An Intelligent Predictive Maintenance Approach Based on End-of-Line Test Logfiles in the Automotive Industry

David Vicêncio[1,2,3](✉), Hugo Silva[2], Salviano Soares[1,4], Vítor Filipe[1,3], and António Valente[1,3]

[1] University of Trás-os-Montes and Alto Douro, Vila Real, Portugal
{salblues,vfilipe,avalente}@utad.pt
[2] Continental Advanced Antenna, Sociedade Unipessoal, Lda., Vila Real, Portugal
{david.2.vicencio,hugo.2.silva}@continental.com
[3] INESC TEC, Porto, Portugal
[4] IEETA, Aveiro, Portugal
https://www.inesctec.pt/
http://www.ieeta.pt/

Abstract. Through technological advents from Industry 4.0 and the Internet of Things, as well as new Big Data solutions, predictive maintenance begins to play a strategic role in the increasing operational performance of any industrial facility. Equipment failures can be very costly and have catastrophic consequences. In its basic concept, Predictive maintenance allows minimizing equipment faults or service disruptions, presenting promising cost savings. This paper presents a data-driven approach, based on multiple-instance learning, to predict malfunctions in End-of-Line Testing Systems through the extraction of operational logs, which, while not designed to predict failures, contains valid information regarding their operational mode over time. For the case study performed, a real-life dataset was used containing thousands of log messages, collected in a real automotive industry environment. The insights gained from mining this type of data will be shared in this paper, highlighting the main challenges and benefits, as well as good recommendations, and best practices for the appropriate usage of machine learning techniques and analytics tools that can be implemented in similar industrial environments.

Keywords: End-of-Line Testing Systems (EOL) · Service monitoring and log operation data · Industry 4.0 · Predictive Maintenance (PdM) · Machine learning (ML)

Supported by Continental Advanced Antenna (CAA), Portugal *(a Continental AG Group Company).*

L. Peñalver and L. Parra (Eds.): Industrial IoT 2020, LNICST 365, pp. 121–140, 2021.
https://doi.org/10.1007/978-3-030-71061-3_8

1 Introduction

Nowadays, facing through a so globalized and competitive environment, industries are constantly challenged to look for new ways to differentiate themselves and improve their effectiveness and efficiency, in order to remain competitive and sustainable. Maintenance management policies are a relevant aspect in the evaluation of the production processes of any industrial plant, and, therefore, it has become an interesting and relevant topic of scientific research. The effectiveness and internal efficiency of an industrial plant can be directly influenced by the role of maintenance and its impact in other areas, as *e.g.*, production and quality processes. According to [1], between 15% and 70% of the total production costs comes from maintenance activities. However, the vast majority of industries still opt for outdated and inefficient maintenance policies, which, consequently, leads to a loss in productive quality and efficiency [2]. Today, with the technological advances that have been made in areas such as Computer Science (CS) and Artificial Intelligence (AI), as well as the increased usage of sensors and monitoring systems, Predictive Maintenance (PdM) approaches can be made in any industrial equipment, allowing a forecast relatively reliable about its Remaining Useful Life (RUL). PdM is relatively of interest in the environment of Industry 4.0 and it can increase drastically the efficiency of modern production facilities. The usage of monitoring sensors provides us a set of historical data on the operational state of industrial equipment, as well as corrective actions that may have occurred. This information is extremely valuable for PdM actions or models, which use both historical and operational data to make diagnostics and predictions, as presented by [1–3].

The possibilities of PdM applications connected with other parts of the production process are increasing, now even more in the context of Industry 4.0 and the Internet of Things (IoT) [3–5]. Generally, PdM models run through prior knowledge about the equipment's normal operation. This knowledge can be acquired through the installation of monitoring sensors for recording relevant signals of degradation, as *e.g.*, vibration, temperature, or voltage. Sometimes, the inclusion of sensors or major hardware upgrades is impractical on industrial equipment, due to expensive implementation costs, effort, or regulatory hurdles. In alternative, we can obtain knowledge about the functioning regions of a device by studying its logs. For example, in the case of an End-of-Line (EOL) testing system, all device operations, from measurement, test environment, signal analysis or calibration, are controlled through different applications. These applications can produce tons of logs during their operation. Theoretically, it's possible to understand the normal operating regions of industrial equipment through the study of its logs. The extraction and analysis of such information through the usage of Analytical methods can help in the preventive detection of system disruptions. The usage of logfiles as a way to prevent equipment malfunctions has not yet been fully explored and is still posing computational challenges given its complex structure and its huge amount of data [5–7]. Later on, in the next sections, as our main contributions, we will focus on the way we formulate our problem, from data gathering to the algorithm selection and results, through a case study developed in a real automotive industry facility (CAA).

This paper is structured in seven sections, according to the following: Sect. 1 gives a brief introduction to the main topic, exploring the principal goals intended; By Sect. 2, we will address the topic with some recent publications, where other works, recently presented by other authors dealing with PdM will be presented and discussed; In Sect. 3, a review of the state of the art is made, trying to fit the topic with some European Norms (EN) and recent publications, that support this paper; The methodology applied in this study is described in Sect. 4; The case study experiments, through the usage of a real operational dataset, obtained from an EOL testing system, are described in Sect. 5; Principal conclusion are shared on Sect. 6, as well as the list of future work in this actuation area is explored.

