Chapter 7

Gaussian Mixture Models and Their Applications to Fault Diagnosis in Electric Motors

Chapter details

Chapter DOI:

https://doi.org/10.4322/978-65-86503-88-3.c07

Chapter suggested citation / reference style:

Ribeiro Junior, Ronny F., et al. (2022). "Gaussian Mixture Models and Their Applications to Fault Diagnosis in Electric Motors". *In* Jorge, Ariosto B., et al. (Eds.) *Uncertainty Modeling: Fundamental Concepts and Models*, Vol. III, UnB, Brasilia, DF, Brazil, pp. 195–220. Book series in Discrete Models, Inverse Methods, & Uncertainty Modeling in Structural Integrity.

P.S.: DOI may be included at the end of citation, for completeness.

Book details

Book: Uncertainty Modeling: Fundamental Concepts and Models

Edited by: Jorge, Ariosto B., Anflor, Carla T. M., Gomes, Guilherme F., & Carneiro, Sergio H. S.

Volume III of Book Series in:

Discrete Models, Inverse Methods, & Uncertainty Modeling in Structural Integrity

Published by: UnB City: Brasilia, DF, Brazil Year: 2022

DOI: https://doi.org/10.4322/978-65-86503-88-3

Gaussian Mixture Models and Their Applications to Fault Diagnosis in Electric Motors

Ronny Francis Ribeiro Junior[1*], Fabricio Alves de Almeida[2],

Ariosto Bretanha Jorge[3] and Guilherme Ferreira Gomes[4]

[1]Mechanical Engineering Institute, Federal University of Itajubá (UNIFEI), Itajubá, Brazil – ronnyjr@unifei.edu.br
[2]Institute of Electrical Systems and Energy, Federal University of Itajubá (UNIFEI), Itajubá, Brazil – fabricio-almeida@unifei.edu.br
[3]Post-Graduate Program - Integrity of Engineering Materials, University of Brasilia, Brazil. E-mail: ariosto.b.jorge@gmail.com
[4]Mechanical Engineering Institute, Federal University of Itajubá (UNIFEI), Itajubá, Brazil – guilhermefergom@unifei.edu.br;

*Corresponding author.

Abstract

This chapter presents the theoretical and practical aspects of Gaussian mixture models method. The introduction section describes the motivation for using the Gaussian mixture model. Next, the historical development, concepts, parameters and equations about Gaussian mixture model are presented. Following, a case study on fault classification in electric motors using Gaussian mixture model is described. Finally, the conclusion section ends the chapter.

Keywords: Gaussian mixture model; fault; electric motors; artificial intelligence; vibration; structural health monitoring.

1 Introduction

Clustering is a useful technique for finding structure in a dataset. From the statistical point of view, clustering methods may be divided into probability model-based approaches and nonparametric approaches. The probability model-based approach assumes that the data set follows a mixture model of probability (Yang *et al.*, 2012).

So, any model that represents a dataset with a mix of two or more probability distributions is known as a mixture model. Mixture models aid in cluster identification by fitting a finite number of probability distributions to the data and repeatedly adjusting the parameters of those distributions until they best match the underlying data. Mixture models also aid in describing the presence of subpopulations within an aggregate population that do not require an observed data set to identify the subpopulation to which an individual observation belongs (known as

unsupervised learning in machine learning). In SHM, mixture models help to identify whether there is more than one failure mode in the data.

In real applications, many datasets may be described by Gaussian distributions. As a result, it is reasonable and straightforward to infer that the clusters are derived from various Gaussian distributions. In other words, the dataset is attempted to be modeled as a blend of various Gaussian distributions. Therefore, Gaussian mixture modeling, which fits Gaussian (or normal) distributions to data, is the most frequent type of mixture modeling.

By the end of this chapter, we hope you'll have an understanding of how the Gaussian mixture model works. We will also present an example of how to apply the GMM to a real case to help the reader to understand how to use mixture model for clustering problems.

2 Gaussian Mixture Model

Gaussian mixture models (GMMs) are statistical models that fit data to a set of stochastic models. These models, as GMMs, fit Gaussian distributions to a collection of data in order to split the data into discrete clusters. One of the benefits of mixed model clustering is that it is a soft clustering strategy (Rhys, 2020). Because it is a soft method, the likelihood of each example belonging to a cluster may be quantified. As a result, more information may be obtained to make better judgments. For example, if certain data has 49 percent of belonging to cluster A and 51 percent of belonging to cluster B, how confident can we declare that it belongs to cluster A or B? As a result, each Gaussian in the mixture model represents a possible cluster. Once our Gaussian mixture best matches the data, we can compute the likelihood of each case belonging to each cluster and allocate instances to the most likely cluster.

2.1 Historical Development

The first instance of Gaussian mixture models being used is usually attributed to Karl Pearson (1894). Pearson noted non-normal patterns of forehead to body length ratios in female shore crab populations. Therefore, the author used not only one normal to model the data, but two. This decision was motivated by the suspicion that asymmetries in the histograms of these length ratio could indicate an evolutionary divergence. By adjusting the mixture's parameters, Pearson could make a model that is able to identify two distinct species of shore crab. Despite being successful, the mathematical formulation for the period was an extremely difficult challenge, making the method not widely used.

Subsequently, Wolfe (1963,1965,1970) put a lot of effort into developing Gaussian mixture models for clustering. In his work, Wolfe developed an interactive scheme to estimate the parameters of Gaussians, something innovative for the period. As the main challenge of mixing models is to estimate the parameters of the Gaussians, this step was extremely important for the development of the method.

Finally, the mixture model methods were popularized by Duda and Hard (1973). The authors in their work introduced the maximum likelihood estimation (MLE) technique. The MLE uses an observed data to estimate the parameters of a probability distribution (Myung, 2003). With this technique allied to the beginning of modern computing, it was possible to develop the mixture models.

2.2 Parameters of Gaussian Mixture Models

To define a one-dimensional Gaussian distribution, two parameters (mean and variance) are required. Before beginning the process, it is required to define the number of clusters that are expected to be identified within the dataset; if this number is unknown, it is feasible to approximate it using various approaches that will be detailed in section 2.5.

For now, let's suppose the dataset contains only two distinct clusters (*k* and *j*). The GMM begins by picking random values for the mean and variance of the two Gaussian distributions. The likelihood of belonging to one of the two clusters is then estimated. The Bayes' theorem is the most often used method for performing this computation (Rhys, 2020). Using Bayes' theorem, we get (Kolmogorov & Bharucha-Reid, 2018):

$$p(k|x) = \frac{p(x|k)p(k)}{p(x)} \tag{1}$$

In this case, *p(k|x)* represents the probability of case *x* belonging to Gaussian *k*; *p(x|k)* represents the probability of observing case *x* if you sampled from Gaussian *k*; *p(k)* represents the probability of a randomly selected case belonging to Gaussian *k*; and *p(x)* represents the probability of drawing case *x* if you sampled from the entire mixture model as a whole. The evidence, *p(x)*, is thus the probability of pulling case *x* from either Gaussian distribution.

However, in the case study we need to calculate the probability of belonging to either cluster. therefore, we need to add to the formula the probability of each event occurring independently. thus, Equation 1 can be rewritten for the case of two cluster (*k* and *j*), as shown in Equation 2 and 3 (Rhys, 2020):

$$p(k|x_i) = \frac{p(x_i|k)p(k)}{p(x_i|k)p(k) + p(x_i|j)p(j)} \tag{2}$$

$$p(j|x_i) = \frac{p(x_i|j)p(j)}{p(x_i|k)p(k) + p(x_i|j)p(j)} \tag{3}$$

All that remains now is to compute the likelihood and prior. The Likelihood function of the Gaussian distribution shows us the relative likelihood of drawing a case with a certain value from a Gaussian distribution with a specific mean and variance combination. Equation 4 depicts the probability density function for the Gaussian distribution *k*:

$$p(x_i|k) = \frac{1}{\sqrt{2\Pi\sigma_k^2}} e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}} \tag{4}$$

Where, μ_k and are the man and variance of the *k* Gaussian.

At the start of the process, the prior probabilities, like the Gaussian means and variances, are generated at random. However, for subsequent calculations, the iterative procedure of this

method must be followed. Expectation Maximization is used for this. Figure 1 shows the convergence process of a one-dimensional GMM.

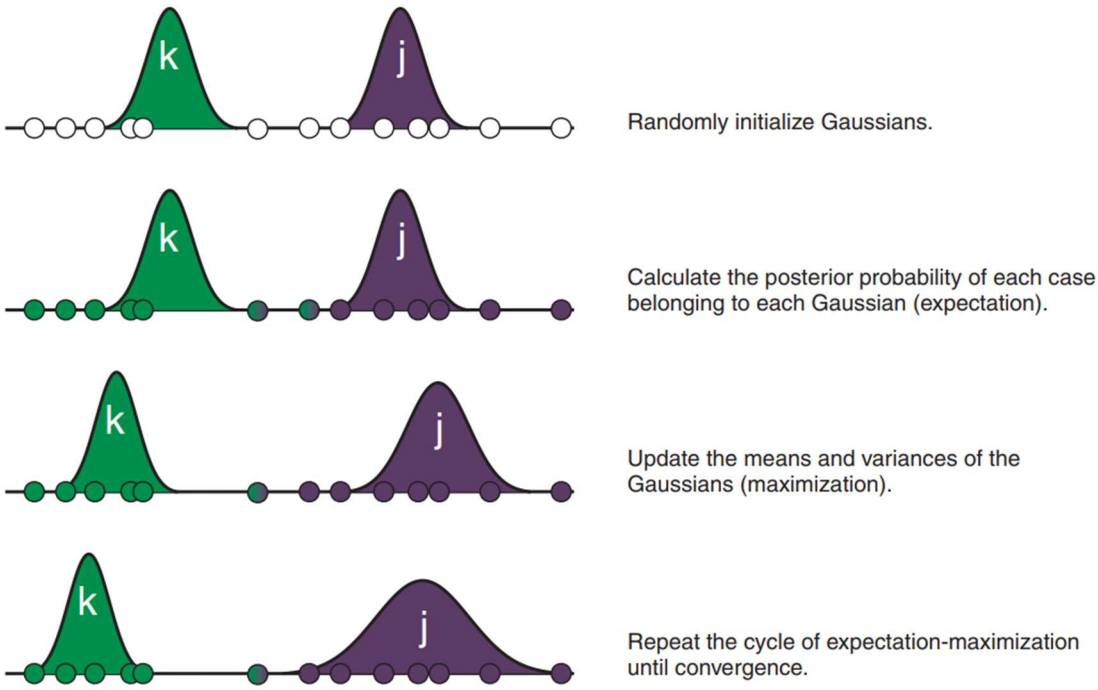

Figure 1 – Convergence process of a one-dimensional Gaussian mixture model (adapted from Rhys, 2020).

