DINAME 2023 - Proceedings of the XIX International Symposium on Dynamic Problems of Mechanics,
A.S. De Paula, A.T. Fabro and M.R. Machado (Editors), ABCM.
Pirenópolis, Go, Brazil, February 26th to March 3rd, 2023

Evaluation of a methodology for low-speed bearing fault diagnosis

Thiago B. Costa [1], João L. L. Soares [1], Walter S. Sousa [1], Alexandre L. A. Mesquita [1], André L. A. Mesquita [1], Elton P. de Souza [1] and Danilo S. Braga [2]

[1] Federal University of Pará, Av. Brasília, s/n - Vila Permanente - 68.455-901 - Tucuruí, PA.

[2] Dynamox, Rod. SC 401, km 01, nº 600, Rua Parque Tecnológico Alfa, Florianópolis - SC, 88030-909.

Abstract: Rolling element bearings are critical components because they often represent high incidence of rotating machinery failures. In the last decades, some researchers have proposed methods of monitoring, detection, diagnosis and prognosis based on artificial intelligence tools, seeking for more reliable decision-making in predictive maintenance. In this sense, application of machine learning has been investigated for fault diagnosis of low-speed bearings due to the monitoring of their condition be more challenger than high-speed bearings, and traditional methods become limited. Hence, this work proposes evaluate the MLE (Maximal Lyapunov Exponent) and the Hjorth's parameters as predictors extracted from the vibration signals measured in a test rig at 60 rpm with different frequency range and sampling rate. First, those features and others time domain features were ranked according to their significance estimated by Welch's t-test statistic value, i.e., the capability to maximize the class separation. Secondly, the best predictors were selected to create a table of superior features, and thus another matrix composed by worse features. Therefore, Principal Component Analysis (PCA) was implemented in order to reduce the dimensionality of data. Finally, the datasets (superior and inferior features tables) were used to train a Support Vector Machine (SVM) algorithm, whose results were excellent (as expected) for superior predictors MLE and Hjorth parameters.

Keywords: Vibration analysis, Low-speed bearing, Fault diagnosis, MLE, Hjorth's parameters, Welch's t-test, SVM.

INTRODUCTION

Approximately 42% of rotating machinery failures is associated to fault bearings. For this reason, monitoring bearing condition with attempt to predict failures leads to reduction of unexpected failure and allows scheduled maintenance, seeking for ensure safety and cost reduction (Hemmer et al, 2020). Low-speed slew bearings are commonly applied in excavators, cranes, loading and unloading machines, transporting machines, material processing machines, military products and medical research equipment (Caesarendra et al, 2013). Jin et al (2021) reported the principal failure modes of slew bearing are plastic deformation, cracks, fractures, wear, improper mounting and lubricant failure, which can affect different parts as inner race, outer race, rolling element, connect bolts or sealing ring (Liu and Zhang, 2020).

Hemmer (2020) reported that bearing faults may be detected and diagnosed by vibration analysis, acoustic emission, temperature, ultrasound and lubricant. In case of vibration monitoring, Caesarendra and Tjahjowidodo (2017) claimed traditional methods tend to be limited to low-speed bearing due to low energy and the non-linear and chaotic characteristic of the signal. So, how to monitor low speed bearings with efficiency using vibration analysis? Hence, to overcome this issue, applications of machine learning techniques have been proposed as seen in Saufi et al (2019) and Sandoval (2021). In this sense, this work proposes the assessment of quality of Hjorth's parameters (Activity, Mobility and Complexity) combined with Maximal Lyapunov Exponent as features to diagnose the condition of low-speed bearing based on statistic tools. Moreover, these features were compared with others time domain features of vibration signal seeking for those could potentially improve the performance of SVM classification learner model.

Hence, the main objective was the selection of better features to be used in a classification learner algorithm for the diagnosis of low-speed bearing condition, which consists in two states: normal and failure. For this purpose, measurements of vibration data were acquired from experimental test bench. Then, from the vibration signals were extracted the following characteristics: Root Mean Square, Peak Value, Shape Factor, Kurtosis, Skewness, Impulse Factor, Crest Factor, Clearance Factor, Activity, Mobility, Complexity and Maximal Lyapunov Exponent. After that, the Welch's t-test statistic and Principal Component Analysis were successively used to evaluate the features as their capability to separate the two classes and their degree of importance to explain the data's variance, respectively. At the end, the better and the worse features were employed in SVM algorithm and, according to the results, the better features presented a superior outcome.

METHODOLOGY

The method utilizes experimental vibration data of low-speed bearing, and it worth noting this type of signals are more complex to analyze than that high-speed vibration data, primarily due to the low energy of vibration (Caesarendra and Tjahjowidodo, 2017). Moreover, another common problem is this type of vibration signals be more sensitive to noise,

and often presents low value of SNR (signal-to-noise ratio). However, the test rig used in this work is free from sources of noise as gearbox reducers, severe unbalance and misalignment, generators, pumps, brake systems, etc.

Feature extraction

Time domain features

A temporal series consists of a sequence of discrete numerical values in time domain, and a digital vibration signal is a pertinent instance. Considering a discrete time series *s(i), i = 1,2,3, …, N*, where *N* is the length of *s(i)*, some common time domain features of vibration signal are calculated using the following equations. Where s_{RMS} is Root Mean Square, σ_s is Standard Deviation, s_{SF} is Shape Factor, s_{Kurt} is Kurtosis, s_{Skew} is Skewness, s_{IF} is Impulse Factor, s_{CF} is Crest Factor, s_{ClF} is Clearance Factor, A_0 is Activity, M_0 is Mobility and C_0 is Complexity (Caesarendra, 2015, Mathworks, 2022, Päivinen et al, 2005).

$$s_{Peak}=\max|s_i| \quad (1) \qquad s_{RMS}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}|s_i|^2} \quad (2) \qquad s_{SF}=\frac{s_{RMS}}{\frac{1}{N}\sum_{i=1}^{N}|s_i|} \quad (3)$$

$$s_{Kurt}=\frac{\frac{1}{N}\sum_{i=1}^{N}(s_i-\bar{s})^4}{\left[\frac{1}{N}\sum_{i=1}^{N}(s_i-\bar{s})^2\right]^2} \quad (4) \qquad s_{Skew}=\frac{\frac{1}{N}\sum_{i=1}^{N}(s_i-\bar{s})^3}{\left[\frac{1}{N}\sum_{i=1}^{N}(s_i-\bar{s})^2\right]^{3/2}} \quad (5) \qquad s_{IF}=\frac{s_{Peak}}{\frac{1}{N}\sum_{i=1}^{N}|s_i|} \quad (6)$$

$$s_{CF}=\frac{s_{Peak}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}s_i^2}} \quad (7) \qquad s_{ClF}=\frac{s_{Peak}}{\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|s_i|}\right)^2} \quad (8) \qquad A_0=\sigma_s^2 \quad (9)$$

$$M_0=\frac{\sigma_{s'}}{\sigma_s} \quad (10) \qquad C_0=\frac{\sigma_{s''}/\sigma_{s'}}{\sigma_{s'}/\sigma_s} \quad (11)$$

Where Peak Value (s_{Peak}) corresponds to maximum absolute value of the signal. The Root Mean Square (s_{RMS}) is the root mean of squared absolute value of each item of the signal, it is used in this case because the simple mean returns an inaccurate value around zero. Shape Factor (s_{SF}) is the ratio between s_{RMS} and mean of absolute values of $\mathbf{s}_i$, and refers to a value that is affected by an object's shape but is independent of its dimensions. Kurtosis (s_{Kurt}) measures the degree of flatness of the probability density function near its center. Skewness (s_{Skew}) measures the asymmetry of probability density function of the vibration signal. Impulse Factor (s_{IF}) compares the height of a peak to the mean level of the signal. Crest Factor (s_{CF}) is a measure of how much impact and it is appropriate for "spiky signals". Clearance Factor (s_{ClF}) is s_{Peak} divided by the squared mean value of the square roots of the absolute amplitudes, whose magnitude is maximal for healthy bearings and reduces progressively with failure development. (Caesarendra, 2015, Mathworks, 2022).

Equations 9, 10 and 11 represents the Hjorth's parameters. The first one, Activity (A_0), means the variance, so it quantifies the dispersion of a signal or a data set around their mean reference. The second one, Mobility (M_0), is defined as the square root of the ratio of the activity of the first derivative of the signal to the activity of the original signal. The first derivative of the signal is given by $s'(j)=s(j+1)-s(j)$, $j=1, \cdots, N-1$. The third parameter, Complexity (C_0), is defined as the ratio of mobility of the first derivative of the signal to the mobility of the signal itself (Päivinen et al, 2005).