2 Related Works

This section reviews the related work on data-driven approaches, based on multiple-instance learning to predict equipment or service disruptions. For the majority of the scientific papers published recently, the theme of PdM is approached as a way of assisting in the attentive detection of equipment failures, which are generally being monitored in critical systems. PdM uses a variety of approaches and ML techniques to study data in order to identify and interpret abnormal operating patterns. As *e.g.*, Z. Li *et al.* [8], attempt an approach to predict the RUL of an aircraft engine based on advanced ensemble algorithms, as *e.g.*, Random Forests (RF), Classification and Regression Tree (CART), Recurrent Neural Networks (RNN) and Relevance Vector Machine (RVM). In his study, a variable selection approach using RF was used to determine an optimal set of variables, whose approximation served as the basis for this paper. Also, Z. Weiting [4], conduct a comprehensive survey of PdM in industrial equipment. This publication, propose a PdM scheme for automatic washing equipment and compares different models and algorithms accuracy from the aspects of ML and Deep Learning (DL) techniques. The study was conducted collecting data from four type of sensors that should represent the equipment performance (*i.e.*, vibration, temperature, electrical signal and rotating speed). One of the main challenges of predictive models, used in maintenance purposes, is that they are based on the assumption that there are certain contexts in the equipment lifetime where the failure rate is increasing. The vast majority of PdM studies that have been made, focus on the application in mechanical components or induction motors, whose specificities are directly influenced by several known factors, as *e.g.*, increasing operating temperature or vibrations. Also, J. Yan et al. [1], tried to predict the RUL of key components of a CNC machining center, where the input variables collected were vibration, acoustical signals, and power data. With the increasing of IoT systems and the more frequently usage of sensors, it's becoming easy to collect pre-treated information about the status of any device over a certain time-frame. In industrial operational context, there are patterns in which the failure probability does not increase, but remains constant during the equipment lifetime, and therefore the equipment can fail at any time, as *e.g.*, electrical and electronic components. There is also the case where the direct factors or variables that may influence the operating regions of an equipment are unknown, or even the case where it's impractical to place sensors to collect information. One of the notable previous works dealing with industrial log analysis to predict anomalies is that of R. Sipos *et al.* [9], in which main contributions served us as a guide in the methodology chosen. However, in our case, the impossibility to calculate the RUL of EOL systems due to the insincerity of data coming from the in-service maintenance database (DB), lead us to do a different approach. Our approach can then be applied in industrial cases, where the collection of logging data is available and whose analysis of such information represents a viable low cost alternative compared to the installation of sensors, or even in the impracticality of installation. Log data can be used to understand operational regions of any equipment and understand common patterns. The conduct of our study was based on unsupervised clustering techniques to analyse operational logfiles and then use Principal Component Analysis (PCA) algorithms to identify the variables, that we assume, with the greatest impact on possible anomalies.

3 State of the Art Review

In this section, a review of the state of the art will be carried out, through the approximation to the main European Standards (Std.), serving as support for the framework and justification of the problem in the next sections.

3.1 Maintenance Approach Classification

Maintenance has suffered different evolutions from the first corrective approaches to nowadays more complex predictive applications. So, according to the principal European Norms (EN), the NP EN 13306:2007 Std. [10], the theoretical and principal foundations of maintenance approaches could be divided, essentially, according to three different categories, *i.e.*, corrective, preventive and predictive maintenance. As a summary, Fig. 1 shows a high-level maintenance approach classification.

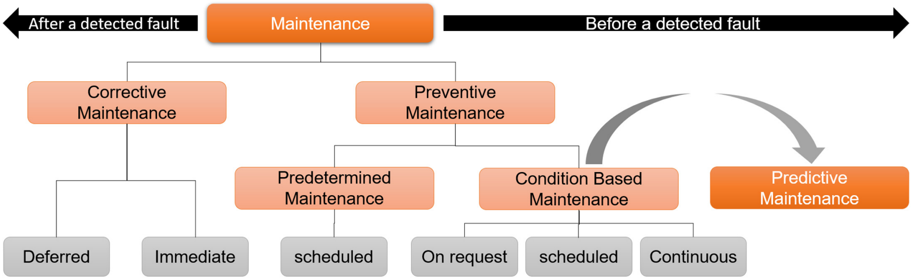


Fig. 1. Maintenance approach classification [2].

1. **Corrective Maintenance:** Taking as reference what is shown in Fig. 1, we can say that the corrective maintenance could be classified according two different categories, *i.e.*, immediate and differed. According to [10], and regarding the first category, it is said that this is an "**emergency maintenance** that is made immediately after a detection of a state of fail, to avoid unacceptable consequences". It is also intended as reactive maintenance. Regarding the second category, the **differed maintenance** is the maintenance that is not immediately made after the detection of a state of fail, but is retarded according to the policies of maintenance established.
2. **Preventive Maintenance:** This type of maintenance could be classified in two main categories, *i.e.*, predetermined maintenance and condition based-based maintenance (see Fig. 1). Regarding the **predetermined maintenance**, it's defined by [10] as a "preventive maintenance carried out at pre-established time intervals or according to a maximum number of uses allowed, but without prior control of the condition". Regarding the **condition-based maintenance**, it's defined by [10] as a "preventive maintenance which includes a combination of condition monitoring and/or inspection and/or testing, analysis, and the ensuing maintenance actions". This technique is fundamentally used to collect data from equipment in order to execute a better repair approach. Some techniques used in this type of maintenance are, as *e.g.*, the monitoring of vibrations, performance, temperature, humidity, tribology, or even visual inspections and tracking of Key Process Indicators (KPI) for health management to discover trends that lead to abnormal operating conditions.
3. **Predictive Maintenance:** As represented schematically in the maintenance approach classification diagram (see Fig. 1), and according to [10], PdM it's defined as a "condition-based maintenance carried out following a forecast derived from repeated analysis or known characteristics and evaluation of the significant parameters of the degradation of the item". This type of maintenance is a sub-classification of the condition-based maintenance. PdM uses a

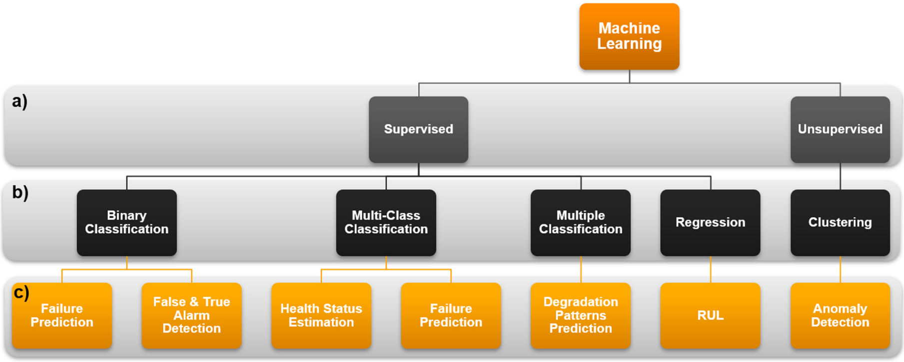


Fig. 2. ML techniques for PdM approaches: (a) ML approaches, (b) ML Techniques and (c) PdM information [2].

set of different ML techniques to learn with historical and recent operational data and make accurate prediction failures in specific equipment. The application of PdM is possible when any equipment has a "symptom" that can characterize the system failure.