2.3 Expectation-maximization

Maximum likelihood estimation is a method of estimating a dataset's density by exploring across probability distributions and their parameters. When there are variables that interact with those in the dataset but are hidden or not observed, maximum likelihood becomes intractable (Murphy, 2012). These variables are also called latent variables.

The expectation-maximization algorithm is a method for calculating maximum likelihood when latent variables are present. This is accomplished by first estimating the values of the latent variables (E-step), then optimizing the model (M-step), and finally repeating these two stages until convergence is achieved. It is a powerful and versatile method that is often used in density estimation with missing data, such as clustering algorithms like the Gaussian mixture model (Hastie *et al*., 2009).

The EM method may be characterized by the following equations given a set of observable data (U), a set of latent data (J), and a vector of unknown parameters (θ), along with likelihood function $L(\theta, U, J) = p(U, J \mid \theta)$ (Moon, 1996; Dellaert, 2002).

Expectation step (E step)

$$Q(\Theta \mid \Theta^{(t)}) = E_{J|U,\Theta^{(t)}}[logL(\Theta; U, J)] \tag{5}$$

Maximization step (M step)

$$\Theta^{(t+1)} = \underset{\Theta}{argmax}\ Q(\Theta|\Theta^{(t)}) \tag{6}$$

The algorithm's iterative procedure begins with a random value of θ. The probability of each potential value of J is then computed, given θ. The new θ parameters are estimated using these values. The procedure is repeated until the algorithm achieves convergence. A maximum number of iterations or a stopping condition is typically employed to determine an algorithm's convergence. There are various alternatives for the halting criterion. The most common is to halt the algorithm due to a lack of progress in the likelihood function (Mc Nicholas, 2016), which means to stop the process when:

$$l^{k+1} - l^{k} < \varepsilon \tag{7}$$

Where l is the likelihood function value, and k denotes the number of the iteration. The value of ε is defined for the sake of the problem under consideration.

2.4 Multivariate clustering problems

All the development in the previous sections has been with a one-dimensional GMM in mind. However, most of the time, we will come across datasets with multiple variables. Therefore, there is a version of GMM for multiple variables. One-dimensional GMMs are defined by two parameters: mean and variance. Multivariate GMM is defined by two new parameters: the centroid and the covariance matrix.

In a nutshell, the centroid is a vector of means that represents an average for each dimension in the dataset. The covariance matrix is a square matrix whose elements are the covariance between the dataset's variables. The covariance matrix is often a non-standardized number. However, you may normalize this matrix by dividing the value by the product of the standard deviations of the variables.

Therefore, the equations change to the multivariate case. In general, what happens is a generalization of the one-dimensional case to more dimensions. Equations 8 to 10 describe the calculation of the centroid, covariance matrix and likelihood for the multivariate case (Rhys, 2020). Figure 2 shows the GMM algorithm for a bivariate case.

$$\mu_{k,a} = \sum_{i=1}^{n} \left(\frac{p(k|xi)}{n \times p(k)}\right)_{x_{i,a}} \tag{8}$$

$$(\sigma_k)_{a,b} = \sum_{i=1}^{n} \left(\frac{p(k|xi)}{n \times p(k)}\right)(x_{i,a} - \mu_{k,a})(x_{i,b} - \mu_{k,b}) \tag{9}$$

$$p(x_i|k) = \frac{1}{\sqrt{(2\pi|\sum_k|)}} e^{-0.5(\vec{x_i}-\vec{\mu_k})^T \sum_k^{-1} (\vec{x_i}-\vec{\mu_k})} \tag{10}$$

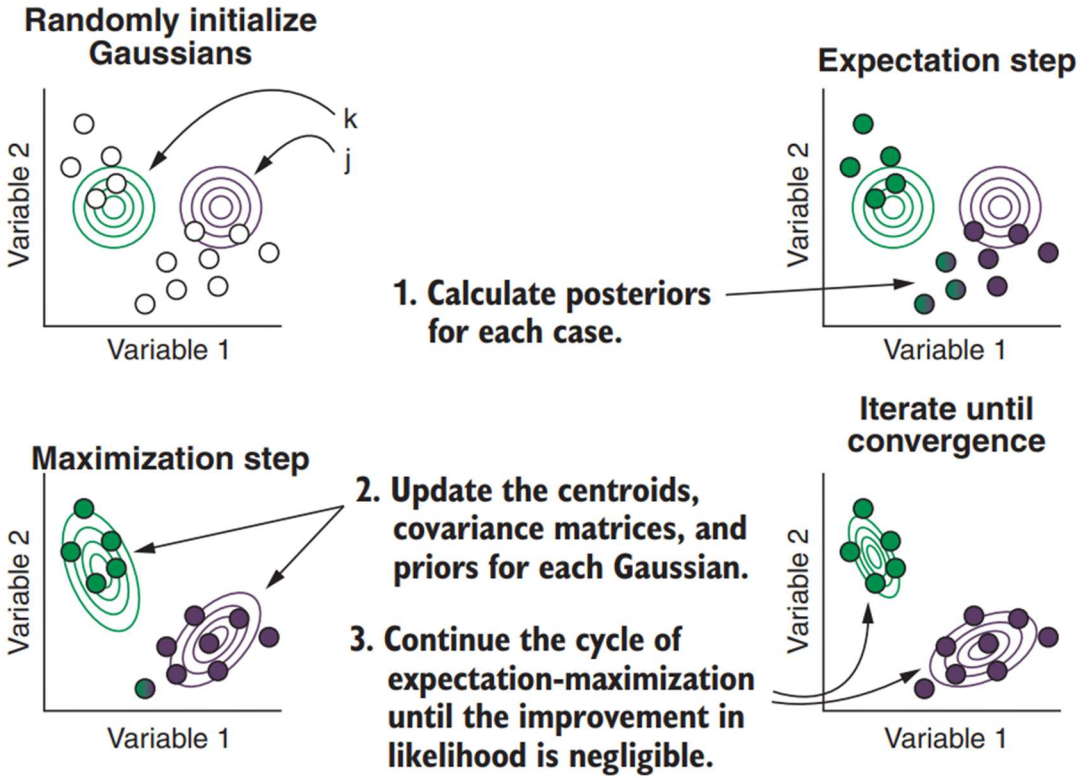

Figure 2 – Gaussian mixture model algorithm for a bivariate case (adapted from Rhys, 2020).

2.5 Number of components in Gaussian Mixture Model

Choosing the number of components simply selects the number of Gaussians for your model. However, when you have no prior knowledge of your dataset, it is necessary to use techniques to assist in this choice, i.e., model selection. In short, a fitting process is performed with several models on a given dataset, and one is chosen among all the others. It is worth mentioning that this is an alternative when there is no information about the dataset. If you know that there are three different types of clusters and the model selection indicates a different number, it is recommended that you use the three and not the one given by the model. There are several ways to choose the number of components in a GMM; even the trial-and-error method can be used. However, there are more optimized methods for this task. In this section, the main methods used for model selection will be discussed.

2.5.1 Akaike Information Criterion

Hirotugu Akaike created the Akaike information criterion (AIC) in 1971 (Akaike, 1974). AIC is a statistical estimator that helps in the selection of the optimal model for a given dataset. The lower the AIC score for a dataset, the better the fit for the data (Profillidis & Botzoris, 2018). Because it is a comparative method, it can only compare the score value for the same dataset. The AIC formula is defined by Equation 11 (Hastie et al., 2009).

$$AIC = -\frac{2}{N} \cdot LL + 2\frac{k}{N} \tag{11}$$

Where N is the number of examples in the training dataset, LL is the log-likelihood of the model on the training dataset, and k is the number of parameters in the model.

2.5.2 Bayesian Information Criterion

The Bayesian information criterion (BIC) was developed by Gideon Schwarz in 1978 (Schwarz, 1978). Like the AIC, the BIC is a statistical estimator for selecting a model that best fits the dataset. The main difference between BIC and AIC is that BIC penalizes more complex models more sharply. This makes simpler models tend to have a lower score, making them more likely to be chosen. BIC is mathematically defined by Equation 12 (Hastie *et al*., 2009).

$$BIC = -\frac{2}{N} \cdot LL + log(N) \cdot k \tag{12}$$

Where *log( )* has the base-*e* called the natural logarithm, LL is the log-likelihood of the model, N is the number of examples in the training dataset, and k is the number of parameters in the model.

2.5.3 Minimum description length

The Minimum description length (MDL) concept is a strong inductive inference approach that serves as the foundation for statistical modeling, pattern recognition, and machine learning. The MDL idea was first suggested in 1978 by Rissanen (Rissanen, 1978). The notion is that with the assistance of a model, every data set can be suitably encoded, and that the more we compress the data by eliminating redundancy from it, the more we find underlying regularities in the data. As a result, code length is directly connected to the model's generalization capabilities, and the model that gives the shortest description of the data should be chosen (Myung, 2001). Like the other two models, the lowest score is chosen. MDL is calculated as Equation 13 (Hastie *et al*., 2009).

$$MDL = -log(P(\theta)) - log(P(y|\theta)) \tag{13}$$

Where *log( )* has the base-e called the natural logarithm, θ is the model parameters, y is the target values.

