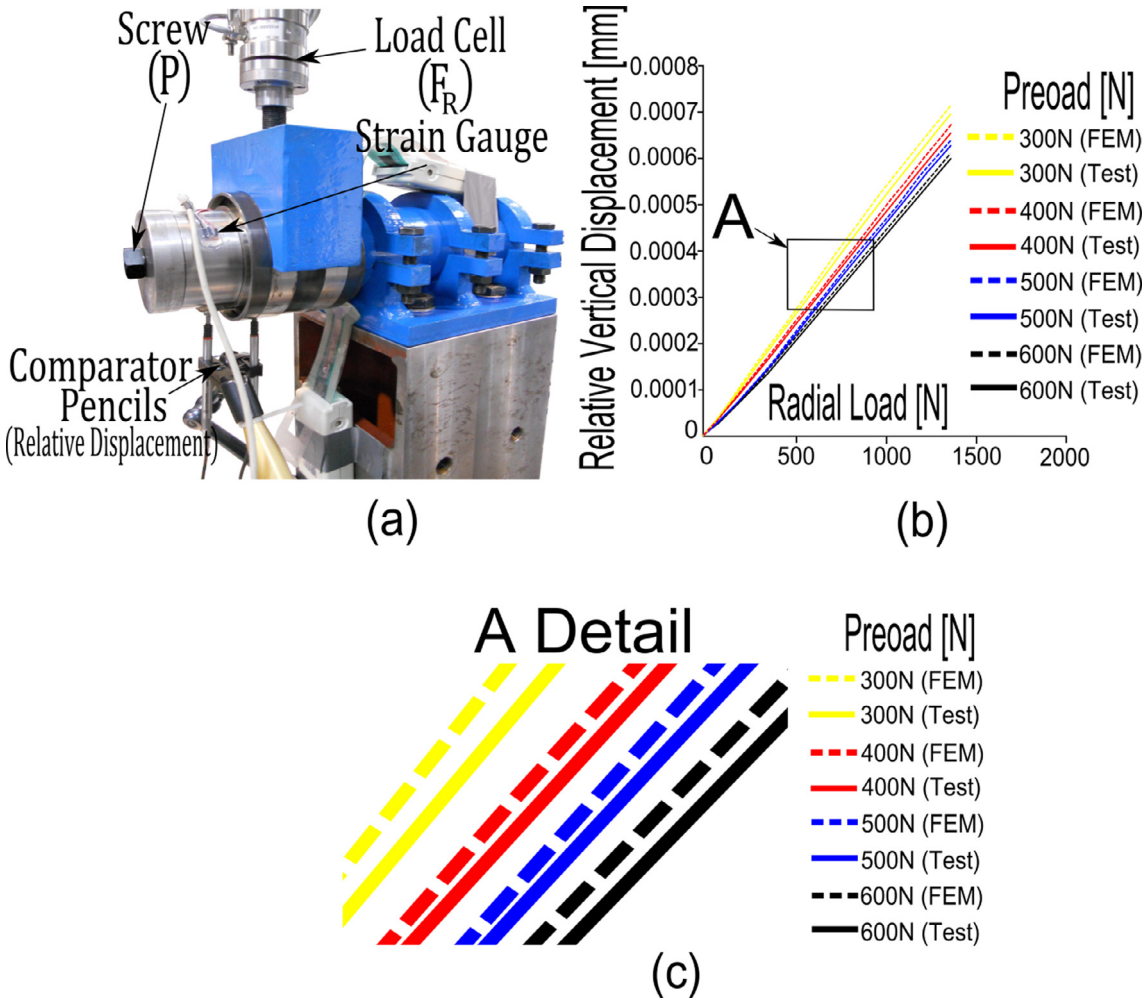

Fig. 3. (a) TRB test bench: comparator pencil, strain gauge and load cell assembly. (b) FE model's relative displacement and experimental data at various values of *P*. (c) Detail of the relative displacement. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Table 3
Selection of DoE parameters, ranges, levels according of each factor and the 3k full factorial method.

Factors	Units	Levels		
		0	1	2
P	N	8000	9000	10,000
F_r	N	60,000	80,000	100,000
F_a	N	−200	200	600
T	N mm	−100,000	0	100,000

an orthogonal change of the variables' initial space to create a set of variables that is known as the Principal Components (PCs). The orthogonal projection captures most the variance in the data and transforms a set of variables that may be correlated into a set of components that are not correlated linearly. Also, to finally apply the proposed classification methodology, a clustering analysis of the data from the proposed DoE was conducted to determine if the input variables can easily define the classification problem. In this case a model-based clustering based on finite normal mixture modeling [33,34] was applied for the automated search of groups of related observations in a dataset.

2.5. Supervised classification methods

The proposed approach used several supervised classification techniques to classify the presenting conditions of the double-row TRB by the four categories that were studied (Table 5). To implement the process, the dataset containing the selected features that define de presetting conditions (P, F_a, F_r and T) was used as input. Along with the building and training stage, feature reduction and parameter tuning were undertaken in order to simplify the models and achieve more robust classification models. Then, the models from this stage were tested to determine their real generalization capacity. The data selected for that test had not been used during the training. The models' performance test results were compared to identify and select the most accurate models. Then, the results were validated to ensure that the model addressed all aspects of the problem without any overtraining.

The entire classification process was performed according to the following methodology. The dataset was separated into two groups. The first, according to the planed DoE, was called the training dataset (81 samples). The second, which was formed by some random instances within the ranges that were proposed on the DoE, was called testing dataset (32 samples). The training dataset was selected to build and train the models using repeated ten-fold cross-validation. This training process was repeated 50 times as some methods (e.g., ANNs or SVM) use random initial values in the learning process. As a result, it was necessary to repeat the process several times for greater accuracy of results. During this training, each algorithms most significant parameters were tuned for better results. For each algorithm, the most accurate models were selected for testing. At this stage, the accuracy of the models that had been selected was tested. This employed the

Table 4
Design matrix and the load combinations to simulate.