3.2 Development Techniques of Data-Driven Models for Maintenance Purposes

The first step of the development phase of a learning model is to understand the business requirements or the specific application context in a more comprehensive way. According to [9,11,12], we can classify different ML approaches in order to obtain a specific PdM result. This information is shown in Fig. 2. The basic steps of a data-driven approach for PdM applications are discussed in this section.

1. **Data Collection and Transformation:** The collection of data is the process of gathering data from relevant sources regarding the analytical problem, as *e.g.*, raw data from monitoring sensors or even operational logfiles. The principal of predictive modelling is to learn patterns from historical data, based on statistical trends, and try to predict future outcomes based on these observations. Collected data can be very complex and the efficiency of a predictive model is directly influenced by the quality of the input data. So, before applying any learning model, it's important to explore it first. The exploration of data can lead us to understand some hidden trends in the data. According to [9,11,12], we can divide data transformation methods in two main tasks, *i.e.*, **processing** and **data analysis**.

2. **Algorithm Selection:** In its concept, according to [9,11,12], "predictive modelling is a mathematical approach to create a statistical model to forecast future behaviour based on input test data". Predictive Modelling can be classified in three main phases, *i.e.*, algorithm selection, training model and evaluation. As shown in Fig. 2, ML techniques are essentially classified in two main approaches: (a) **supervised learning**, where the information about the occurrence of failures is present in the modelling dataset; and (b) **unsupervised learning**, where there is no historical data of maintenance operations. Supervised learning techniques can be grouped in two main areas: (i) **classification methods**, which are generally used in PdM to classify groups of normal and abnormal operation, *i.e.*, Decision Trees, RF, Nearest Neighbors (NN) or Support Vector Machine (SVM); and (ii) **regression methods**, as *e.g.*, LASSO Regression, are usually employed to predict the RUL of equipment. On the other hand, unsupervised learning models, combining **clustering techniques** are valid arguments to detect patterns, behaviours or deviations in what are considered the regions of normal operation. Some unsupervised learning techniques are, as *e.g.*, K-means and PCA.
3. **Model Training and Evaluation:** In order to test the performance of a built model, according to [2,3], there are two types of evaluation metrics capable of quantify its applicability: (a) **offline**, that is generally employed on training datasets, using historical data to obtain performance metrics like F1-score, precision and recall; and (b) **online**, that is used to estimate business metrics through real-time data in deployed models. According to [2,3], a good practice should be to split data sampling into training and cross-validation datasets. In the case of models that learn from very unbalanced datasets, the confusion matrix is also applied as a way to understand, with more detail, the percentage of correct and incorrect classifications. Some confusion matrix insights are, as *e.g.*, the AUC-ROC Curve, F1-score, precision, recall and accuracy.

3.3 Ensemble Learning-Based Predictive Modelling

According to [8,13–15], ensemble learning methods could be defined as "meta-algorithms that combine multiple base learners into a single predictive model in order to improve prediction performance". Ensemble learning methods can be classified into two main categories, *i.e.*, parallel and sequential ensemble methods. The parallel ensemble techniques, as *e.g.*, PCA or RF, generates parallel base learners independently and then average their predictions or take a weighted sum of their predictions. **Parallel ensemble methods** can be implemented, for *e.g.*, using a Dempster–Shafer framework [16] or a Bayesian [17] where the weights are interpreted as probabilities. In the other hand, **sequential ensemble methods** such as AdaBoost construct base learners sequentially and then reduce the bias of the combined base learners. In our case study, the development of the ensemble learning algorithm was based in PCA. This framework is sequenced in four phases: (a) variable selection; (b) model training; (c) model validation; and (d) test. A greater weight will be assigned to the base learner with better performance.

3.4 End-of-Line (EOL) Testing Systems

The development of this paper was made in CAA, an industrial automotive facility, dedicated and specialized in the development of automotive intelligent antennas. Therefore, considering the example of the automotive industry where this case study was developed, in all the Final Assembly (FA) area, there are test equipments, or machines, that measures all the functionalities of each antenna produced at the plant. These equipments, End-of-Line (EOL) Testing Systems (see Fig. 3), are responsible for one of the most crucial parts of the production process, once they perform the screening of defective parts, or, whose measurements do not fall within the limits considered as accepted for the final customer. In this way, the least number of EOL faults, interventions or unavailability, is crucial for the manufacturing process.

Fig. 3. Example of a real EOL test system station. EOL MT460 Motortester [18].

Essentially, we can say that EOL maintenance is accomplished considering: (a) the maximum number of cycle operations recommended by the manufacturer - preventive maintenance; or (b), when physical anomalies are detected in the test equipment - corrective maintenance. In the second case, this could lead to a considerable increase in the number of parts evaluated incorrectly, increasing the First Time Quality (FTQ), as well as an increase in the number of secondary evaluations, and, in the last case, the possibility of shipping defective parts to the customer.

There are some factors that can lead to the predictive detection of EOL anomalies: (a) temperature increase; (b) number of EOL cycle operations;

(c) impurity detection on the test needles; (d) detection of patterns or deviations in the curves from the electrical test signals originated by the antennas; (e) Angle of contact with the test needles.