2.5.4 Silhouette Coefficient

Silhouette coefficient is a method to estimate distance between cluster, proposed by Peter Rousseeuw in 1987 (Rousseeuw, 1987). The silhouette coefficient compares the similarity of an object to its own cluster (cohesion) to other clusters (separation). The silhouette has a value between -1 and +1, with a high value indicating that the object is well fitted to its own cluster and a low value indicating that the object is poorly adapted to surrounding clusters. The

clustering setup is useful if the majority of the items have a high value. If a large number of points have a low or negative value, the clustering configuration may have too many or too few clusters. The coefficient is defined for each sample and is defined by Equation 14.

$$s = \frac{b - a}{max(a, b)} \tag{14}$$

where a is the mean distance between a sample and all other points in the same class and b is the mean distance between a sample and all other points in the next nearest cluster.

3 Case study: Motor Fault Classification

In this section we will present a practical case where the Gaussian mixture model was used to cluster, identify, and diagnose six different failures in electric motor. We hope that this example will help to better understand the usefulness of the GMM.

3.1 Experimental Setup

All tests were carried out at PS Soluções in Itajuba (Brazil). The test bench (Figure 1) employed a 0.5 HP induction motor that was directly linked to the SpectraQuest machine failure simulator (Ribeiro Junior et al., 2020). Accelerometer 1 detects vibration in the y direction, whereas accelerometer 2 detects vibration in the x direction. Table 1 also illustrates the motor's basic characteristics at 60 and 50 Hz.

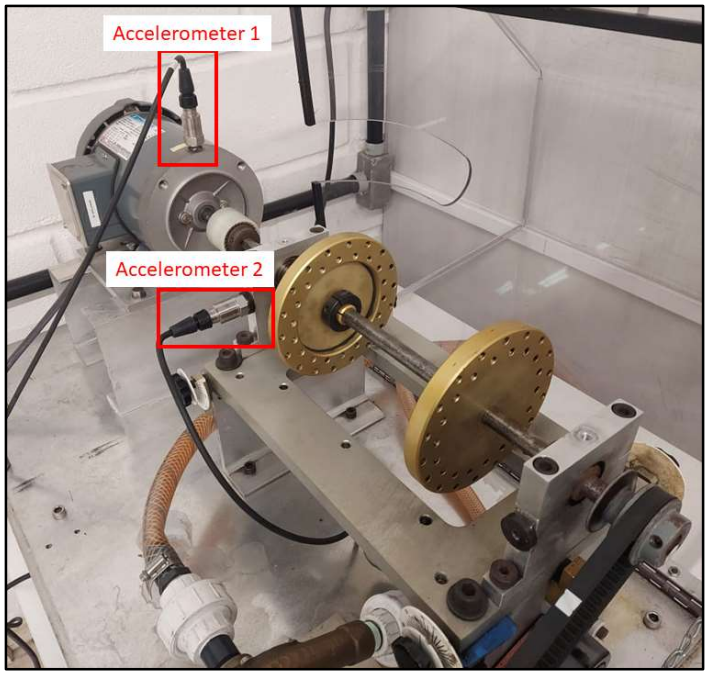

Figure 3. Experimental setup (adapted from Ribeiro Junior et al., 2022).

Table 1. Marathon motor nameplate (adapted from Ribeiro Junior et al., 2022).

Parameter	60 Hz	50 Hz
Power (HP)	0.50	0.33
Rotation (RPM)	3450	2850
Voltage (V)	208-230/460	190/380
Amperage (A)	2.1-2.2/1.1	2.0/1.0
Service Factor	1.15	1.15

The experimental bench is made up of five electric motors that may test up to seven operational conditions. These conditions include normal, bent shaft, broken bar, misalignment, mechanical looseness, bearing problem, and imbalanced (Figure 4).

- **Normal** condition is the electric motor in a healthy state, without any fault.
- **Bent shaft** refers to the motor whose shaft is slightly bent (approximately 4 mm), inducing a scratch between the rotor and the stator as shown in Figure 4(a). This fault mainly occurs due to overloads and coupling conditions.
- **Broken bar** refers to the motor whose squirrel cage rotor bar is broken. This fault is more common on bigger electric motors. As shown in Figure 4(b) this fault is simulated by drilling 6 holes with 2 mm diameter in the rotor.
- **Misalignment** refers to the motor whose shaft is misaligned. This fault occurs due to bad installation or overloads. Figure 4(c) shows how this defect is induced. On the bench, it is possible to generate up to 3 mm of misalignment
- **Mechanical looseness** refers to the motor whose screws are loosening. This fault commonly occurs due to bad installation but high vibration may cause this fault. Figure 4(d) shows which screws are loosened.
- **Bearing fault** refers to the motor whose bearing is damaged. This is the most common fault in electric motors. Figure 4(e) shows the bearing fault in the outer race. The bearing used is SKF-6203.
- **Unbalanced** refers to the motor whose load is unbalanced. This fault is very common in electric motors, being the reason for other faults. Figure 4(f) shows how this fault is induced.

These conditions cover the most frequent electric motor failures. Several investigations confirm that the selected failures are the most frequently defects in electric motors (Randal, 2011; Plante et al., 2016; Zarei et al, 2014; Choudhary et al., 2019; ABRAMAN, 2011).

In all, four rounds of tests with all operational conditions were done, gathering 28 signals from each accelerometer. To properly analyze the faults, the defects were removed and installed in different motors for each run.

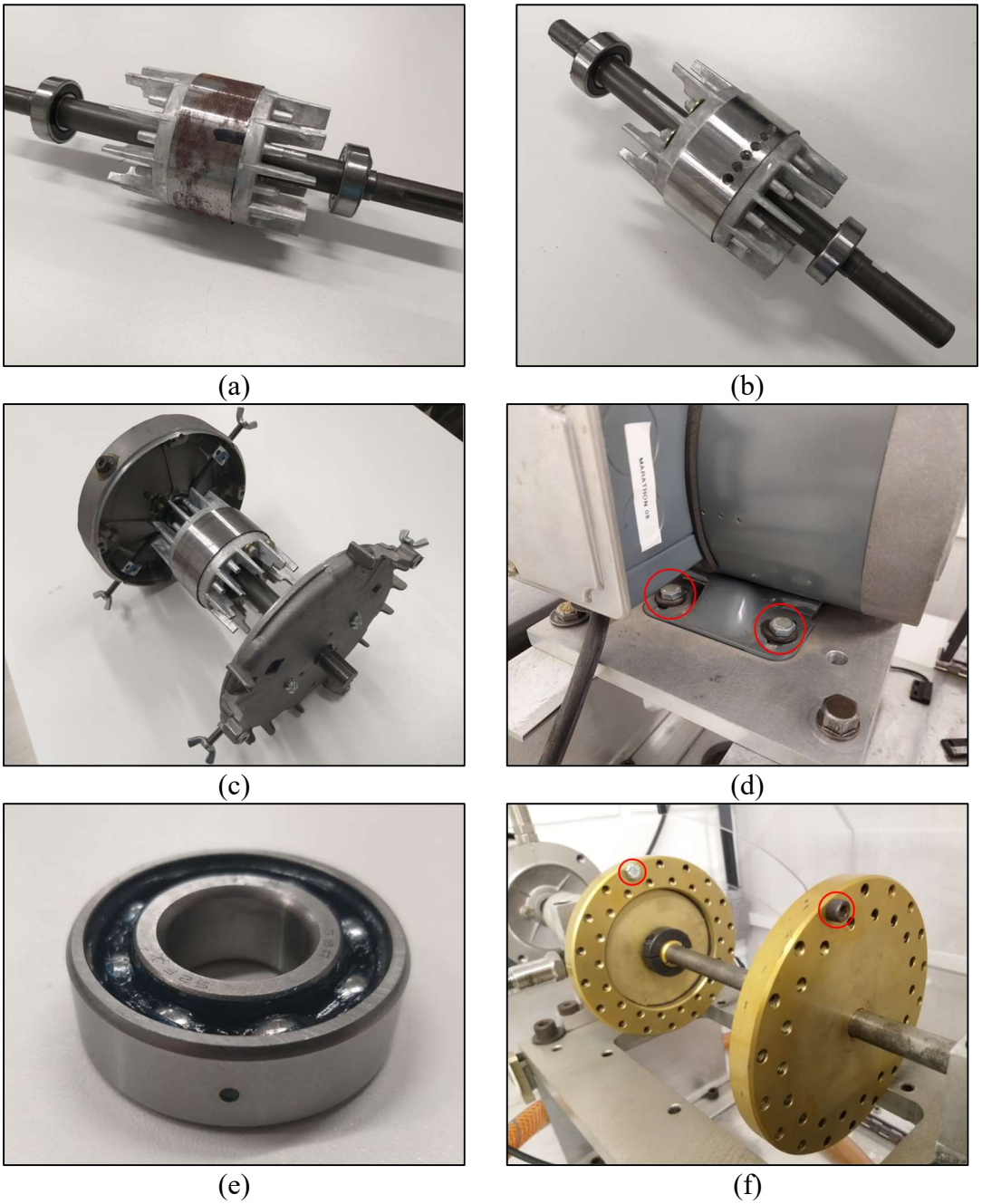

Figure 4. Fault types: (a) Bent shaft, (b) Broken bar, (c) Misalignment, (d) Mechanical looseness, (e) Bearing fault, (f) Unbalanced (adapted from Ribeiro Junior et al., 2022).

3.2 Vibration signal acquisition

The IMI uniaxial accelerometer records the two vibration signals. The accelerometer has a frequency range of ±3 dB, a measuring range of ±50 g, a sensitivity of 100 mV/g, and resonance at a frequency of 25 kHz. Because the frequency of the recorded signals is low and the engine mass is significantly more than the mass of the accelerometer, these facts make the sensor suitable for vibration measurement. A threaded base links the first accelerometer to the motor. The second is also kept in place by a threaded base that is put into the bearing housing. Araldite was used to bind all threaded bases. The acquisition system (Preditor®) gathers 262144 sample points at a sampling frequency of 11718.5 Hz and an acquisition window of 22.37 seconds.

