



A traditional machine learning and deep learning approaches comparison for predictive maintenance.

**Roberto Oliveira Vita**



Dissertation

Master in Data Analytics



Supervised by

**João Gama** – Faculdade de Economia da Universidade do Porto



2019 – 2020



## Abstract

The data-driven models using machine learning and deep learning algorithms has been widely applied to predictive maintenance. Bearing is a critical component of rotating machinery and its failure is one of the most frequent reasons for machine breakdown. Vibration signals is often used for the diagnosis and the prognosis of bearings. Research studies are preferring deep learning to machine learning approaches to deal with the challenges of PdM applications related to the data characteristics and the demand for processing IoT data. This dissertation compares a traditional machine learning approach and deep learning model in PdM tasks considering the IEEE PHM 2012 prognostic challenge as application case. A model using statistical recurrent unit had similar performance than a model composed of principal component analysis and support vector regression. Additionally, a health stage assessment approach using DBSCAN and support vector machine failed to improve model's performance for RUL prediction.

**Key-words:** Bearing, Deep Learning, Diagnostic, Feature Extraction, Industry 4.0, Machine Learning, Predictive Maintenance, Prognostic, Remaining Useful Life, Vibration.

# List of Contents

|                                                              |                                     |
|--------------------------------------------------------------|-------------------------------------|
| Abstract.....                                                | i                                   |
| List of Contents .....                                       | ii                                  |
| List of Tables .....                                         | iv                                  |
| List of Figures .....                                        | v                                   |
| List of Contractions .....                                   | vii                                 |
| 1 Introduction .....                                         | 1                                   |
| 2 Predictive Maintenance Background.....                     | 3                                   |
| 2.1 PdM Process.....                                         | 3                                   |
| 2.2 Data Analytics for PdM.....                              | 8                                   |
| 2.2.1 Data Preparation .....                                 | 8                                   |
| 2.2.2 Modelling.....                                         | 13                                  |
| 2.3 New Industrial Environment and IoT - Deep Learning ..... | 17                                  |
| 2.3.1 Deep Learning applications using vibration data .....  | 18                                  |
| 3 PHM 2012 Challenge – Benchmark.....                        | 20                                  |
| 3.1 PHM 2012 Challenge Winner.....                           | 20                                  |
| 3.2 Recent Papers.....                                       | 20                                  |
| 3.3 Other related works.....                                 | 23                                  |
| 4 Proposed Models – Selected Methods.....                    | 25                                  |
| 4.1 Wavelets – Wavelet Packet Decomposition .....            | 25                                  |
| 4.2 Principal Components Analysis.....                       | 26                                  |
| 4.3 Autoencoder.....                                         | <b>Error! Bookmark not defined.</b> |
| 4.4 DBSCAN and Support Vector Machine.....                   | 27                                  |
| 4.5 Recurrent Neural Network.....                            | 28                                  |

|       |                                                                                                                                              |    |
|-------|----------------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.6   | Support Vector Regression.....                                                                                                               | 29 |
| 5     | Methodology .....                                                                                                                            | 30 |
| 5.1   | Research .....                                                                                                                               | 30 |
| 5.2   | Application.....                                                                                                                             | 30 |
| 5.3   | Case Study and Data Source.....                                                                                                              | 31 |
| 5.3.1 | Objectives and Experiments .....                                                                                                             | 32 |
| 5.3.2 | Evaluation metric.....                                                                                                                       | 33 |
| 5.3.3 | Tools.....                                                                                                                                   | 33 |
| 6     | Application .....                                                                                                                            | 34 |
| 6.1   | Signal Processing: Wavelet Packet Decomposition .....                                                                                        | 34 |
| 6.2   | Model 1 and its variants .....                                                                                                               | 35 |
| 6.2.1 | Health Indicator: Principal Component Analysis .....                                                                                         | 35 |
| 6.2.2 | Diagnosis: Health Stages - DBSCAN and Classifier .....                                                                                       | 37 |
| 6.2.3 | Remaining Useful Life: Support Vector Regression .....                                                                                       | 39 |
| 6.3   | Model 2.....                                                                                                                                 | 42 |
| 6.3.1 | Health Indicator: Recorrent Neural Network – Statistical Recurrent Unit ...                                                                  | 42 |
| 6.3.2 | Remaining Useful Life .....                                                                                                                  | 43 |
| 6.4   | Results and comparisons .....                                                                                                                | 44 |
| 6.4.1 | Question 1: Does the diagnosis procedure of health assessment improves models performance?.....                                              | 44 |
| 6.4.2 | Question 2: Does the model using deep learning techniques outperforms the model using traditional methods? Compare Model 1 with Model 2..... | 45 |
| 7     | Conclusion.....                                                                                                                              | 46 |
|       | References.....                                                                                                                              | 48 |
|       | Annex 1 – 3 Health Stages Analysis.....                                                                                                      | 54 |
|       | Annex 2 – Work Style.....                                                                                                                    | 55 |

## List of Tables

|                                                                                                                              |    |
|------------------------------------------------------------------------------------------------------------------------------|----|
| Table 1: Data Analytics sub processes in PdM domain.....                                                                     | 7  |
| Table 2: Summary of the metrics for evaluating prognostics HIs. Adapted from (Lei et al., 2018).....                         | 12 |
| Table 3: Deep Learning groups and models types. ....                                                                         | 17 |
| Table 4: Deep Learning works in PdM using vibration data. Quantity identified in (L. Zhang et al., 2019). ....               | 18 |
| Table 5: Last applied research with results for PHM 2012 challenge data set in SCOPUS ..                                     | 20 |
| Table 6: Proposed Models - Selected Methods .....                                                                            | 25 |
| Table 7: IEEE PHM 2012 challenge dataset. (Nectoux et al., 2012) .....                                                       | 31 |
| Table 8: Models .....                                                                                                        | 32 |
| Table 9: PCA – Features Eliminated - Kaiser, Meyer, Olkin Measure of Sampling Adequacy (KMO) .....                           | 35 |
| Table 10: PCA - features matrix correlation.....                                                                             | 35 |
| Table 11: DBSCAN Results - Quantity of normal and anomaly samples. ....                                                      | 37 |
| Table 12: SVM Classifier evaluation. Bearing 1_1 results in the left amd Bearing 1_2 in the right .....                      | 38 |
| Table 13: Summary of Scores .....                                                                                            | 44 |
| Table 14: DBSCAN Results - Quantity of normal and anomaly samples. Bearing 2_1 in the left and Bearing 2_2 in the right..... | 54 |

## List of Figures

|                                                                                                                                                                                                                                                                           |               |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| Figure 1: PdM Processes - Diagnostic and Prognostic. Adapted from (Sikorska, Hodkiewicz, & Ma, 2011).....                                                                                                                                                                 | 4             |
| Figure 2: P-F Curve (extracted from <a href="https://www.maintworld.com/Applications/Maximizing-the-P-F-Interval-Through-Condition-Based-Maintenance">https://www.maintworld.com/Applications/Maximizing-the-P-F-Interval-Through-Condition-Based-Maintenance</a> ) ..... | 5             |
| Figure 3: Health Stages - (a) One-stage, (b) Two-stages and (c) Three-stages (Lei et al., 2018). .....                                                                                                                                                                    | 6             |
| Figure 4: Summary of statistical Feature in time domain and frequency domain (Zhang et al., 2019). .....                                                                                                                                                                  | 10            |
| Figure 5: Health Indicators: (a) ideal HI and (b) not ideal HI. (L. Liao et al., 2016) .....                                                                                                                                                                              | 11            |
| Figure 6: Weibull Failure Rate Function (WFRF) - RMS, Kurtosis and RMSEE smoothing. Adapted from (Ben Ali et al., 2015).....                                                                                                                                              | 11            |
| Figure 7: Deep autoencoder feature extraction approach (Mishra et al., 2019) .....                                                                                                                                                                                        | 12            |
| Figure 8: Comparison of diagnostic techniques for rotating machinery. Adapted from (R. Liu et al., 2018).....                                                                                                                                                             | 15            |
| Figure 9: Difference between ML and end-to-end DL. (W. Zhang et al., 2019) .....                                                                                                                                                                                          | 18            |
| Figure 10: Wavelet Packet Decomposition ilustration. (G. Liu et al., 2019) .....                                                                                                                                                                                          | 26            |
| Figure 11: Convolutional auto-encoder (CAE) architecture. (Sadoughi et al., 2019) ....                                                                                                                                                                                    | <b>Error!</b> |
| <b>Bookmark not defined.</b>                                                                                                                                                                                                                                              |               |
| Figure 12: RNN - SRU operation and its behavior. (Ribeiro De Miranda et al., 2019) .....                                                                                                                                                                                  | 29            |
| Figure 13: (a) CRISP-DM Methodolgy and (b) PdM Process (Khan & Yairi, 2018).....                                                                                                                                                                                          | 30            |
| Figure 14: PRONOSTIA experiment bench. (Nectoux et al., 2012).....                                                                                                                                                                                                        | 31            |
| Figure 15: Scoring function of a RUL estimates according to its percent error. (Nectoux et al., 2012) .....                                                                                                                                                               | 33            |
| Figure 16: Raw signal, WPD and Smoothed RMS illustration. ....                                                                                                                                                                                                            | 34            |

|                                                                                                         |    |
|---------------------------------------------------------------------------------------------------------|----|
| Figure 17: PCA - % Variance Explained .....                                                             | 36 |
| Figure 18: Health Indicators - Bearing B2_1 .....                                                       | 36 |
| Figure 19: Health Indicators – Bearing B2_2.....                                                        | 37 |
| Figure 20: Clusters - Health Stages .....                                                               | 38 |
| Figure 21: RUL of bearing 2_3. In the right is the model that considers the HS identification.<br>..... | 39 |
| Figure 22: RUL of bearing 2_4. In the right is the model that considers the HS identification.<br>..... | 40 |
| Figure 23: RUL of bearing 2_5. In the right is the model that considers the HS identification.<br>..... | 40 |
| Figure 24: RUL of bearing 2_6. In the right is the model that considers the HS identification.<br>..... | 41 |
| Figure 25: RUL of bearing 2_7. In the right is the model that considers the HS identification.<br>..... | 41 |
| Figure 26: Boxplot: comparing model 1.1 and model 1.2 score performance.....                            | 45 |
| Figure 28: Clusters - Health Stages .....                                                               | 54 |

## **List of Contractions**

ANN - Artificial Neural Networks

AE - Auto Encoder

BRNN - Bidirectional Rnn

CBM - Condition Based Maintenance

CM - Condition Monitoring

CAE - Convolutional Auto-Encoder

CNN - Convolutional Neural Network

CRISP-DM - Cross-Industry Process For Data Mining

DL - Deep Learning

DE - Differential Evolution

EMD - Empirical Mode Decomposition

FMEA - Failure Modes And Effects Analysis

GRU - Gated Recurrent Units

GAN - Generative Adversarial Network

HI - Health Indicator

HS - Health Stage

IOT - Internet Of Things

LSTM - Long Short Term Memory

ML - Machine Learning

MLP - Multilayer Perceptron

NN - Neural Networks

P-F CURVE - Potential Failure To Failure Curve

PDM - Predictive Maintenance

PVM - Preventive Maintenance

PCA - Principal Component Analysis

PHM - Prognostic Health Management

RF - Random Forest

RELU - Rectified Linear Unit

RNN - Recurrent Neural Network

RUL - Remaining Useful Life

RBN - Restricted Boltzmann Machine

RMS - Root Mean Square

RMSEE - Root Mean Square Entropy Estimator

R2F - Run-To-Failure

SOM - Self-Organizing Feature Maps

STFT - Short Time Fourier Transform

SRU - Statistical Recurrent Unit

SVM - Support Vector Machine

SVR - Support Vector Regression

WPD - Wavelet Packet Decomposition

WT - Wavelet Transform

# 1 Introduction

Products in all kind of applications are subjected to a degradation process, which could happen in an early stage or in the end of the product's lifecycle. This degradation process depends on the product's material, on the operation's nature such as the period and form of usage, environment, amount of effort and stress placed and even on randomness. In industrial context, this process can be related the final product or equipment, and its consequences interferes customer satisfaction, processes operations and even safety conditions, all of them critical aspects for business survival.

As long as equipment reliability is a concern, different maintenance practices have been implemented in order to maintain an adequate level of the asset's operational readiness. These practices are delineated regarding three main strategies in maintenance management (Jardine, Lin, & Banjevic, 2006). The "Run-to-failure" (R2F) strategy considers maintenance activities only when breakdown occurs. The preventive maintenance (PvM) strategy is often based on life usage (time) or level of usage (cycles) of the asset. The predictive maintenance (PdM) strategy considers asset evidences such as temperature and cycle-time in order to decide when maintenance interventions are required. The PdM strategy is also known as Condition Based Maintenance (CBM), or as part of the Prognostic Health Management (PHM) because its maintenance policies are created based on condition information, which indicates the health of the monitored assets.

The selection of the best maintenance strategy regards particular context of each company. However, it is known that often R2F and PvM approaches imply in higher system unavailability and could represent a major expense on organizational budget and therefore the CBM has become a more efficient maintenance strategy (Jardine et al., 2006), mainly because it attempts to avoid unnecessary maintenance tasks by taking maintenance actions only when there is evidence of abnormal behaviors of a physical asset (Peng, Dong, & Zuo, 2010). Usually, the reactive maintenance costs that are a result of unplanned equipment failure come out to be six to eight times costlier than preventive or predictive maintenance. The scheduled maintenance procedures are mainly done based on the maintenance manual that the equipment manufacturer provides. However, the disadvantage of the preventive maintenance method is that the maintenance procedure scheduled specified by

manufacturers is fixed based on time and running hours and fails to consider the real-time degradation of the equipment's working condition (Kohli, 2018).

Moreover, CBM has received greater attention as improvements in the digital technologies have been developed, along with the adoption of a data based culture on companies most of them pushed by the Industry 4.0. Indeed, these improvements raised again the discussion and development of artificial intelligence in the context of PdM. "Data-driven methods have become the most effective solution to address smart manufacturing and industrial big data, especially for performing health perception" (W. Zhang, Yang, & Wang, 2019).

In line with data science methodologies, CBM consists of three main steps: data acquisition, data processing and decision-making (Jardine et al., 2006). In addition, It is a proactive process, which requires the development of a predictive model that helps maintenance managers to plan better economical maintenance schedule by providing useful information such as the remaining useful life (RUL) and reliability of equipment and even trigger alarm when maintenance is required (Peng et al., 2010)

In the middle of several types of equipment, the rolling element bearings can be highlighted as a very common component, present in almost all rotative machine. More than 40% of motor failures are related to bearing faults. Many effective methods have been proposed and applied for bearing CBM maintenance strategies, such as vibration analysis, temperature analysis and oil analysis. Among all these methods, vibration analysis is the most convincing one and numerous indicators derived from vibration signals are implemented as criteria in traditional fault diagnosis. (B. Zhang, Zhang, & Li, 2019)

The present dissertation aims to study the state-of-art of machine learning and deep learning in the context of predictive maintenance by considering a literature review and a case study for application purpose. The case study considers the 2012 IEEE PHM Challenge problem of bearing component for PdM application.

This project proposal is divided into seven sections. The present section is dedicated to introduce the subject, motivation and contents of the work. Section 2, Section 3 and Section 4 present the literature review regarding PdM processing and Machine Learning & Deep Learning techniques. Section 5 describes the methodology and the presents the case application. Section 6 describes the application case using the PHM 2012 challenge data set. The last section presents the conclusions and future research.

## **2 Predictive Maintenance Background**

This chapter includes a literature review of pertinent material about predictive maintenance and data analytics in the context of condition based maintenance. The main databases used in this research were Scopus and “Web of Science”.

Firstly, this section aims to understand the concepts, principles and applications of predictive maintenance. Secondly, a review of the main methodologies and activities in data analytics in the context of predictive maintenance were studied in order to prepare the basis for a practical application in a following section of this research.

The literature about PdM is abundant and covers diverse aspects of management fields. There is a huge amount of quantitative analysis’s studies in the context of CBM, most of them focused on the statistical perspective. The machine learning and deep learning approaches are not recent, however have lately received a lot of attention with the boom of digital technologies implementation and the Industry 4.0 movement.

### **2.1 PdM Process**

Predictive maintenance, positioned in the dissertation as a condition based maintenance, is a maintenance strategy that monitors equipment data in real time in order to make maintenance decision based on condition monitoring information (Vachtsevanos, Lewis, Hess, & Wu, 2006). CBM methods take advantage of measured characteristics of a particular system or component in order to assess health conditions and make predictions (Schwabacher & Goebel, 2007). Health prognostics is one of the major tasks in CBM, which aims to predict the remaining useful life of machinery based on the historical and on-going degradation trends (Lei et al., 2018). The condition of a system is quantified by obtaining data from one or several data sources, periodically or continually (Peng et al., 2010).

CBM attempts to avoid unnecessary maintenance tasks by taking maintenance actions only when there is evidence of abnormal behaviors. It is a proactive process that requires the development of a analytical model able to trigger alarm or actions for corresponding maintenance (Peng et al., 2010). It is an effective strategy to reduce maintenance cost and improving the reliability of machinery.