However, as most of these equipments are working 24×7, it's impractical to place physical sensors, capable of collecting information that can be considered useful for the preventive detection of malfunctions. Thus, the analysis structure of this paper will be based on the collection of information present in the logfiles, produced during their operation.

4 Problem Description

In this section we will describe, in detail, the methodology applied, the logging structure, the PdM approach and the requirements take in consideration on this paper.

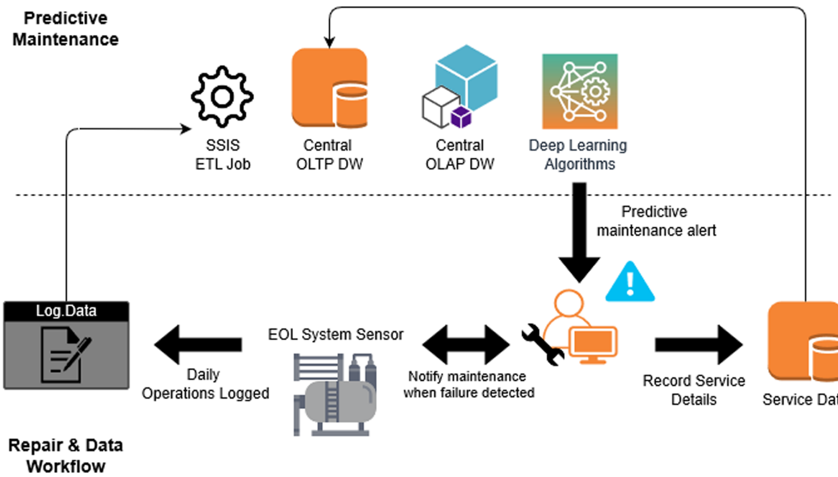


Fig. 4. EOL operation life cycle and PdM workflow interactions.

4.1 Log Data Description

During the normal life cycle service, the EOL operates normally until a detection of a state of fail. When the problem is identified in the FA area by the operator, the maintenance service team is notified and then a repair service is created/scheduled in the maintenance service DB. After the intervention, the maintenance technician updates the scheduled service details, such as component consumption, time spent, repair description, and then closes the service intervention. During the whole process, the EOL still records all component testing information into logfiles. This life cycle (bottom part of Fig. 4) gets repeated over and over for all the FA area in more than fifty equipment units. The dataset used in this study contains all the information present in the logfiles, collected in the past recent years[1]. Logfiles recorded during the EOL's operation (see Fig. 5), are, essentially, a collection of measure points, recorded during the test of each antenna produced. So, essentially, for each antenna produced on the plant, there's a logfile with its measurement results. In detail, EOL's logfile are constituted by: (a) a timestamp (indicating when an antenna was tested); (b) the partnumber and serial number of the antenna tested; (c) the EOL number; (d) test specification, name and results; and (e) the antenna final electrical result (indicating if the antenna is according to the std., and then, can be shipped to the client). In the EOL's normal operation, for each different antenna model (partnumber), there's different test specifications, and each EOL can test more than one antenna partnumber.

Logfiles are differential in some aspects, they are temporal, and can be seen both as symbolic sequences and as numeric time series, with some specific deviations from test signals over some window, *e.g.* time or days. EOL logfiles can give us important feedback, when for a specific measurement there's deviation considering the all population. The old-fashioned PdM approach in this domain, is to manually create PdM patterns for a specific electrical test, based on the boolean combination of a few relevant samples. This approach is heavily time consuming and experience-based, but it's representative that component failures (*e.g.* test needles) can be predicted by checking daily logs for patterns, consisting in the EOL's electrical curve results.

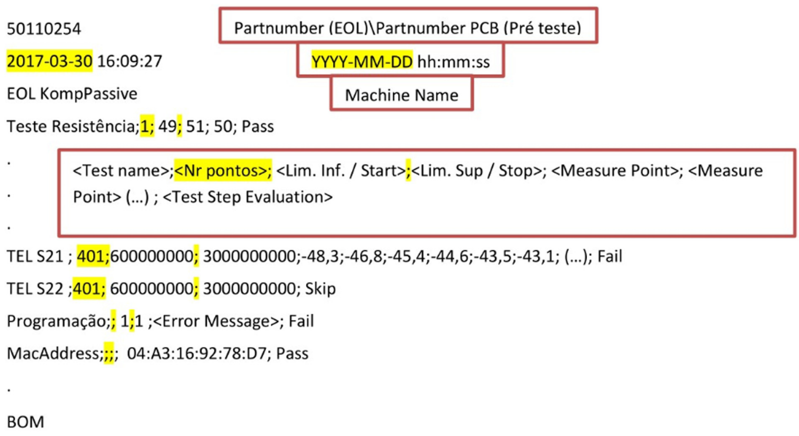


Fig. 5. Example of a real std. logfile from an EOL testing operation.

[1] To demonstrate this approach, the authors used datasets from in-service EOL devices, nevertheless, this methodology can also be applied to other research fields, as *e.g.*, IT infrastructures or any industrial equipment, in which such sort of logs could be collected.

4.2 PdM Approach

The PdM approach adopted in this study can be seen in the top part of Fig. 4. A central DB is at the core of the PdM workflow. This central DB combines the information that was originated from the EOL's operation, as well as the information recorded about the history of disruptions and maintenance support from the in-service DB. The ML module processes data from the central DB. The analytics methodology used in this paper was based on the approach made by [9], and it can be classified into four different stages: (a) data acquisition and transformation; (b) model building; (c) model evaluation; and (d) monitoring. In order to build a model, the analytics module first loads the relevant data from the central DB, extracts predictive features (see Sect. 5), forms and represents the data in a matrix format for learning algorithms. Learning algorithms can then build the model and pass it for evaluation. To do this evaluation, the model needs to be compared against historical known data and scored against Key Performance Indicators (KPI). After evaluation, the model needs to be revised or trained until its acceptable performance is considered to be placed in production environment. In the monitoring stage, the system pulls the new daily logfiles from the EOL system and predict a component failure using a model. If, the predicted score exceeds a predefined threshold, an alert will be sent to the maintenance support in order to make some visual inspections locally in the equipment suspected of malfunction.