Nonlinear Feature (MLE – Maximal Lyapunov Exponent)

Caesarendra (2015) developed the algorithm applied in this work, and it was roughly based on Rosenstein's method proposed in 1993. The first step involves reconstructing the attractor dynamics from a single *N*-points time series, $\{x_1, x_2, \ldots, x_N\}$, using the method of delays. Then, the reconstructed trajectory $\mathbf{X}$ can be expressed as a matrix where each row is a phase-space vector, i.e., $\mathbf{X} = (\boldsymbol{X}_1\ \boldsymbol{X}_2 \ldots \boldsymbol{X}_M)^T$. Where $\boldsymbol{X}_i$ represents the state of the system at discrete time *i*, where each $\boldsymbol{X}_i$ is given by $\boldsymbol{X}_i = (x_i\ x_{i+J} \ldots x_{i+(m-1)J})$. In this case, *J* is the lag or reconstruction delay, and *m* is the embedding dimension. Thus, $\mathbf{X}$ is a matrix with $M \times m$ dimension, and the constants *N*, *m*, *M* and *J* are related in $M=N-(m-1)J$. For determination of parameter *J*, Caesarendra (2015) utilizes the Fast Fourier Transform to find the dominant frequency of the vibration signal, then calculate the inverse of dominant frequency and finally multiply by *N*. For estimation of *m* Rosenstein suggests values below Takens' theorem, i.e., $m < 2N$.

After reconstructing the phase-space matrix, the algorithm locates the nearest neighbor of each point on the trajectory. According to Rosenstein, Collins and De Luca (1993), the nearest neighbor, $\mathbf{X}_{\hat{j}}$, is found by searching for the point that minimizes the distance to the particular reference point, $\mathbf{X}_j$. Thus, the initial Euclidian distance between those two points can be expressed by $d_j(0) = \min_{\boldsymbol{X}_{\hat{j}}} \|\boldsymbol{X}_j - \boldsymbol{X}_{\hat{j}}\|$, subject to the constraint that nearest neighbors have a temporal separation greater than the mean period of the time series. For Mehdizadeh (2019), the constraint considers distances greater than mean period $\left(\mu = \lceil f_s/\bar{f}_p \rceil\right)$ avoid to pick points of same trajectory. Analytically, that means $|j-\hat{j}| > \left\lceil f_s/\bar{f}_p \right\rceil$, where f_s is sampling rate, $\bar{f}_p$ is the mean normalized frequency of the power spectrum (can also be the median frequency of the magnitude spectrum), and $\lceil . \rceil$ represents a ceil function. Then, the maximal Lyapunov exponent is estimated as the mean rate of separation of the nearest neighbors (Mathworks, 2022).

Caesarendra's algorithm utilized the method of *k* iteration proposed by Sato et al (1987) to improve the accuracy of the measured distance. Finally, the iteration is limited to $|M-k| \geq j$, and the maximal Lyapunov exponent (MLE) is accurately computed using a least-square fit equation described by Eq. 12 based on average separation. Where *k* assumes successively all values between the Expansion Range from k_{min} to k_{max}. The Fig. 1 illustrate a flowchart that presents a sequential step to be followed in order to construct an algorithm to compute MLE.

$$\lambda(i) = \frac{f_s}{(M-k)} \sum_{j=1}^{M-k} \ln d_j(i+k) \tag{12}$$

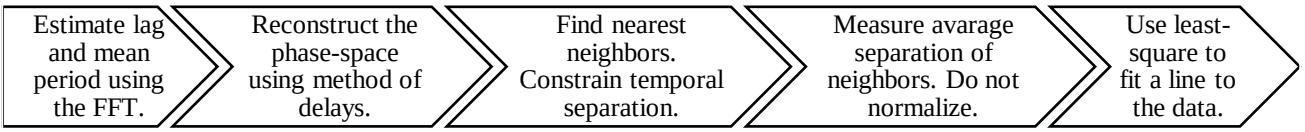


Figure 1 – Flowchart for computation of MLE algorithm [Adapted from Rosenstein, Collins and De Luca (1993)].

Welch's *t*-test statistic

According to Zimmerman and Zumbo (1993), in order to compare the means of two independent groups, the Welch's *t*-test is considered better than Student *t*-test when the populations don't have the same variance and size. Additionally, it is worth noting that when Welch's *t*-test is applied to groups with same variance and size, the result is equal to Student *t*-test. For this work, the groups were assumed to have unequal variances and follow a normal distribution.

Thus, the statistical significance value for each calculated feature in order to evaluate their capability to maximize class separation is estimated by Eq. 13 of Welch's *t*-test statistic as following (Shri and Sriraam, 2017):

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{\sigma_i^2}{N_i} + \frac{\sigma_j^2}{N_j}}} \tag{13}$$

where *i* and *j* correspond both groups evaluated, $\bar{x}$ corresponds to sample mean of each group, σ^2 are sample variances and *N* is the length of each group.

PCA (Principal Component Analysis)

Principal Component Analysis (PCA) is perhaps the most popular statistical analysis technique. It consists of a non-parametric statistical technique that analyzes a representative data table of samples defined by inter-correlated variables. The technique aims to extract the most relevant information from the data by representing them in a new set of orthogonal variables called principal components. So, it is a technique that can be used to reduce the dimensionality of a dataset (Abdi and Williams, 2010; Cadavid; Lawrence; Ruzmaikin, 2007).

According to Bro and Smilde (2014), the Principal Component Analysis can be geometrically understood as a rotation of the initial system of coordinate, in such a way the new system become capable to explain better the variance of the data with fewer number of new coordinates. Then, the principal component is a new axis that must have the larger representativity of the dataset's variance, and the remaining information is represented by others perpendicular axes. The projection of each point of the data on principal axes is called *scores*, and the projection of each variable coordinate on the same principal components is *loadings*. Hence, *scores* represent the information of the data on the principal subspace, and *loadings* consists of how much an original variable is correlated to the principal axes.

Wold, Esbensen and Geladi (1987) shows that the algebra behind the calculations of *loadings* is based on eigenvectors and eigenvalues. First, it is necessary to compute the covariance matrix of a dataset, $\mathbf{X}$, multiplying the transpose of the centered data, $\bar{\mathbf{X}}$, for itself, i.e., $\boldsymbol{C}_X = \bar{\boldsymbol{X}}^T\bar{\boldsymbol{X}}$. Then, the vector of loadings is the eigenvector, $\mathbf{v}$, of the covariance matrix $\mathbf{C}_X$, such that:

$$\mathbf{C}_X\mathbf{v} = \lambda\mathbf{v} \tag{14}$$

where λ is the eigenvalue of $\mathbf{C}_X$ associated to $\mathbf{v}$. Thus, there will be an eigenvalue of covariance matrix for each component adopted, and the greater the magnitude of eigenvalue, it means that more information the respective component represents. So, the principal component has the largest λ value associated to the vector $\mathbf{v}$, the next component has the second greater eigenvalue, and so on.

Finally, each element of the eigenvector, $\mathbf{v}$, represents the loading of respective original variable or predictor in relation to certain component. In order to organize the eigenvectors in decrescent order of importance, it is necessary to obtain

every eigenvalue, and then find the majors. A simple way to calculate the eigenvalues is $|\mathbf{C}_X - \lambda\mathbf{I}| = 0$, where $\mathbf{I}$ is the identity matrix, and the percentual of explained variance for each λ is: explained (%) = $100\lambda/\mathrm{tr}(\mathbf{C}_X)$.

Bro and Smilde (2014) also presents a way to represent the lost information in terms of an error matrix. Hence, considering that the matrix of *scores*, $\mathbf{t}$, is given by the product between $\mathbf{X}$ and the matrix of eigenvector in descendance order of significance $\mathbf{V}$, i.e., $\mathbf{t} = \mathbf{XV}$, the non-captured information is given by $\mathbf{E} = \mathbf{X} - \mathbf{tv}^{\mathbf{T}}$.

SVM (Support Vector Machine)

A Support Vector Machine (SVM) is a supervised learning algorithm used for many classification and regression problems. The objective of the SVM algorithm is to find a hyperplane that, to the best degree possible, separates data points of one class from those of another class. "Best" is defined as the hyperplane with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points. Support vectors refer to a subset of the training observations that identify the location of the separating hyperplane (Mathworks, 2022).