Run	P [N]	F_r [N]	F_a [N]	T [N·mm]	Run	P [N]	F_r [N]	F_a [N]	T [N·mm]	Run	P [N]	F_r [N]	F_a [N]	T [N·mm]
1	8000	60,000	−200	−100,000	28	9000	60,000	−200	−100,000	55	10,000	60,000	−200	−100,000
2	8000	60,000	−200	0	29	9000	60,000	−200	0	56	10,000	60,000	−200	0
3	8000	60,000	−200	100,000	30	9000	60,000	−200	100,000	57	10,000	60,000	−200	100,000
4	8000	60,000	200	−100,000	31	9000	60,000	200	−100,000	58	10,000	60,000	200	−100,000
5	8000	60,000	200	0	32	9000	60,000	200	0	59	10,000	60,000	200	0
6	8000	60,000	200	100,000	33	9000	60,000	200	100,000	60	10,000	60,000	200	100,000
7	8000	60,000	600	−100,000	34	9000	60,000	600	−100,000	61	10,000	60,000	600	−100,000
8	8000	60,000	600	0	35	9000	60,000	600	0	62	10,000	60,000	600	0
9	8000	60,000	600	100,000	36	9000	60,000	600	100,000	63	10,000	60,000	600	100,000
10	8000	80,000	−200	−100,000	37	9000	80,000	−200	−100,000	64	10,000	80,000	−200	−100,000
11	8000	80,000	−200	0	38	9000	80,000	−200	0	65	10,000	80,000	−200	0
12	8000	80,000	−200	100,000	39	9000	80,000	−200	100,000	66	10,000	80,000	−200	100,000
13	8000	80,000	200	−100,000	40	9000	80,000	200	−100,000	67	10,000	80,000	200	−100,000
14	8000	80,000	200	0	41	9000	80,000	200	0	68	10,000	80,000	200	0
15	8000	80,000	200	100,000	42	9000	80,000	200	100,000	69	10,000	80,000	200	100,000
16	8000	80,000	600	−100,000	43	9000	80,000	600	−100,000	70	10,000	80,000	600	−100,000
17	8000	80,000	600	0	44	9000	80,000	600	0	71	10,000	80,000	600	0
18	8000	80,000	600	100,000	45	9000	80,000	600	100,000	72	10,000	80,000	600	100,000
19	8000	100,000	−200	−100,000	46	9000	100,000	−200	−100,000	73	10,000	100,000	−200	−100,000
20	8000	100,000	−200	0	47	9000	100,000	−200	0	74	10,000	100,000	−200	0
21	8000	100,000	−200	100,000	48	9000	100,000	−200	100,000	75	10,000	100,000	−200	100,000
22	8000	100,000	200	−100,000	49	9000	100,000	200	−100,000	76	10,000	100,000	200	−100,000
23	8000	100,000	200	0	50	9000	100,000	200	0	77	10,000	100,000	200	0
24	8000	100,000	200	100,000	51	9000	100,000	200	100,000	78	10,000	100,000	200	100,000
25	8000	100,000	600	−100,000	52	9000	100,000	600	−100,000	79	10,000	100,000	600	−100,000
26	8000	100,000	600	0	53	9000	100,000	600	0	80	10,000	100,000	600	0
27	8000	100,000	600	100,000	54	9000	100,000	600	100,000	81	10,000	100,000	600	100,000

Table 5
Applied classification techniques.

Methodology category	Name of the classification technique	References
Linear classification methods	Linear Discriminant Analysis (LDA)	[35,36]
	Partial Least Squares discriminant analysis (PLS)	[37,38]
	Nearest Shrunken Centroids (NSC)	[39,40]
Nonlinear classification methods	Mixture Discriminant Analysis (MDA)	[41]
	Flexible Discriminant Analysis (FDA)	[42–44]
	Feed-forward Artificial Neural Networks (ANN)	[45]
	K-nearest neighbors (KNN)	[45]
	Support Vector Machines (with Linear kernel (SVM-L), with Polynomial kernel (SVM-P), and with Radial Basic Function kernel (SVM-RBF))	[46–48]
Classification trees and rule-based methods	Basic Classification Trees (BCT)	[49]
	C 5.0 (C50)	[50]
	C 4.5 (C45)	[51]
	Bagged CART (BCART)	[52]
	Random Forest (RF)	[53]
	Evolutionary learning of globally optimal classification Tree (evTree)	[54]

dataset that had not been used for training. This testing enabled the selection of the most accurate models, as their accuracy had been determined with instances unused previously. After selecting, for each technique, the most accurate models, the process was validated. The validation showed that the selected models address successfully all aspects and also demonstrated their classification performance. Finally, the most accurate model was selected based on these results. It best classifies the optimal presetting values and ensures correct working conditions.

2.6. Criteria of robustness

A model's robustness is commonly accepted as a useful parameter in the selection of models. It is normally verified by its overall classification accuracy at a 95% confidence level as an objective function. Other criteria also are considered when overall classification accuracies are similar (i.e., sensitivity, un-weighted kappa statistic, specificity, detection rate, prevalence, positive prediction value, negative detection prevalence, prediction value and balanced accuracy) (Table 6).