3.3 Procedures of the Proposed Method

The primary step is to gather all vibration signals from various operational circumstances. For each operational state, we gathered four signals, 75 percent for training and 25 percent for testing. Following that, the signals are divided into smaller sizes, each with 512 sample points. Finally, we employ RMS and Skewness statistics to extract features from vibration signals in order to improve classification results. The RSM and Skewness measures were chosen because of their significance and capacity to aggregate information, as proved by Ribeiro Junior et al (2020). Other statistics employed were Kurtosis, SRA, Peak-to-Peak, Crest factor, and Shape factor. However, none of them performed as well as the duo of RMS and Skewness. The main parameter of GMM is the covariance matrix, being possible to choose 4 different types:

- Full: each component has its own general covariance matrix
- Tied: all components share the same general covariance matrix
- Diag: each component has its own diagonal covariance matrix
- Spherical: each component has its own single variance

Because each data set can have a best-fitting parameter, the optimal form of covariance will be picked. The average of two random variables' degrees of dependency or linear numerical interrelationship is their covariance or variance. Choosing the best covariance type for your dataset can help the GMM model perform better.

Another method is to begin the process supervised, that is, with the mean of each distribution as the initial weight. This reduces the number of iterations required for optimizers to converge while also avoiding incorrect convergence. Because the model was begun supervised, the number of mixing components will be limited to two (healthy and a faulty motor). The number of iterations was limited to 10,000 with a10^{-6} convergence criterion.

The Mahalanobis distance is then computed. The Mahalanobis distance is the distance between two points in multivariate space. Variables in a normal Euclidean space are represented by axes drawn at right angles to each other. To measure the distance between two points, use a ruler. For uncorrelated data, the Euclidean distance equals the Mahalanobis distance. When two or more variables are connected, the axes are no longer at right angles, prohibiting ruler measurements. Figure 5 shows a comparison between the Mahalanobis distance and the Euclidian distance in a two-dimensional space. The elongated cluster's Euclidean distance to its centroid is greater than that of the round cluster. The Mahalanobis distance to the centroid of an elongated cluster, on the other hand, is lower than that of a spherical one. Furthermore, if you have more than three variables, you will be unable to plot them at all in normal 3D space. The Mahalanobis distance, which quantifies distances between points, including correlated points, solves this measurement difficulty. The Mahalanobis distance is a measure of distance relative to the centroid, which may be thought of as the overall mean for the Mahalanobis distance is a measure of distance relative to the centroid, which is the general mean for multivariate data. The centroid in multivariate space is the point at which all meanings from all variables coincide. The bigger the Mahalanobis distance, the greater the distance between the data point and the centroid. The formula version utilized is described by Equation 15 (Hamil et.al, 2016):

$$d_i = [(x_i - \bar{x})^T C^{-1} (x_i - \bar{x})]^{0.5} \quad (15)$$

Where $\boldsymbol{x_i}$ is the observation point, $\bar{\boldsymbol{x}}$ is the mean of the reference cluster (normal condition) and $\boldsymbol{C^{-1}}$ is the inverse covarian matrix of independent variables.

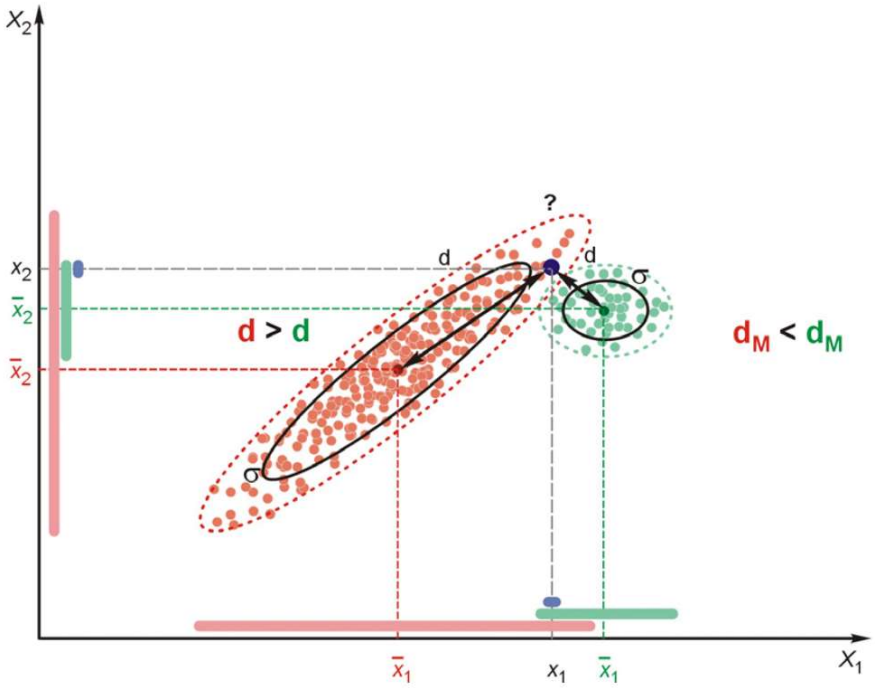

Figure 5. Comparison between Mahalanobis distance and Euclidian distance (adapted from Woźniak et al., 2019).

All of the features in the proposed method were created in Python using the Scikit-learn library. Figure 6 depicts the overall flowchart of the approach followed in this study.

(Intentionally left blank)

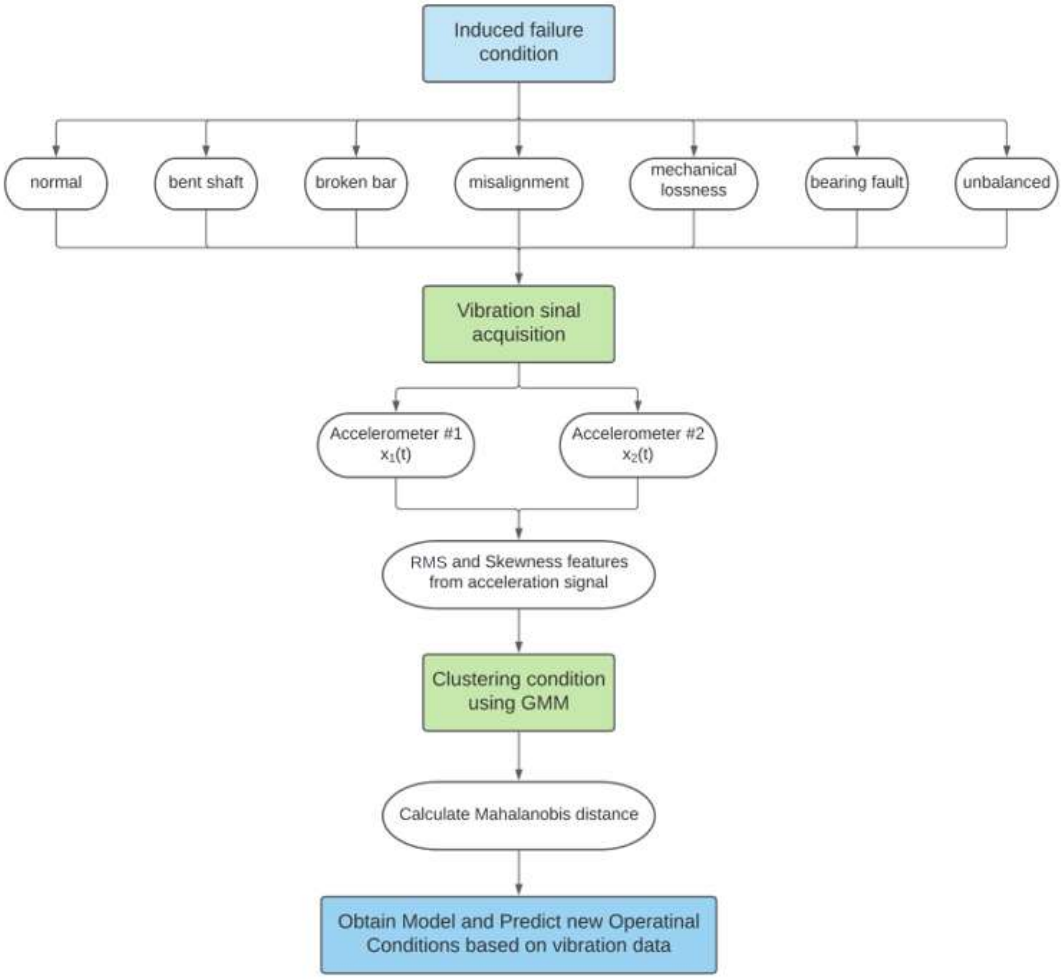

Figure 6. General flowchart of the motor operation condition clustering using GMM (adapted from Ribeiro Junior, 2022).

3.4 Results and Discussion

In this section, we will discuss and exhibit the primary and most important findings from the vibration signal analysis in seven distinct operating conditions (six faults) in electric motors. Figure 7 compares the time domain vibration signals for the six fault scenarios to the accelerometer 1 normal signal. Figure 8 shows the same findings when accelerometer 2 is used. In certain ways, it is obvious that the signal varies under fault circumstances, each with its own distinct pattern. This is especially true for accelerometer 1 and operating circumstances characterized by bearing failure and a lack of mechanical looseness. As a result, only accelerometer signal 1 will be used as a GMM input. Figure 9 also provides the scatter plot data for the accelerometer signal. Except for the mechanical loss working state (Figure 9(d)), when probable nonlinearities are present, a dot pattern is found in all conditions.