Mathematically, considering a dataset $(\boldsymbol{x}_i,\boldsymbol{y}_i)$, $\boldsymbol{x}_i \in R^d$, $\boldsymbol{y}_i \in \{-1,+1\}$, $i=1,2,\ldots,n$. Then, the classification function for a non-linear separable data point can be summarized as follows (Li et al, 2013):

$$f(x)=\mathrm{sgn}\left\{\sum_{i=1}^{L}\alpha_i \mathbf{y}_i K(\mathbf{x}_i\cdot\mathbf{x}_j)+b\right\} \tag{17}$$

where $K(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)$ is the kernel function, b is the bias, L is the number of support vectors, and α_i is the non-zero Lagrange coefficients that satisfies the condition $\alpha_i\{\boldsymbol{y}_i[(\boldsymbol{w}\cdot\boldsymbol{x}_i)+b]-1+\xi_i\}=0$. Where $\boldsymbol{w}$ is the weight vector, and ξ is the slack factor. Those last both is estimated through optimization $\min_i \frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{n}\xi_i$ (C is the penalty) subjected to $\boldsymbol{y}_i[(\boldsymbol{w}\cdot\boldsymbol{x}_i)+b]\geq 1-\xi_i$.

The slack factor is a parameter that allows some misclassification for non-linear separable data, as the penalty is a regularization parameter used to stablish the degree of tolerance of misclassifications. The Kernel functions map the data to a different, often higher dimensional space with the expectation that the classes are easier to separate after this transformation, potentially simplifying a complex non-linear decision boundary to linear one in the higher dimensional, mapped feature space (Bishop and Nasrabadi, 2006, Mathworks, 2022).

Data acquisition

The test rig is presented in Fig. 2 where can be seen the acquisition system utilized NK820. Upon (vertical) the bearing housing (yellow color) is the accelerometer NK30 43649 with sensitivity of 101 mV. Additionally, a WEG CFW08 frequency inverter is associated to a 3 hp three-phase induction motor AC with 3 poles, but the disc at the extremity of the shaft was no longer used during the tests.

Furthermore, Tab. 1 shows the parameters set to data acquisition, where "D" means "dataset". There are four datasets obtained at shaft speed of 60 rpm. The main difference between them is the frequency range and the sampling rate.

Figure 2 – Test rig and acquisition system.

Table 1 – Parameters adjusted for vibration signals acquisition.

Parameters	D1	D2	D3	D4
Shaft speed (rpm)	60	60	60	60
Frequency range (Hz)	0.1 – 500	0.1 – 1000	0.1 – 2000	0.1 – 5000
Sampling rate (kHz)	1.25	2.5	5.0	12.5
Number of points	2048	4096	8192	16384
Time step (ms)	0.8	0.4	0.2	0.08
Time period (s)	1.6376	1.6380	1.6382	1.3106

Therefore, 250 observations were acquired for each dataset (D1, D2, D3 and D4), 125 belong to a good bearing (normal) and others 125 refer to a bearing with a simulated fault in outer race as soft scratches. Those both bearing tested were GBR 6202z. Figure 3 shows the scalograms relative to normal and fault instances of the first vibration signal measured for each condition.

Scalograms are generally represented as diagrams that gives a visual perspective of a wavelet transform of a signal, having axes for time, frequency, and magnitude (that last one is represented in color graduation). In examples of Fig. 3, a result initially noted is that the maximum energy level is always reached by fault observations. Another conclusion is that signals are apparently stationary, except for a softly unstable case seen for D3 and D4 in fault examples. Furthermore, all examples shows that the most energy levels are about frequencies of 100 Hz, 500 Hz, 1.9 kHz and 4 kHz.

The scalograms of Fig. 3 were plotted for four datasets and the purpose was just to present some acquired vibration signals in time-frequency domain. Hence, each scalogram was obtained through continuous wavelet transform of the signal. And it was based on the analytic Morse wavelet with the symmetry parameter equals to 3, the time-bandwidth product equals to 60, and the voices per octave equals to 10. Finally, the minimum and maximum amplitude scales were determined based on the energy spread of the wavelet in frequency and time (Mathworks, 2022).

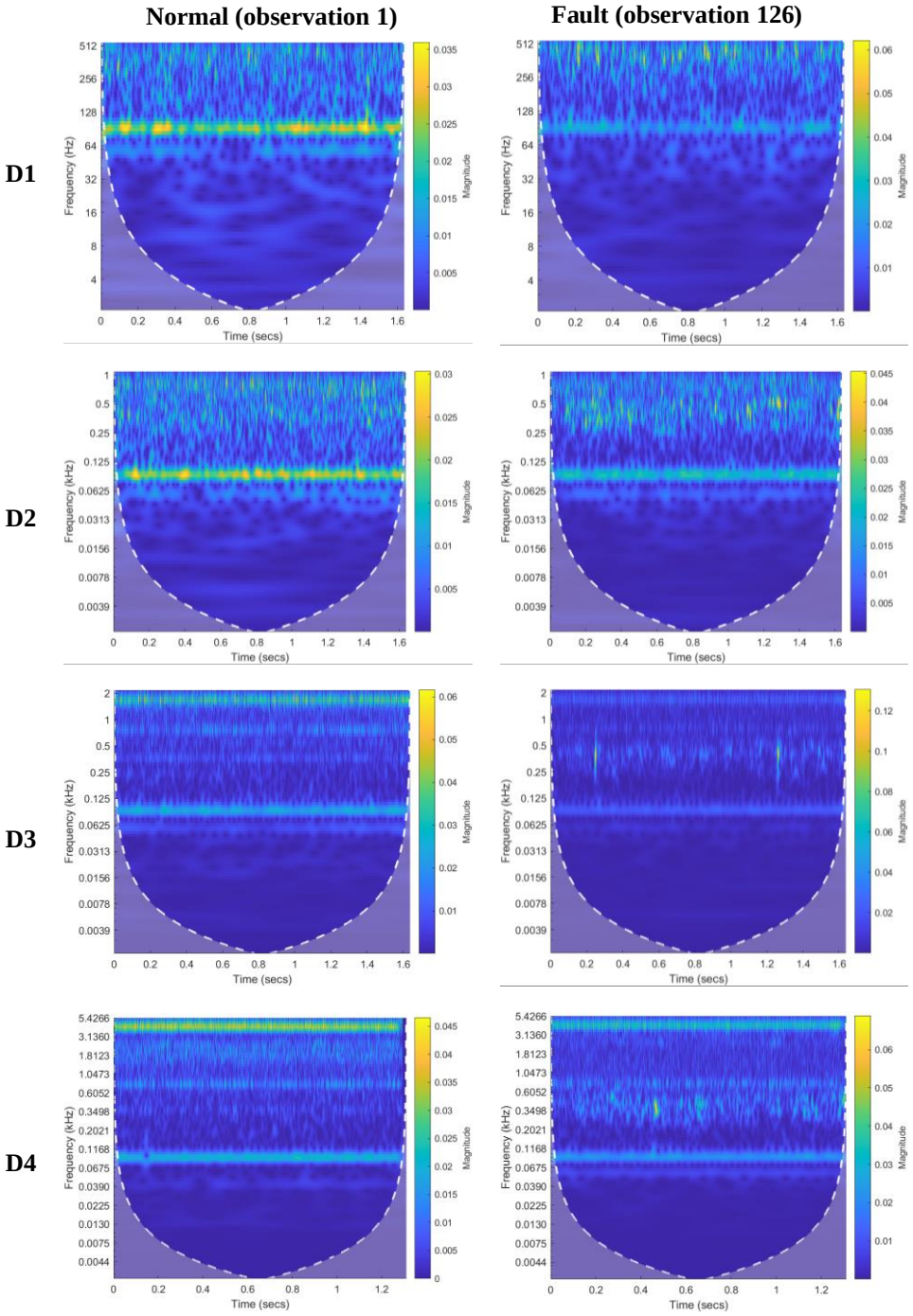


Figure 3 – Scalograms of some vibration signals measured.

RESULT AND DISCUSSION

The computation of the time domain features is simpler when compared with the algorithm necessary to the calculation of MLE. In other words, the implementation of Eq.1 – 11 is relatively easier to do. However, the computation of MLE requires precisely definition of parameters as mean period μ, embedding dimension m, reconstructing delay (lag) J, and the number of iteration k.

Among those parameters, the lag is considered an open problem, even it could be estimated based on the dominant frequency of the signal, this work selected a value of $J=1$. Adopting $J=1$ allowed to find a more satisfactory results, such as Rosenstein, Collins and De Luca (1993). About the others parameters, the embedding dimension of the phase-space was considered as $m = 50$, the mean period is particular to each signal and was calculated by $\mu = \lceil f_s/\bar{f}_p \rceil$, and the Expansion Range was defined between $k_{min}=1$ and $k_{max}=50$ (where k_{max} must be between μ and m).

Figure 4 shows an example of calculation of MLE for a normal observation from D3 using the least-square method to fit a line to the data (blue curve), where the slope of the red dashed line is proportional to the Largest Lyapunov Exponent.