R statistical software environment v3.4.1 was used to program the

Table 6
Metrics used to optimize the classification models' accuracy and to select the best of them.

$$Accuracy = \frac{TP + TN}{TI}$$

$$k = \frac{p_a + p_e}{1 + p_e}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{FP + TN}$$

$$Prevalence = \frac{TP + FN}{TP + FP + FN + TN}$$

$$PPV = \frac{Sensitivity \cdot Prevalence}{(Sensitivity \cdot Prevalence + (1 - Specificity) \cdot (1 - Prevalence))}$$

$$NPV = \frac{Specificity \cdot (1 - Prevalence)}{((1 - Sensitivity) \cdot Prevalence + Specificity \cdot (1 - Prevalence))}$$

$$DetectionRate = \frac{TP}{TP + FP + FN + TN}$$

$$DetectionPrevalence = \frac{TP + FP}{TP + FP + FN + TN}$$

$$BalancedAccuracy = \frac{Sensitivity + Specificity}{2}$$

TI (number of instances)
TP (instance that has the condition and tests positive for it)
TN (instance that does not have the condition and tests negative for it)
FP (instance that does not have the condition, but tests positive for it)
FN (instance that has the condition, but tests negative for it)
p_a (proportion of instances in agreement)
p_e (proportion of instances in agreement due to chance)

Table 7
Results obtained from the 81 FE simulations based on the DoE defined on Table 4.

Run	S_1[%]	S_2[%]	α_{max}	$\Delta\delta$ [mm]	Run	S_1[%]	S_2[%]	α_{max}	$\Delta\delta$ [mm]	Run	S_1[%]	S_2[%]	α_{max}	$\Delta\delta$ [mm]
1	33.368	29.973	0.00064	0.0216	28	37.237	36.349	0.00064	0.0214	55	40.095	39.721	0.00067	0.0217
2	28.819	34.251	0.00062	0.0196	29	32.783	41.148	0.00065	0.0194	56	36.277	44.336	0.00068	0.0198
3	24.691	39.368	0.00064	0.0197	30	28.818	46.626	0.00067	0.0197	57	32.078	49.855	0.00070	0.0197
4	34.542	28.883	0.00063	0.0223	31	38.237	35.442	0.00065	0.0223	58	41.410	38.927	0.00068	0.0222
5	30.070	33.236	0.00063	0.0199	32	30.702	32.049	0.00064	0.0204	59	36.938	43.358	0.00069	0.0198
6	25.724	38.746	0.00065	0.0202	33	29.737	45.829	0.00068	0.0204	60	33.386	49.021	0.00071	0.0207
7	35.979	27.774	0.00063	0.0233	34	39.228	34.584	0.00067	0.0228	61	42.395	38.058	0.00070	0.0228
8	31.032	31.847	0.00064	0.0453	35	34.621	39.044	0.00067	0.0207	62	37.781	42.475	0.00070	0.0199
9	27.063	37.589	0.00067	0.0209	36	30.961	44.905	0.00069	0.0209	63	34.336	48.155	0.00073	0.0212
10	25.401	22.042	0.00109	0.0278	37	28.691	27.787	0.00076	0.0274	64	31.357	31.101	0.00079	0.0285
11	20.995	25.767	0.00074	0.0259	38	27.700	34.715	0.00080	0.0254	65	27.989	34.852	0.00080	0.0261
12	17.486	29.891	0.00076	0.0260	39	21.779	36.088	0.00079	0.0259	66	24.648	39.292	0.00082	0.0264
13	25.622	21.102	0.00076	0.0289	40	29.628	26.867	0.00075	0.0291	67	32.049	30.580	0.00080	0.0287
14	21.524	24.590	0.00075	0.0276	41	25.558	30.451	0.00087	0.0266	68	28.526	33.806	0.00081	0.0266
15	18.472	28.814	0.00077	0.0270	42	22.731	35.120	0.00080	0.0272	69	25.673	38.382	0.00083	0.0271
16	26.710	20.013	0.00076	0.0306	43	30.506	25.831	0.00078	0.0296	70	33.029	29.626	0.00082	0.0298
17	22.729	23.862	0.00077	0.0279	44	26.402	29.662	0.00079	0.0268	71	29.398	32.944	0.00082	0.0272
18	20.100	28.150	0.00079	0.0275	45	23.529	34.042	0.00081	0.0273	72	26.585	37.462	0.00084	0.0278
19	20.536	17.675	0.00089	0.0355	46	21.845	21.420	0.00088	0.0343	73	24.745	24.660	0.00091	0.0345
20	17.793	19.625	0.00087	0.0337	47	19.089	24.682	0.00089	0.0326	74	22.104	27.705	0.00086	0.0324
21	15.770	23.149	0.00089	0.0339	48	16.892	28.344	0.00091	0.0327	75	19.227	30.687	0.00131	0.0329
22	20.883	17.467	0.00089	0.0359	49	22.811	20.450	0.00089	0.0358	76	25.934	23.269	0.00130	0.0357
23	18.153	18.821	0.00087	0.0338	50	19.924	23.016	0.00126	0.0327	77	22.805	25.925	0.00129	0.0332
24	16.135	22.304	0.00089	0.0341	51	17.417	27.714	0.00092	0.0331	78	19.850	30.466	0.00095	0.0336
25	21.603	17.355	0.00125	0.0366	52	23.878	20.233	0.00091	0.0367	79	26.653	23.163	0.00094	0.0361
26	18.659	18.228	0.00124	0.0338	53	20.560	22.985	0.00092	0.0342	80	23.428	26.153	0.00094	0.0339
27	16.433	21.854	0.00128	0.0343	54	18.111	26.592	0.00094	0.0346	81	20.864	29.842	0.00097	0.0344

proposed methodologies and to develop the classification models [55].

3. Results and discussion

3.1. Results from the FE models

Once the FE model had been validated, an automatic procedure conducted the 81 FE simulations according to the proposed DoE of Table 4. The results of these FE simulations are shown in Table 7 (S_1, S_2, α_{max} and $\Delta\delta$).