In the case of R2F policies, the data related to a maintenance cycle (the production activity between two successive failure events) are available and therefore supervised approaches can readily be adopted. When PvM policies are currently in place, the full maintenance cycle may not be observable given the fact that maintenance is generally performed in advance of any potential failure, and hence only unsupervised learning approaches are feasible. (Susto, Schirru, Pampuri, McLoone, & Beghi, 2015). In both cases, condition monitoring (CM) data is an important source of information. Here the CM data is defined in a broad sense of any data, which may have a connection with the fault identification and to the estimation of the RUL such as sensors, operational data, performance data, environmental information, and degradation signals. Such data are very versatile, such as vibration, oil analysis, temperature, pressure, humidity, loading of equipment, weather data, etc (Si, Wang, Hu, & Zhou, 2011).

The CBM program presents two main abilities: diagnostics and prognostics, see Figure 1. The diagnostics focus is on three different sub processes: fault detection, fault isolation and fault identification. It aims to identify when, where and why a system presents the abnormal condition, respectively. Diagnostics involves identifying and quantifying the damage that has occurred and is thus retrospective in nature and the prognostics aims to predict the damage that is yet to occur, the occurrence of a fault or a degradation process (Sikorska, Hodkiewicz, & Ma, 2011). In addition, CBM can even consider the prescriptive ability by suggesting specific actions when facing the occurrence of such events (Schwabacher & Goebel, 2007).

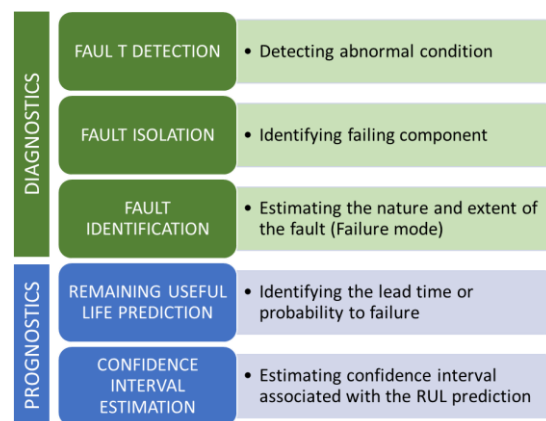


Figure 1: PdM Processes - Diagnostic and Prognostic. Adapted from (Sikorska, Hodkiewicz, & Ma, 2011).

The literature presents the prognostic process as dependent of the diagnostic process. Prognostics uses diagnostic results to assess the current health state of a system and also uses some outputs such as the fault indicator and degradation rate to model and predict the fault

progression and calculate the remaining useful life (Vichare & Pecht, 2006). Although the prognostic's process adds more value to an organization compared to the diagnostic one, it is not always possible to implement it because some faults or failures are not predictable (Jardine et al., 2006) or because the available data do not allow this kind of development.

To better understand the diagnostics and prognostics processes, it is desirable to perceive the behaviour of the monitored system from a healthy state until failure occurs (Sikorska et al., 2011). It involves two main concepts: the failure mode and the degradation process. The failure mode refers to a potential cause of a failure. It is important because firstly the PdM activities are organized to cope with a pre-defined type of failure and because each failure mode may have different triggers to initiate the failure (Susto et al., 2015) and henceforth have different patterns of degradation (Sikorska et al., 2011). The most common methodology to investigate and learn about failures is the Failure Modes and Effects Analysis (FMEA), which analyse the causes of failures, their frequency and impacts. The FMEA study, developed in-sync with domain experts, is crucial to prioritize failures and to identify the most relevant condition variables to be monitored for the PdM application (Vichare & Pecht, 2006). The failures patterns are most analysed on the reliability centred maintenance by using the “potential failure to failure” curve (P-F curve), which graphically represents the condition variable behaviour during the interval between a potential failure and the real functional failure (Jardine et al., 2006) see Figure 2.

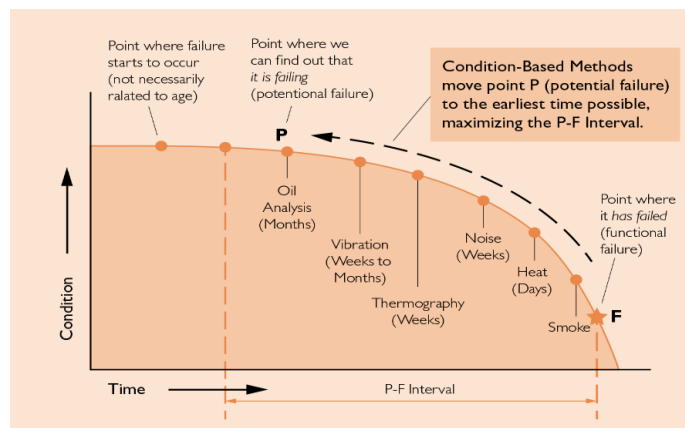


Figure 2: P-F Curve (extracted from <https://www.maintworld.com/Applications/Maximizing-the-P-F-Interval-Through-Condition-Based-Maintenance>)

With this perspective, (Lei et al., 2018) highlights two concepts: health indicator (HI) and health stage (HS), and incorporate the HI construction and the HS identification as steps in

the general methodology of data science for PdM. From the monitored data, Health Indicators are constructed using signal processing and/or artificial intelligent techniques in order to capture the health condition of an equipment and then to expose its degradation process. By analyzing the Health Indicator during lifetime of an equipment, it is possible to identify different stages of health condition, the health stages, and the degradation process. Depending on the equipment, the failure mode and HI construction technique, the HS can be unique with a continuous degradation rate or can be divided into two or more different HS according to the change of their degradation trends, see Figure 3. Note that different models can be assigned to each HS and the RUL prediction is traditionally based on the HS where the HI present an ideal degradation trend until the failure (Lei et al., 2018).

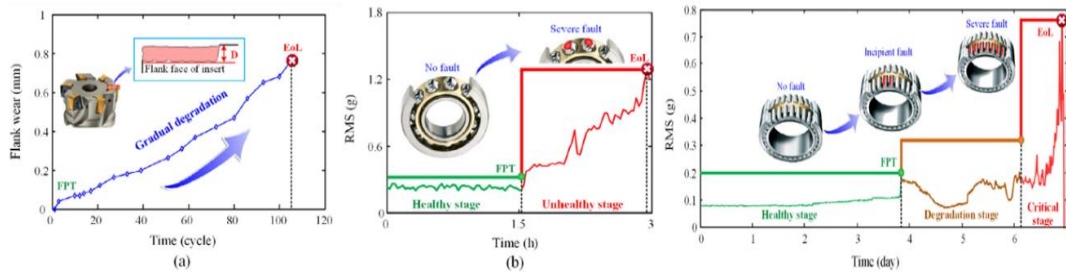


Figure 3: Health Stages - (a) One-stage, (b) Two-stages and (c) Three-stages (Lei et al., 2018).

The most recent PdM techniques are data-driven methods and consider multivariable problems. Although, Sikorska et al. (2011) by following the ISO13381-1 standard suggests that all approaches to implement prognostics applications must answer 6 main questions:

- (a) Is a component in the degraded state?
- (b) Which failure mode has initiated the degradation?
- (c) How severe is the degradation (i.e. 'where' the component is on the degradation curve)?
- (d) How quickly is degradation expected to progress from current state to functional failure?
- (e) What novel events will change (e.g. accelerate, retard) this expected degradation?
- (f) How may other factors (e.g. the type of model, noise) affect our estimate of RUL?

In line with data science methodologies, CBM consists of three main steps: data acquisition, data processing and decision-making (Jardine et al., 2006). The Table 1 summarizes the topics in data analytics activities involved in the PdM process. It considers the steps presented by

Lei et al. (2018): 1) Data Acquisition, 2) Health Indicator construction, 3) Health Stage division and 4) RUL prediction; and by W. Zhang et al. (2019): 1) data acquisition and pre-processing, 2) feature engineering, and 3) model training and predicting.

|  |                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <p><b>1) Data acquisition and pre-processing</b></p> <ul style="list-style-type: none"> <li>• Event Data and Condition Monitoring Data</li> <li>• Value Type, Waveform or Multidimensional Data (Jardine et al., 2006)</li> <li>• Collection sequence (real-time, frequency, samples) and resolution (Baruah and Chinnam, 2005)</li> <li>• Single sensory or multisensory (Lei et al., 2018)</li> <li>• Normalization</li> </ul> |
|  | <p><b>2) Feature engineering, extraction, concatenation, and selection (HI construction)</b></p> <ul style="list-style-type: none"> <li>• Signal Processing ((Z. Liu, Zuo, &amp; Qin, 2016)</li> <li>• Statistical Features (W. Zhang et al., 2019)</li> <li>• Dimensionality Reduction</li> <li>• Features Selection</li> <li>• Criteria to Evaluate HI (Lei et al., 2018)</li> </ul>                                           |
|  | <p><b>3) Modelling: Training, Detecting and Predicting</b></p> <ul style="list-style-type: none"> <li>• Diagnostic: Un/Semi-supervised models and Classification Problem (Pecht, 2009)</li> <li>• Prognostic: Regression Problem (Pecht, 2009) and Similarity-based models (XXXX)</li> <li>• Prognostic: Remaining useful life (RUL) or Probability of Failure (Jardine et al., 2006)</li> </ul>                                 |

Table 1: Data Analytics sub processes in PdM domain.

Lei et al. (2018) describe a general equipment health prognostic procedure initiating by gathering measured data, such as vibration signals and/or temperature from sensors in order to monitor the condition of the equipment. These monitoring signals contain lots of information as well as measurement noise. To transform them in representative index of health condition, some HIs are constructed by using signal processing, machine learning and deep learning techniques. The HI behavior represents the varying degradation evolution that allows to partition the equipment lifetime in suitable health stages. Finally, the HS, degradation trend, and the RUL can be estimated (Lei et al., 2018).

## **2.2 Data Analytics for PdM**

Due to the diversity of maintenance issues, PdM developments are often fit to a particular problem (Susto et al., 2015) and thus present a vast set of combination of data analytics techniques that can be applied to PdM applications.

The applications of PdM in the literature vary regarding its purpose, (diagnostics or prognostics), and also depends on the nature of the data, the collection method (Si et al., 2011) and the source of data such as from Enterprise Resource Planning systems (Kohli, 2018), or Supervisory Control and Data Acquisition devices (Gonzalez, Reder, & Melero, 2016), or IoT sensors. It is also possible to create PdM applications based on already built-in alarm system of equipment's Distribute Control System (Zhu, Wang, Li, Gao, & Zhao, 2016).

Jardine et al. (2006) classify the PdM data as event data and condition monitoring data. The event data refers to information of what happened (e.g., installation, breakdown, etc.) and/or what was done (e.g., minor repair, preventive maintenance, oil change, etc.) to the monitored asset. Condition monitoring data are the measurements related to the health condition/state of the asset and is subdivided into three categories:

- Value type: Data collected at a specific time epoch as a discrete value. For example: oil analysis data, temperature and humidity are all value type data.
- Waveform type: Data collected at a specific time epoch for as a time series. For example, vibration data and acoustic data are waveform type.
- Multidimensional type: Data collected at a specific time epoch as a multidimensional data. The most common multidimensional data are image data.

### **2.2.1 Data Preparation**

#### **2.2.1.1 Data Processing**

Different type of data processing technique can be applied depending on the application purpose and on the data type. Data analysis for event data is commonly related with probability approaches, which study the time between events also known as reliability analysis. When considering the CM data, the value type data analysis focus on their structure correlations. Multivariate analysis techniques, mainly dimensionality reduction ones, such as

principal component analysis (PCA) and independent component analysis or trend analysis techniques such as regression analysis and time series model are commonly used to process value type data. Data processing for waveform and multidimensional data is also called signal processing. Various signal processing techniques have been developed to analyse and interpret waveform and multidimensional data to extract useful information for further diagnostic and prognostic purpose, which can be divided in time domain, frequency domain or time-frequency domain analysis (Jardine et al., 2006).

The main idea of time series modelling, in time-domain analysis, is to fit the waveform data to a parametric time series model and extract features based on this parametric model. The popular models used in the literature are the autoregressive model and the autoregressive moving average model (Jardine et al., 2006). Frequency-domain analysis is based on the transformed signal in frequency domain. The most widely used conventional analysis is the spectrum analysis by means of Fast Fourier Transform. The most commonly used tool in spectrum analysis is power spectrum (Jardine et al., 2006). One limitation of frequency-domain analysis is its inability to handle non-stationary waveform signals, which are very common when machinery faults occur. Indeed, Time domain or frequency domain processing are not able to deal with non-stationary signals. For real world PdM applications it is preferable to work with time-frequency domain, once it considers the frequency over the time. (G. Liu, Zhao, Li, & Zhang, 2019). For time-frequency domain, it is common to use Short Time Fourier Transform (STFT), Wavelet Transform (WT) or Empirical Mode Decomposition (EMD) (L. Zhang et al., 2019). In the Verstraete, Ferrada, Droguett, Meruane, and Modarres (2017) study, a Convolutional Neural Network model using WT outperformed the same model with STFT and EMD techniques.

The data transformation methods of waveform and multidimensional data often end up by using statistical feature as extracted variables to be used in the next data processing task or in modelling techniques. W. Zhang et al. (2019) summarize the most common statistical features used in the time domain and frequency domain works, see Figure 4.

| No. | Name               | Equation                                                                                                              | No. | Name             | Equation                                                         | No. | Name                         | Equation                                                              |
|-----|--------------------|-----------------------------------------------------------------------------------------------------------------------|-----|------------------|------------------------------------------------------------------|-----|------------------------------|-----------------------------------------------------------------------|
| 1   | Maximum            | $s_1 = \max(x_i)$                                                                                                     | 8   | Mean square      | $s_8 = \frac{1}{N} \sum_{i=1}^N x_i^2$                           | 15  | Variance frequency           | $F_2 = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^2$                    |
| 2   | Minimum            | $s_2 = \min(x_i)$                                                                                                     | 9   | Root mean square | $s_9 = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$                    | 16  | Standard deviation frequency | $F_3 = \sqrt{\frac{\sum_{j=1}^N ((f_j - F_1)X_j)}{\sum_{j=1}^N X_j}}$ |
| 3   | Median             | $s_3 = \begin{cases} x_{(N+1)/2}, N \text{ is odd} \\ \frac{x_{N/2} + x_{(N+1)/2}}{2}, N \text{ is even} \end{cases}$ | 10  | Skewness         | $s_{10} = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - s_5)^3}{s_7^3}$  | 17  | Root mean square frequency   | $F_4 = \sqrt{\frac{\sum_{j=1}^N (f_j^2 X_j)}{\sum_{j=1}^N X_j}}$      |
| 4   | Peak-to-peak       | $s_4 = \max x_i  - \min x_i $                                                                                         | 11  | Kurtosis         | $s_{11} = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - s_5)^4}{s_7^4}$  | 18  | Spectral skewness            | $F_5 = \sum_{j=1}^N \left( \frac{f_j - F_1}{F_3} \right)^3 S(f_j)$    |
| 5   | Mean               | $s_5 = \frac{1}{N} \sum_{i=1}^N x_i$                                                                                  | 12  | Skewness factor  | $s_{12} = \frac{1}{N} \sum_{i=1}^N \frac{x_i^3}{(\sqrt{s_8})^3}$ | 19  | Spectral kurtosis            | $F_6 = \sum_{j=1}^N \left( \frac{f_j - F_1}{F_3} \right)^4 S(f_j)$    |
| 6   | Variance           | $s_6 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$                                                                    | 13  | Kurtosis factor  | $s_{13} = \frac{1}{N} \sum_{i=1}^N \frac{x_i^4}{(\sqrt{s_8})^4}$ | 20  | Spectral power               | $F_7 = \sum_{j=1}^N (f_j)^3 S(f_j)$                                   |
| 7   | Standard deviation | $s_7 = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$                                                             | 14  | Mean frequency   | $F_1 = \frac{1}{N} \sum_{j=1}^N f_j$                             | 21  | Wavelet energy               | $F_8 = \sum_{j=1}^N \omega t_b^2(j)/N$                                |

Figure 4: Summary of statistical Feature in time domain and frequency domain (Zhang et al., 2019).

Regarding diagnostic and prognostic analysis for bearing applications, the root mean square (RMS) and Kurtosis features are the most used. However, in real applications, these statistical features present relevant fluctuations, which implies inadequacy for prognosis (Ben Ali, Chebel-Morello, Saidi, Malinowski, & Fnaiech, 2015). In order to improve the quality of predictive approaches, authors use to propose new methods to find better indicators for bearing degradations. For example Sutrisno, Oh, Vasan, and Pecht (2012) developed a signal processing method that took advantage of both time and frequency domain features and Ben Ali et al. (2015) used the RMS Entropy Estimator (RMSEE). Additionally, RUL prediction based on combined traditional and cumulative indexes is better, such as using accumulative features in the time-domain (C. Liu & Zhang, 2019) and frequency-domain (Cui, Ren, Wang, & Zhang, 2019).