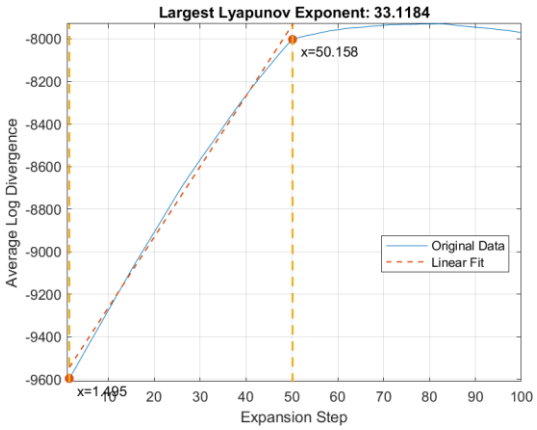


Figure 4 – An example of estimating the Largest Lyapunov Exponent based on least-square method.

The features extracted were ranked as a method for feature selection based on their significance estimated by Welch's *t*-test statistic value. Then, Fig. 5 shows that Hjorth's parameters, Maximal Lyapunov Exponent (MLE) and Root Mean Square (RMS) presented a good capacity to maximize class separation – except MLE in D2.

Additionally, while time domain predictors had the best significance for D1 and D2, MLE presented the highest values of statistic Welch's *t*-test for D3 and D4 (primarily for D4). So, it seems how much larger are number of samples, sampling rate and frequency range of the signal, the greater the distance between classes for MLE calculation.

Moreover, the increase of separability for D3 and D4 only occurs to MLE, as the others time domain features showed relatively low significance. In order to present the features values, this paper brings an Appendix containing histograms, where the two classes were compared in a plot based on probabilistic distribution.

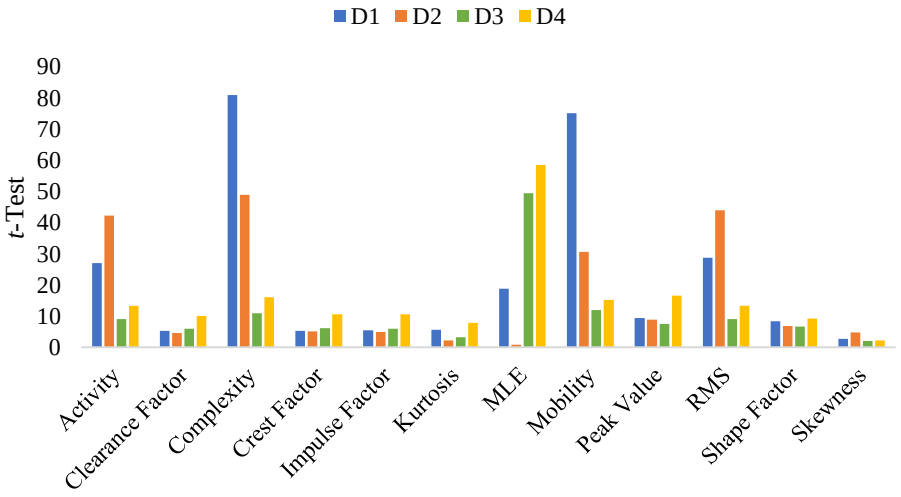


Figure 5 – Ranking of features.

Therefore, the predictors with larger separability formed a superior feature table: MLE, RMS, Activity, Mobility and Complexity. And the predictors with lower separability formed an inferior feature table: Clearance Factor, Crest Factor,

Impulse Factor, Kurtosis, Skewness and Shape Factor. Then, a dimensionality reduction procedure was implemented to both datasets using Principal Component Analysis (PCA).

The objective consisted of represent the data in a plane, so only two components were held. Then, Fig. 6 shows the biplots for superior features, in which the label "-1" corresponds to group of "fault" elements and "+1" represents a "normal" condition. The loadings of predictors correspond to projection of each predictor on PC1 and PC2. This means that MLE have the greater loading on PC2 in D2 and D3, but in D4 it is more significant on PC1.

In every biplots, Activity and RMS seems to have high correlation and the same loadings. Hence, it was an expected result because the Activity is affected by simple mean value, which is about zero, turning Activity equivalent to RMS. Finally, considering the explained information by PC1 and PC2, a valid conclusion is time domain features are more relevant for class separation when vibration signals are acquired at low sampling rate and short frequency range, while MLE is more significant for otherwise.

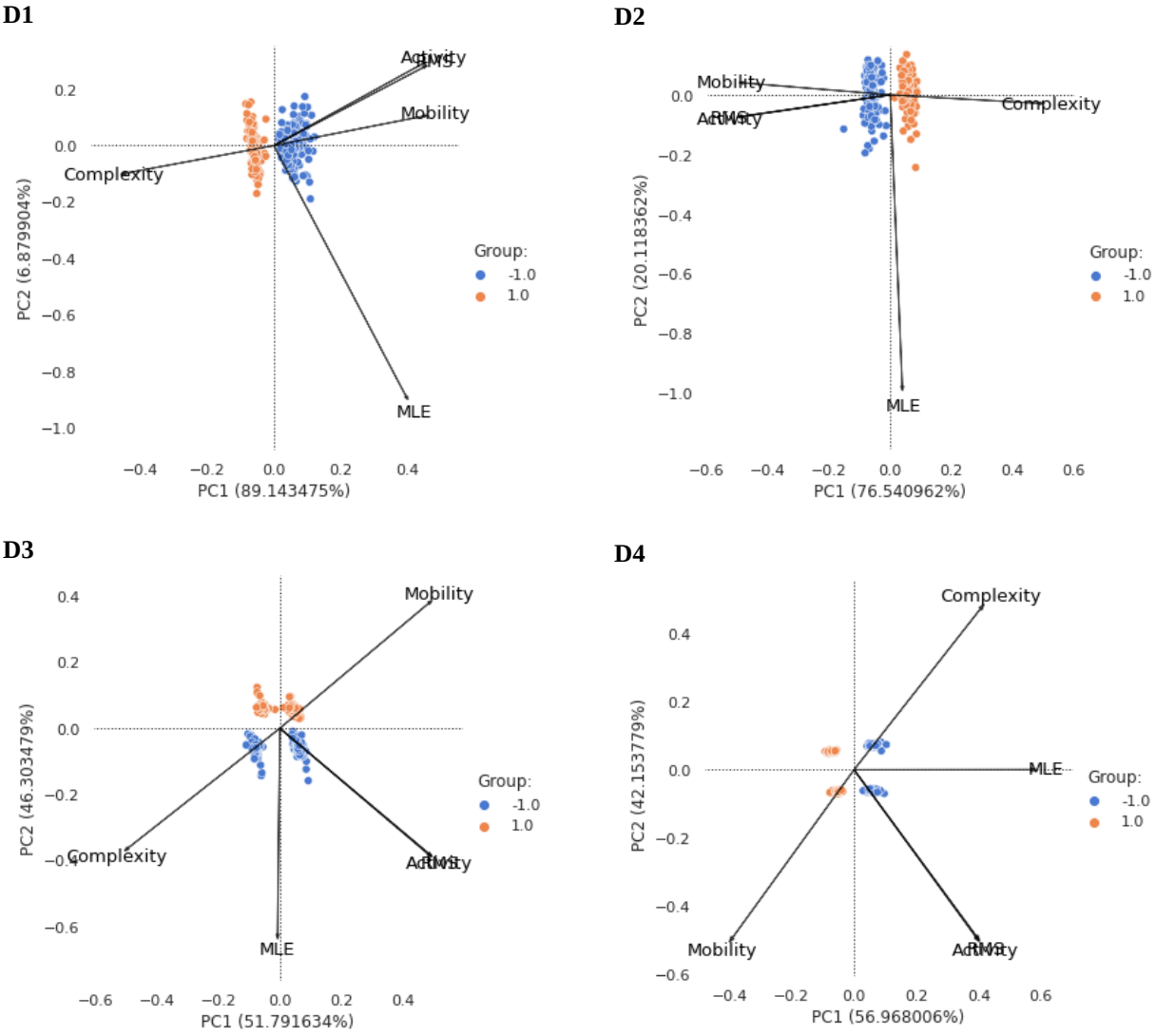


Figure 6 – Biplots.

In order to verify the impact of separability level of the predictors, the low dimensional datasets were applied to Support Vector Machine (SVM) algorithm. The kernel function applied was *linear* and the penalty was *C=1* for all SVM models. Moreover, 20% of dataset was reserved to test the trained algorithm, i.e., 200 individuals were used train the supervised learner process, and the remaining 50 observations were preserved for test stage. However, the 200 observations used as train data were subject to a Cross-Validation with 5 folds.

As expected, the models trained with superior features presents better results than the models trained with inferior features. Table 2 and Table 3 expose the accuracy (%) for Cross-Validation procedure and for test process with new data, respectively. The primarily reason for perfect accuracy performed by model based on superior features is the data to be linearly separable, and this fact is clearly observed in Fig. 7.