5 Case Study

This work was conducted in a complex intelligent manufacturing system, CAA, in order to detect patterns in test measurements coming from EOL systems, in an attempt to predict eventual failures. More information about EOL systems can be found in Sect. 3. The proposed case study can be divided essentially in two different parts, but are complementary to each other. The first part of the project consists in the data acquisition and archive solution of the logfiles, produced during the EOL's operation. This solution aims to fill one of the existing fails in maintaining the history of the test curves, necessary for the analysis and development of new products, as well as in the case of complaints from customers for specific antennas. This part is essential for the next step, once that the archive DB was used in order to develop ML techniques for the EOL PdM.

5.1 Methodology and Materials

The range of Software (Sw) used in this work were chosen taking into account the Sw licensing available, and in compliance, according to the Std. defined by the company. In this sense, regarding the data acquisition, the Sw used for the Extractions, Transformations and data Loads (ETL), was the Microsoft SQL Server Integration Services (SSIS) with SQL Server Data Tools (SSDT) 2015, integrated with data storage, performed in a SQL Server 2016 (RTM-GDR) DB.

Regarding the reporting development, the v13 of QlikSense was the Sw chosen. Lastly, for data analytics and machine learning algorithms, the Sw used was MATLAB R2019b. Both solutions are running in different servers, with the same characteristics, this is, Windows Server 2016 with 64 GB of Ram and Intel(R) Xeon(R) CPU ES-2667 v4. The methodology approach and problem description where already addressed on this paper, and can be found in Sect. 4 and Fig. 4.

5.2 Data Acquisition and Preparation

The data acquisition methodology implemented in this case study, can be found in Fig. 6. Logfiles are collected from two different workflows. According to Fig. 6, in (1), the EOL testing system directly stores the antenna test results (raw data) on a primary Staging Area (SA) DB, avoiding the usage of text (.txt) files. This data acquisition is fully integrated with the traceability system, the Manufacturing Execution System (MES). For this purpose, specific .dll methods were developed, in order to get data automatically from the EOL's operation. On the other hand, for the FA workcenters with no MES traceability system, the .txt logfiles are imported by SSIS into the primary SA DB (2). In (3,4), a Daily Job is collecting the last 24h of information (stored on different SQL Server instances) into the Archive DB, and cleaning redundant info from the SA DB. Data should be cleaned, transformed and stored into a Data Warehouse (DW) containing the last 3 week's info. In (5), the treated info is imported from the DW DB into .Qvd files, for in-memory reports; (6) Qlik Report available for usage; (7) Web App for Archive Solution. Search for period or antenna should be available. User Alert (via e-mail) after solution reloaded. (8) MES DataSErver does all the validations on production environment and inserts the results at PMSE DB.

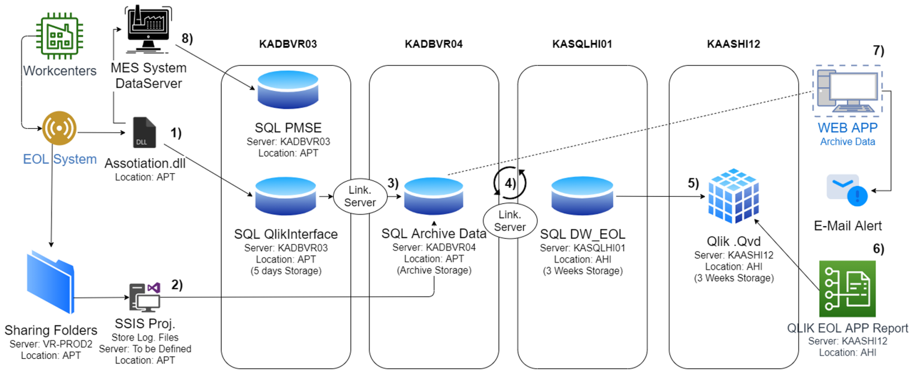


Fig. 6. EOL logfile data acquisition, data storage and archive solution diagram.

5.3 Dataset Overview and Principal Component Analysis

After the first phase of the project complete, it was intended to find a dataset that could fit the overall behaviour of the EOL on the plant. The dataset overview that was chosen for this use case, can be consulted on Table 1. For this analysis, the EOL18 was the chosen one, and the dataset overall includes the collection of information regarding the operation of this equipment, from October 2019 to April 2020. During this period, this equipment performed the validation test for 12 different antenna partnumbers, covering a total of 162.744 produced antennas (see Table 1). It should also be noticed that the overall value for the First Time Quality (FTQ) indicator, for this equipment, reached values above 95%. Achieving, by that, a good ratio between the number of parts rejected during the manufacturing process, versus the total number of pieces attempts.

Table 1. Dataset overview for EOL18, total tested parts by partnumber and FTQ.

Station	Partnumber	Total tested parts $[inUn.]$	FTQ $[in\%]$
EOL18	011031	23.838	99.0
EOL18	011032	22.902	97.9
EOL18	011038	5.624	94.9
EOL18	011039	14.287	98.0
EOL18	011065	11	81.8
EOL18	011066	10	80.0
EOL18	020013	112	95.5
EOL18	020014	2	50.0
EOL18	226944014	110	99.1
EOL18	50010990	1.417	95.4
EOL18	50010991	94.151	93.7
EOL18	52512577	281	94.3
Total		**162.744**	**95.5**

In the model training stage, not all the variables are useful. On the contrary, some variables may even reduce prediction accuracy, because these variables may not be correlated to the degradation behaviour of EOL test systems. To select the most effective variables, PCA mathematical procedures were used to measure the importance of measurement variables with respect to their performance on prediction accuracy. More information about PCA and RF can be consulted on Sect. 3. The list of variables considered in the dataset used in this case study are presented on Table 2.