#### 2.2.1.2 Feature extraction, dimensionality reduction and feature selection

By processing these data, it is expected to extract useful information from raw signals. Indeed, the preparation tasks is not only concerned to transform the raw data into adequate formats but also to extract and select useful features to represent the equipment health (W. Zhang et al., 2019), the “health indicator”.

The HI construction is a critical and challenge task in PdM, because it refers to a quantitative indicator of an equipment’s status associated with a specific maintenance issue (Susto et al., 2015). The HI influences the predictive model complexity and the prediction performance (Lei et al., 2018). Although many extracted features present no trend until a moment very

close to failure, it is desirable to compute a good monotonic trend feature to achieve higher performances (L. Liao, Jin, & Pavel, 2016), see in Figure 5 (a) ideal HI and (b) not ideal HI.

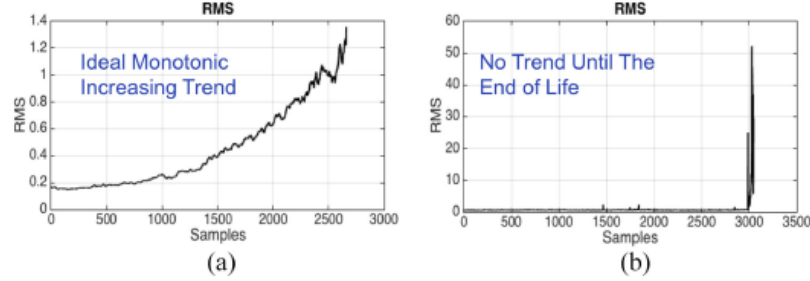


Figure 5: Health Indicators: (a) ideal HI and (b) not ideal HI. (L. Liao et al., 2016)

In order to reduce fluctuations areas, Ben Ali et al. (2015) presents an approach for smoothing the extracted RMS, Kurtosis and RMSEE measurement from bearing operation by fitting them to a modified Weibull failure rate function called Universal failure rate function. Figure 6 shows the fitting of the extracted features.

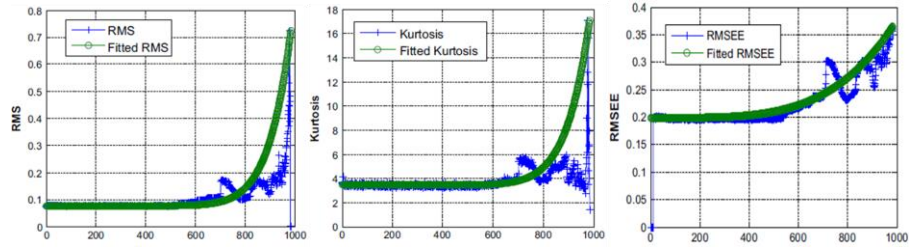


Figure 6: Weibull Failure Rate Function (WFRF) - RMS, Kurtosis and RMSEE smoothing. Adapted from (Ben Ali et al., 2015)

Lei et al. (2018) summarize critical criteria and metrics for evaluating HIs on prognostics applications, see Table 2. The main theoretical basis for these criteria refer that in real applications, the equipment's degradation processes are irreversible, (a fault component is unable to recover without repair), indicating a Monotonicity criteria. The degradation process and operation conditions add noisy to the HI (Robustness). It is expected a correlation between degradation trend of a HI and the operating time (Trendability). Usually, equipment presents different HSs during the lifetime (Identifiability). Usually, one metric is not enough to select a suitable HI for the task of RUL prediction. Operators may hope to select HIs which hold various properties or make a trade-off among different properties. Z. Liu et al. (2016) proposed a hybrid metric for selecting prognostic HIs from the properties of both identifiability and consistency.

| Categories                                          | Names                             |
|-----------------------------------------------------|-----------------------------------|
| Metrics depending on a single HI                    | • Monotonicity                    |
| Metrics depending on a HI and time                  | • Robustness<br>• Trendability    |
| Metrics depending on a HI and stage sequence        | • Identifiability                 |
| Metrics depending on multiple HIs<br>Hybrid metrics | • Consistency<br>• Hybrid metrics |

Table 2: Summary of the metrics for evaluating prognostics HIs. Adapted from (Lei et al., 2018).

Data-driven methodology for HI constructions is based upon statistical and learning techniques, most of which originated from the theory of pattern recognition. Statistical approaches include multivariate statistical methods (e.g., static and dynamic principal components analysis, linear and quadratic discriminant, partial least square, canonical variety analysis, and learning vector quantization), state space models (e.g., Bayesian networks, hidden Markov models, and hidden semi-Markov models)), and regressive models. The AI approaches are most represented by ANN and its variants (e.g., polynomial neural networks, dynamic wavelet neural networks, self-organizing feature maps (SOM), multilayer perceptron (MLP) neural network) (Peng et al., 2010). The majority of applications focus on developing a unidimensional HI. It can be a drawback when dealing with multiple failure modes (Hinchi & Tkiouat, 2018).

Besides the traditional process to extract features, recent studies are most relying on the AI to construct HI in autonomous approach. Mishra, Krogerus, and Huhtala (2019) developed a generic model for automated feature extraction and fault detection in the health state monitoring of elevator systems, using a five layer deep autoencoder (Figure 7) including input, output, encoder, decoder and representation layers. They argue that the automated feature extraction approach does not require any prior domain knowledge, and provides dimensionality reduction and is robust against overfitting characteristics.

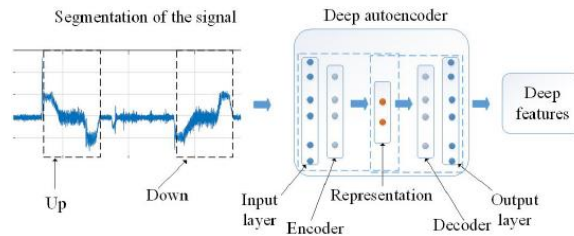


Fig. 2. Deep autoencoder feature extraction approach

Figure 7: Deep autoencoder feature extraction approach (Mishra et al., 2019)

Shao, Jiang, Wang, and Zhao (2017) highlight the strength of Auto Encoder (AE) in feature extraction and feature learning. The paper describes how to further improve the feature learning ability with reduced influence of background noises by stacking Deep AE (noise reduction) and contractive AE (enhanced feature recognition), called deep feature fusion method (Rieger, Regier, Stengel, & Clarke, 2019).

### **2.2.2 Modelling**

PdM approaches can be classified into three categories: 1) model-based prognosis; 2) knowledge-based prognosis; and 3) data-driven prognosis. The present dissertation is focused on the data-driven approach, which has been applied to industrial manufacturing mainly by using machine learning (ML) and deep learning algorithms (W. Zhang et al., 2019).

Considering the availability of the target variable, which can be equipment-status or quality or performance indicators, different situations can occur: (i) complete labels, (ii) partially missing labels and (iii) missing labels (Zschech, Heinrich, Bink, & Neufeld, 2019). This situation indicates which type of model is suitable to be applied. Supervised learning is dependent of the availability of historic data containing the labelling of dependent variables. In the absence of target data, the unsupervised techniques are applied.

#### **2.2.2.1 Diagnosis**

When dealing with diagnosis process, the most common approaches rely on unsupervised methods (such as clustering) or supervised methods (more specifically, classification methods), once the main objective is to distinguish the normal and abnormal conditions of a system or to identify its health stages. The most used target variable is the maintenance intervention data that can indicate when a maintenance intervention was necessary. Alternatively, one can use the judgement of experts to indicate the different health stages, or even use active learning techniques to self-label the data.

Machine fault diagnostics, also known by pattern recognition, is a procedure of mapping the information obtained in the available data to machine faults analysis. This task can be performed manually with data analysis and auxiliary graphical tools. Therefore, automatic pattern recognition is also a reality. This can be achieved by classification of signals based on the information and/or features extracted from the signals (Jardine et al., 2006). Reddy, Sarkar, Venugopalan, and Giering (2016) use a multi modal deep encoder for anomaly

detection in aircraft data. Ben Ali et al. (2015) divided the bearing degradation processes into 7 health stages by analyzing the changes RMS, Kustosis and RMSEE features. AI classifiers are also applied to multi-stage identification, such as Artificial Neural Networks, Support Vector Machine, and neural fuzzy (Lei et al., 2018).

In the following sections, different machine fault diagnostic approaches are discussed.

#### **2.2.2.1.1 Diagnosis: Unsupervised and Semi-Supervised Methods**

Unsupervised methods aim to explore and detect hidden patterns in data and are most used in anomaly detection perspectives. It is also possible to use them to create artificial labels to be used in supervised algorithms. Rather than assuming a closed world environment that comprises a fixed number of classes, modern pattern recognition systems need to recognize outliers, that is, identify anomalies or discover entirely new classes that are not currently part of the established model (BAHIA).

The anomaly or novelty detection methods are useful classifiers for this kind of applications and are used when the unknown event is considered rare compared to the normal condition. Currently applications are able to update automatically, by incorporating new events as normal condition in a dynamic way (Ding, Li, Belatreche, & Maguire, 2014).

Examples of PdM applications using some of these techniques can be found in Amruthnath and Gupta (2018) where a combined model using PCA, Cluster Analysis (K-Means and C-Means Clustering) and Gaussian mixture model. Baruah and Chinnam (2005) combine PCA for dimensionality reduction and clustering technique to identify and predict health stages. L. Liao et al. (2016) applied self-organizing maps for assessing health condition in a prognostic application of a bearing component.

This type of technique is the most used in real world due the lack of labeling data. Despite of monitoring data availability, the correct failure point is normally missed (Cui et al., 2019).

#### **2.2.2.1.2 Diagnostis: Supervised (Classification)**

Carvalho et al. (2019) identified the most used traditional ML algorithms in PdM for classification applications: logistic regression, support vector machine (SVM), decision tree, random forest (RF) and neural networks (NN). Canizo, Onieva, Conde, Charramendieta, and Trujillo (2017) employed RF to generate predictive models for monitoring of wind

turbines. To do this, status data (activated and deactivated alarms) and operational data (about the performance of the wind turbines) are employed to design the RF model. Su and Huang (2018) propose a realtime predictive fault detection system, called “HDPass”, to perform hard disk drive faults (using RF). Praveenkumar, Saimurugan, Krishnakumar, and Ramachandran (2014) propose a SVM model to identify failures in automotive transmission boxes.

R. Liu, Yang, Zio, and Chen (2018) surveyed diverse AI techniques for rotating machinery diagnosis, summarizing the theoretical background and industrial application points of view regarding k-NN-based, Naive Bayes-based, SVM-based, ANN-based and deep learning-based techniques, see Figure 8. They suggest that AI techniques will continue to be attractive and powerful for rotating machinery fault diagnosis.

List of AI algorithms and their advantages and limitations.

| Algorithm     | Advantages                                                                                                                                                                    | Limitations                                                                                                       |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| k-NN          | 1. Mature theory and easy to implement<br>2. Can be used for both classification and regression                                                                               | 1. Large computation<br>2. Needs lots of storage space<br>3. The selection of $k$ influence rightarrow too much   |
| Naive Bayes   | 1. Robust to missing values<br>2. Requires little storage space<br>3. Good physical explanation ability                                                                       | 1. Strong prior assumptions<br>2. Combinatorial explosion and computation problem<br>3. Need of prior probability |
| SVM           | 1. High classification accuracy<br>2. Can deal with high-dimensional features                                                                                                 | 1. Low efficiency for big data<br>2. No physical meaning                                                          |
| ANN           | 1. High classification accuracy<br>2. Good approximation of complex nonlinear function                                                                                        | 1. Many parameters and easy to over-fitting<br>2. No physical meaning<br>3. Training procedure cannot be seen     |
| Deep learning | 1. Learning features and recognizing fault automatically<br>2. Learning more complex structure from data due to the deep architecture<br>3. Do not need the feature extractor | 1. Needs of large samples<br>2. No physical meaning<br>3. Training for long time                                  |

Performance comparison of AI algorithms.

|                          | ANN  | SVM  | k-NN | Naive Bayes | Deep learning |
|--------------------------|------|------|------|-------------|---------------|
| Accuracy in general      | ***  | **** | **   | *           | ****          |
| Classification speed     | **** | **** | *    | ****        | **            |
| Robustness to noise      | **   | **   | *    | ***         | ****          |
| Dealing with overfitting | *    | **   | ***  | ***         | ***           |
| Physical explanation     | *    | *    | ***  | ****        | *             |
| Robustness to parameters | *    | *    | ***  | ****        | **            |

Figure 8: Comparison of diagnostic techniques for rotating machinery. Adapted from (R. Liu et al., 2018).

### 2.2.2.2 Prognostics

The prognostics processes focus on the prediction of the future health condition, which is often indicated by the expected remaining useful life of a system’s component, or by the failure probability given a future interval of time. It relies on supervised models that often uses the time-to-failure indicator as the label.

There are two main prediction types in prognostics. Both are based on the current system condition and past operation profile but generate a different type of output. The first most widely approach aims at predicting how much time is left before a failure (RUL). The second has the objective of estimating the probability that a system performs without a fault or a failure up a specific interval of time (e.g., until the next inspection) (Jardine et al., 2006).

The statistical model-based approaches, also named as empirical model-based approaches, estimate the RUL of machinery by establishing statistical models based on empirical knowledge, and generally present the RUL prediction result as a conditional PDF depending on observations. In these approaches, RUL prediction models are constructed by fitting available observations into random coefficient models or stochastic process models under a probabilistic method, without relying on any physics or principles. The main techniques refer to AR models, Random coefficient models, Wiener process models, Gamma process models, Inverse Gaussian process models, Markov models and Proportional hazards models. Fang, Zhou, and Gebraeel (2015) proposed a RUL prediction approach based on a Wiener process model in cases of missing data. Park and Padgett (2005) developed a gamma process model based on a general cumulative damage approach. H. Liao, Zhao, and Guo (2006) employed a proportional hazard model to predict the RUL of bearings. Similar to Wiener process models, Gamma process models are able to describe the temporal variability of degradation processes. To improve the flexibility of hidden Markov models for representing complicated state transition processes, the hidden semi-Markov model was also applied in the field of machinery prognostics (Baruah & Chinnam, 2005).

ANN, fuzzy logic systems, support vector regression (SVR), and Gaussian processes for regression purposes are the technique highlighted for AI approach (Lei et al., 2018). Xiao, Chen, Zhang, and Liu (2017) used a FFNN to conduct multi-step ahead prediction for the bearing health states. Benkedjouh, Medjaher, Zerhouni, and Rechak (2013) used the SVR to map the HIs into nonlinear regressions for RUL prediction of machinery. Fumeo, Oneto, and Anguita (2015) developed an online SVR model for the RUL prediction of bearings by optimizing the trade-off between the accuracy and the computing efficiency.

## 2.3 New Industrial Environment and IoT - Deep Learning

Rieger et al. (2019) published a review of adequate technique for dealing with the new Industrial environment with internet of things (IoT) applications demands: self-adaptive algorithms able to continuously learning and improving their models in a real-time manner, considering continuous flowing data streams and particularities of data reording. They grouped the list of Deep Learning (DL) models mentioned by Mohammadi, Al-Fuqaha, Sorour, and Guizani (2018), regarding IoT application:

| <b>Generative approaches</b>                         | <b>Discriminative approaches</b>                | <b>Hybrid</b>                        |
|------------------------------------------------------|-------------------------------------------------|--------------------------------------|
| Auto-encoder (AE) and its variants                   | Recurrent Neural Network (RNN) and its variants | Generative Adversarial Network (GAN) |
| Restricted Boltzmann Machine (RBN) and its variantes | Convolutional Neural Network (CNN)              | Ladder Net                           |

Table 3: Deep Learning groups and models types.

This categorisation mainly refer to the underlying learning method whereas generative approaches basically follow the principle of unsupervised learning and discriminative approaches follow the principle supervised learning. Beside the definition of the required number of layers (complexity) the underlying learning method is a decisive factor for the selection of a DL approach (Rieger et al., 2019).

DL is applied to a vast types of input data, including complex data, and for different purposes such as for fault detection using anomaly detection or binary classification abilities, for fault diagnosis commonly using soft-max layer to perform multiclass classification, and for prognosis using a regression layer (L. Zhang et al., 2019).

Rengasamy, Morvan, and Figueredo (2018) highlight that deep learning models often outperform model-based and statistical machine learning techniques. Architectures such as AE, RBN, RNN, CNN and some of their hybrid forms are currently being employed for diagnosis and prognosis in parts of the aircraft, such as aeroengines, fuel systems and actuators, with important contribution in the anomaly detection, feature extraction and feature reduction tasks.