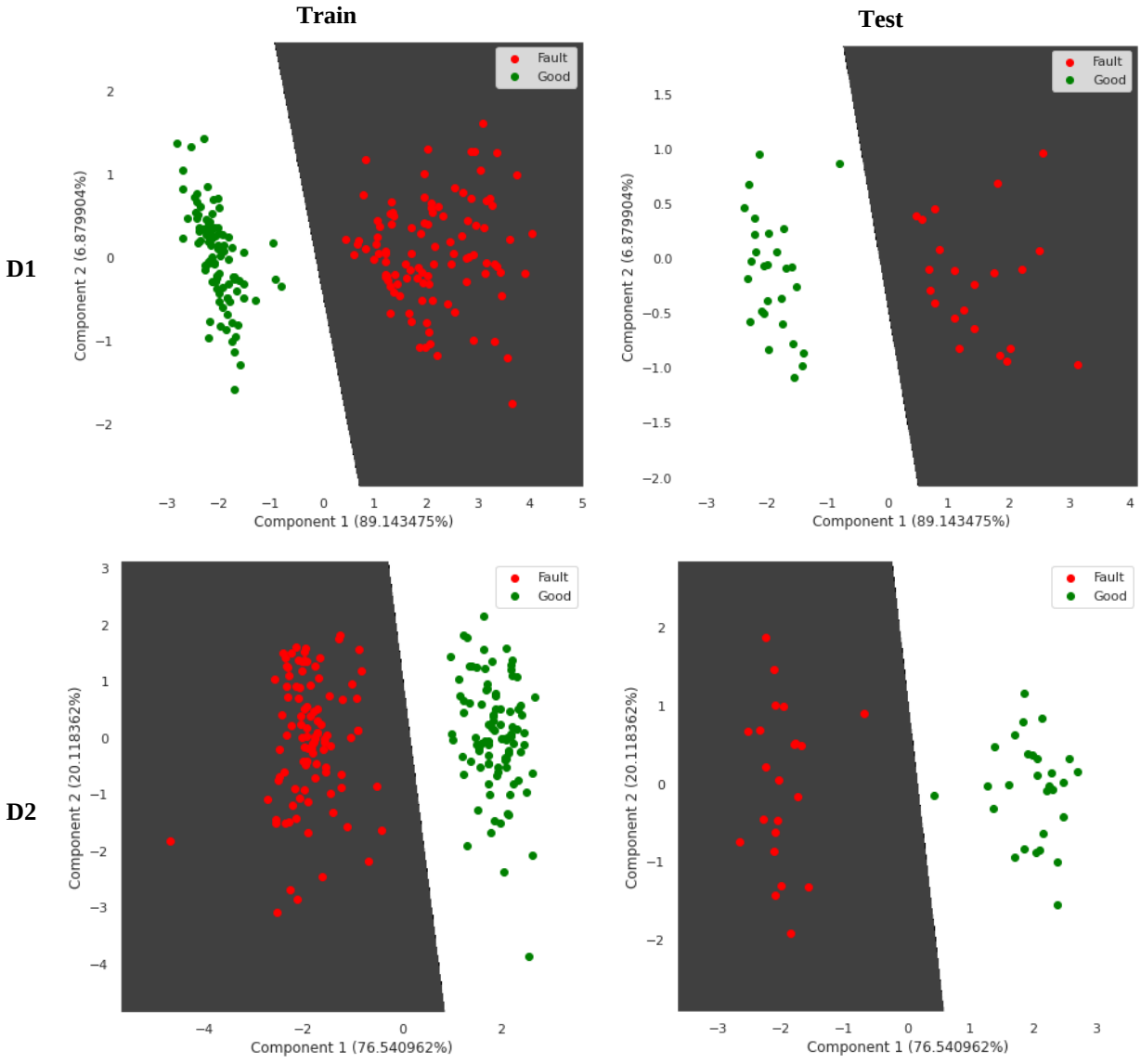
The plots at left side of Fig. 7 correspond to the definition of an optimal hyperplane according to the training data, where the green and red points are good and fault individuals, respectively. The hyperplane split the plot into two regions: “white” for good samples and “black” for fault examples. When the ideal hyperplane was determined, it was used (the plots at right side) to perform predictions using test data. Likewise, Fig. 8 corresponds to the same representation of Fig. 7, but this time related to training and testing models based on inferior features. So, through Fig. 8 is clearly observed there are misclassifications due to overlapping classes.

Table 2 – Accuracy (%) of Linear SVM model by 5-folds Cross-Validation for train dataset.

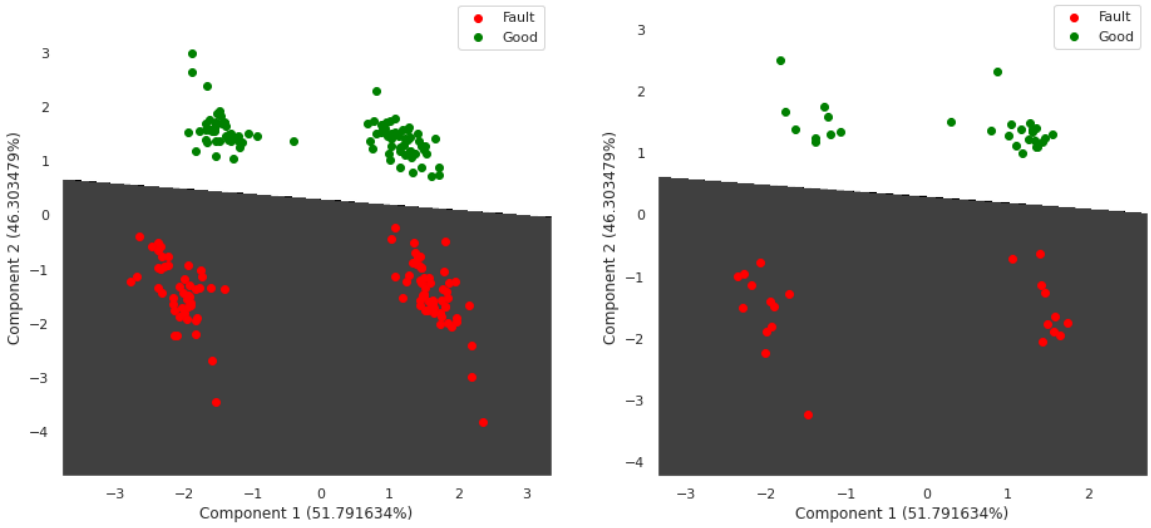
Model based on:	D1	D2	D3	D4
Inferior features	74	83	69,5	92,5
Superior features	100	100	100	100

Table 3 – Accuracy (%) of trained Linear SVM model performing predictions for new data.

Model based on:	D1	D2	D3	D4
Inferior features	72	86	74	98
Superior features	100	100	100	100