Fig. 4 shows the distribution of the dataset generated according to the applied loads (Table 1) and the geometry obtained from the FE simulations (Table 7). Fig. 4(a) shows a distribution of the dataset on a 3D chart made up of α_{max} and S_1 and S_2. Fig. 4(b) shows another distribution of the dataset on a 3D chart, but in this case consisting of $\Delta\delta$ and S_1 and S_2. In both figures, one can see the structures of patterns that are related to the conditions of excellent, good, poor and bad.

Additionally, 32 new FE models were used to test the classification procedure with a new combination of loads (Table 8). Similarly, an automatic procedure was used to run the 32 new FE simulations. The results of these FE simulations appear in Table 9.

3.2. Multivariate analysis and unsupervised classification

Four proposed statistical tests of homogeneity of variances were conducted for the dataset's four variables. The statistics, as well as the *p*-values, were calculated for all (Table 10). In addition, a correlation analysis was conducted. This concluded that, for the present working conditions, four features are significant, although F_r is the most important feature.

Later, the dataset was introduced in a PCA and the results showed that all new PCs explained approximately the same proportion of variance of the data (Table 11) than the original features. For this reason, it was considered that using the original features was a better option, since working with known attributes instead of linear combination of them will build models easier to understand.

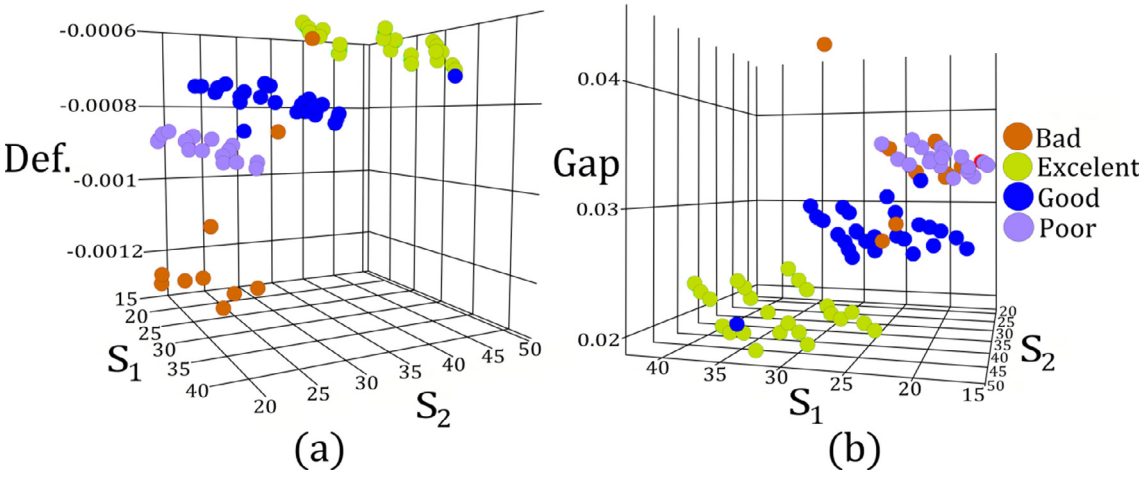

Fig. 4. Distribution of the dataset made up of: (a) α_{max}, S_1 and S_2. (b) $\Delta\delta$, S_1 and S_2.

Table 8
New combination of loads to simulate.

Run	P [N]	F_r [N]	F_a [N]	T [N·mm]	Run	P [N]	F_r [N]	F_a [N]	T [N·mm]	Run	P [N]	F_r [N]	F_a [N]	T [N·mm]
1	8470	65,350	220	37,890	12	9312	86,228	325	31,141	23	9043	72,340	245	20,345
2	8475	100,000	−150	−90,000	13	8656	73,117	62	−34,416	24	8000	60,345	−106	−20,450
3	8530	100,000	−150	−90,000	14	8084	61,682	−167	−91,589	25	8345	61,456	365	−24,570
4	8630	97,900	−300	−12,530	15	8206	96,782	304	97,985	26	8155	97,750	−59	−97,345
5	9050	99,500	−195	88,085	16	8245	98,989	−102	−96,789	27	8989	70,325	256	72,000
6	9835	90,480	245	−93,315	17	8325	99,800	−195	−98,350	28	8870	62,450	−90	25,096
7	9909	98,181	564	90,901	18	8326	95,880	458	96,350	29	8786	95,890	456	−96,894
8	98,181	564	90,901	98,181	19	8525	99,880	568	99,350	30	9234	97,780	468	−95,870
9	99,771	596	98,854	99,771	20	8825	99,830	195	98,350	31	8231	99,875	534	−93,450
10	95,582	512	77,908	95,582	21	8003	99,679	−199	−98,769	32	8578	98,756	−186	95,860
11	87,712	354	38,561	87,712	22	9875	60,535	179	75,000	–	–	–	–	–

Table 9
Results of the FE simulations using inputs from Table 8.