It is possible to construct end-to-end solutions for PdM application with DL (L. Liao et al., 2016; W. Zhang et al., 2019). These approaches avoid the manual feature engineering process and define a unique architecture for data preparation and modelling tasks, as illustrated in Figure 9. Moreover, the end-to-end learning allows DL models to be generic, allowing a single DL model to be used in different PdM applications easily. (L. Zhang et al., 2019)

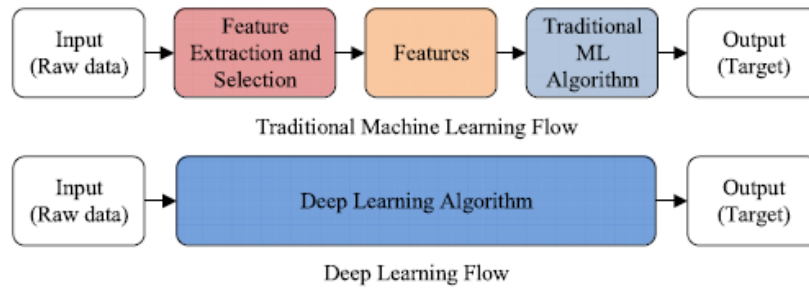


Figure 9: Difference between ML and end-to-end DL. (W. Zhang et al., 2019)

By analysing 27 applied research using the CWRU dataset L. Zhang et al. (2019) indicates that there is no significant difference between the traditional processing techniques and the end-to-end learning paradigm regarding model accuracy. The results suggest a preference on end-to-end approach once it avoids the manual and exhaustive work.

### 2.3.1 Deep Learning applications using vibration data

The survey research of L. Zhang et al. (2019) identifies several works that use DL for PdM using vibration data. Considering unsupervised applications, AE is the method most used. For supervised learning, CNN is prevalent in all kind of applications: fault detection, health stage identification and prognosis. See Table 4.

|     | Unsupervised Fault Detection | Supervised Fault Detection | Supervised HS Identification | Prognostic |
|-----|------------------------------|----------------------------|------------------------------|------------|
| RNN | 2                            | -                          | 5                            | 3          |
| CNN | -                            | 4                          | 30                           | 6          |
| RBM | 1                            | -                          | 5                            | -          |
| AE  | 4                            | 1                          | 17                           | -          |

Table 4: Deep Learning works in PdM using vibration data. Quantity identified in (L. Zhang et al., 2019).

Autoencoder model appears mainly as a substitute method of traditional handcrafted approaches for feature extraction. Luo, Wang, Liu, Li, and Peng (2018) indicate that features extracted from a stack AE is more accurate and stable compared with traditional time domain, frequency domain and time-frequency domain, when discriminating impulse from

non-impulse responses in vibration data. Principi, Rossetti, Squartini, and Piazza (2019) realized that AEs with a fully connected layer or a Long Short Term Memory (LSTM) layer presented best accuracy compared with the convolutional layer and a traditional one-class SVM algorithm. Lu et al. (2018) used a sequenced model with AE for feature extraction, with a transition feature pool to a LSTM network. Roy, Bose, Kar, Gopalakrishnan, and Basu (2019) achieved 100% of accuracy for classifying bearing health states of NASA bearing dataset using a traditional autoencoder and an Extreme Machine Learning network. Abdeljaber, Avci, Kiranyaz, Gabbouj, and Inman (2017) used a end-to-end 1D CNN model to detect structural damage using vibration signals. Janssens et al. (2016) combined a CNN with random forest classifier to identify bearing faults from the vibration signal.

### 3 PHM 2012 Challenge – Benchmark

The literature presents a variety of approaches to copy with the IEEE PHM 2012 prognostic challenge. In addition to the approaches already mentioned in the body of the previous sections, the following paragraphs describes some approaches for diagnosis and prognosis of the bearings tests surveyed in SCOPUS database.

#### 3.1 PHM 2012 Challenge Winner

The overall best performance in the PHM 2012 challenge presented a total score of 0.28. The authors proposed a feature using cumulative values of the vibration signal. They used an integral operation and did not used machine learning techniques (Porotsky & Bluvband, 2012).

#### 3.2 Recent Papers

The search used the following key words ("Predictive Maintenance" OR "PdM" OR "Condition based maintenance" OR "CBM") AND ("prognostic" OR "PHM" OR "FEMTO" OR "PRONOSTIA") AND ("bearing" OR "vibration"), and also it restricted years from 2018 to 2020. The search resulted in 48 works. Among those, 12 papers respected the research criterias: being specific of PHM 12 Challenge; and having open access to the paper document. Seven studies presented their application results see Table 5.

| Year | Author                                | Title                                                                                                                            | Er (B 2) | Average Er   | Average Score |
|------|---------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|----------|--------------|---------------|
| 2020 | Cheng, C., Ding, J., Zhang, Y.        | A Koopman operator approach for machinery health monitoring and prediction with noisy and low-dimensional industrial time series | 58,29%   | -            | -             |
| 2020 | Chen, Y., Peng, G., Zhu, Z., Li, S.   | A novel deep learning method based on attention mechanism for bearing remaining useful life prediction                           | 19,50%   | 2,62%        | 0,44          |
| 2019 | Wang, F.-K., Mamo, T.                 | Hybrid approach for remaining useful life prediction of ball bearings                                                            | 14,83%   | 16,53%       | 0,61          |
| 2019 | Cui, J., Ren, L., Wang, X., Zhang, L. | Pairwise comparison learning based bearing health quantitative modeling and its application in service life prediction           | -        | 5,12%        | -             |
| 2019 | Sadoughi, M., Lu, H., Hu, C.          | A deep learning approach for failure prognostics of rolling element bearings                                                     | 15,34%   | 28,66%       | 0,56          |
| 2019 | Liu, C., Zhang, C.                    | An elman artificial neural network for remaining useful life prediction                                                          | 1444,33% | -<br>413,56% | -             |
| 2018 | Hinchi, A.Z., Tkouat, M.              | Rolling element bearing remaining useful life estimation based on a convolutional long-short-Term memory network                 | 20,36%   | 6,42%        | 0,38          |

Table 5: Last applied research with results for PHM 2012 challenge data set in SCOPUS

The best scored model is a hybrid approach, which combines an exponential weighted moving average control chart for anomaly detection and a machine learning model, the support vector regression with differential evolution (DE) algorithm to estimate the RULs. DE algorithm optimizes SVR hyperparameters. The approach considers three steps: HI construction, anomaly detection, and RUL prediction. The RUL prediction considers the data from the detected anomaly to the end of life. The HI construction uses the root mean square (RMS) and ABS\_5 extracted from the horizontal and vertical accelerometers. The authors suggest a future research using features in time-frequency domain and deep learning models (F. K. Wang & Mamo, 2019).

The second beset model also uses a three step approach: Feature extraction, anomaly detection and rul prediction. It begins with a butterfly filtering technique in the raw vibration data. Then, a convolutional auto-encoder is applied as a robust feature learning strategy that considers RMS, peak-to-peak and kurtosis features. The early fault detection is ensured by a similarity-based method. While a LSTM component estimates the rul. Additionally, the authors applied a data augmentation technique to improve robustness DL methods (Sadoughi, Lu, & Hu, 2019).

Similar to the other two methods, (Chen et al., 2020) relies on the three steps approach. Firstly, five band-pass energy values of frequency spectrum are extracted as features. A RNN based on encoder–decoder with attention mechanism is used to calculate HI values. Finally, the RUL prediction is obtained by a linear regression. (Chen et al., 2020)

Only one paper did not feed the model with extracted features from signal processing techniques. A combination of a convolutional layer for feature extraction, with an LSTM for encoding the temporal representation, allows the neural network to automatically encode the degradation information and predict the RUL of the bearings (Hinch & Tkouat, 2018).

Cui et al. (2019) proposes an extreme learning machine network structure with an extra hidden layer for output enhancement. Both hidden layers are used on different levels to extract degradation process from a multidimensional frequency-domain feature (Frequency Spectrum Partition Summation). In the sequence of the health state quantitative assessment, a polynomial function is applied to fit the rul curve. The authors used only data sets of bearings within the same condition to train their model.

There is an approach that relies on a parallel processing for diagnosis and prognosis prediction. The proposed framework is a driven spectral decomposition framework (denoted as Koopman-CBM). The higher-order dynamic mode decomposition extracts dynamical features from noisy signals in the form of one dimensional time-series. The obtained de-noised model characteristics (i.e., eigenvalues, eigenfunctions, Koopman operator) is effective in predicting the system health stages. Simultaneously, the rul prediction is modeled by the Koopman operator as the linear predictor associated with the nonlinear dynamics (Cheng, Ding, & Zhang, 2020). C. Liu and Zhang (2019) indicate best performance when using traditional and cumulative time domain indexes rather than only traditional time domain indexes for RUL prediction using an elman neural network.

Others papers, mentioned below, which did not mentioned their results presented different approaches for HI construction, diagnosis and prognosis tasks. For example, G. Liu et al. (2019) propose a model using wavelet packet decomposition, AR model and entropy with time parameter to perform the process of degradation. Then, fuzzy c-means clustering method to evaluate the degradation state and warn the failure of bearings. The authors do not present RUL predictions. They focused on degradation monitoring. The method is able to monitor the bearing in real time and get the threshold adaptively. Lim, Kim, Seo, and Choi (2020) proposed a method based on the weighted correlation of the bearing fault frequencies over cycles, which shows a better overall monotonic increase over time compared with rms and kurtosis features. It uses a combination of anomaly detection and prognosis steps. Additionally, they demonstrated a model generalization ability by applying it to the three types of bearings.

A separate learning method was proposed by Kundu, Chopra, and Lad (2019). They used the clustering and change point detection algorithm for identifying multiple failure behaviors. Individual models are developed for single or multiple failure using General Log-Linear-Weibull distribution. Results show that identification of failure behavior helps in accurate prediction of RUL. Additionally, by evaluating different extracted features, they selected RMS and Kustosis as best features accoding to a Chi-square test.

### 3.3 Other related works

Complementing these studies, Lei et al. (2018) mentioned different applied research that uses the same dataset. 17 papers are related to the HI construction from vibration signals using signal processing techniques and/or AI techniques. 8 papers focus on HS division. 12 are dedicated to RUL prediction model development.

L. Liao et al. (2016) develop feature extractions techniques. It compares the average RUL from four methods by using raw features, PCA, original RBM, Bayesian RBM, and the enhanced RBM. The prognosability and trendability for health values of enhanced RBM are higher than the other four, and thus presenting better performance for RUL prediction.

Ben Ali et al. (2015) by analyzing several applications and benchmark studies of prognosis based on vibration signals, suggest that the Multi Layer feed-forward neural networks based on multi-valued neurons have demonstrated their ability to extract complex dynamic patterns from time series for mid-term and long-term predictions as exemplified in (Fink, Zio, & Weidmann, 2014).

Kerdsiri and Gullayanon (2017) present a fault detection algorithm using a logistic regression method to assess the health state of the system. This regression utilizes the node energy of wavelet packet with Daubechies40 at level 5 of decomposition, kurtosis and crest factor as key features. The resulting signals are processed using the Savitzky-Golay smoothing algorithm before determining the time location where the defects occur with a confidence interval of 99%.

Z. Liu et al. (2016) proposes a data-driven method of RUL prediction based on multiple health state assessment for rolling element bearings. Instead of finding a unique RUL prediction model, the life cycle of bearings is clustered into three health states: the normal state, the degradation state, and the failure state. A local RUL prediction model is separately built in each health state. In RUL prediction of bearings, the methods using a unique regression model may be hard to represent the entire history and easily over fit the inconsistent patterns in some features. Therefore, instead of looking for an overall regression model, this paper proposes a RUL prediction method based on multiple health state assessment. It includes four process steps: raw data collection, feature calculation, health state assessment, and RUL prediction modeling. They proposed a hybrid approach

consisting of both unsupervised learning and supervised learning to estimate the health state of a bearing. The unsupervised learning with PCA and fuzzy cmeans is used to automatically extract knowledge about health state labels of all the time points in the training phase. With the provided label knowledge, the supervised learning is employed to build a health state assessment model. Support vector machine is the technique implemented in both, the health state assessment (classification) and RUL prediction modeling (regression).

## 4 Proposed Models – Selected Methods

This section aims to describe the main techniques used for PdM application using the PHM 2012 challenge. Different models will be designed with some of these techniques for comparison purposes. Wavelet packet decomposition, principal component analysis and convolutional auto-encoder (CAE) are techniques selected to extract features and construct health indicators. DBSCAN and support vector machine were identified for diagnosis tasks. For regression, this study selected the support vector regression. Additionally, a RNN is applied, see **Error! Reference source not found..**

| Health Index                 |                                       |                               | Health Stage Identification           | Rul Prediction            |
|------------------------------|---------------------------------------|-------------------------------|---------------------------------------|---------------------------|
| Signal Processing            | Features                              | Dimensionality Reduction      | Anomaly Detection or Classification   | Regressão                 |
| Wavelet Packet Decomposition | Energy<br>RMS<br>Kurtosis<br>Skewness | Principal Components Analysis | DBSCAN<br>&<br>Support Vector Machine | Support Vector Regression |
|                              |                                       | Recurrent Neural Network      |                                       |                           |

Table 6: Proposed Models - Selected Methods

### 4.1 Wavelets – Wavelet Packet Decomposition

The task of constructing reliable HI aims to extract most revealing features from equipment for ensuring accurate health monitoring for either diagnosis or prognosis. There is no unanimity on the best way to extract information among the conventional signal processing techniques or data-driven feature methods (Sadoughi et al., 2019). This study selected the Wavelet signal processing to extract time-frequency features from sensor signals.

As presented in previous section, the most common methods used for vibration analysis, a non-stationary signal, are the Short Time Fourier Transform (STFT), Empirical Mode Decomposition (EMD), Wavelet Transform (WT) or Wavelet Packet Decomposition (WPD). Experiments STFT, EMD and WT, demonstrated a superiority of the WT (Verstraete et al., 2017). WT outperforms STFT because it reduces computational effort, and uses a flexible time window size. Nevertheless, WT presents a drawback of only decomposing low frequency signals, reducing the time-frequency resolution. The WPD presents better resolution in both low and high frequency signals, improving WT (G. Liu et al., 2019).

Wavelet packet decomposition technique is also called subband tree because it decomposes a signal time window in different frequency levels, the packet. It uses parse tree to indicate wavelet packet. Signals can be decomposed into  $n$  levels. At each level, the band is partitioned in two. The Figure 10 illustrates a signal decomposition in 3 levels resulting in the division of 8 frequency sub-bands, the packets. (G. Liu et al., 2019)

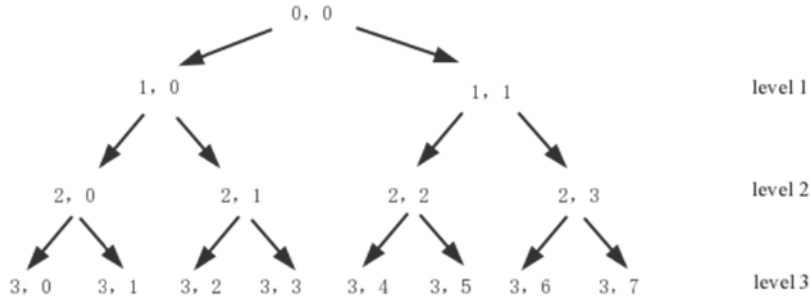


Figure 10: Wavelet Packet Decomposition illustration. (G. Liu et al., 2019)

Commonly, the extracted information are summarized in statistical features. The present study captures the RMS, kurtosis, skewness, and sum square (Energy). RMS is used to detect the unbalance in rotating machinery. Kurtosis and skewness represent the time series distribution of the signal in the time domain. Kurtosis describes the impulsive shape of the signal and measures the peakedness of the signal and skewness measures the asymmetry of the probability of a signal about its mean. These features were mentioned as relevant in various applications using the PHM 2012 data set, such as (Kundu et al., 2019) that indicate the RMS and Kustosis as best features.

## 4.2 Principal Components Analysis

PCA is implemented when the objective is to reduce dimensionality or to fuse features. PCA processes multi-dimensional features from systems in order to outline their operation state (H. Wang, Ni, Chen, & Qu, 2020). For its application, it is important to observe a high correlation among the selected features from signal processing (at least higher than 0.3). Orthogonal rotation is the most used when the goal is to reduce dimensionality, with Varimax method [Hair]. H. Wang, Yu, and Guo (2020) used the threshold of 85% of explained variability for selecting the number of PCA principal components. The resultant components are used to monitoring equipment health status.

### 4.3 DBSCAN and Support Vector Machine

In RUL prediction application, it is stinctive trying to implement regression models using all life-time of equipments without considering different degradation processes. However, a degradation process does not fit a linear trend and can present multiple phases. Based on that, it is recommended to perform fault detection before prognostic tasks (L. Zhang et al., 2019).

As there is no labelling data indicating the health status of equipment during its running time, unsupervised learning techniques are useful to identify patterns on data. Clustering is an algorithm that groups similar data into the same group and forms different groups with dissimilar characteristics. The present study uses the DBSCAN clustering in order to identify different patterns in the vibration signal during the bearing experiments.

DBSCAN was used for fault diagnosis in different works such as for bearing analysis (Li, Wang, Huang, & Li, 2020). Dudik, Kurosu, Coyle, and Sejdić (2015) showed that the DBSCAN outperformed k-means for vibration signal clustering. It is a density-based spatial clustering algorithm, which forms cluster of arbitrary shapes and behaves well with noises. The clustering results depend mainly on two parameters: the search space of core object (eps) and the density threshold (Minpoints). If eps is too small, the model may classify useful data as noise. Inversely, it may assign noise data into a cluster (Li et al., 2020).