D3



D4

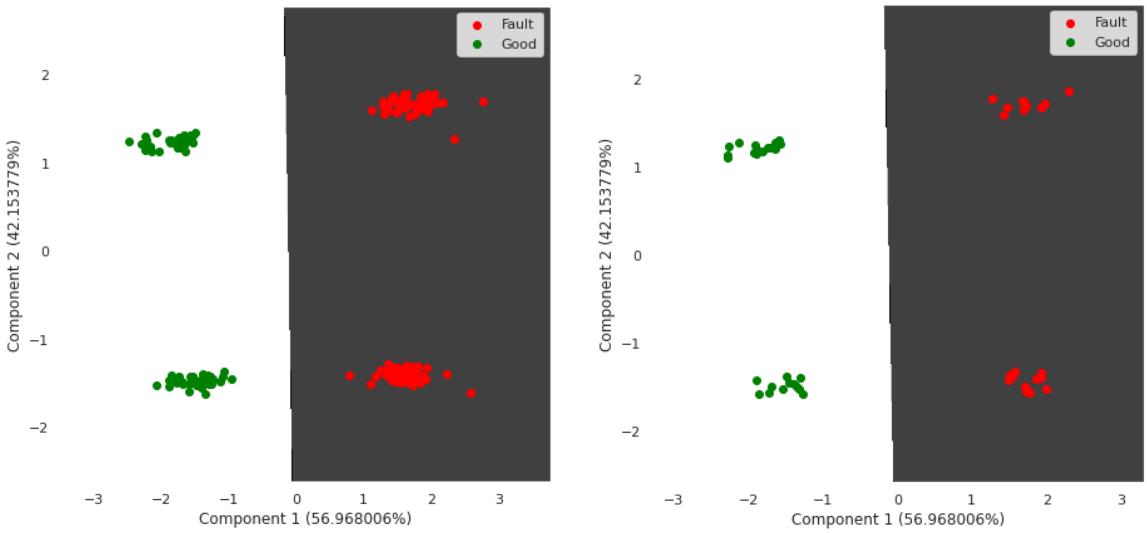
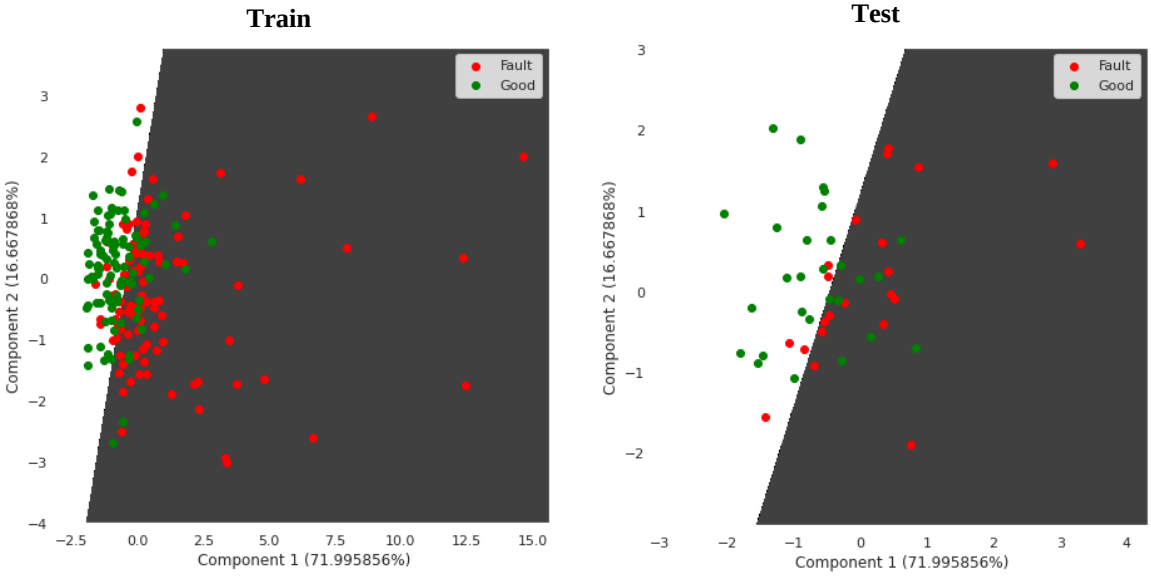


Figure 7 – Representation of hyperplanes defined by SVM models based on superior features.

D1



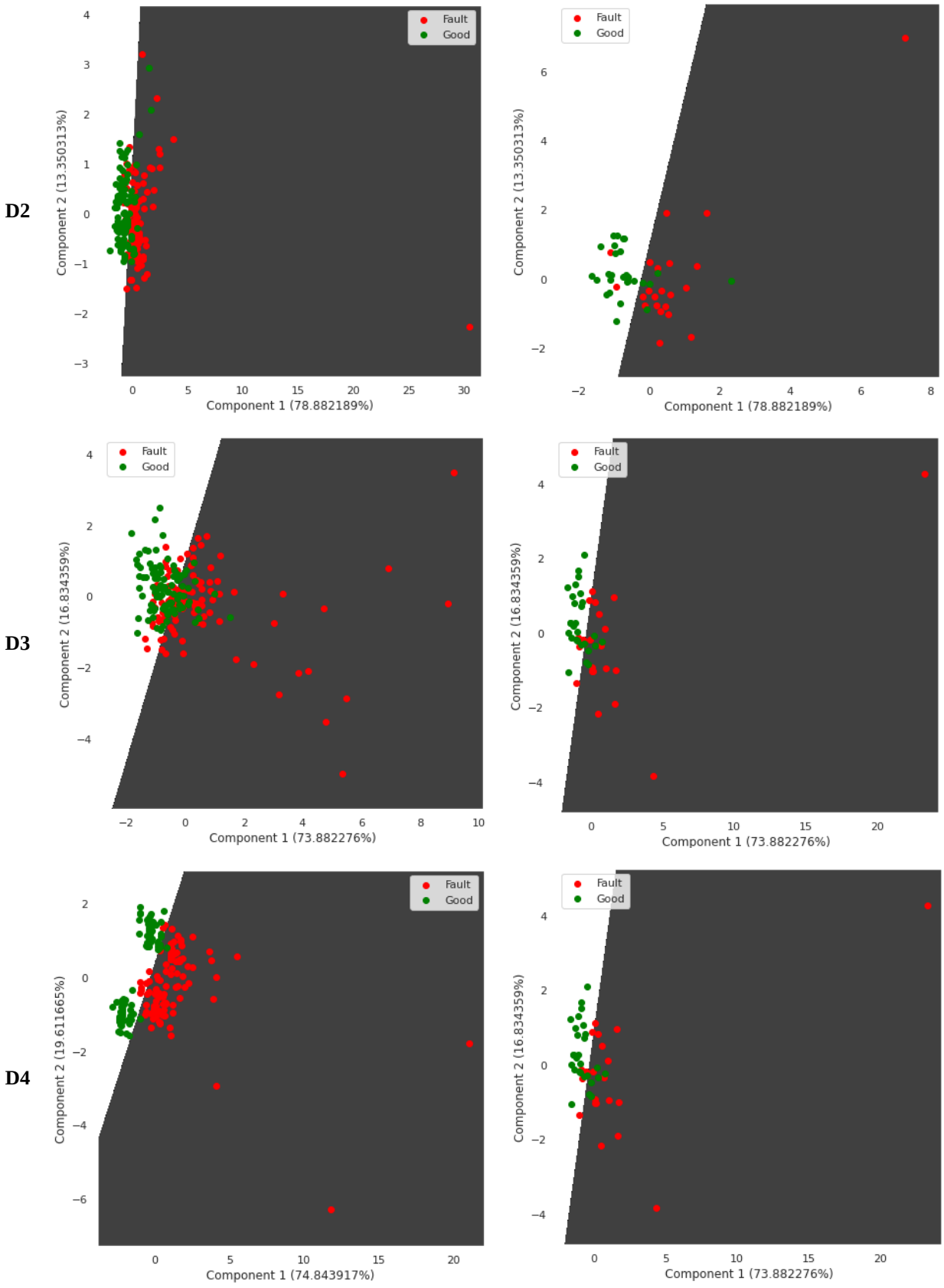


Figure 8 – Representation of hyperplanes defined by SVM models based on inferior features.

CONCLUSION

The results showed the association of MLE, RMS and Hjorth's parameters, were good features to highlight the accuracy of SVM classifier in determination of condition of low-speed rolling element bearing at 60 rpm. The differentiation between normal and outer race fault conditions were accurately enhanced, even varying the frequency

range, the number of points (samples) and sampling rate during data acquisition of the vibration signals. It is worth mentioning that even when MLE presented extremely low significance in D2, the model accuracy was maximal, probably to the high *t*-test statistic value of the others predictors.

On the other hand, the methods used to feature selection showed a coherent result, when Welch's *t*-test indicates the more separable features, which posteriorly turn the low dimensional dataset linearly separable to SVM. Moreover, changes in sampling rate and frequency range during data acquisition induced substantial differences in features values, affecting the separability of classes. Finally, the lack of others sources of noise in the test rig allows to measure a quasi-noiseless vibration signal. Therefore, the maximal accuracy of the SVM results may be possibly favored by a high SNR value of the signal, what perhaps would not occur in a noisier scenario.

ACKNOWLEDGMENTS

The authors thank Dynamox SA for all the support.

REFERENCES

Abdi, H., Williams, and L. J., 2010, "Principal component analysis", Wiley interdisciplinary reviews: computational statistics, Vol. 2, No. 4, pp. 433-459.

Bishop, C. M. and Nasrabadi, N. M., 2006, "Pattern recognition and machine learning", New York: springer.

Bro, R., and Smilde, A. K., 2014, "Principal component analysis", Analytical methods, Vol. 6, No. 9, pp. 2812-2831.

Caesarendra, W. and Tjahjowidodo, T., 2017, "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. Machines", Vol. 5, No. 4, 21 p, https://doi.org/10.3390/machines5040021.

Caesarendra, W. et al, 2013, "Condition monitoring of naturally damaged slow speed slewing bearing based on ensemble empirical mode decomposition", Journal of Mechanical Science and Technology, Vol. 27, No. 8, pp. 2253-2262.

Caesarendra, W., 2015, "Vibration and acoustic emission-based condition monitoring and prognostic methods for very low speed slew bearing", Doctor of Philosophy thesis, School of Mechanical, Materials and Mechatronic Engineering, University of Wollongong, https://ro.uow.edu.au/theses/4479.