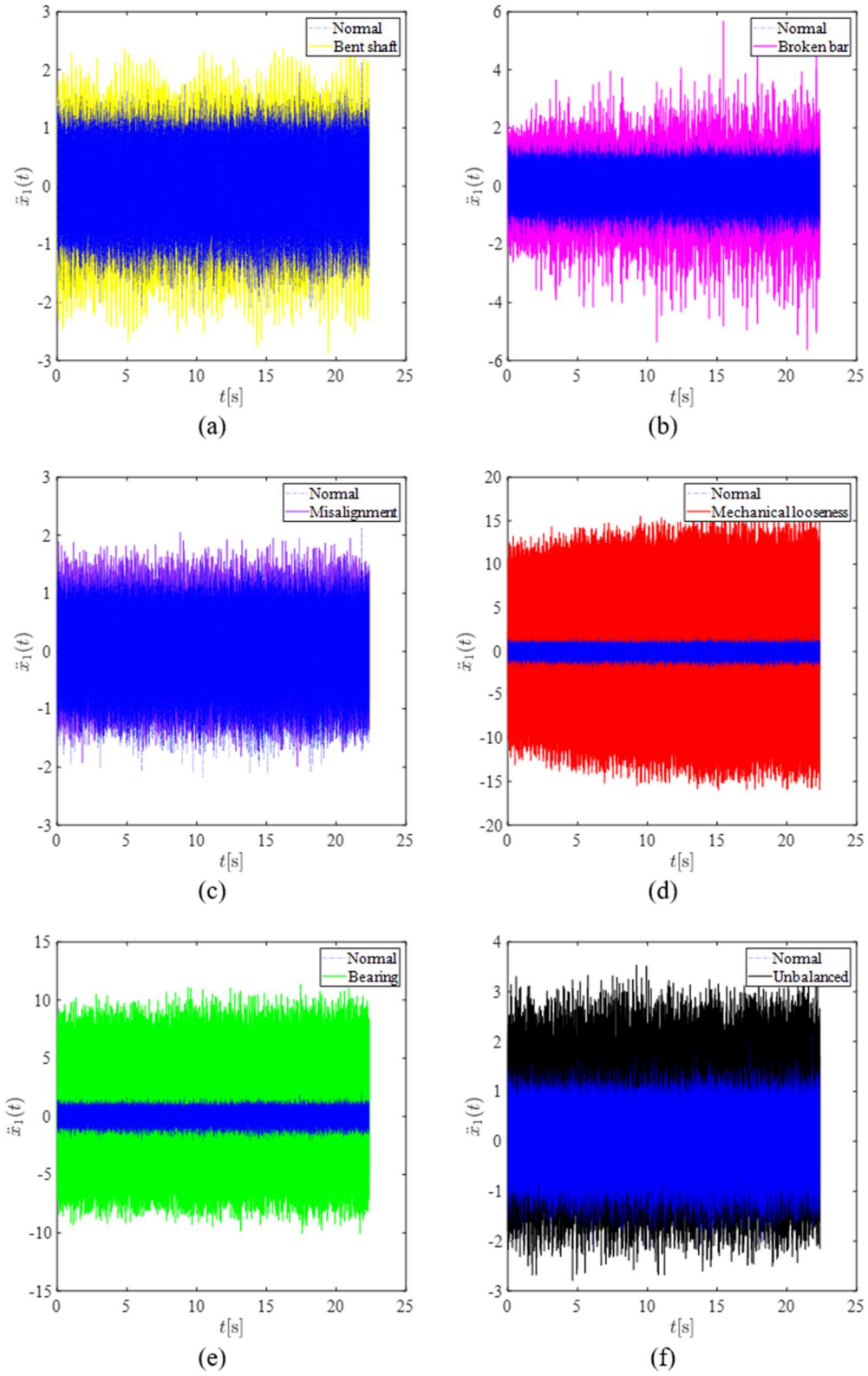

Figure 7. Vibrations signal for the six different fault conditions considering the vibration accelerometer #1 (adapted from Ribeiro Junior, 2022).

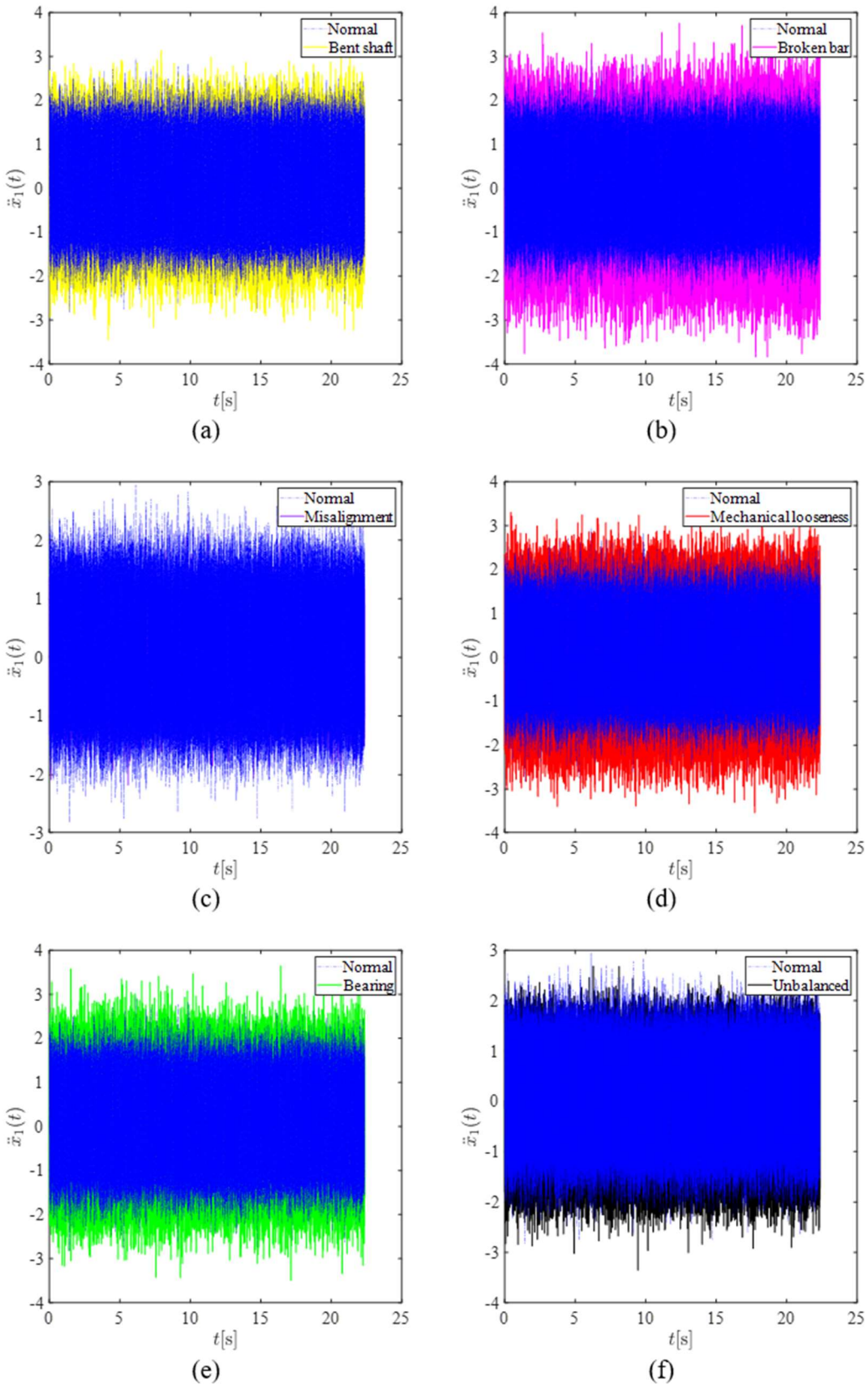

Figure 8. Vibrations signal for the six different fault conditions considering the vibration accelerometer #2 (adapted from Ribeiro Junior, 2022).

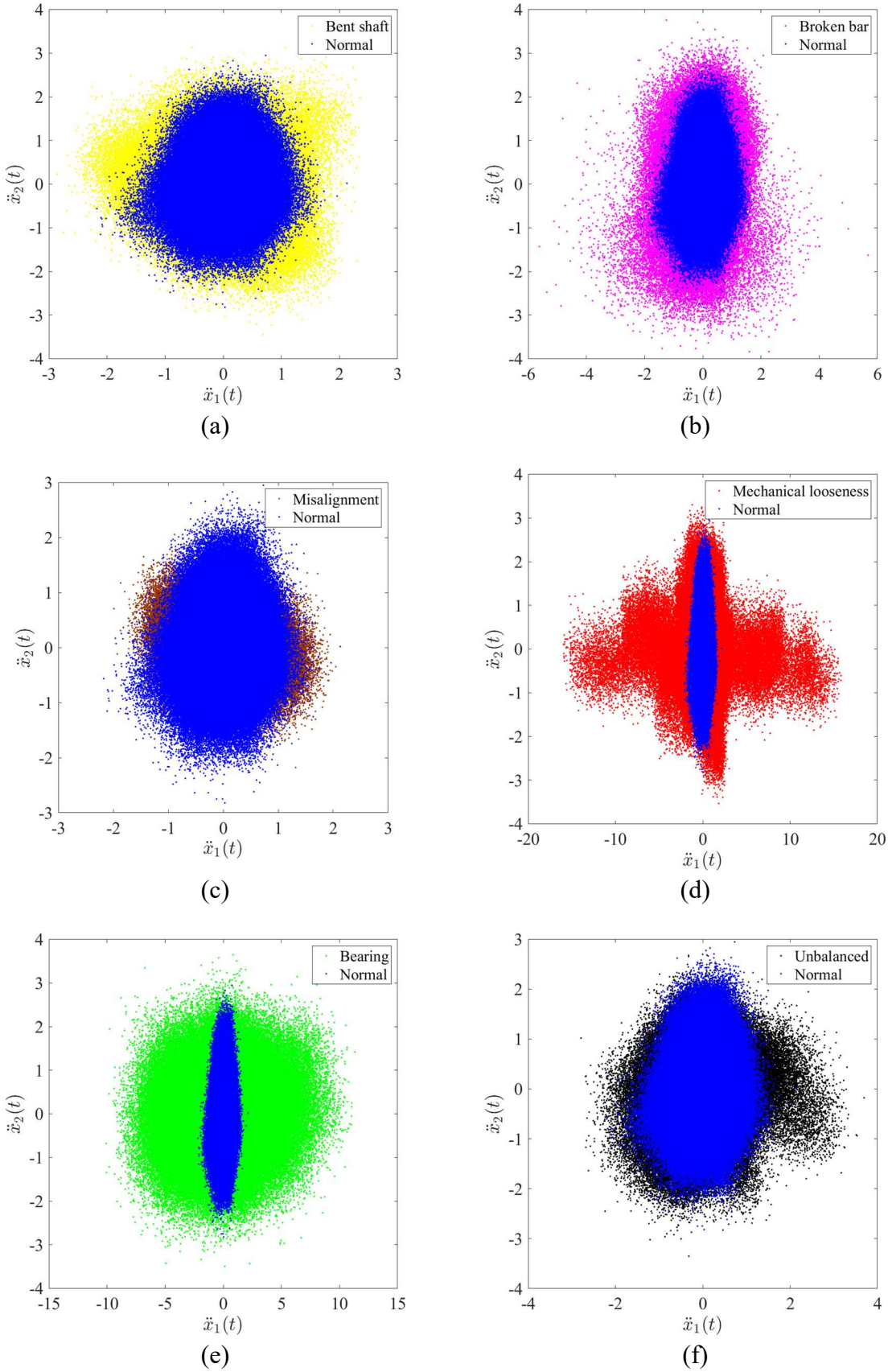

Figure 9. Scatter vibration signal for the six different fault conditions considering the vibration considering both the accelerometers #1 and #2 (adapted from Ribeiro Junior, 2022).