It is hard to set eps and Minpoints parameters. Ester, Kriegel, Sander, and Xu (1996) suggested finding the elbow on the k-nn distances graph in order to define eps value. Li et al. (2020) proposed an adaptable method for finding reasonable values for eps, using Minpoints as the division of the number of observations per 25.

This paper considers the approach adopted by (Rousopoulou et al., 2019), which combines DBSCAN and SVM for identifying faults using acoustic signals. The extracted features from previous analysis are fitted to DBSCAN, each signal sample of a bearing is assigned to a cluster. The number of clusters is estimated automatically. The parameters of DBSCAN are selected after a few trials. With the information of DBSCAN analysis, a SVM model is trained for online predictions.

In the present study, the quantity of clusters is interpreted as different health stages.

## 4.4 Recurrent Neural Network

The RNN is able to encode temporal information in sequential data. It uses neurons connections, which allows hidden layer to receive a signal from the input at current time  $t$ , as well as from the output of the hidden state at previous time. RNN uses the same parameters over all steps, differently from traditional neural networks. Variants of RNN are: Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), Bidirectional RNN (BRNN) and Statistical Recurrent Unit (SRU). They can vary in power, flexibility and complexity depending on the application (L. Zhang et al., 2019).

The present paper is based in Ribeiro De Miranda, De Andrade Barbosa, Scolari Conceicao, and Soares Alcala (2019). They propose a RNN model based on a SRU (Oliva, Poczos, & Schneider, 2017) for RUL prediction. The model was used to extract hidden patterns from multi sensors time series of equipment with several faults and degradations behaviors. The SRU method operates with simple moving averages of summary statistics to extract features from temporal data with long-term dependencies. It uses the Rectified Linear Unit (ReLU) activation function (Nair & Hinton, 2010) and an exponential moving average for solving the vanishing gradient problem (Bengio, Simard, & Frasconi, 1994), contrasting LSTM that uses gates (Hochreiter & Schmidhuber, 1997).

The network architecture has a RNN model and an output layer. It uses the back-propagation method for updating the synaptic weights. The optimization operation run in mini-batches of 512 and the AMSgrad method (Reddi, Kale, & Kumar, 2018). The summary recurrent and multi-scaled statistics are calculated as (Oliva et al., 2017):

$$\begin{aligned} R &= f(S_{t-1}W_r + b_r), \\ I &= f([R|\hat{X}_t]W_i + b_i), \\ S_{t-1} &= \text{reshape}(S_{t-1}, L, d, k), \\ I &= \text{reshape}(I, L, 1, k), \\ A &= \text{reshape}(a, 1, d, 1), \\ S_t &= AS_{t-1} + (1 - A)I, \\ S_t &= \text{reshape}(S_t, L, kd), \\ O &= f(S_tW_o + b_o), \end{aligned}$$

$f(.) = \max(., 0)$  represents a ReLU activation function.  $S \in R^{L \times kd}$  is the recurrent statistics matrix. A concatenation matrix operation is indicated as  $[.|.].$   $\hat{X}t \in R^{L \times N}$  is a slice of input matrix  $D_b$  in an instant  $t$ .  $reshape(., ., ., .)$  is a tensor reshape approach.  $A \in R^{l \times d \times 1}$  is a vector a after reshape.  $R \in R^{L \times c}$ ,  $I \in R^{L \times k}$ ,  $W_r \in R^{kd \times c}$ ,  $W_i \in R^{(N+c) \times k}$  and  $W_o \in R^{kd \times kd}$  are the recurrent, input, and output synaptic weight matrices, respectively. They are initialized by a random uniform distribution.  $b_r \in R^{1 \times c}$ ,  $b_i \in R^{1 \times k}$ ,  $b_o \in R^{1 \times kd}$  are vectors of biases from recurrent, input, and output synaptic weight matrices. The bias are initialized as zero.  $c$  is the number of recurrent states, and  $O \in R^{L \times kd}$  is the output of a recurrent unit. The temporal action of the model is illustrated in Figure 1Figure 11.

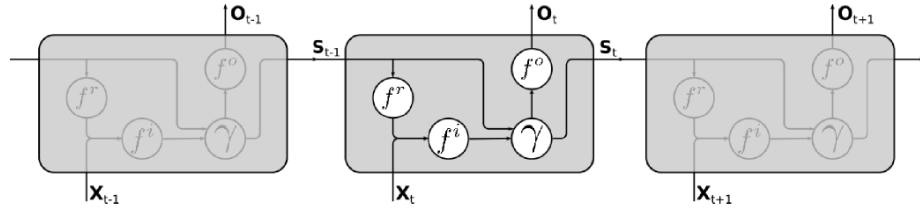


Figure 11: RNN - SRU operation and its behavior. (Ribeiro De Miranda et al., 2019)

The output operation is computed by:

$$o = \sigma \left( \left( \sum_{u=1}^{kd} o_{l,u} w \right) + \beta \right), \text{ with } l = 1, \dots, L$$

$o$  is a vector with the predicted outputs, which depend on the synaptic weights of the output layer ( $w \in R^{kd \times l}$ , initialized with Xavier uniform initializer (Glorot & Bengio, 2010)) and on past outputs  $O_{l,u}$ .  $\beta$  is a bias initialized to zero,  $\sigma$  is the sigmoid activation function. The back-propagation algorithm is used to train the model.

## 4.5 Support Vector Regression

Support vector regression is a type of SVM. In this study, it is applied to predict the remaining useful life of the bearings. It searches optimal hyperplane that achieve lowest errors between training observations and the optimal hyperplane. Finding the optimal hyperplane is equivalent to solving a convex programming problem (Zhou, Zhang, & Wang, 2016). It is wide used in PdM applications, mainly because of its capacity to deal with nonlinear regression cases.

## 5 Methodology

### 5.1 Research

The present research study initiated by the literature review in three different stages. In chapter 2, a review of the general topics of PdM. Chapter three focused on applied research specifically to the PHM 2012 challenge. In the fourth chapter, a review of the main data analytics techniques used in the application section (section 6) of this dissertation.

### 5.2 Application

This dissertation pursues the cross-industry process for data mining (CRISP-DM) which presents a 6 steps approach to planning and execute a data mining project. On the CRISP-DM method, the author encompasses the PdM process in order to adapt it for data mining projects with predictive maintenance purpose considering 5 main activities: Signal processing; Feature extraction; Classification task.; RUL estimation and Evaluation. Therefore, The “deployment” phase of CRISP-DM is out of scope of this study. See Figure 12.

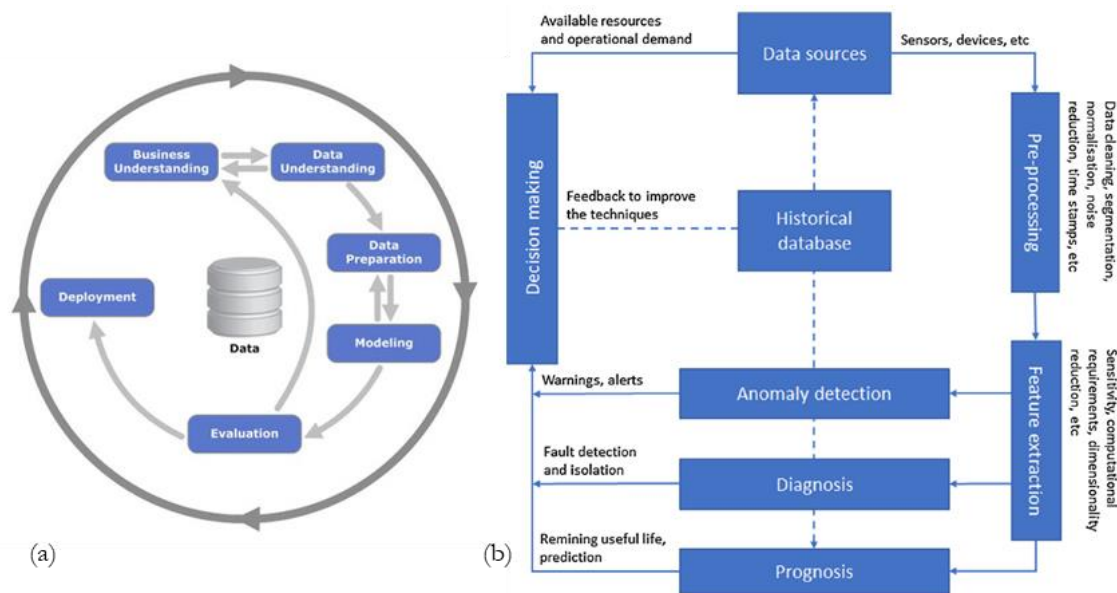


Figure 12: (a) CRISP-DM Methodology and (b) PdM Process (Khan & Yairi, 2018)

### 5.3 Case Study and Data Source

The present dissertation uses the public dataset of bearing experimentation used in the 2012 IEEE PHM Challenge problem of RUL prediction. The dataset was generated in the PRONOSTIA experimentation bench (see Figure 13), which was designed by FEMTO-ST Institute, for testing and validating bearings diagnostic, and prognostic applications. Different from the simulation degradation dataset, this is a real dataset where the bearings naturally degraded without being seeded a fault in advance. (Lei et al., 2018)

This dataset contains 17 run-to-failure data of rolling element bearing. Two accelerometers and a thermocouple were used to capture the vibration signals and the temperatures of the bearings. The bearing useful life is considered to end when the amplitude of the vibration signal exceeds 20 g.

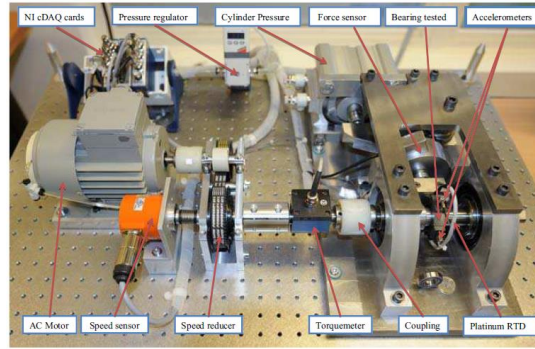


Figure 13: PRONOSTIA experiment bench. (Nectoux et al., 2012)

The experiments were conducted under three working conditions varying with loads and speeds. Working condition 1: 4000N radial load and 1800 r/min speed. Working condition 2: 4200N and 1650 r/min speed. Working condition 3: 5000N and 1500 r/min speed. All the 17 datasets are summarized in Table 7, indicated by bearing  $i\_j$ , where  $i$  denotes the index of the working condition and  $j$  denotes the case index.

| Data sets    | Operating Conditions |              |              |
|--------------|----------------------|--------------|--------------|
|              | Conditions 1         | Conditions 2 | Conditions 3 |
| Learning set | Bearing1_1           | Bearing2_1   | Bearing3_1   |
|              | Bearing1_2           | Bearing2_2   | Bearing3_2   |
| Test set     | Bearing1_3           | Bearing2_3   | Bearing3_3   |
|              | Bearing1_4           | Bearing2_4   |              |
|              | Bearing1_5           | Bearing2_5   |              |
|              | Bearing1_6           | Bearing2_6   |              |
|              | Bearing1_7           | Bearing2_7   |              |
|              |                      |              |              |

Table 7: IEEE PHM 2012 challenge dataset. (Nectoux et al., 2012)

Among the 17 datasets, the six training datasets are provided with complete run-to-failure data, while the rest ones are provided with only censored bearing life data. For each dataset, the vibration data were collected every 10 s with a time duration of 0.1 s at a sampling frequency of 25,600 Hz from two sensors: a vertical accelerometer and a horizontal accelerometer. As the temperature data are not consistently available for all the 17 datasets, we do not include them in the proposed framework. (Sutrisno et al., 2012)

### 5.3.1 Objectives and Experiments

The objective is to study different approaches regarding the PdM considering machine learning and deep learning techniques, and compare their performance using the IEEE PHM 2012 prognostic challenge.

Three combinations of techniques were implemented, as indicated in

| Model | Health Index      |                          | Health Stage Identification | Rul Prediction | Training_Set            |
|-------|-------------------|--------------------------|-----------------------------|----------------|-------------------------|
|       | Signal Processing | Dimensionality Reduction | Diagnosis                   | Prognosis      |                         |
| 1.1   | WPD               | PCA                      | No                          | SVR            | Bearing2_1 & Bearing2_1 |
| 1.2   | WPD               | PCA                      | DBSCAN + SVM                | SVR            | Bearing2_1 & Bearing2_1 |
| 2     | WPD               | RNN - SRU                |                             | SVR            | Bearing2_1 & Bearing2_1 |

Table 8. The first model applies traditional machine learning techniques such as PCA and SVR. A variant of this first model considers the DBSCAN for fault classification. The last model uses a RNN based in the SRU for predicting the RUL.

| Model | Health Index      |                          | Health Stage Identification | Rul Prediction | Training_Set            |
|-------|-------------------|--------------------------|-----------------------------|----------------|-------------------------|
|       | Signal Processing | Dimensionality Reduction | Diagnosis                   | Prognosis      |                         |
| 1.1   | WPD               | PCA                      | No                          | SVR            | Bearing2_1 & Bearing2_1 |
| 1.2   | WPD               | PCA                      | DBSCAN + SVM                | SVR            | Bearing2_1 & Bearing2_1 |
| 2     | WPD               | RNN - SRU                |                             | SVR            | Bearing2_1 & Bearing2_1 |

Table 8: Models

The models are trained using data sets of the experiments in condition 2 (BearingB2\_1 and 2\_2) and tested using each test set in the same operation condition (Bearings2\_3, Bearings2\_4, Bearings2\_5, Bearings2\_6 and Bearings2\_7).

The results of them are compared with the aims to answer three questions:

Question 1: Does the diagnosis procedure of health assessment improves models performance?

Question 2: Does the model using deep learning techniques outperforms the model using traditional methods?

When comparing two models, the t-student test is applied.

### 5.3.2 Evaluation metric

The performance of the proposed model will be evaluated using the same metric applied in the IEEE PHM 2012 prognostic challenge, described in (Nectoux et al., 2012). It is based on their RUL results, converted into predictions percent errors. Let note  $\widehat{RUL}_i$  and  $ActRUL_i$  respectively the estimated RUL, and the actual RUL to be predicted (where  $i$  states for the test bearings). The percent error on experiment  $i$  is defined by:

$$\%Er_i = 100 \times \frac{ActRUL_i - \widehat{RUL}_i}{ActRUL_i}$$

Underestimates and overestimates will not be considered in the same manner: good performance of estimates relates to early predictions of RUL (i.e. cases where  $\%Er_i > 0$ ), with deduction to early removal, and more severe deductions for RUL estimates that exceed actual component RUL (i.e. cases where  $\%Er_i < 0$ ). The score of accuracy of a RUL estimates for experiment  $i$  is thereby defined as follows:

$$A_i = \begin{cases} \exp^{-\ln(0.5) \cdot (Er_i/5)} & \text{if } Er_i \leq 0 \\ \exp^{+\ln(0.5) \cdot (Er_i/20)} & \text{if } Er_i > 0 \end{cases}$$

Figure 14 depicts the evolution of this scoring function

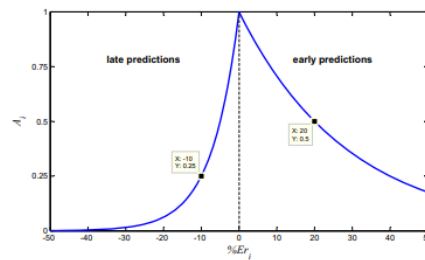


Figure 14: Scoring function of a RUL estimates according to its percent error. (Nectoux et al., 2012)

The final score of all RUL estimates will be defined as the mean of all experiment's score:

$$Score = \frac{1}{11} \sum_{i=1}^{11} (A_i) \quad (3)$$

### 5.3.3 Tools

This work was conducted using R programming language and mainly the caret, e1071, fpc, dbscan, keras and tensorflow libraries. Part of the analysis used python and the pyTorch framework.

## 6 Application

This section presents the main results obtained using the techniques described for the models. The last subsection is dedicated to model results presentation and comparisons.

### 6.1 Signal Processing: Wavelet Packet Decomposition

The vibration data were loaded and feeded in the wavelet packet decomposition function. The approach used the Daubechies db4 as the mother wavelet function, with three decomposition levels. It results in 8 packets frequency decomposition per vibration sample. The in Figure 15 illustrates the raw signal data from horizontal accelerometer of the bearing2\_1 during its whole life time, its decomposed packet values and a smooth curve from the decomposed packets.