Hemmer, M. et al, 2020, "Health Indicator for Low-speed Axial Bearings using Variational Autoencoders", IEEE Access, Vol. 8, pp. 35842-35852, https://doi.org/10.1109/ACCESS.2020.2974942.

Hemmer, M., 2020, "Condition Monitoring Methods for Large, Low-speed Bearings", Doctoral thesis, University of Agder, Oslo, https://hdl.handle.net/11250/2653864.

Jin, X. et al, 2021, "Failure prediction, monitoring and diagnosis methods for slewing bearings of large-scale wind turbine: A review", Measurement, Vol. 172, p. 108855, https://doi.org/10.1016/j.measurement.2020.108855.

Li, W. et al, 2013, "Design of online monitoring and fault diagnosis system for belt conveyors based on wavelet packet decomposition and support vector machine", Advances in Mechanical Engineering, Vol. 5, p. 797183.

Liu, Z. and Zhang, L., 2020, "A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings", Measurement, Vol. 149, p. 107002, https://doi.org/10.1016/j.measurement.2019.107002.

Mathworks, 2022, The Math Works, Inc., MATLAB (Version 2022a), Help Center, Documentation, [lyapunovExponent], Natick, Massachusetts, United States. https://www.mathworks.com/help/predmaint/ref/lyapunovexponent.html.

Mathworks, 2022, The Math Works, Inc., MATLAB (Version 2022a), Help Center, Documentation, [Signal Features], Natick, Massachusetts, United States. https://www.mathworks.com/help/predmaint/ug/signal-features.html.

Mathworks, 2022, The Math Works, Inc., MATLAB (Version 2022a), Help Center, Documentation, [Continuous 1-D Wavelet Transform], Natick, Massachusetts, United States. https://www.mathworks.com/help/wavelet/ref/cwt.html.

Mathworks, 2022, The Math Works, Inc., MATLAB (Version 2022a), Help Center, Documentation, [Support Vector Machine (SVM)], Natick, Massachusetts, United States. https://www.mathworks.com/discovery/support-vector-machine.html.

Mehdizadeh, S., 2019, "A robust method to estimate the largest lyapunov exponent of noisy signals: a revision to the rosenstein's algorithm", Journal of biomechanics, Vol. 85, pp. 84-91.

Päivinen, N. et al, 2005, "Epileptic seizure detection: A nonlinear viewpoint", Computer methods and programs in biomedicine, Vol. 79, No 2, pp. 151-159, https://doi.org/10.1016/j.cmpb.2005.04.006.

Rosenstein, M. T., Collins, J. J., and De Luca, C. J.,1993, "A practical method for calculating largest Lyapunov exponents from small data sets", Physica D: Nonlinear Phenomena, Vol. 65, No 1-2, pp. 117-134.

Sandoval Núñez, D.A., 2021, "Diagnosis of low-speed bearings via vibration-based entropy indicators and acoustic emissions". Doctoral thesis, UPC, Department of Civil and Environmental Engineering, http://hdl.handle.net/2117/363068.

Saufi, S. R. et al, 2019, "Low-speed bearing fault diagnosis based on ArSSAE model using acoustic emission and vibration signals", IEEE Access, Vol. 7, pp. 46885-46897.

Shri, T. P. and Sriraam, N., 2017, "Comparison of t-test ranking with PCA and SEPCOR feature selection for wake and stage 1 sleep pattern recognition in multichannel electroencephalograms", Biomedical Signal Processing and Control, Vol. 31, pp. 499-512.

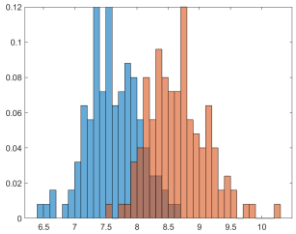
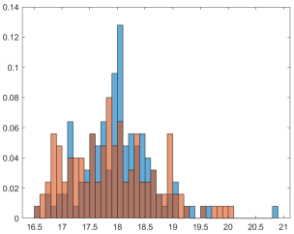
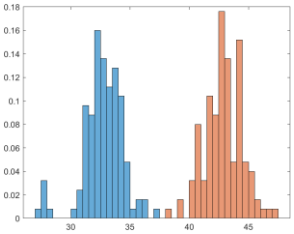
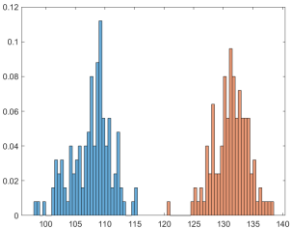
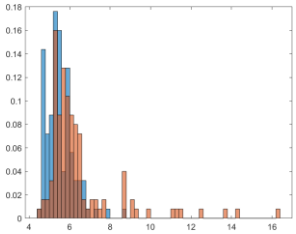
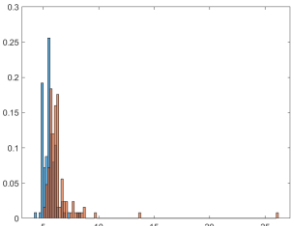
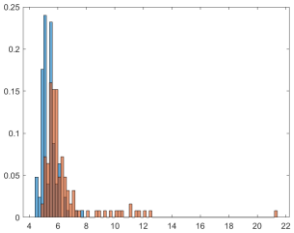
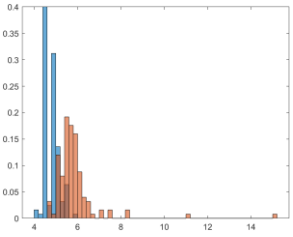
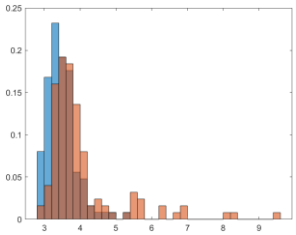
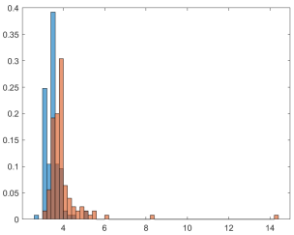
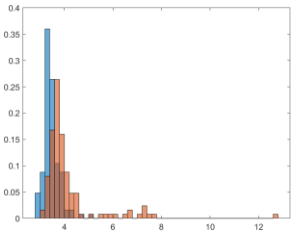
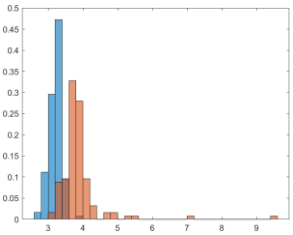
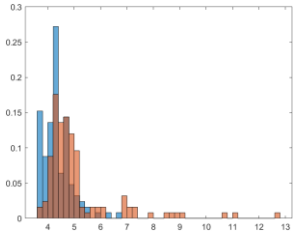
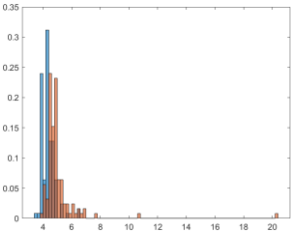
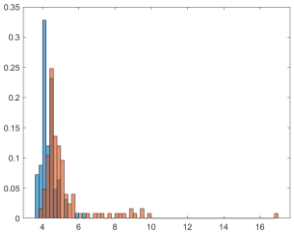
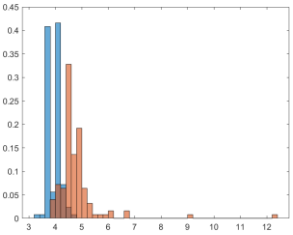
Wold, S., Esbensen, K., and Geladi, P., 1987, "Principal component analysis", Chemometrics and intelligent laboratory systems, Vol. 2, No. 1-3, pp. 37-52.