Run	S_1[%]	S_2[%]	α_{max}	Δδ [mm]	Run	S_1[%]	S_2[%]	α_{max}	Δδ [mm]	Run	S_1[%]	S_2[%]	α_{max}	Δδ [mm]
1	27.744	34.421	−0.00069	0.02255	12	24.511	30.404	−0.00084	0.02951	23	28.531	35.002	−0.00075	0.02497
2	20.796	18.501	−0.00090	0.03485	13	27.829	28.855	−0.00073	0.02597	24	29.996	33.298	−0.00063	0.02116
3	20.878	18.622	−0.00090	0.03464	14	32.371	29.648	−0.00065	0.02217	25	31.971	32.714	−0.00066	0.02234
4	18.491	22.756	−0.00089	0.03352	15	16.710	23.713	−0.00088	0.03307	26	20.971	17.947	−0.00087	0.03375
5	17.263	28.390	−0.00090	0.03252	16	20.779	17.947	−0.00089	0.03430	27	26.996	38.243	−0.00074	0.02376
6	28.200	26.395	−0.00086	0.03269	17	20.756	18.037	−0.00089	0.03426	28	31.793	41.361	−0.00067	0.02033
7	21.133	29.818	−0.00095	0.03347	18	17.256	23.866	−0.00089	0.03349	29	23.461	18.961	−0.00088	0.03522
8	20.956	29.867	−0.00096	0.03359	19	17.203	23.465	−0.00093	0.03491	30	25.036	21.574	−0.00090	0.03499
9	22.249	30.085	−0.00093	0.03284	20	17.165	24.819	−0.00092	0.03413	31	21.504	17.779	−0.00089	0.03557
10	17.913	37.385	−0.00088	0.02945	21	20.483	17.559	−0.00088	0.03473	32	16.589	25.409	−0.00089	0.03309
11	27.744	34.421	−0.00069	0.02255	22	33.965	47.088	−0.00071	0.02057	–	–	–	–	–

Table 10
Results of tests of homogeneity of variance for each variable.

Attribute	ANOVA p-value	Bartlett p-value	Brown–Forsyth p-value	Fligner–Killeen p-value
P	0.5601	0.9190	0.6608	0.9190
F_r	8.7122e−32	0.4094	0.1798	0.4094
F_a	0.0528	0.9072	0.8667	0.9072
T	0.9942	0.9887	0.5258	0.9887

Table 11
Importance of the PCs.

	PC1	PC2	PC3	PC4
Standard deviation	0.4230	0.4085	0.3756	0.3404
Proportion of variance	0.2969	0.2769	0.2341	0.1922
Cumulative proportion	0.2969	0.5737	0.8078	1.0000

From this point, an unsupervised clustering was done to obtain four groups of classification according to the four original features. The Gaussian finite mixture model that was fitted by the EM algorithm was used to determine if and how the aggrupation of the data could define four clusters (equal to the number of factors on the class), based on the DoE information and similarity measures of their instances. This unsupervised algorithm was evaluated for multivariate mixture (Fig. 5) showing that the Bayesian Information Criterion (BIC) increased, in absolute value, significantly at the same time as the number of clusters. The results of this analysis showed that the data cannot be grouped into four groups according to the classes with low errors using simple clustering techniques. The best BIC was −343.45 for one cluster and for the next three multivariate mixtures: EII, VII, EEI.

3.3. Results of the classification techniques

At this point, because the feature reduction of variables using PCA and the use of cluster techniques to make classification groups did not improve the method, supervised classification was performed with all of the original features. Several models were built by the proposed method and trained with the training dataset. The algorithm's defining parameters at this stage were tuned to improve the accuracy. The techniques applied that were used are listed in Table 12. The table also gives the parameters that had been tuned during the analysis, the tuning range, and the values that were chosen for the final models. A ten-fold cross-validation process that had been repeated 50 times determinated the models' classification accuracy. The main criterion to analyze this accuracy was the overall accuracy. However, if similar

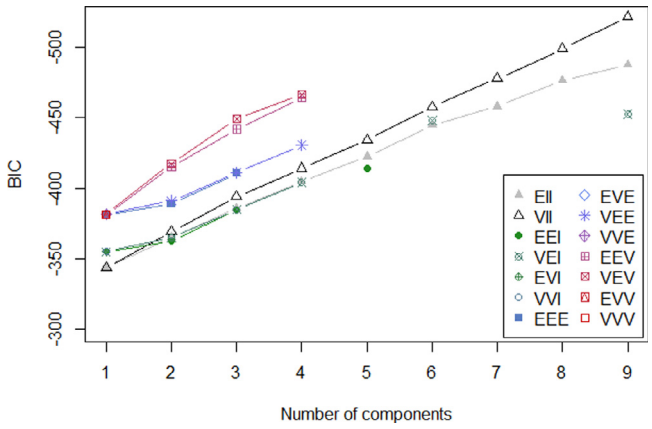

Fig. 5. Relationship of the number of clusters to the BIC. Several multivariate mixtures were applied. They were: EII (spherical, equal volume), VII (spherical, unequal volume), EEI (diagonal, equal volume and shape) VEI (diagonal, varying volume, equal shape), EVI (diagonal, equal volume, varying shape), VVI (diagonal, varying volume and shape), EEE (ellipsoidal, equal volume, shape, and orientation), EVE (ellipsoidal, equal volume and orientation), VEE (ellipsoidal, equal shape and orientation), VVE (ellipsoidal, equal orientation), EEV (ellipsoidal, equal volume and equal shape), VEV (ellipsoidal, equal shape) EVV (ellipsoidal, equal volume), and VVV (ellipsoidal, varying volume, shape, and orientation).

Table 12
Tuning of the most significant parameters for each technique, grid of studied values and the final selection.

Technique	Tuned parameters	Studied values	Selection
LDA	No tuning		
PLS	**ncomp** (number of components used in the estimation of variance)	From 1 to 5	2.0
NSC	**threshold** (centroid shrinkage value)	From 1 to 31	1.0
MDA	**subclasses** (number of subclasses in each class)	From 1 to 9	2.0
FDA	**degree** (max. degree of interaction, Friedman's mi)	From 1 to 6	2.0
	nprune (max. number of terms in the pruned model (including intercept))	From 1 to 21	5.0
ANN	**size** (neurons in the only hidden layer)	From 1 to 21	4.0
	decay (parameter to decrease the learning rate of the optimization function)	From 0.001 to 0.1	0.1
KNN	**k** (number of neighbors considered)	From 1 to 21	3.0
SVM-L	**C** (cost of constraints violation)	From 0.2 to 12	0.2
SVM-P	**C** (cost of constraints violation)	From 0.1 to 3	0.2
	degree (degree of the polynomial kernel function)	From 1 to 6	1.0
SVM-RBF	**C** (cost of constraints violation)	From 0.1 to 7	5.4
	sigma (inverse kernel width used by the Gaussian kernel)	From 0.1 to 0.7	0.1
BCT	**cp** or complexity parameter (any split that does not decrease the overall lack of fit by a factor of cp is not attempted)	From 0 to 3	0.0
C50	**trials** (number of boosting iterations)	From 1 to 150	1.0
	model (tree model or decomposed into a rule-based model)	Tree, rules	Rules
	winnow (feature selector)	Yes, no	No
C45	**C** (confidence threshold)	From 0 to 2	0.26
BCART	No tuning		
RF	**mtry** (randomly sampled variables as candidates for each split)	From 1 to 150	51.0
evTree	**α** or complexity parameter (any split that does not decrease the overall lack of fit by a factor of α is not attempted)	From 0 to 4	0.4