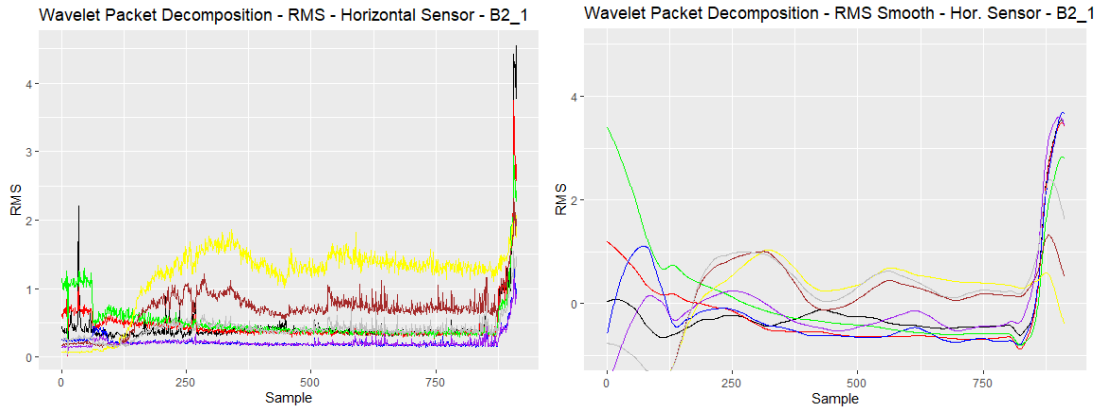


Figure 15: WPD and Smoothed RMS illustration.

The raw signal, in time domain, indicates an increasing trend of signal over the time. Meanwhile, the wavelet packets indicate changes only at the end of the curve.

In the end of the signal processing Energy, RMS, Kurtosis and Skewness were extracted from the raw signal from each sub-band signal and then smoothened using locally weighted regression (LOESS) (Cleveland & Devlin, 1988). The smoothed values are determined using the neighboring data points within with a span = 0.1, similar to Duong et al. (2018) work. In total, 64 features are obtained. 8 features for each statistical metrics from each sensor, the vertical and horizontal accelerometer.

## 6.2 Model 1 and its variants

### 6.2.1 Health Indicator: Principal Component Analysis

The PCA was applied in all 64 extracted features with all 1708 observation from training data sets of Bearing 2\_1 and Bearing 2\_2. It means a proportion of 28 observations per variable, respecting a threshold of 5 indicated by [Hair]. After two iterations, 4 features were eliminated because did not respect the KMO threshold of 0.5, see

| A1_KTS_2 | A1_KTS_5 | A2_KTS_1 | A2_KTS_4 |
|----------|----------|----------|----------|
| 0.44     | 0.49     | 0.39     | 0,45     |

Table 9.

| Kaiser, Meyer, Olkin Measure of Sampling Adequacy (KMO) |          |          |          |
|---------------------------------------------------------|----------|----------|----------|
| A1_KTS_2                                                | A1_KTS_5 | A2_KTS_1 | A2_KTS_4 |
| 0.44                                                    | 0.49     | 0.39     | 0,45     |

Table 9: PCA – Features Eliminated - Kaiser, Meyer, Olkin Measure of Sampling Adequacy (KMO)

The remaining 76 features presented an overall MSA of 0.91. Despite the high value of MSA, only 35,6% of the variables have the absolute correlation higher than 0.3 between them. The correlation matrix in Figure 10 gives an idea of this proportion.

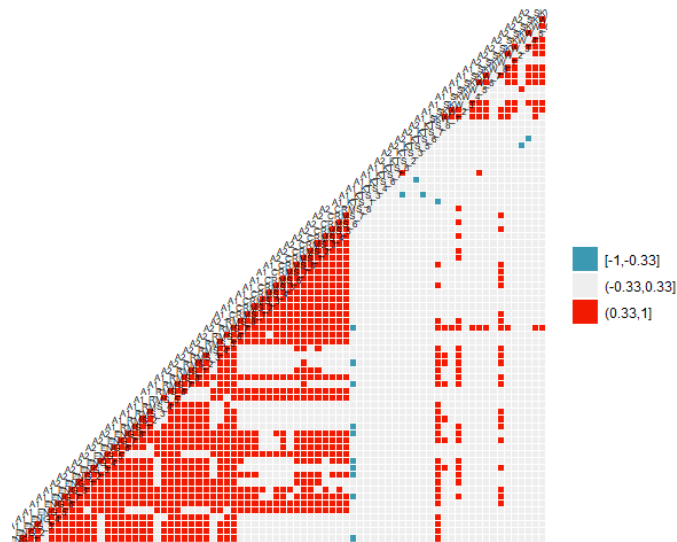


Table 10: PCA - features matrix correlation.

The number of principal components was identified using a threshold of 85% of cumulative explained variance as suggested by H. Wang, Yu, et al. (2020). 7 principal components explain 85,9% of the total variance, see Figure 16.

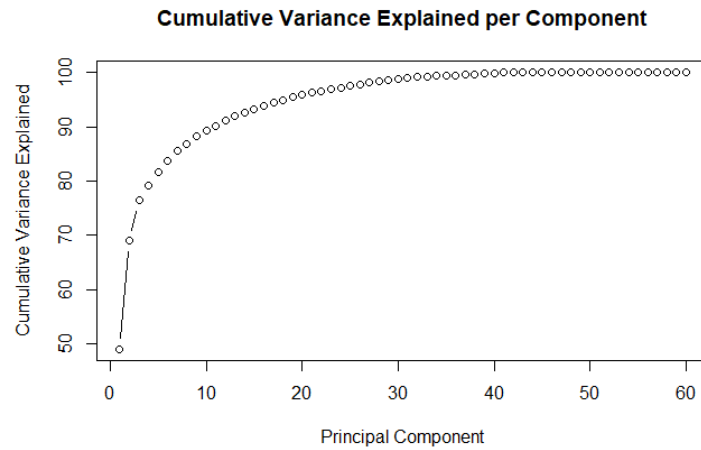


Figure 16: PCA - % Variance Explained

Setting the quantity of principal component equal to 1 is useful when it is desirable to interpret the degradation process using a unique variable. However, it can lose useful information. Here, 1 component explains 47,7% of total variance. Figure 18 and Figure 18 show the principal components for Bearing2\_1 and Bearing 2\_2 respectively. Left plots have the health indicators represented by three principal components and the right plots show a unique principal component as the HI.

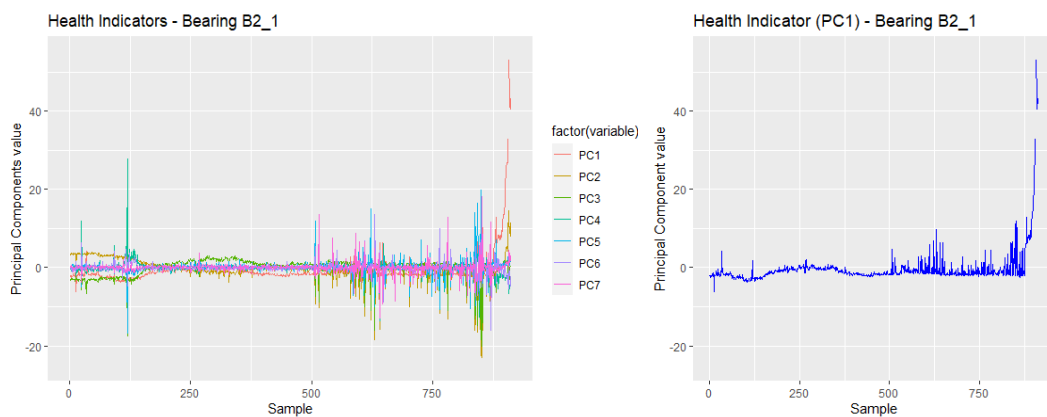


Figure 17: Health Indicators - Bearing B2\_1

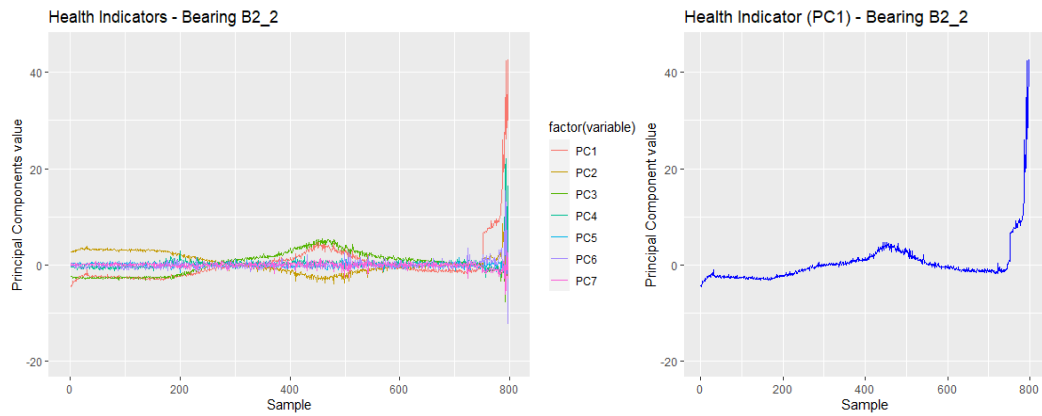


Figure 18: Health Indicators – Bearing B2\_2

In both bearings, all features appear to oscillate around an almost constant mean until final stages of bearing's life. It is noticeable a fluctuation in the beginning and in the end of the curve. The fast growth in the tail of the curve characterizes the behavior close to the fault.

### 6.2.2 Diagnosis: Health Stages - DBSCAN and Classifier

The DBSCAN is used here to identify different health stages in the bearing life cycle. The cluster analysis is performed using the Health Indicators, extracted from the PCA. Before analysis, the HI from training and test data sets were mean normalized.

The cluster analysis was performed individually for each data set in the training set (Bearing 2\_1 and Bearing 2\_2) using the 7 principal components. The DBSCAN used the MinPoints equal to the number of observations per bearing divided per 25 as suggested by Li et al. (2020). After some essays, the eps parameter was set on 0.28 and 0.12, respectively.

Each bearing was organized in two clusters, indicated in Table 11. The bearing 2\_1 has 214 samples assigned as “Cluster 0”, 17.3%. The bearing 2\_2 has 92 observations samples assigned as “Cluster 0”, 11.5%. It is an unbalanced quantity of classes, with more samples grouped in the cluster 1.

| Clusters |   | Bearing 2_1 | Bearing 2_2 |
|----------|---|-------------|-------------|
| DBSCAN   | 1 | 810         | 705         |
|          | 0 | 159         | 92          |

Table 11: DBSCAN Results - Quantity of normal and anomaly samples.

These clusters suggest two different health stages for the bearing life, see Figure 19. Colored points reflect the different clusters regarding the health indicators curve. Clearly, the cluster 1 indicated by the color blue represents a stable phase of the operation. The cluster 0, frequent in the end of the bearings life, is characterized by critic change in the curves. Cluster 0 is interpreted as an anomaly condition that precedes a bearing fault. However, Bearing 2\_1 presents an unexpected behavior in early phase. Additional study using three health stages is described in Annex 1 – 3 Health Stages .



Figure 19: Clusters - Health Stages

The cluster results became labels to train a classifier model that uses the extracted features from PCA as input. The classifier relied on support vector machine model using a Radial Kernel. It was trained using a cross-validation method with  $K=10$ . The data sets Bearing1\_1 and Bearing 1\_2, not used to train the SVM model, were used for testing the classifier model. Additionally, an oversampling technique was applied in order to deal with the class unbalance problem. The final training set changed from 1708 observations and a proportion of 24% for the “Cluster 0” to a total of 3500 samples, 51% of them of “Cluster 0”. The confusion matrixes computed from DBSCAN groupment and the classifier, using the bearing 1\_1 and bearing 1\_2 as test sets are shown in Table 12.

|                       |   | Bearing1_1 |      | Bearing1_2 |     | Overall Performance |     |
|-----------------------|---|------------|------|------------|-----|---------------------|-----|
|                       |   | Cluster    |      | Cluster    |     | Accuracy            | 82% |
|                       |   | 0          | 1    | 0          | 1   | Precision           | 74% |
| Classifier Prediction | 0 | 42         | 0    | 173        | 74  | Recall              | 27% |
|                       | 1 | 303        | 2458 | 277        | 347 | Specificity         | 97% |

Table 12: SVM Classifier evaluation. Bearing 1\_1 results in the left and Bearing 1\_2 in the right

The model presented an accuracy of 82%. 89,1% for bearing set 1\_1 and 59,7% for bearing 1\_2. The model identified correctly 100% of the samples in normal condition in bearing 1\_1 and 82% in bearing 1\_2. The precision of the model in Bearing 1\_1 was 100%. However, the sensitivity stayed in a low level, 12% and 38% respectively. The model ability of detecting anomaly performed under expected. Despite of the results, the study continues with this approach, once the model has high specificity (97%) and a precision of 74%, suggesting that the model could be a differential to identify anomaly conditions.

### 6.2.3 Remaining Useful Life: Support Vector Regression

This section presents the procedures used to predict the bearing remaining useful life in two approaches. One model is based only in the health indicators (principal components). The other considers the health stage assessment in order to make the predictions.

Both approaches feed the principal components into a SVR model in order to predict the RUL. The train process uses the time-to-failure of the experiments bearing 2\_1 and bearing 2\_2 as the label variable.

The SVR model parameters were set equal for both approaches. It is a SVM eps-regression, with Radial Kernel, gamma = 0.33, epsilon = 0.1, and cost = 1. The predictions of the test bearings are shown in the Figure 20 to Figure 24.

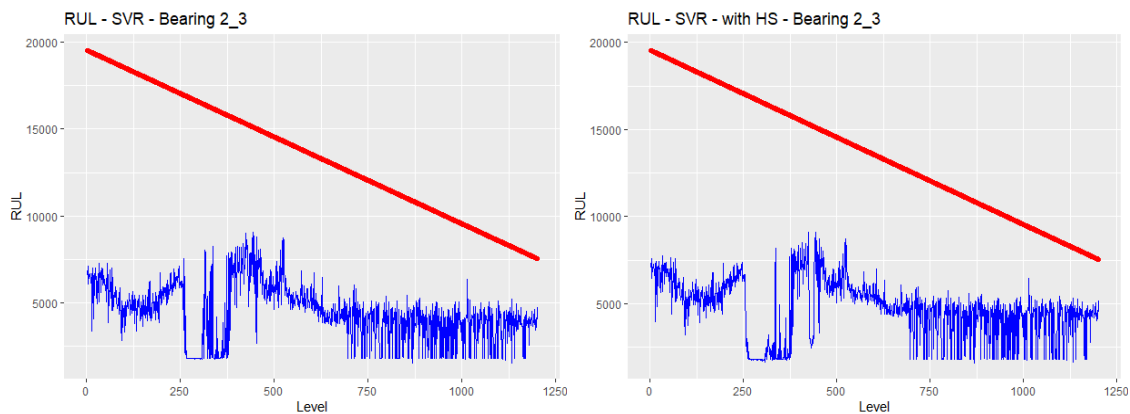


Figure 20: RUL of bearing 2\_3. In the right is the model that considers the HS identification.

The predictions of bearing 2\_3 using the model without consider the HS presented an average score of 0.21 with a standard deviation of 0.13. The model that considers the HS identification achieved an average score of 0.21 with a std of 0.11. Evaluating only the last prediction, it presents a score of 0.70 and 0.61 respectively. Despite the high score in the

last prediction, the RUL curve indicates a poor model. It does not present a period with high correlation with the time.

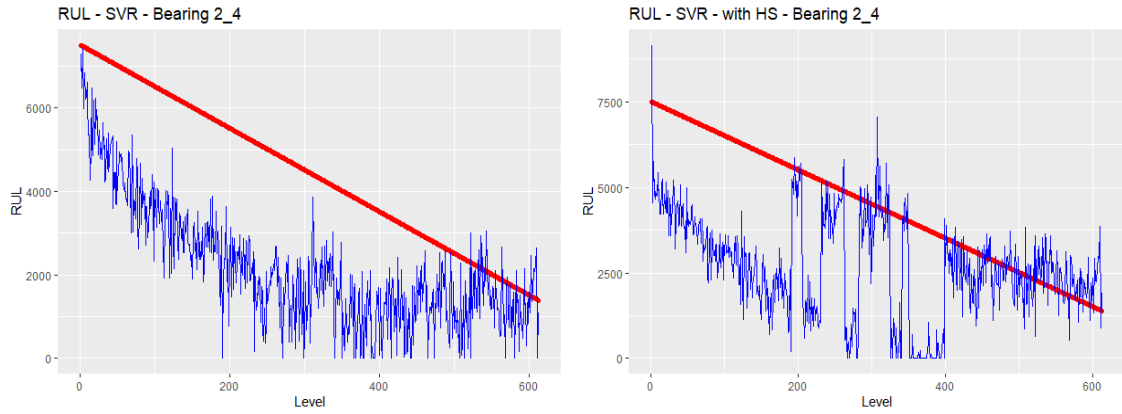


Figure 21: RUL of bearing 2\_4. In the right is the model that considers the HS identification.

The RUL curve of this second experiment shows a better behavior. The RUL predictions of bearing 2\_4 using the model without considering the HS presented an average score of 0.21 with a std of 0.18. The model that considers the HS identification achieved an average score rate of 0.28 with std equal to 0.27.