Table 2. List of variables considered in the case study dataset.

Var. ID	Symbol	Description	Unit
1	RTR	Resistance Test Results	(Ω)
2	TTR	Test Time Results	(s)
3	RKECR	RKE Curr. Results	(A)
4	RKETXFR	RKE TX Freq ECE Results	(Hz)
5	RKETXPR	RKE TX Power Results	(db)

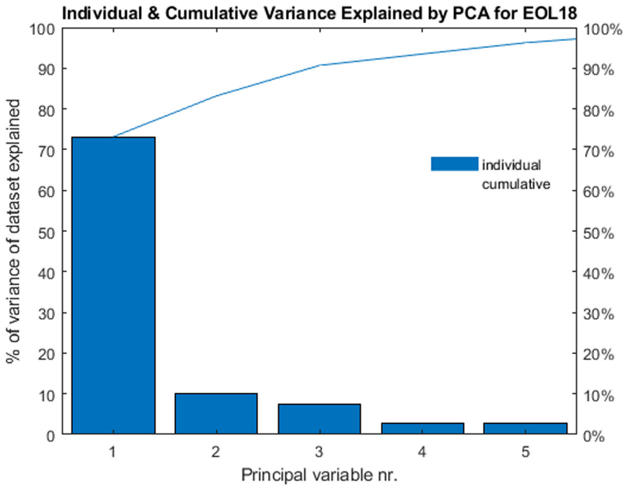


Fig. 7. Individual and cumulative variance on principal variables explained by PCA for EOL18.

Figure 7 illustrates the variable importance. Based on this criterion, the most important variable is the RTR (the variable importance is above 70%). The least important variable is RKETXPR (the variable importance is lower than 5%). In this study, only the three variables with the highest percentage of importance were considered on the analysis, covering a percentage, representative of the all dataset, above 90%.

As shown in Fig. 7, the measurement variables, including RTR, TTR and RKECR, were selected for the predictive models. This result is partially consistent to the methods applied on [1,6], where three measurement variables were selected according to the lognormal distributions of the measurement data.

Figure 8 shows the correlation between the three principal variables, resulted from PCA methods. Notice that only five groups of different partnumber where considered on the analysis, covering the majority of the test measurements performed by the EOL18. So, according to Table 1, only the partnumber: 01103; 011032; 011038; 011039 and 50010991 where chosen, corresponding to data1 until data5, respectively, on Fig. 8. Analysing the Fig. 8, it can be seen that the

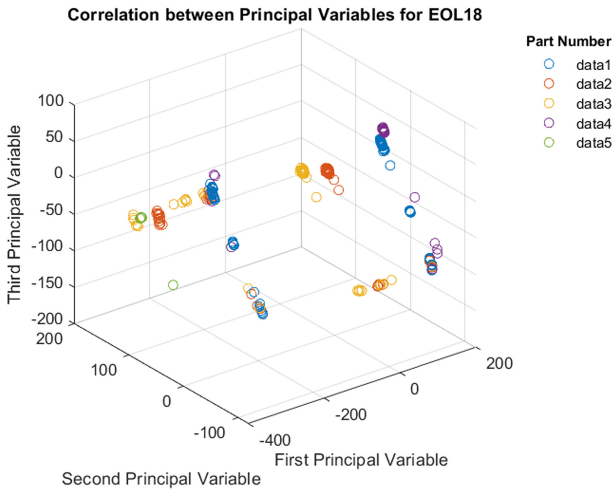


Fig. 8. Correlation between three principal variables for EOL18, considering five different datasets, representing five different antenna partnumber.

behaviour is quite similar for each of the dataset samples. The normal functioning region is in a cluster with a larger number of samples, with some clusters forming around it, with a less concentrated population, moving from the normal operating region. This type of behaviour could be better understood through a two-dimensional analysis, considering only the first two principal variables from the dataset (RTR and TTR).

Figure 9 and Fig. 10 shows how the collection data evolved over time, regarding the principal variables, RTR and TTR, respectively. Notice that, analysing Fig. 9, it's possible to verify, in spite of the noise that is noticed, a pattern in all the analysed partnumber (*i.e.* 011031, 011032, 011038 and 011039). Analysing the RTR variable, its possible to see a peak and a constant increase in the monitored values. This constant increasing in the monitored values may be representative of a malfunction, as well as wear of parts, or damage, caused by improper operation or maintenance.

5.4 EOL Operation Regions Classification and RUL

As planned in Sect. 4 and described in Fig. 4, it was intended on this scope, to collect data of in-service DB, with precious information about the history of faults and maintenance schedules. However, this analysis has become unsuccessful, due to the uncertainty of the records collected and the inaccuracy of the historical data, since they are manual input data. In this sense, without the inclusion of this data in the ML algorithm, it became impossible to implement and detect the RUL of the equipments under analysis. Thus, according to Fig. 2, a different approach was taken into account, as it is not possible to calculate the

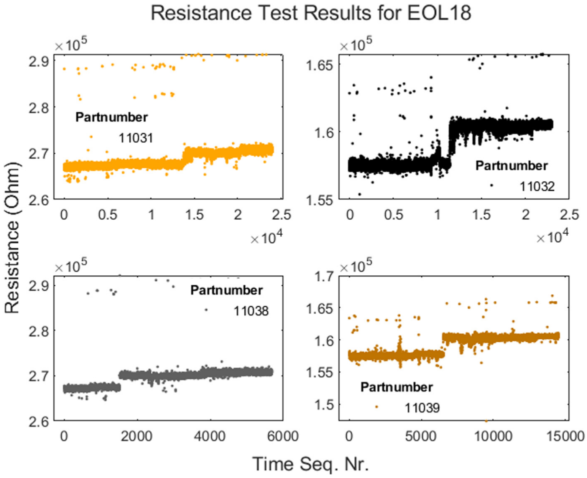


Fig. 9. Resistance test results discretization for EOL18, considering four different samples.