The second step of machine learning was carried out after an exploration phase of vibration data in various operating situations (caused failures). An unsupervised learning method was required during this step. Because it is a flexible approach characterized as "soft clustering," a Gaussian mixing model (GMM) was employed. In other terms, a GMM is a probabilistic model in which all data points are created by a finite number of Gaussian distributions with unknown parameters.

Figures 10–15 illustrate the training results for the GMM algorithm under all operational circumstances. All covariance matrix parameters are displayed in the findings (full, tied, diag, and spherical). For all figures, the · represents training data and the X represents test data.

Some clustering results are more segregated (distant) than others. The clusters are closer together in the case of misalignment. The groups of bent and broken bars are far enough apart yet share a border. In the other cases, the generated groupings are more separated. Because GMM is a soft algorithm, its usage is justified at this stage. When the various forms of covariance matrix are analyzed, the full or diagonal covariance matrix settings are the best for the dataset, with the full option slightly better. This setup has an accuracy of more than 90% for all defects assessed. The findings are comparable to earlier researches (Wang et al., 2009; Yan et al., 2017; Hong et al., 2019; Pani et al., 2020).

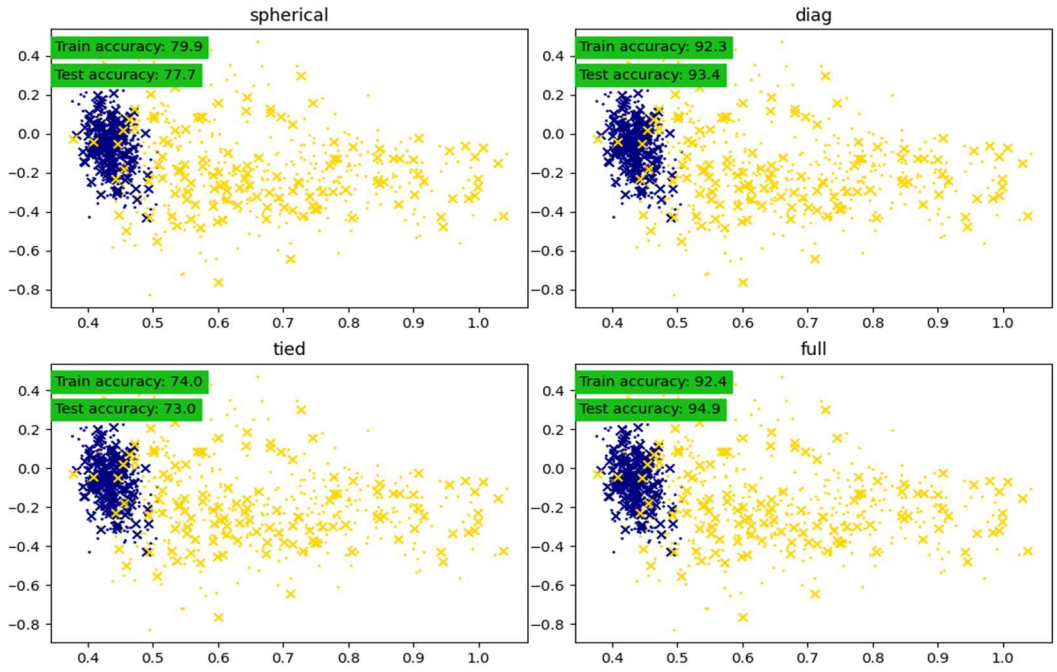

Figure 10. Gaussian mixture model results (RSM *vs.* Skewness) for the Bent and Normal conditions (adapted from Ribeiro Junior, 2022).

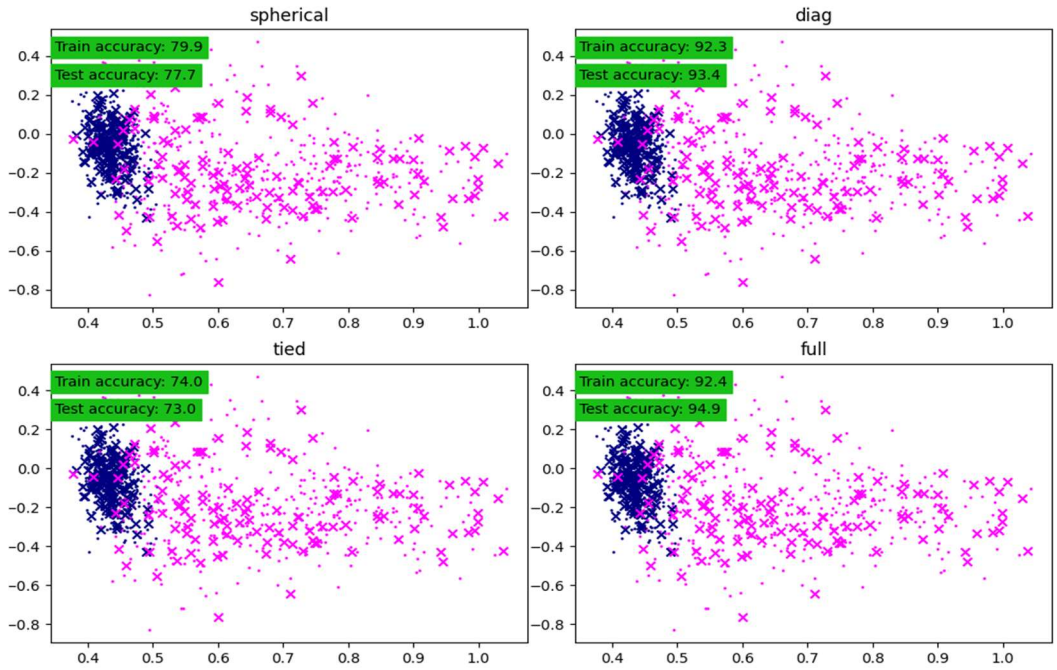

Figure 11. Gaussian mixture model results (RSM *vs.* Skewness) for the Broken bar and Normal conditions (adapted from Ribeiro Junior, 2022).

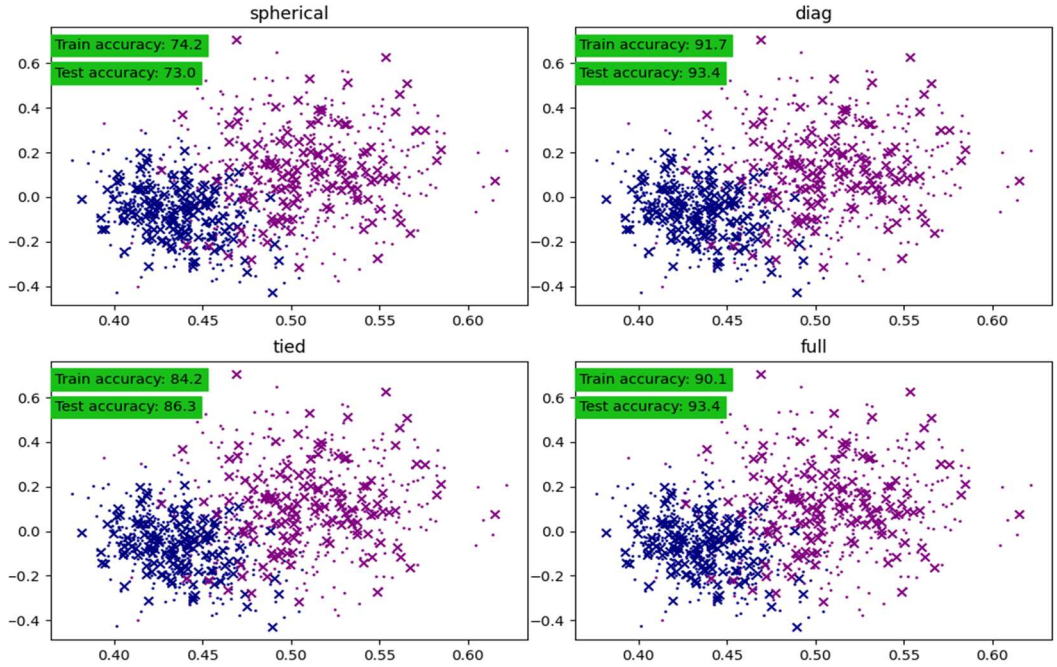

Figure 12. Gaussian mixture model results (RSM *vs.* Skewness) for the Misalignment and Normal conditions (adapted from Ribeiro Junior, 2022).