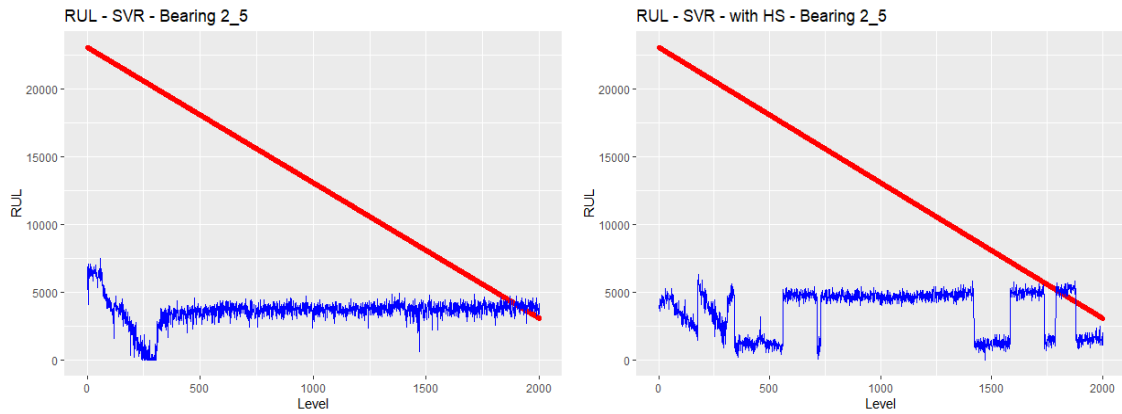


Figure 22: RUL of bearing 2\_5. In the right is the model that considers the HS identification.

The model was not able to explain the RUL variation with the selected health indicators for the experiment with bearing 2\_5. The average score was 0.15 with an std of 0.16 for the model without diagnosis task. The model that considers the HS identification achieved a score of 0.14 with a std of 0.15. The effect of the health stage classification can be verified in the right plot when the curve seems to change as a stairway



Figure 23: RUL of bearing 2\_6. In the right is the model that considers the HS identification.

The predictions of bearing 2\_6 using the model without consider the HS presented a score of 0.34 with a std of 0.24. The model with health stage assessment scored 0.29 in average with a standard deviation of 0.33.



Figure 24: RUL of bearing 2\_7. In the right is the model that considers the HS identification.

The worst results were obtained with the experiment bearing 2\_7. The models scored 0.0015 and 0.0004, respectively. Both with a standard deviation of 0.024.

It is clear in the graphs that the model was able to capture the variation of the RUL using the selected health indicators only to the experiments with the bearings 2\_4 and 2\_6. It suggest that the failure mode of these bearings are similar to one of the Bearing 2\_1 and 2\_2.

A summary with score results is presented in the Table 13 and the comparisons between the evaluations are discussed in the section 6.4.

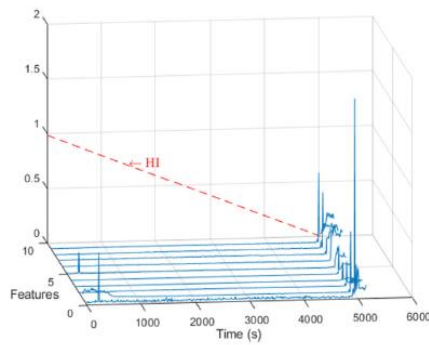
## 6.3 Model 2

The RNN applied in this study is based on the SRU and is illustrated in Figure 11. The SRU model is developed using the pyTorch framework. It has one recurrent layer with 32 nodes. The initialization method is described in the section 4.4. The training process uses the back-propagation method to refresh the network weights. The gradients are determined using the Adam optimization method, with mini-batches of size 2 and learning rate of 0.001. MSE was used as the loss function.

### 6.3.1 Health Indicator: Recorcent Neural Network – Statistical Recurrent Unit

The model uses 64 features extracted by the WPD process from the raw singal data. The features are normalized and then structured in 3 dimensions before to feed the SRU model. The data dimentionis indicates the batch, the time-window and the variables. The batch specify each training bearing, and the time window represents the maximal lifetime of the training bearings. In this case, time-window is equal to 911 (total of samples from the bearing 2\_1). The bearing 2\_2, with 797 samples, is complemented with observation equal to 0. The padding operation of the model ignores these null samples.

The SRU model estimates the RUL based in a health indicator with a linear trend until the fault. It is considered that the bearing suffer a deterioration in constant rate during all life cycle. It is obtained by the proportion of actual RUL ( $RUL_t$ ) divided by the RUL at the beginning of life ( $RUL_0$ ) (Chen et al., 2020):



$$HI_t = \frac{RUL_t}{RUL_0}$$

Figure 25: RNN - HI. Comparison between traditional features (Chen, Peng, Zhu, & Li, 2020)

For each instant  $t$  in the training process, the input data is associated with the HI at the same time. The HI represents the degradation percentage of the bearing at time  $t$ . For instance, a

bearing in the training set with 2000s of whole life is assessed in the point of 1000 second, then the HI is equal to 0.5 (Guo, Li, Jia, Lei, & Lin, 2017).

### 6.3.2 Remaining Useful Life

The model estimates the RUL based in the actual state and in past data. The first prediction of the model uses all historic data of train data set. From the second prediction ahead, the model increments information from past observation (in form of variable prediction) as input to the RUL estimation. The remaining useful life is calculated by the multiplication of the actual accumulated time life of the bearing the HI computed by the model.

Health indicators of the bearings are shown in the Figure 26 and Figure 27. The bearing operation initiation when the orange curve starts. The green curve represents the predicted HI. The orange and blue lines indicate the real RUL.

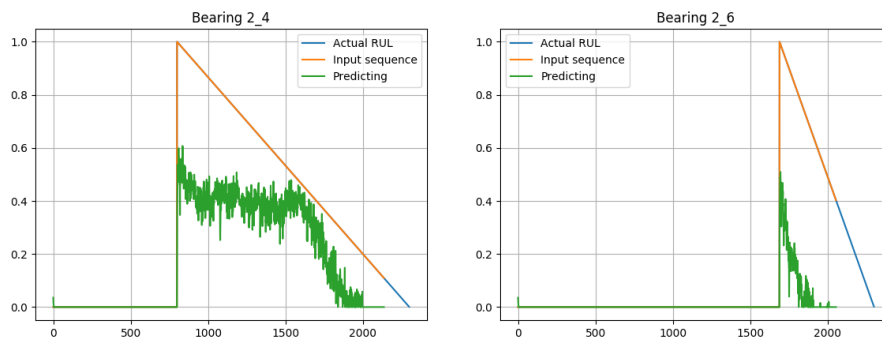


Figure 26: RNN - Health Indicator - Bearings 2\_4 and 2\_6

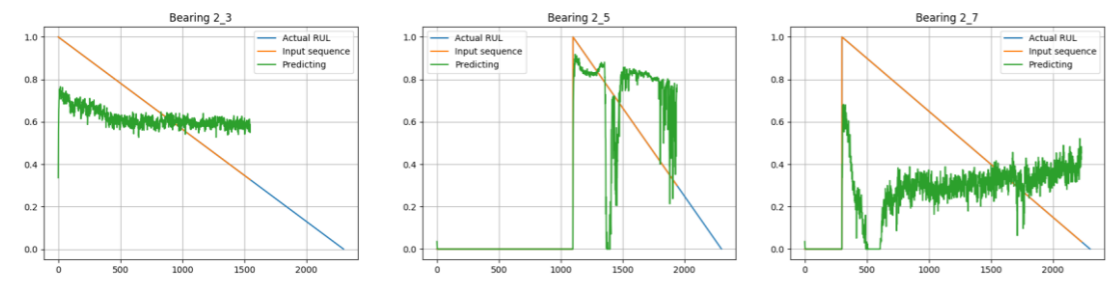


Figure 27: RNN - Health Indicator – Bearings 2\_3, 2\_5 and 2\_7

In general, the model presented similar behavior when compared to model 1. It was not able to follow the true RUL of the bearings. The predictions for the Bearing 2\_4 was the unique that approached the real RUL. The prediction curve of Bearing 2\_3 presented a good trend,

however, with higher values than the real RUL. The other 3 tests were not reasonable and do not present the expected

## 6.4 Results and comparisons

Table 13 shows the models performance, measured in terms of score, for each test experiment. There are three information per model: the score of the last sample in the test set; the average score that considers all the samples in the test experiment; and the standard deviation the the sample scores. “N” indicates the quantity of samples in each test experiment.

| Bearings       | N    | Model 1.1  |             |           | Model 1.2  |             |           | Model 2    |             |           |
|----------------|------|------------|-------------|-----------|------------|-------------|-----------|------------|-------------|-----------|
|                |      | Last Score | Avg Score   | Std Score | Last Score | Avg Score   | Std Score | Last Score | Avg Score   | Std Score |
| Bearing 2_3    | 1202 | 0,7        | <b>0,21</b> | 0,13      | 0,61       | <b>0,21</b> | 0,11      | 0,02       | <b>0,34</b> | 0,20      |
| Bearing 2_4    | 612  | 0,47       | <b>0,21</b> | 0,18      | 0,01       | <b>0,28</b> | 0,27      | 0,03       | <b>0,03</b> | 0,02      |
| Bearing 2_5    | 2002 | 0,01       | <b>0,15</b> | 0,16      | 0,31       | <b>0,14</b> | 0,15      | 0,00       | <b>0,17</b> | 0,18      |
| Bearing 2_6    | 572  | 0,57       | <b>0,34</b> | 0,24      | 0,00       | <b>0,29</b> | 0,33      | 0,00       | <b>0,23</b> | 0,19      |
| Bearing 2_7    | 172  | 0,00       | <b>0,00</b> | 0,01      | 0,00       | <b>0,00</b> | 0,00      | 0,43       | <b>0,29</b> | 0,21      |
| <b>Average</b> |      | 0,35       | <b>0,19</b> | 0,18      | 0,19       | <b>0,19</b> | 0,2       | 0,02       | <b>0,21</b> | 0,20      |

Table 13: Summary of Scores

### 6.4.1 Question 1: Does the diagnosis procedure of health assessment improves models performance?

In order to answer this question, Model 1.1 and Model 1.2 were compared in terms of score performance. Model 1.1 and Model 1.2 used the same techniques excepting that Model 1.2 incorporated a health assessment term in the regression model. Figure 28 shows the score distribution of all testing samples, and the scores distributions per testing bearing.

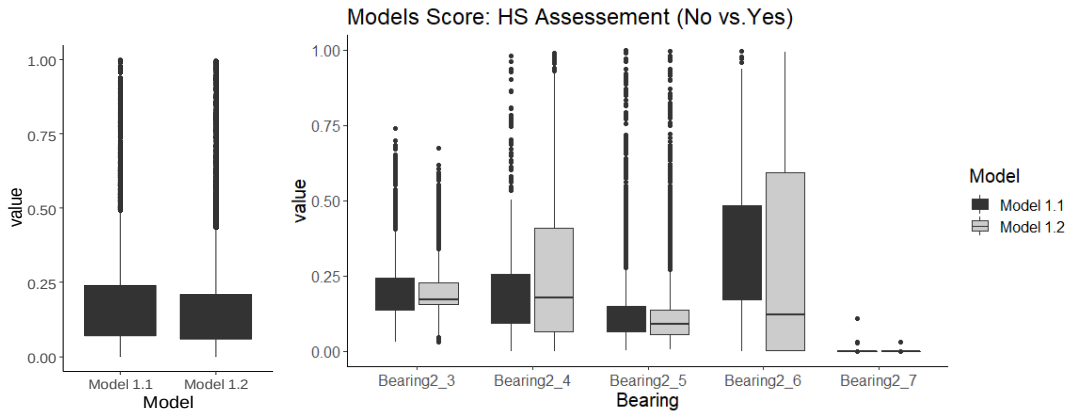


Figure 28: Boxplot: comparing model 1.1 and model 1.2 score performance.

The model without health assessment ability presents an average score of 0.19 with 0.18 standard deviation. The other one has an overall score of 0.19 with 0.2 standard deviation. The Bartlett test resulted in Bartlett's K-squared = 70.935 and a p-value of  $2e-16$ , suggesting to reject the null hypothesis of equal variance between in the scores of both models. Considering it, the mean score of the models were compared using a two sided t-student test with 95% of confidence level. The result of  $t = 0.51682$  and p-value = 0.60 indicates that both approaches present the same score mean. Thus, the health assessment procedure did not improve model 1.1 performance. Similar conclusion is true when comparing both models using only Bearing2\_4 and Bearing2\_6 experiments.

#### 6.4.2 Question 2: Does the model using deep learning techniques outperforms the model using traditional methods?

The model 2 is compared with model 1.1 with the aims to answer the question 2. The model 2 presents an average score of 0.21 with 0.18 standard deviation. First, the equality of variances between both models is tested using the Bartlett. The test resulted in Bartlett's K- suggesting to reject the null hypothesis of equal variance between in the scores of both models. Considering it, the mean score of the models were compared using a two sided t-student test with 95% of confidence level indicates that both approaches present the same score mean. Thus, both models performed with the same performance.

## 7 Conclusion

The data-driven PdM using machine learning and deep learning (DL) algorithms has been widely applied to industrial manufacturing. Recently, a huge quantity of papers were published in conferences and academic journals approaching theory and practical works on the diagnosis and prognostics domain. Research works on data-driven fault diagnosis and prognostics can be divided into five stages: Signal processing, Feature Extraction, Anomaly Detection, Fault Classification and Prognostic. Many researchers have studied the applications of different time domain, frequency domain, and time-frequency domain signal processing techniques. The features extraction is a very important task in PdM application and usually are developed ad-hoc for specific problem, imposing difficulty to generalize for problems of the same nature. Deep learning (DL) has emerged as a promising technique to face different challenges of PdM. It can automatically extract and select representative features from raw signal data and construct complex input output relationships between signals and fault types with its multi-layer nonlinear modelling capacity. Even more, it can create a complete end-to-end pipeline, avoiding manual and time-consuming activities. More recently, DL models, including deep belief network (DBN), convolutional neural network (CNN), recurrent neural network (RNN) and so on, have been widely applied on fault diagnosis. Moreover, PdM applications can benefit from DL, especially because of the fact that high complex, non-linear and unlabeled (or partially labeled) data is the normal case.

This study used the vibration data sets from the PHM 2012 Challenge for testing two approaches. One relying on traditional machine learning techniques and other using deep learning methods. Wavelet packet decomposition was used for signal processing for both methods, which extracted RMS, Energy, Kurtosis and Skewness features. The model based on traditional machine learning techniques fuses features using principal component analysis. The resultant principal components feed a support vector regression for RUL estimation. The deep-learning based model used a convolutional auto-encoder and the statistical recurrent unit techniques for extract useful information considering temporal association for RUL prediction. The experiments results that both modes SRU and PCA with SVM model had same performance.

Additionally, a health stage assessment procedure using DBSCAN and support vector machine was implemented for diagnosis purposes. The experiments demonstrated that the

inclusion of this ability in the the model using traditional machine learning techniques did not improve its score performance.

In general, the models were not able to explain the remaining useful life variation. It can be caused because of the low quantity of training data. This study used only two run to failure data sets to train the predictive models (Bearing 2\_1 and 2\_2). Indeed, the models could fit in a low degree two of the five test sets (Bearing 2\_4 and 2\_6). Other important aspect regards the failure mode and the selected features extracted from signal processing. Probably the Bearings 2\_4 and 2\_6 had similar fault type than occurred in Bearing 2\_1 and/or 2\_2, and the selected features were relevant to identify this type of fault. It is possible to achieve better performance using the presented models if trained with a greater amount of data that includes several failure modes. Auto-Encoders models could also enhance the ability of extracting useful features.