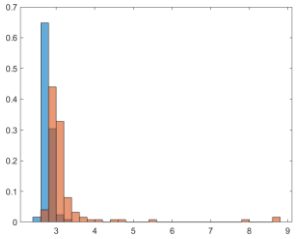
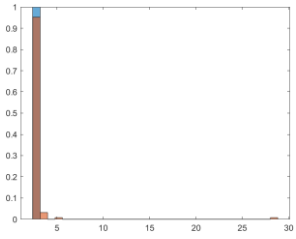
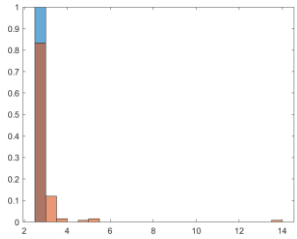
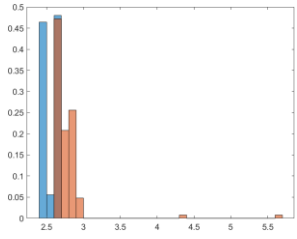
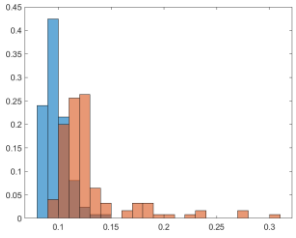
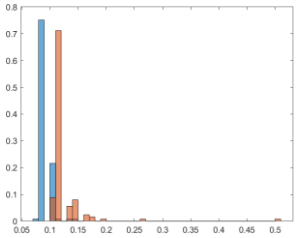
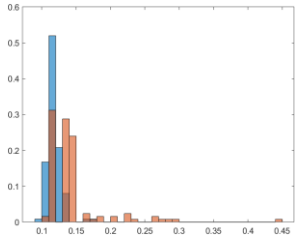
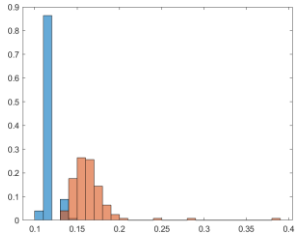
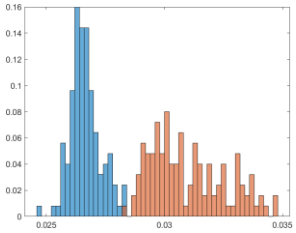
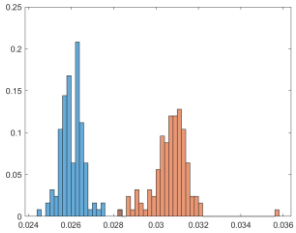
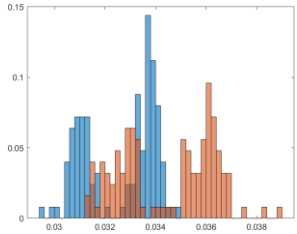
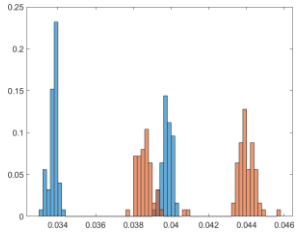
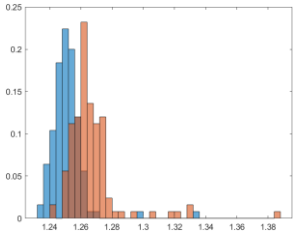
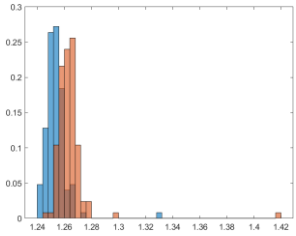
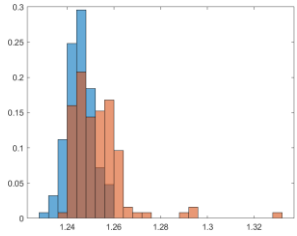
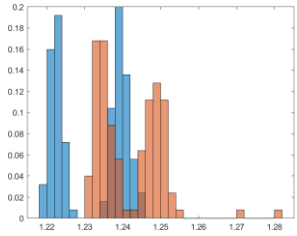
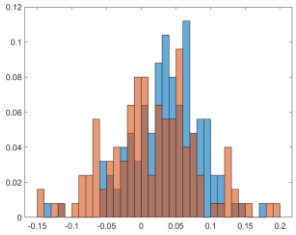
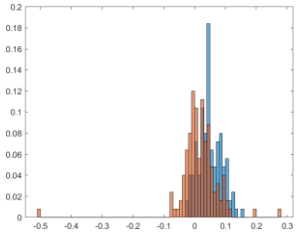
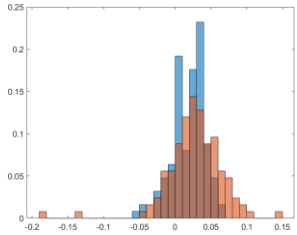
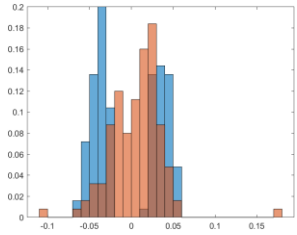
Zimmerman, D. W., and Zumbo, B. D., 1993, "Rank transformations and the power of the Student *t*-test and Welch t-test for non-normal populations with unequal variances", Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, Vol. 47, No. 3, pp .523–539.

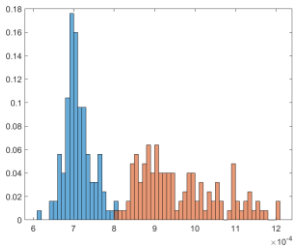
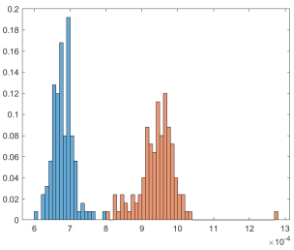
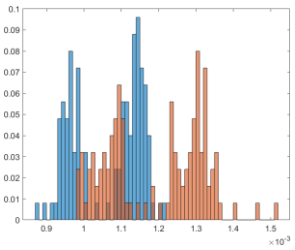
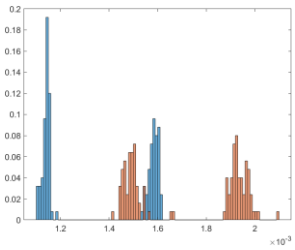
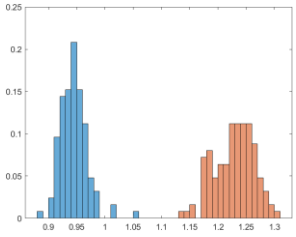
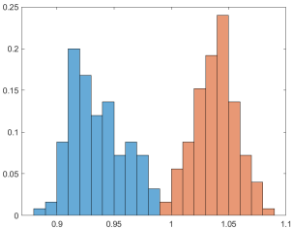
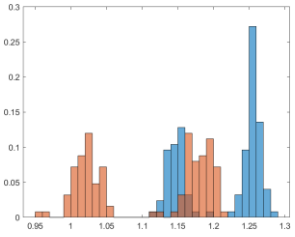
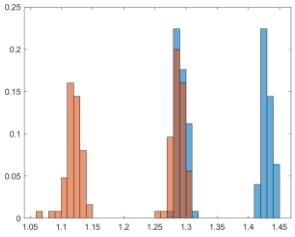
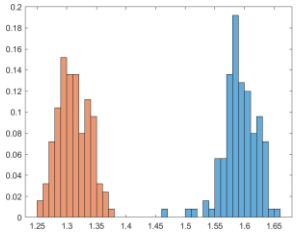
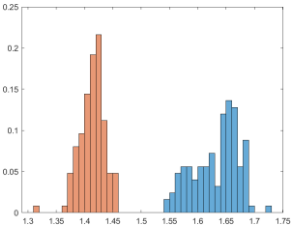
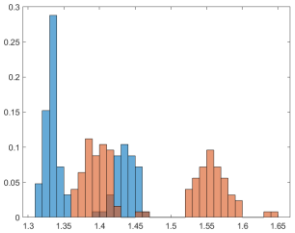
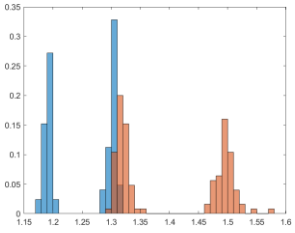
RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.

APPENDIX – Histograms of Feature Tables [Legend: normal (blue) and fault (red) conditions].

D1	D2	D3	D4
MLE			
(0.1)	(0.1)	(0.5)	0.5)
			
Clearance Factor			
(0.2)	(0.2)	(0.2)	(0.2)
			
Crest Factor			
(0.2)	(0.2)	(0.2)	(0.2)
			
Impulse Factor			
(0.2)	(0.2)	(0.2)	(0.2)
			

D1	D2	D3	D4
Kurtosis			
(0.2)	(0.8)	(0.5)	(0.1)
			
Peak Value			
(0.01)	(0.01)	(0.01)	(0.01)
			
RMS			
($2x10^{-4}$)	($2x10^{-4}$)	($2x10^{-4}$)	($2x10^{-4}$)
			
Shape Factor			
($4x10^{-3}$)	($4x10^{-3}$)	($4x10^{-3}$)	($2x10^{-3}$)
			
Skewness			
(0.01)	(0.01)	(0.01)	(0.01)
			

D1	D2	D3	D4
Activity			
($1x10^{-5}$) 	($1x10^{-5}$) 	($1x10^{-5}$) 	($1x10^{-5}$) 
Mobility			
(0.01) 	(0.01) 	(0.01) 	(0.01) 
Complexity			
(0.01) 	(0.01) 	(0.01) 	(0.01) 

The number inside parentheses means the bin width value for the histogram plots. Additionally, the type of normalization applied to the feature data was the relative probability (which is represented by vertical axis), such that the sum of the bar heights is less than or equals to 1.