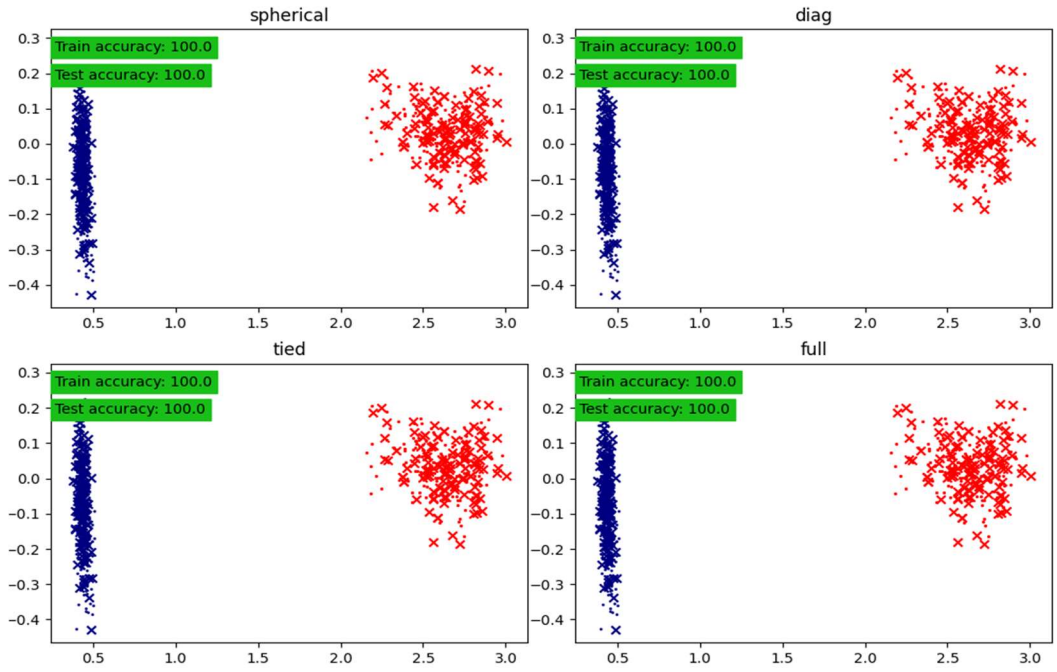

Figure 13. Gaussian mixture model results (RSM *vs.* Skewness) for the Mechanical looseness and Normal conditions (adapted from Ribeiro Junior, 2022).

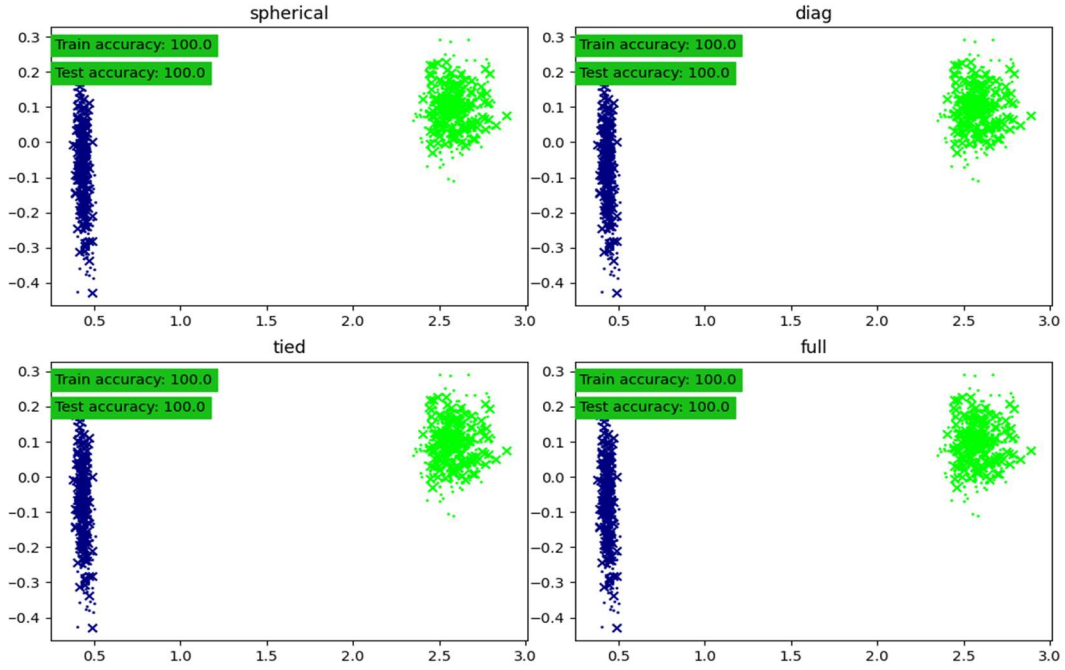

Figure 14. Gaussian mixture model results (RSM *vs.* Skewness) for the Bearing and Normal conditions (adapted from Ribeiro Junior, 2022).

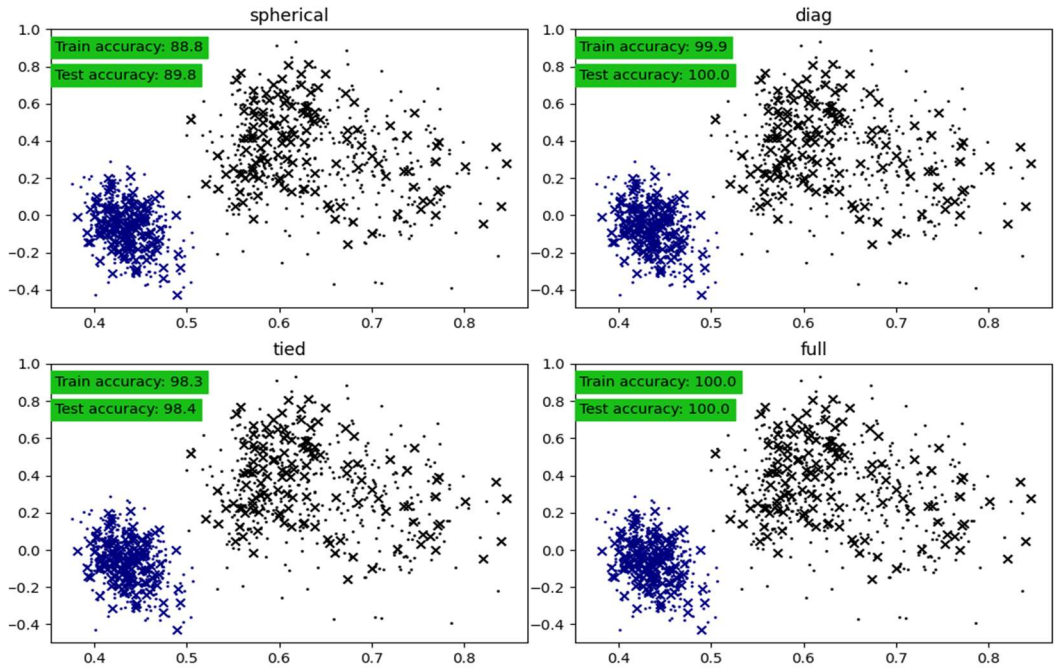

Figure 15. Gaussian mixture model results (RSM *vs.* Skewness) for the Unbalanced and Normal conditions (adapted from Ribeiro Junior, 2022).

The Mahalanobis distance measure was used to determine how far apart the clusters are. The Mahalanobis distances of the Gaussian distributions adapted to the usual operating state are shown in Figure 16. It was discovered that the mechanical looseness situation produced a certain pattern of distances. Figure 17 organizes all of the distances. It should be noted that mechanical looseness and bearing faults are the most extreme operation condition. This suggests that these are the most easily identified defects. Furthermore, the Mahalanobis distance pattern may be used to distinguish one fault from another, i.e., identify which fault it is. This is not achievable with GMM alone since it simply detects the presence of a fault. Figure 17 shows that each defect has a pattern and various distance value ranges. As a result of integrating the GMM with the Mahalanobis distance, it is able to not only identify but also discriminate between faults.

(Intentionally left blank)

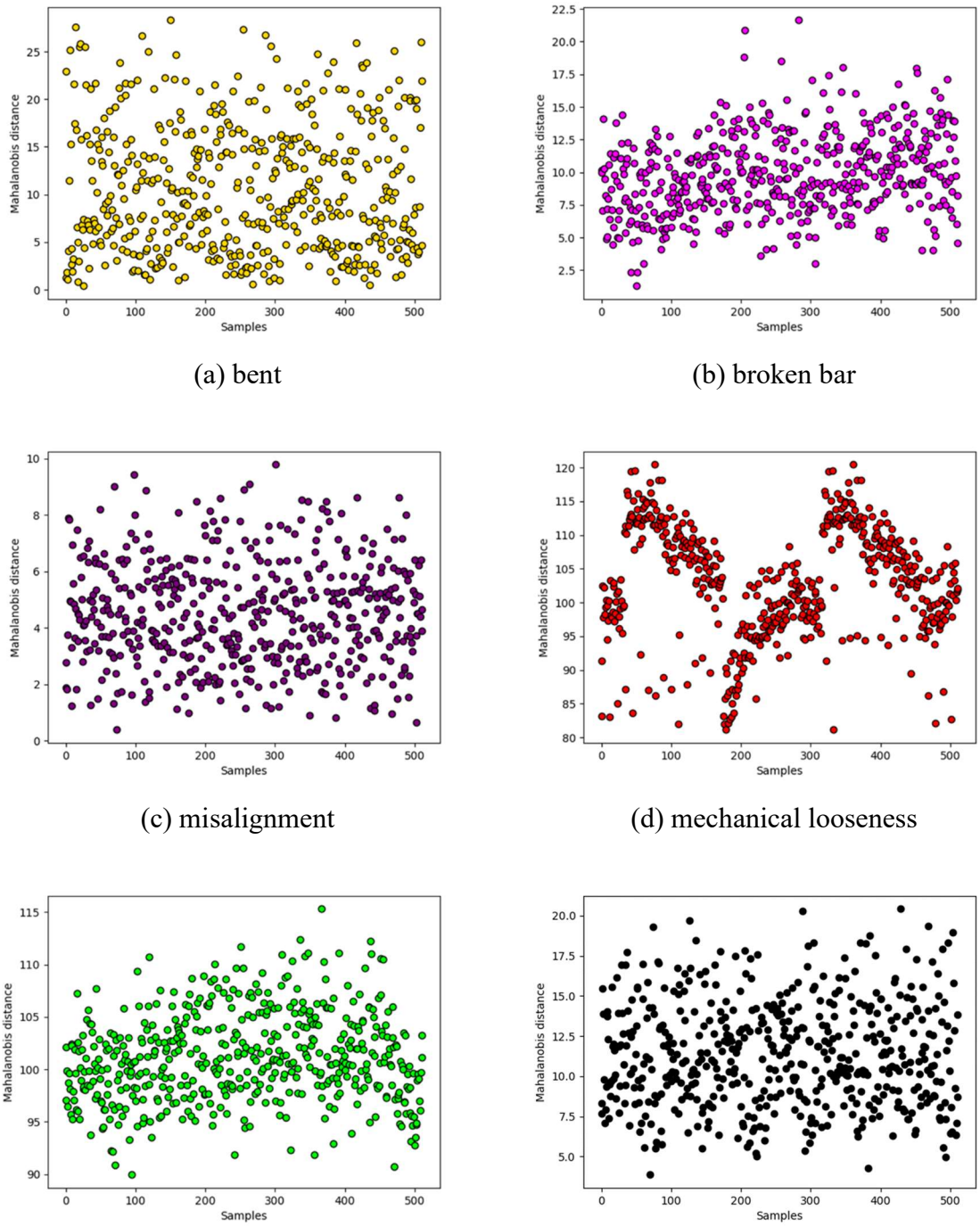

Figure 16. Mahalanobis distance for the different induced operational conditions (adapted from Ribeiro Junior, 2022).