error percentages raised any doubt, the other robustness criteria were also studied.

For each technique, the most accurate models in the training stage was then tested. The testing dataset was used to determine the real generalization capacity of each. Results of the training and testing stages appear in Table 13. These results show that some methods like ANN and SVM from the nonlinear classification methods or C45 and evTree from the classification trees and rule-based methods produced really accurate results during the training stage. However, when new instances were applied to the models, the accuracy was much lower. This reduction in the accuracy shows that, during the training stage, there was overfitting of some models, and that the use of a testing dataset proved to be very useful for analyzing the real generalization capacity of the models.

The model that has all original features and is based on FDA provided the greatest accuracy in classifying optimal presetting conditions at these stages. However, differences in accuracy among different nonlinear techniques on the basis of discriminant analysis were not too significant. Therefore it was decided to use the other proposed criteria to determine which FDA was the most accurate (e.g., the un-weighted kappa statistic, whose value was greater using FDA that using MDA).

Table 14
Confusion matrix from the model based on FDA during the testing stage.

	Bad	Excellent	Good	Poor
Bad	3	0	1	5
Excellent	0	6	0	0
Good	0	0	5	1
Poor	1	0	0	10

Table 14 shows the confusion matrix of the testing dataset using the selected FDA model. This matrix indicates that the prediction of the class ‘excellent’, the class that is requested to ensure correct working conditions for the TRB, was quite accurate. This prediction shows how every class labeled as ‘excellent’ was predicted as ‘excellent’. However, instances labeled as ‘poor’ were the more difficult to classify since six of the sixteen instances were wrongly predicted. Classes ‘bad’ and ‘good’ predicted all of the instances as labeled, except one. Also, Table 15 shows more statistics of the performance of the classification model.

Although the total accuracy when the testing dataset was used for the FDA model was 75%, when these results were analyzed, it is seen that this accuracy can be improved. In the case in which the four classes

Table 13
Results of the training and testing stages using all dataset features. The most accurate models are indicated by bold letters.

Classification technique	Training		Testing		
	Accuracy (%)	Kappa	Accuracy (%)	95% confidence interval	Kappa
LDA	83.84	0.7762	68.75	(49.99, 83.88)	0.5676
PLS	68.89	0.5757	53.12	(34.74, 70.91)	0.2749
NSC	63.14	0.4531	37.50	(21.10, 56.31)	0.2308
MDA	84.20	0.7814	75.00	(56.60, 88.54)	0.6224
FDA	**84.93**	**0.7907**	**75.00**	**(56.60, 88.54)**	**0.6541**
ANN	85.03	0.7927	68.75	(49.99, 83.88)	0.5733
KNN	84.48	0.7835	59.38	(40.64, 76.30)	0.4124
SVM-L	85.01	0.7923	65.62	(46.81, 81.43)	0.5319
SVM-P	85.01	0.7923	65.62	(46.81, 81.43)	0.5319
SVM-RBF	86.69	0.8145	68.75	(49.99, 83.88)	0.5604
BCT	81.71	0.7463	71.88	(53.25, 86.25)	0.5909
C50	84.85	0.7897	43.75	(26.36, 62.34)	0.2519
C45	88.27	0.8372	43.75	(26.36, 62.34)	0.2519
BCART	81.45	0.7421	56.25	(37.66, 73.64)	0.3829
RF	82.38	0.7556	56.25	(37.66, 73.64)	0.3829
evTree	85.64	0.8011	34.38	(18.57, 53.19)	0.1724

Table 15
Statistics for the model based on FDA during the testing stage.

	Bad	Excellent	Good	Poor
Sensitivity	0.7500	1.0000	0.8333	0.6250
Specificity	0.7857	1.0000	0.9615	0.9375
PPV	0.3333	1.0000	0.8333	0.9091
NPV	0.9565	1.0000	0.9615	0.7143
Prevalence	0.1250	0.1875	0.1875	0.5000
Detection rate	0.0937	0.1875	0.1562	0.3125
Detection prevalence	0.2812	0.1875	0.1875	0.3438
Balanced accuracy	0.7678	1.0000	0.8974	0.7812

Table 16
Confusion matrix from the model based on FDA during the testing stage for only two classes 'Bad–Poor' and 'Good–Excellent'.

	Bad–Poor	Good–Excellent
Bad–Poor	19	1
Good–Excellent	1	11

Table 17
Relationship of the three dimensions obtained of the model to the original features.

	θ_1	θ_2	θ_3
Bias	0.2855322	−1.1668836	−0.3981208
$h(0.5 - P)$	−2.7839394	3.5875880	−0.0334087
$h(P - 0.5)$	−0.7762908	2.1148651	5.9943328
$F_r \cdot h(P - 0.5)$	1.2848951	0.5387261	−3.1010467
$h(P - 0.5) \cdot F_a$	2.1682605	1.8530735	−3.6390169

Table 18
Prediction results for each and all of the dimensions. (GCV = Generalized Cross-Validation; RSS = Residual Sum-of-Squares; GRSq = Generalized *R*-Squared; and RSq = *R*-Squared).