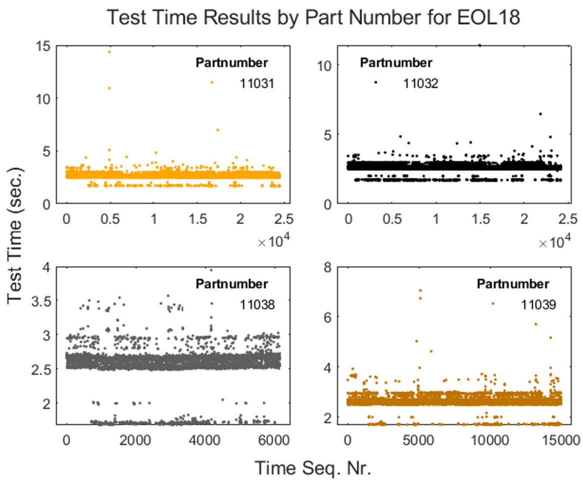


Fig. 10. Operation test time results discretization for EOL18, considering four different samples.

RUL, but the detection of anomalies through the usage of Unsupervised Learning algorithms. Through the analysis of the correlation between the first two principal variables resulted from PCA, it was possible to trace the normal and degradation regions of operation on EOL18. Notice that for the construction of Fig. 11, only the analysis of the results obtained for a partnumber (data1) were taken into account, in which it's possible to see the formation of different clusters moving away from the normal region of operation.

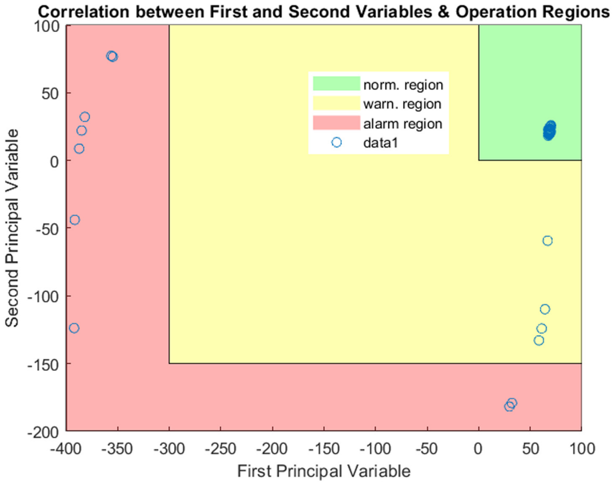
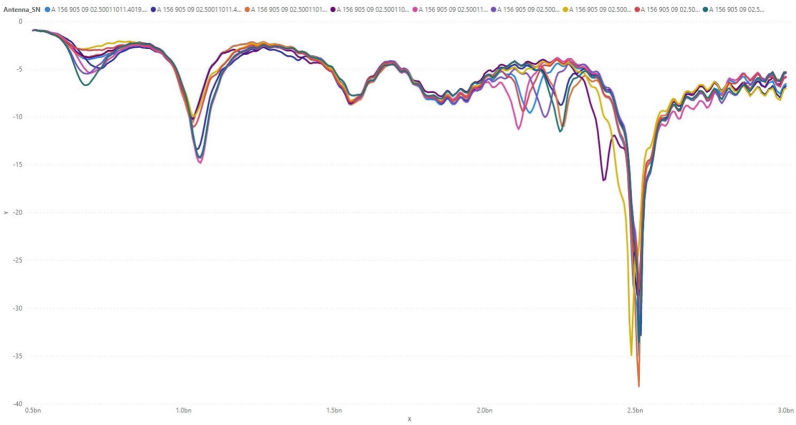


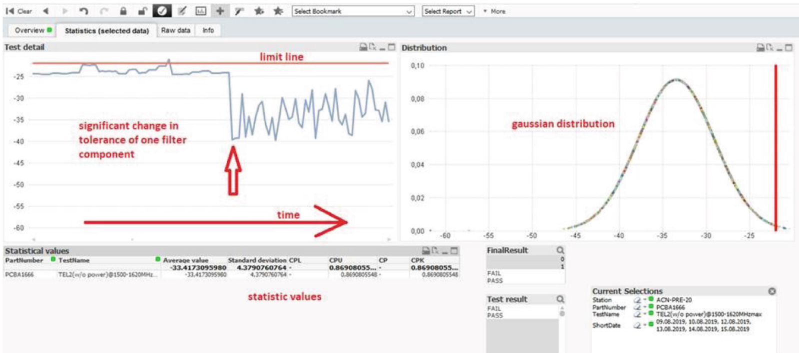
Fig. 11. Correlation between the first two principal variables for dataset1, on EOL18, and different operation regions (*i.e.*: normal, warning and alarm regions).

5.5 EOL Testing Curves Dashboard and Statistical Measurements Information

With the automation made with the data acquisition, it was possible to reconstruct the test signal curves, resultant for each antenna. Despite not being the main part of the project, this phase made it possible to fill one of the main flaws in the storage of information of this type, necessary not only for the development and continuous improvement of products, but also in the archive of information, and made consultation available always when needed (in the case of customer complaints about a particular product). Considering specific regions of the electrical test curve results for each antenna, proved to be a relevant aspect for the PdM analysis. Despite the test measurements, shown in Fig. 12(b), be within the stipulated limits, the analysis of specific points of the curves may indicate that after some time, there may have been some deterioration in the test needles. The analysis of patterns at specific points or regions on the measurement curves, can then be applied as a way to teach the PdM algorithm. Figure 12 shows the final result of the dashboard implemented in QlikView. It was possible to consult the test curves for a specific antenna serial number, Fig. 12(a), as well as check some trends in measurements during the operation time, Fig. 12(b).



(a) *E.g.* of the final result for the reconstruction of the electrical test curves for the antennas tested in the EOL28.



(b) Significant change in tolerance of one filter component during time operation.

Fig. 12. Final *e.g.* of the dashboard developed for the EOL electrical test result consultation. Notice that it's possible to filter data by partnumber, EOL, test name or date.