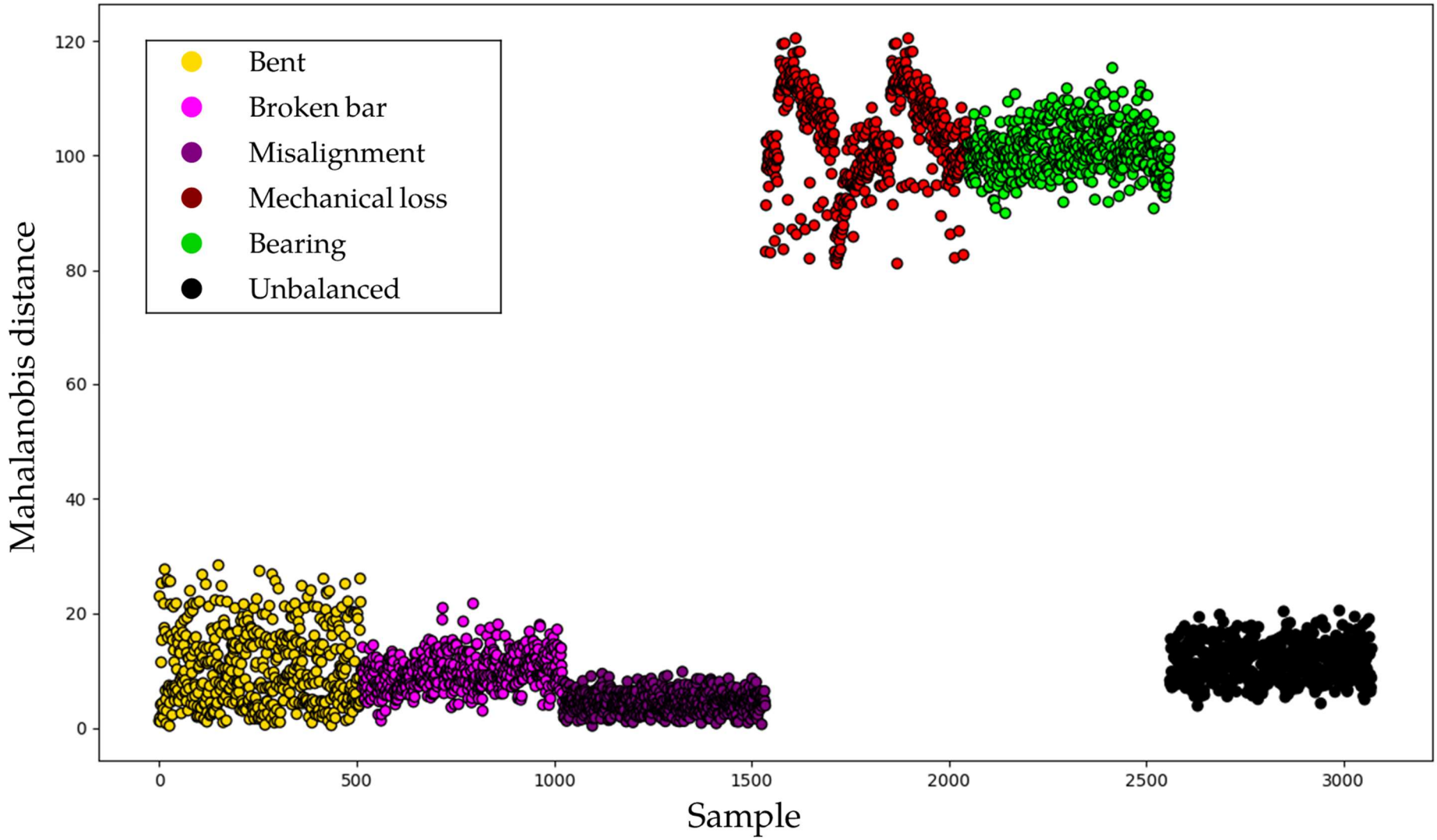

Figure 17. Mahalanobis distance for all the evaluated and induced operational condition (adapted from Ribeiro Junior, 2022).

4 Chapter Conclusion

In this chapter we present the theory of the Gaussian mixture model as well as its effectiveness in clustering and classifying data. In addition, the GMM theory was applied to a real problem. By applying the GMM to a real case, it was possible to demonstrate and discuss the strength of the applied method, as well as the challenges in using it. Also, as it is an unsupervised learning algorithm, this method does not need human support to carry out the identification of faults. Finally, we believe that with this chapter the readers have understood how the GMM works as well as ways to apply the method to other engineering problems.

Acknowledgements

The authors would like to acknowledge the financial support from the Brazilian agency CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – 150117/2021-3), and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais - APQ-00385-18). Also, the authors would like to acknowledge the support from PS Soluções, that provided the experimental laboratory to carry out this Chapter.

References

Abraman, A. situação da Manutenção no Brasil. In: Anais do 26º Congresso Brasileiro de Manutenção. Curitiba: Abraman, 2011.

Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.

Choudhary, A., Goyal, D., Shimi, S. L., & Akula, A. (2019). Condition monitoring and fault diagnosis of induction motors: A review. Archives of Computational Methods in Engineering, 26(4), 1221-1238.

Dellaert, F. (2002). The expectation maximization algorithm. Georgia Institute of Technology.

Duda, R. O., & Hart, P. E. (1973). Pattern classification and scene analysis (Vol. 3, pp. 731-739). New York: Wiley.

Hamill, P., Giordano, M., Ward, C., Giles, D., & Holben, B. (2016). An AERONET-based aerosol classification using the Mahalanobis distance. Atmospheric Environment, 140, 213-233.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

Hong, Y., Kim, M., Lee, H., Park, J. J., & Lee, D. (2019). Early fault diagnosis and classification of ball bearing using enhanced kurtogram and Gaussian mixture model. IEEE Transactions on Instrumentation and Measurement, 68(12), 4746-4755.

Kolmogorov, A. N., & Bharucha-Reid, A. T. (2018). Foundations of the theory of probability: Second English Edition. Courier Dover Publications.

McNicholas, P. D. (2016). Mixture model-based classification. Chapman and Hall/CRC.

Moon, T. K. (1996). The expectation-maximization algorithm. IEEE Signal processing magazine, 13(6), 47-60.

Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.

Myung, I. J. (2001). Computational approaches to model evaluation.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. Journal of mathematical Psychology, 47(1), 90-100.

Panić, B., Klemenc, J., & Nagode, M. (2020). Gaussian Mixture Model Based Classification Revisited: Application to the Bearing Fault Classification. Strojniski Vestnik/Journal of Mechanical Engineering, 66(4).

Pearson, K. (1894). Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London. A, 185, 71-110.

Plante, T., Stanley, L., Nejadpak, A., & Yang, C. X. (2016, September). Rotating machine fault detection using principal component analysis of vibration signal. In 2016 ieee autotestcon (pp. 1-7).

Profillidis, V. A., & Botzoris, G. N. (2018). Modeling of transport demand: Analyzing, calculating, and forecasting transport demand. Elsevier.

Randal, R.B. Vibration-based Condition Monitoring: Industrial, aerospace and automotive applications. John Wiley & Sons, 2011.

Rhys, H. (2020). Machine Learning with R, the tidyverse, and mlr. Simon and Schuster.

Ribeiro Junior, R. F., dos Santos Areias, I. A., Campos, M. M., Teixeira, C. E., da Silva, L. E. B., & Gomes, G. F. (2022). Fault Detection and Diagnosis in Electric Motors Using Convolution Neural Network and Short-Time Fourier Transform. Journal of Vibration Engineering & Technologies, 1-12.

Ribeiro Junior, R.F., de Almeida, F.A. & Gomes, G.F. Fault classification in three-phase motors based on vibration signal analysis and artificial neural networks. Neural Comput & Applic 32, 15171–15189 (2020).

Ribeiro Junior, R. F., Machine learning-based fault detection and diagnosis in electric motors (2022).

Rissanen, J. (1978). Modeling by shortest data description. Automatica, 14(5), 465-471.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.

Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 461-464.

Wang, G. F., Li, Y. B., & Luo, Z. G. (2009). Fault classification of rolling bearing based on reconstructed phase space and Gaussian mixture model. Journal of Sound and Vibration, 323(3-5), 1077-1089.

Wolfe, J. H. (1963). Object cluster analysis of social areas (Doctoral dissertation, University of California).

Wolfe, J. H. (1965). A computer program for the maximum likelihood analysis of types. Technical Bulletin 65-15, U.S. Naval Personnel Research Activity.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. Multivariate Behavioral Research 5, 329–350

Woźniak, M., Gałązka-Friedman, J., Duda, P., Jakubowska, M., Rzepecka, P., & Karwowski, Ł. (2019). Application of Mössbauer spectroscopy, multidimensional discriminant analysis, and Mahalanobis distance for classification of equilibrated ordinary chondrites. Meteoritics & Planetary Science, 54(8), 1828-1839.

Yan, H. C., Zhou, J. H., & Pang, C. K. (2017). Gaussian mixture model using semisupervised learning for probabilistic fault diagnosis under new data categories. IEEE Transactions on Instrumentation and Measurement, 66(4), 723-733.

Yang, M. S., Lai, C. Y., & Lin, C. Y. (2012). A robust EM clustering algorithm for Gaussian mixture models. Pattern Recognition, 45(11), 3950-3961.