	GCV	RSS	GRSq	RSq
θ_1	0.3556398	21.51401	0.6530872	0.7343949
θ_2	0.3698497	22.37362	0.6392260	0.7237824
θ_3	0.5230440	31.64093	0.4897909	0.6093712
All	1.2485335	75.52857	0.5940347	0.6891828

that define the problem were grouped into just two classes, one class was defined by the two original classes ('bad' and 'poor') that produce the TRB works under unsuitable operating conditions, and another class was defined by the other two original classes ('good' and 'excellent') that produce the TRB works under suitable operating conditions. In this case, the confusion matrix, which appears in Table 16, indicates that only two instances were predicted as a class that differs from its label. That means that the total accuracy of the model can be improved to 93.75% when classes that are related are merged to reduce the number of classes.

FDA is based on the linear technique LDA, but with an extra penalization strategy. This model uses an elastic-net strategy [56]. L_1 penalties have, in effect, eliminated predictors, whereas an L_2 penalty decreases the coefficients of discriminant functions. The dimension of discriminant space on the constructed model is equal to 3 (θ_1, θ_2, θ_3), and the percent of variance that are explained for the model in each of the three dimensions are 66%, 29.46%, and 4.54%. Also, the relationship of the three dimensions obtained to the original features are shown in Table 17, which shows that, of the four original features, only *P*, F_r and F_a are used and *T* is not used. Where $h(\cdot)$, shown in Eq. (5), is the hinge function. Table 18 shows the prediction results for each and all of the dimensions of the discriminant space.

$$h(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{5}$$

4. Conclusions

This work validates a two-step method to apply FEM and data mining processes based on data analysis and machine learning techniques. The goal was to create a model that could automatically classify a combination of loads that define the presetting conditions of a double-row TRB in relation to the steadiness of its working conditions. For this purpose, a three-dimensional FE model of a double-row TRB was generated successfully to simulate the TRB. The results of simulations and experiments were in agreement. Based on this model and following the proposed DoE, a dataset with the combination of the loads and the class that defines the working conditions was performed. A subsequent statistical analysis of the data was conducted. It included an exploratory analysis, as well as variance and correlation analyses in order to eliminate variables of no significance. Also, several classification models were built. They were optimized by feature reduction, as well as parameter tuning techniques. To determine the performance of the method, it was necessary to test and validate the optimized models with new instances. The results showed that some methods generated an overfitting in defining the space of possibilities, but new instances that was not used when generating the models reduced its performance. Finally, it is seen that the most accurate model was within the group of nonlinear classification methods. More specifically, it was based on flexible discriminant analysis, and produced very accurate results, a 93.75%, whereas the initial four classes were reduced to only two. The results concluded that combining FEM and machine learning techniques can be used successfully in determining the device's working conditions.

Acknowledgments

The authors wish to express their appreciation for the financial support of the Basque Government through KK-2016/00029-ABADE and KK-2016/00032-MESALIQ2 projects.

References

[1] Harris TA, Kotzalas MN. Essential concepts of bearing technology: Rolling bearing analysis. CRC Press; 2006.
[2] Nagatomo T, Takahashi K, Okamura Y, Kigawa T, Noguchi S. Effects of load distribution on life of radial roller bearings. J Tribol 2012;134(2):021101.
[3] Feng Q, Prinja NK. NAFEMS benchmark tests for finite element modelling of contact, gapping and sliding. NAFEMS Report R0081; 2001. https://www.nafems.org.
[4] Zhang X P, Ahmed H, Yao Z. Multi-body contact modeling and statistical experimental validation for hub-bearing unit. Tribol Int 2003;36(7):505–10.
[5] Lostado R, Fernandez Martinez R, Mac Donald BJ. Determination of the contact stresses in double-row tapered roller bearings using the finite element method, experimental analysis and analytical models. J Mech Sci Technol 2015;29(11):4645–56.
[6] Lostado R, Martinez De Pison FJ, Pernia A, Alba F, Blanco J. Combining regression trees and the finite element method to define stress models of highly non-linear mechanical systems. J Strain Anal Eng Des 2009;44(6):491–502.
[7] Lostado R, Escribano Garcia R, Fernandez Martinez R, Illera Cueva M, Mac Donald BJ. Using the finite element method and data mining techniques as an alternative method to determine the maximum load capacity in tapered roller bearings. J Appl Logic 2016;24(A):4–14.
[8] Sugumaran V, Sabareesh GR, Ramachandran KI. Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine. Expert Syst Appl 2008;34(4):3090–8.
[9] Gryllias KC, Antoniadis IA. A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments. Eng Appl Artif Intell 2012;25(2):326–44.
[10] Kankar PK, Sharma SC, Harsha SP. Fault diagnosis of ball bearings using machine learning methods. Expert Syst Appl 2011;38(3):1876–86.
[11] Jiang X, Wu L, Ge M. A novel faults diagnosis method for rolling element bearings based on EWT and ambiguity correlation classifiers. Entropy 2017;19(5):231.
[12] Bishop C. Pattern recognition and machine learning. Springer-Verlag; 2006.
[13] Lostado R, Fernandez Martinez R, Mac Donald BJ, Villanueva PM. Combining soft computing techniques and finite element method for the design and optimization of