6 Conclusion and Future Work

This paper presents a data-driven approach, based on multiple instance learning, for predicting EOL disruptions, through the analysis of equipment operational logfiles, which was applied, and validated, through a use case in a world leader automotive industry for intelligent antenna systems (CAA). This study was conducted with some state-of-the-art ML techniques to build PdM models from an EOL testing operation log dataset. The presented application on EOL testing machines shows that, major breakdowns can be predicted or anticipated, by the identification of different operation regions of these equipments. Thus, many corrective and preventive maintenance operations can be precisely scheduled and resources, for example, "spare parts", can be managed in a better way to be provided according to the needs. It was also clear from this study that maintenance costs can be optimized and, depending on the applicability, costs for installing sensors and creating physical models can be avoided. On the other hand, as well as other studies carried out in the field of PdM based on log analysis, as *e.g.*, C. Gutschi *et al.* [11], this study showed that creating stable data-driven log-based models is very time-consuming and its profitability has to be proven.

The comparison of feature selection techniques showed a significant impact on the results regardless of the applied prediction methods. Although feature selection was not in the major scope of this paper, log-based PdM is highly influenced by the applied feature selection algorithm. Thus, further investigations should also consider features combined with other prediction methods into account. It's also intended, as future work, with the inclusion of data from the equipments' in-service maintenance records, to be able to do another type of approaches, such as supervised learning models, in order to obtain a more approximate method regarding the RUL of the equipments, as described in Sect. 3.

Acknowledgements. This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project UIDB/50014/2020. Special thanks to all the personnel involved, that gave the support necessary to make the publishing of this work possible, in particular to the Continental Advanced Antenna and the University of Trás-os-Montes and Alto Douro.

References

1. Yan, J., Meng, Y., Lu, L., Li, L.: Industrial big data in an Industry 4.0 environment: challenges, schemes, and applications for predictive maintenance. IEEE Access **5**, 23484–23491 (2017). IEEE Special Section on Complex System Health Management Based on Condition Monitoring and Test Data, USA
2. Merkt, O.: On the use of predictive models for improving the quality of Industrial maintenance: an analytical literature review of maintenance strategies. In: 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), Germany, vol. 18, pp. 693–704. ACSIS (2019)
3. Krupitzer, C., et al.: A survey on predictive maintenance for Industry 4.0. Cornell University Computer Science Article, Germany (2020). https://arxiv.org/abs/2002.08224
4. Weiting, Z., Dong, Y., Hongchao, W.: Data-driven methods for predictive maintenance of industrial equipment: a survey. IEEE Syst. J. **13**, 2213–2227 (2019)
5. Cachada, A.: Intelligent and predictive maintenance in manufacturing systems. IPB MSc dissertation, Portugal (2018). http://hdl.handle.net/10198/18301
6. Markiewicz, M., Wielgosz, M., Tabaczynski, W., Konieczny, T., Kowalczyk, L.: Predictive maintenance of induction motors using ultra-low power wireless sensors and compressed recurrent neural networks. IEEE Access **7**, 178891–178902 (2019). IEEE Special Section on Intelligent and Cognitive Techniques for Internet of Things, Poland

7. Ding, H., Yang , L., Yang, Z.: A predictive maintenance method for shearer key parts based on qualitative and quantitative analysis of monitoring data. IEEE Access **7**, 108684–108702 (2019). IEEE Special Section on Advances in Prognostics and System Health Management, USA
8. Li, Z., Goebel, K., Wu, D.: Degradation modeling and remaining useful life prediction of aircraft engines using ensemble learning. J. Eng. Gas Turbines Power Trans. ASME **141**, 041008 (2019)
9. Sipos, R., Fradkin, D., Moerchen, F., Wang, Z.: Log-based predictive maintenance. In: KDD 2014: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, USA, pp. 1867–1876. ACM (2014)
10. European/Portuguese Standard. Norma Portuguesa NP EN 13306:2007 (Registo no. 20294). (Portugal) IPQ, Terminologia da Manutençã (2007). https://biblioteca.isel.pt/cgi-bin/koha/opac-ISBDdetail.pl?biblionumber=20294
11. Gutschi, C., Furian, N., Suschnigg, J., Neubacher, D., Voessner, S.: Log-based predictive maintenance in discrete parts manufacturing. Procedia CIRP **79**, 528–533 (2019). 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering, Italy
12. Kadechkar, A., Moreno-Eguilaz, M., Riba, J.-R., Capelli, F.: Low-cost online contact resistance measurement of power connectors to ease predictive maintenance. IEEE Trans. Instrum. Meas. **5**(12), 4825–4833 (2019)
13. Proto, S., et al.: REDTag: a predictive maintenance framework for parcel delivery services. IEEE Access **8**, 14953–14964 (2020). IEEE Special Section on Intelligent and Cognitive Techniques for Internet of Things, Italy
14. Kennedy, J., Eberhart, R.: Particle swarm optimization. IEEE Access **4**, 1942–1948 (1995). Proceedings of ICNN'95 - International Conference on Neural Networks, Australia
15. Li, Z., Wu, D., Hu, C., Terpenny, J.: An ensemble learning-based prognostic approach with degradation-dependent weights for remaining useful life prediction. Reliab. Eng. Syst. Saf. **184**, 110–122 (2019)
16. Goebel, K., Eklund, N., Bonanni, P.: Fusing competing prediction algorithms for prognostics. In: Accepted for Presentation and Publication in Proceedings of 2005 IEEE Aerospace Conference, USA. AERO, Big Sky, MT, p. 10 (2006). https://doi.org/10.1109/AERO
17. Chen, H.: A multiple model prediction algorithm for CNC machine wear PHM. Int. J. Prognostics Health Manage (2011). http://works.bepress.com/huimin_chen/5/
18. DEUTRONIC. End of Line (EOL) e-motor test system. DEUTRONIC, Modern test concepts for various electro mobility components, Germany (2020). https://www.deutronic.com/testsystem/end-of-line-eol-motor-test-system/