## References

- Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, 388, 154-170. doi:10.1016/j.jsv.2016.10.043
- Amruthnath, N., & Gupta, T. (2018). *A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance*. Paper presented at the 2018 5th International Conference on Industrial Engineering and Applications (ICIEA).
- Baruah, P., & Chinnam, R. B. (2005). HMMs for diagnostics and prognostics in machining processes. *International Journal of Production Research*, 43(6), 1275-1293. doi:10.1080/00207540412331327727
- Ben Ali, J., Chebel-Morello, B., Saidi, L., Malinowski, S., & Fnaiech, F. (2015). Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. *Mechanical Systems and Signal Processing*, 56, 150-172. doi:10.1016/j.ymssp.2014.10.014
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, 5(2), 157-166. doi:10.1109/72.279181
- Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2013). Remaining useful life estimation based on nonlinear feature reduction and support vector regression. *Engineering Applications of Artificial Intelligence*, 26(7), 1751-1760.
- Canizo, M., Onieva, E., Conde, A., Charramendieta, S., & Trujillo, S. (2017). *Real-time predictive maintenance for wind turbines using Big Data frameworks*. Paper presented at the 2017 IEEE International Conference on Prognostics and Health Management, ICPHM 2017.
- Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. d. P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, 106024. doi:<https://doi.org/10.1016/j.cie.2019.106024>
- Chen, Y., Peng, G., Zhu, Z., & Li, S. (2020). A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. *Applied Soft Computing Journal*, 86. doi:10.1016/j.asoc.2019.105919
- Cheng, C., Ding, J., & Zhang, Y. (2020). A Koopman operator approach for machinery health monitoring and prediction with noisy and low-dimensional industrial time series. *Neurocomputing*, 406, 204-214. doi:10.1016/j.neucom.2020.04.005
- Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, 83(403), 596-610.
- Cui, J., Ren, L., Wang, X., & Zhang, L. (2019). Pairwise comparison learning based bearing health quantitative modeling and its application in service life prediction. *Future Generation Computer Systems*, 97, 578-586. doi:10.1016/j.future.2019.03.026

- Ding, X., Li, Y., Belatreche, A., & Maguire, L. P. (2014). An experimental evaluation of novelty detection methods. *Neurocomputing*, 135, 313-327. doi:10.1016/j.neucom.2013.12.002
- Dudik, J. M., Kurosu, A., Coyle, J. L., & Sejdić, E. (2015). A comparative analysis of DBSCAN, K-means, and quadratic variation algorithms for automatic identification of swallows from swallowing accelerometry signals. *Computers in Biology and Medicine*, 59, 10-18. doi:10.1016/j.combiomed.2015.01.007
- Duong, B. P., Khan, S. A., Shon, D., Im, K., Park, J., Lim, D.-S., . . . Kim, J.-M. (2018). A Reliable Health Indicator for Fault Prognosis of Bearings. *Sensors*, 18(11), 3740.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. Paper presented at the Kdd.
- Fang, X., Zhou, R., & Gebraeel, N. (2015). An adaptive functional regression-based prognostic model for applications with missing data. *Reliability Engineering and System Safety*, 133, 266-274. doi:10.1016/j.res.2014.08.013
- Fink, O., Zio, E., & Weidmann, U. (2014). Predicting component reliability and level of degradation with complex-valued neural networks. *Reliability Engineering & System Safety*, 121, 198-206. doi:<https://doi.org/10.1016/j.res.2013.08.004>
- Fumeo, E., Oneto, L., & Anguita, D. (2015). *Condition based maintenance in railway transportation systems based on big data streaming analysis*. Paper presented at the Procedia Computer Science.
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of machine learning research*, 9, 249-256.
- Gonzalez, E., Reder, M., & Melero, J. J. (2016). *SCADA alarms processing for wind turbine component failure detection*. Paper presented at the Journal of Physics: Conference Series.
- Guo, L., Li, N., Jia, F., Lei, Y., & Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, 240, 98-109. doi:<https://doi.org/10.1016/j.neucom.2017.02.045>
- Hinchi, A. Z., & Tkouat, M. (2018). *Rolling element bearing remaining useful life estimation based on a convolutional long-short-Term memory network*. Paper presented at the Procedia Computer Science.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780. doi:10.1162/neco.1997.9.8.1735
- Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., . . . Van Hoecke, S. (2016). Convolutional Neural Network Based Fault Detection for Rotating Machinery. *Journal of Sound and Vibration*, 377, 331-345. doi:10.1016/j.jsv.2016.05.027
- Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510. doi:10.1016/j.ymssp.2005.09.012
- Kerdsiri, T., & Gullayanon, R. (2017). *Early fault detection based on ball bearing vibration analysis using multinomial logistic regression*. Paper presented at the ACM International Conference Proceeding Series.

- Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, 107, 241-265. doi:10.1016/j.ymssp.2017.11.024
- Kohli, M. (2018). Using machine learning algorithms on data residing in SAP ERP application to predict equipment failures. *International Journal of Engineering and Technology(UAE)*, 7(2), 312-319. doi:10.14419/ijet.v7i2.28.12952
- Kundu, P., Chopra, S., & Lad, B. K. (2019). Multiple failure behaviors identification and remaining useful life prediction of ball bearings. *Journal of Intelligent Manufacturing*, 30(4), 1795-1807. doi:10.1007/s10845-017-1357-8
- Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 104, 799-834. doi:<https://doi.org/10.1016/j.ymssp.2017.11.016>
- Li, H., Wang, W., Huang, P., & Li, Q. (2020). Fault diagnosis of rolling bearing using symmetrized dot pattern and density-based clustering. *Measurement: Journal of the International Measurement Confederation*, 152. doi:10.1016/j.measurement.2019.107293
- Liao, H., Zhao, W., & Guo, H. (2006). *Predicting remaining useful life of an individual unit using proportional hazards model and logistic regression model*. Paper presented at the Proceedings - Annual Reliability and Maintainability Symposium.
- Liao, L., Jin, W., & Pavel, R. (2016). Enhanced Restricted Boltzmann Machine with Prognosability Regularization for Prognostics and Health Assessment. *IEEE Transactions on Industrial Electronics*, 63(11), 7076-7083. doi:10.1109/TIE.2016.2586442
- Lim, C., Kim, S., Seo, Y. H., & Choi, J. H. (2020). Feature extraction for bearing prognostics using weighted correlation of fault frequencies over cycles. *Structural Health Monitoring*. doi:10.1177/1475921719900917
- Liu, C., & Zhang, C. (2019) An elman artificial neural network for remaining useful life prediction. In: *Vol. 1102. Communications in Computer and Information Science* (pp. 176-198).
- Liu, G., Zhao, J., Li, H., & Zhang, X. (2019). Bearing degradation assessment based on entropy with time parameter and fuzzy c-means clustering. *Journal of Vibroengineering*, 21(5), 1322-1329. doi:10.21595/jve.2019.20255
- Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 108, 33-47. doi:10.1016/j.ymssp.2018.02.016
- Liu, Z., Zuo, M. J., & Qin, Y. (2016). Remaining useful life prediction of rolling element bearings based on health state assessment. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 230(2), 314-330. doi:10.1177/0954406215590167
- Lu, W., Li, Y., Cheng, Y., Meng, D., Liang, B., & Zhou, P. (2018). Early fault detection approach with deep architectures. *IEEE Transactions on Instrumentation and Measurement*, 67(7), 1679-1689. doi:10.1109/TIM.2018.2800978

- Luo, B., Wang, H., Liu, H., Li, B., & Peng, F. (2018). Early Fault Detection of Machine Tools Based on Deep Learning and Dynamic Identification. *IEEE Transactions on Industrial Electronics*, 66(1), 509-518. doi:10.1109/TIE.2018.2807414
- Mishra, K. M., Krogerus, T. R., & Huhtala, K. J. (2019). *Fault Detection of Elevator Systems Using Deep Autoencoder Feature Extraction*. Paper presented at the Proceedings - International Conference on Research Challenges in Information Science.
- Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT big data and streaming analytics: A survey. *IEEE Communications Surveys and Tutorials*, 20(4), 2923-2960. doi:10.1109/COMST.2018.2844341
- Nair, V., & Hinton, G. E. (2010). *Rectified linear units improve Restricted Boltzmann machines*. Paper presented at the ICML 2010 - Proceedings, 27th International Conference on Machine Learning.
- Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012). *PRONOSTLA: An experimental platform for bearings accelerated degradation tests*. Paper presented at the IEEE International Conference on Prognostics and Health Management, PHM'12.
- Oliva, J. B., Poczos, B., & Schneider, J. (2017). *The statistical recurrent unit*. Paper presented at the 34th International Conference on Machine Learning, ICML 2017.
- Park, C., & Padgett, W. J. (2005). New cumulative damage models for failure using stochastic processes as initial damage. *IEEE Transactions on Reliability*, 54(3), 530-540. doi:10.1109/TR.2005.853278
- Peng, Y., Dong, M., & Zuo, M. J. (2010). Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology*, 50(1), 297-313. doi:10.1007/s00170-009-2482-0
- Porotsky, S., & Bluvband, Z. (2012). *Remaining useful life estimation for systems with non-trendability behaviour*. Paper presented at the PHM 2012 - 2012 IEEE Int. Conf.on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHM Technology and Application, Conference Program.
- Praveenkumar, T., Saimurugan, M., Krishnakumar, P., & Ramachandran, K. I. (2014). *Fault diagnosis of automobile gearbox based on machine learning techniques*. Paper presented at the Procedia Engineering.
- Principi, E., Rossetti, D., Squartini, S., & Piazza, F. (2019). Unsupervised electric motor fault detection by using deep autoencoders. *IEEE/CAA Journal of Automatica Sinica*, 6(2), 441-451. doi:10.1109/JAS.2019.1911393
- Reddi, S. J., Kale, S., & Kumar, S. (2018). *On the convergence of Adam and beyond*. Paper presented at the 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings.
- Reddy, K. K., Sarkar, S., Venugopalan, V., & Giering, M. (2016). *Anomaly detection and fault disambiguation in large flight data: A multi-modal deep auto-encoder approach*. Paper presented at the Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM.

- Rengasamy, D., Morvan, H. P., & Figueredo, G. P. (2018). *Deep Learning Approaches to Aircraft Maintenance, Repair and Overhaul: A Review*. Paper presented at the IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC.
- Ribeiro De Miranda, A., De Andrade Barbosa, T. M. G., Scolari Conceicao, A. G., & Soares Alcala, S. G. (2019). *Recurrent neural network based on statistical recurrent unit for remaining useful life estimation*. Paper presented at the Proceedings - 2019 Brazilian Conference on Intelligent Systems, BRACIS 2019.
- Rieger, T., Regier, S., Stengel, I., & Clarke, N. (2019). *Fast predictive maintenance in Industrial Internet of Things (IIoT) with Deep Learning (DL): A review*. Paper presented at the CEUR Workshop Proceedings.
- Rousopoulou, V., Nizamis, A., Giugliano, L., Haigh, P., Martins, L., Ioannidis, D., & Tzovaras, D. (2019) Data analytics towards predictive maintenance for industrial ovens: A case study based on data analysis of various sensors data. In: *Vol. 349. Lecture Notes in Business Information Processing* (pp. 83-94).
- Roy, M., Bose, S. K., Kar, B., Gopalakrishnan, P. K., & Basu, A. (2019). *A Stacked Autoencoder Neural Network based Automated Feature Extraction Method for Anomaly detection in On-line Condition Monitoring*. Paper presented at the Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence, SSCI 2018.
- Sadoughi, M., Lu, H., & Hu, C. (2019). *A deep learning approach for failure prognostics of rolling element bearings*. Paper presented at the 2019 IEEE International Conference on Prognostics and Health Management, ICPHM 2019.
- Schwabacher, M., & Goebel, K. (2007). *A survey of artificial intelligence for prognostics*. Paper presented at the AAAI Fall Symposium - Technical Report.
- Shao, H., Jiang, H., Wang, F., & Zhao, H. (2017). An enhancement deep feature fusion method for rotating machinery fault diagnosis. *Knowledge-Based Systems, 119*, 200-220. doi:<https://doi.org/10.1016/j.knosys.2016.12.012>
- Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D.-H. (2011). Remaining useful life estimation – A review on the statistical data driven approaches. *European Journal of Operational Research, 213*(1), 1-14. doi:<https://doi.org/10.1016/j.ejor.2010.11.018>
- Sikorska, J. Z., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing, 25*(5), 1803-1836. doi:10.1016/j.ymssp.2010.11.018
- Su, C. J., & Huang, S. F. (2018). Real-time big data analytics for hard disk drive predictive maintenance. *Computers and Electrical Engineering, 71*, 93-101. doi:10.1016/j.compeleceng.2018.07.025
- Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics, 11*(3), 812-820. doi:10.1109/TII.2014.2349359
- Sutrisno, E., Oh, H., Vasan, A. S. S., & Pecht, M. (2012). *Estimation of remaining useful life of ball bearings using data driven methodologies*. Paper presented at the PHM 2012 - 2012 IEEE Int. Conf.on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHM Technology and Application, Conference Program.

- Vachtsevanos, G. J., Lewis, F., Hess, A., & Wu, B. (2006). *Intelligent fault diagnosis and prognosis for engineering systems* (Vol. 456): Wiley Hoboken.
- Verstraete, D., Ferrada, A., Droguett, E. L., Meruane, V., & Modarres, M. (2017). Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Shock and Vibration*, 2017. doi:10.1155/2017/5067651
- Vichare, N. M., & Pecht, M. G. (2006). Prognostics and health management of electronics. *IEEE Transactions on Components and Packaging Technologies*, 29(1), 222-229. doi:10.1109/TCAPT.2006.870387
- Wang, F. K., & Mamo, T. (2019). Hybrid approach for remaining useful life prediction of ball bearings. *Quality and Reliability Engineering International*, 35(7), 2494-2505. doi:10.1002/qre.2538
- Wang, H., Ni, G., Chen, J., & Qu, J. (2020). Research on rolling bearing state health monitoring and life prediction based on PCA and Internet of things with multi-sensor. *Measurement: Journal of the International Measurement Confederation*, 157. doi:10.1016/j.measurement.2020.107657
- Wang, H., Yu, Z., & Guo, L. (2020). *Real-time Online Prediction of Data Driven Bearing Residual Life*. Paper presented at the Journal of Physics: Conference Series.
- Xiao, L., Chen, X., Zhang, X., & Liu, M. (2017). A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition. *Journal of Intelligent Manufacturing*, 28(8), 1893-1914. doi:10.1007/s10845-015-1077-x
- Zhang, B., Zhang, S., & Li, W. (2019). Bearing performance degradation assessment using long short-term memory recurrent network. *Computers in Industry*, 106, 14-29. doi:10.1016/j.compind.2018.12.016
- Zhang, L., Lin, J., Liu, B., Zhang, Z., Yan, X., & Wei, M. (2019). A Review on Deep Learning Applications in Prognostics and Health Management. *IEEE Access*, 7, 162415-162438. doi:10.1109/ACCESS.2019.2950985
- Zhang, W., Yang, D., & Wang, H. (2019). Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. *IEEE Systems Journal*, 13(3), 2213-2227. doi:10.1109/JSYST.2019.2905565
- Zhou, X., Zhang, X., & Wang, B. (2016) Online support vector machine: A survey. In: *Vol. 382. Advances in Intelligent Systems and Computing* (pp. 269-278).
- Zhu, J., Wang, C., Li, C., Gao, X., & Zhao, J. (2016). Dynamic alarm prediction for critical alarms using a probabilistic model. *Chinese Journal of Chemical Engineering*, 24(7), 881-885. doi:<https://doi.org/10.1016/j.cjche.2016.04.017>
- Zschech, P., Heinrich, K., Bink, R., & Neufeld, J. S. (2019). Prognostic Model Development with Missing Labels: A Condition-Based Maintenance Approach Using Machine Learning. *Business and Information Systems Engineering*, 61(3), 327-343. doi:10.1007/s12599-019-00596-1

## Annex 1 – 3 Health Stages Analysis

Three groups were identified on each training data set. The bearing 2\_1 has 102 (11%) of observations assigned as anomaly (Cluster 0), and the remaining samples divided in two groups. The bearing 2\_2 has 10% of possible anomalies.

|        | 0   | 1   | 2   |
|--------|-----|-----|-----|
| border | 102 | 12  | 22  |
| seed   | 0   | 143 | 632 |
| total  | 102 | 155 | 654 |

|        | 0  | 1   | 2   |
|--------|----|-----|-----|
| border | 90 | 20  | 22  |
| seed   | 0  | 190 | 475 |
| total  | 90 | 210 | 497 |

Table 14: DBSCAN Results - Quantity of normal and anomaly samples. Bearing 2\_1 in the left and Bearing 2\_2 in the right.

Both bearings presented the same quantity of clusters, which suggests three different health stages for the bearing life, see Figure 19. Colored rectangles reflect the different clusters regarding the health indicators curve. Clearly, the cluster 1 indicated by the color yellow represents the early phase of bearing life, where the health indicators present a low volatility. The cluster 2, in blue, suggest a stable phase of operation. The cluster 0 frequent in the end of the bearing life, characterized by critic change in the curves may indicates the anomaly condition. Unexpexted, Bearing 2\_2 presents samples on the anomaly cluster in eraly phase.

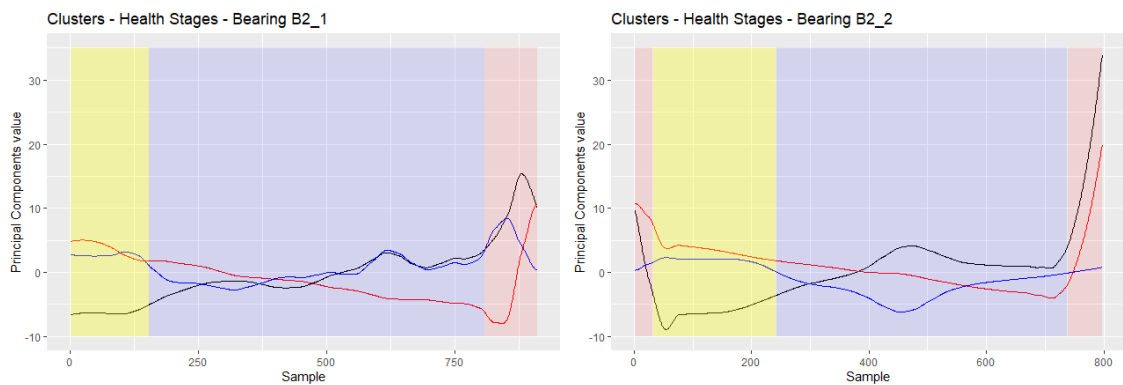


Figure 29: Clusters - Health Stages

## **Annex 2 – Work Style**

Estilo da referência bibliográfica: APA 6th.