complex welded products. Integr Comput-Aided Eng 2015;22(2):153–70.
[14] Fernandez Ceniceros J, Fernandez Martinez R, Fraile Garcia E, Martinez De Pison FJ. Decision support model for one-way floor slab design: a sustainable approach. Autom Constr 2013;35:460–70.
[15] Fernandez Martinez R, Lostado R, Fernandez Ceniceros J, Martinez De Pison Ascacibar FJ. Comparative analysis of learning and meta-learning algorithms for creating models for predicting the probable alcohol level during the ripening of grape berries. Comput Electron Agric 2012;80:54–62.
[16] Eschmann P, Hasbargen L, Weigand K. Ball and roller bearings: Theory, design and application. Munich: R. Oldenbourg; 1985.
[17] Demirhan N, Kanber B. Stress and displacement distributions on cylindrical roller bearing rings using FEM. Mech Based Des Struct Mach 2008;36(1):86–102.
[18] Zavarise G, De Lorenzis L. A modified node-to-segment algorithm passing the contact patch test. Int J Numer Methods Eng 2009;79(4):379–416.
[19] El-Abbasi N, Bathe KJ. Stability and patch test performance of contact discretizations and a new solution algorithm. Comput Struct 2001;79(16):1473–86.
[20] Lostado R, Escribano Garcia R, Fernandez Martinez R. Optimization of operating conditions for a double-row tapered roller bearing. Int J Mech Mater Des 2016;12(3):353–73.
[21] Lundberg G, Sjövall H. Stress and deformation in elastic contacts. Gothenburg, Sweden: Institute of Theory of Elasticity and Strength of Materials, Chalmers University of Technology; 1958. Pub. 4.
[22] Hertz H. On the contact of rigid elastic solids and on hardness. London: MacMillan; 1896. p. 163–83. Miscellaneous Papers.
[23] Kania L. Modelling of rollers in calculation of slewing bearing with the use of finite elements. Mech Mach Theory 2006;41(11):1359–76.
[24] Zienkiewicz OC, Taylor RL. The finite element method: Solid mechanics. UK Butterworth-Heinemann; 2000.
[25] Timken Tapered Roller Bearing Catalog. https://www.timken.com; 2018.
[26] Chambers JM, Hastie TJ. Statistical models in S. Wadsworth & Brooks/Cole; 1992.
[27] Bartlett MS. Properties of sufficiency and statistical tests. Proceedings of the 1937 royal statistical society series A. 160. 1937. p. 268–82.
[28] Brown MB, Forsythe AB. Robust tests for equality of variances. J Am Stat Assoc 1974;69(346):364–7.
[29] Fligner MA, Killeen TJ. Distribution-free two-sample tests for scale. J Am Stat Assoc 1976;71(353):210–3.
[30] Fernandez Martinez R, Martinez De Pison Ascacibar FJ, Pernia Espinoza AV, Lostado Lorza R. Predictive modeling in grape berry weight during maturation process: comparison of data mining, statistical and artificial intelligence techniques. Span J Agric Res 2011;9(4):1156–67.
[31] Abdi H, Williams LJ. Principal component analysis. Wiley Interdiscip Rev Comput Stat 2010;2(4):433–59.
[32] Jolliffe IT. Choosing a subset of principal components or variables. Principal component analysis. Springer; 2002.
[33] Fraley C, Raftery AE, Scrucca L. Normal mixture modeling for model-based clustering, classification, and density estimation 23. Department of Statistics, University of Washington; 2012.
[34] Fraley C, Raftery AE. Model-based clustering, discriminant analysis and density estimation. J Am Stat Assoc 2002;97(458):611–31.
[35] McLachlan GJ. Discriminant analysis and statistical pattern recognition. Wiley Interscience; 2014.
[36] Huberty CJ. Applied discriminant analysis. John Wiley & Sons; 1994.
[37] Liland KH, Indahl UG. Powered partial least squares discriminant analysis. J Chemom 2009;23(1):7–18.
[38] Indahl U. A twist to partial least squares regression. J Chemom 2005;19(1):32–44.
[39] Tibshirani R, Hastie T, Narasimhan B, Chu G. Class prediction by nearest shrunken centroids, with applications to DNA microarrays. Stat Sci 2003;18(1):104–17.
[40] Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci 2002;99(10):6567–72.
[41] Hastie T, Tibshirani R. Discriminant analysis by gaussian mixtures. J R Stat Soc Ser B 1996;58:155–76.
[42] Clemmensen L, Hastie T, Witten D, Ersboll B. Sparse discriminant analysis. Technometrics 2011;53(4):406–13.
[43] Hastie T, Tibshirani R, Friedman J. Support vector machines and flexible discriminants. Elements of statistical learning – Data mining, inference and prediction. Springer; 2009.
[44] Hastie T, Tibshirani R, Buja A. Flexible discriminant analysis by optimal scoring. J Am Stat Assoc 1994;89(428):1255–70.
[45] Ripley BD. Pattern recognition and neural networks. Cambridge: Cambridge University Press; 1996.
[46] Karatzoglou A, Meyer D, Hornik K. Support vector machines in R. J Stat Softw 2006;15(9):1–28.
[47] Platt J. Fast training of support vector machines using sequential minimal optimization. Advances in kernel methods – Support vector learning. MIT Press; 1998.
[48] Vapnik V. Statistical learning theory. John Wiley & Sons; 1998.
[49] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. CRC Press; 1984.
[50] Kuhn M, Johnson K. Applied predictive modeling. Springer; 2013.
[51] Quinlan R. C4.5: Programs for machine learning. San Francisco, USA: Morgan Kaufmann Publishers; 1993.
[52] Breiman L. Bagging predictors. Mach Learn 1996;24(2):123–40.
[53] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
[54] Grubinger T, Zeileis A, Pfeiffer KP. evTree: Evolutionary learning of globally optimal classification and regression trees in R. Research platform empirical and experimental economics. Universitt Innsbruck; 2011.
[55] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017https://www.R-project.org/.
[56] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B (Stat Methodol) 2005;67(2):301–20.