Computers in Industry 151 (2023) 103980

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.sciencedirect.com/journal/computers-in-industry

Digital twins in condition-based maintenance apps: A case study for train axle bearings

Adolfo CRESPO MARQUEZ [a,*], José Antonio MARCOS ALBERCA [b], Antonio J. GUILLÉN LÓPEZ [a], Antonio DE LA FUENTE CARMONA [c]

[a] Department of Industrial Management, School of Engineering, University of Seville, Spain
[b] Patentes Talgo, Smart Maintenance Systems Department, Majadahonda, Madrid, Spain
[c] Ingeman, Association for the Development of Maintenance Engineering, School of Engineering Seville, Spain

ARTICLE INFO

Keywords:
Digital Twin
DT Functionality
DT Features
CBM Apps

ABSTRACT

Digital Twins (DTs) are gaining popularity in the context of the fourth industrial revolution to replicate physical equipment and systems in the digital world. DTs promise increased productivity and sustainable performance by integrating data, models, and decision-support systems. However, before realizing the potential benefits of DTs for maintenance management, several challenges need to be addressed, including a lack of conceptual basis, functional description, and established requirements. Hence, the paper presents, in a practical manner, how to cover this gap in digital configurations for maintenance management, designed to benefit of DTs. The scope of the paper includes the design and implementation of an innovative condition-based maintenance application (CBM App) based on a DT of train axle bearings, and uses a generic framework for digital maintenance management for the functional description of the DT within the CBM App. The paper provides details of the models and algorithms used to build the DT and ensures that recommended features are fulfilled. To test the DT's effectiveness and robustness, the design and framework are implemented in real CBM applications of TALGO, a high-speed train manufacturer. These tools are deemed helpful for easing DT implementation within the CBM App and can be replicated in other operational contexts.

1. Introduction

Nowadays we are experiencing rapid advances in digital technologies, data analytics and artificial intelligence applied to maintenance. These approaches have the potential to transform the way maintenance is managed. The Fourth Industrial Revolution has equipped industry with tools that help generate a deeper understanding of how complex industrial systems behave and perform, thus enabling us to manage them better. In this context, data plays a pivotal role to enhance maintenance management processes. Data can now be extracted, prepared, and recorded, for specific decision-making maintenance processes, automatically. Then, intelligent assets management systems apps (IAMS Apps) support the different decision-making processes organizing the collection and the analysis of data. Despite encouraging developments in digital solutions, several challenges remain to be addressed before the potential opportunities they present can be realized effectively for maintenance. The challenges include a general lack of awareness of which techniques and technologies are suitable to tackle specific maintenance management problems (Marquez et al., 2020).

In this paper, we first use a recently presented framework for digital maintenance management (Crespo Márquez, 2022) with the purpose to describe, rigorously, the functionality of an original Digital Twin (DT) configuration for Condition Based Maintenance (a CBM App), developed in this research for the company TALGO, a well-known high speed train manufacturer. The way to define the configuration and functionality of the tool can be generalized for similar tools.

In Section 3, we describe the configuration of the DT CBM App. In Sections 4 to 6, we offer the readers the details of the set of models used for detection, diagnosis, and prognosis of the considered failure modes of the bearings. The readers can notice that the model for detection is a ML model that was the subject of a previous publication of the team (Crespo Márquez et al., 2020), however the models adopted for failure modes classification (including innovative data transformation strategies) and for the determination of their remaining useful life (RUL) are presented for the first time in this paper.

During this research effort, the design and development team has reviewed the DT features that were mostly recommended in the

* Corresponding author.
E-mail address: adolfo@us.es (A. CRESPO MARQUEZ).

https://doi.org/10.1016/j.compind.2023.103980
Received 15 December 2022; Received in revised form 13 June 2023; Accepted 15 June 2023
Available online 27 June 2023

Nomenclature

ANN	Artificial neural network
BI Apps	Business intelligence applications
CBM	Condition-based maintenance
CBM App	Condition-based maintenance application
DT	Digital twin
ETL	Extraction, transformation and loading
FM	Failure mode
IAMS App	Intelligent assets management systems application
IaaS	Infrastructure as a service
ML	Machine learning
PaaS	Platform as a service
PHM	Prognostic health management
RUL	Remaining useful life

literature for this type of solutions (contributions from 25 different authors). This review is presented in Section 2.3 and in this paper, it is explained how these characteristics were considered for the design of the CBM App solution and system in see Section 8. Before that, special attention is paid to the "Interaction" feature in Section 7.

As summary, the main contributions of the paper are as follows:

1. The design of a DT configuration for Condition Based Maintenance.
2. The description of the functionality of the DT configuration using the framework in (Crespo Márquez, 2022), as presented in Fig. 6.
3. The models for diagnosis/classification and RUL estimations are included. These models are presented in this paper, and are added, in the solution designed, to the ones for detection already presented in (Crespo Márquez et al., 2020).
4. The data transformation strategy to move from temperature data points to temperature cycles data points, reducing considerably the quantity of data required for reasonable diagnosis quality.
5. The selection of the recommended features to be present in these types of solutions/systems. This is done after a review of the literature. An interesting innovation is introduced to present the way the DT CBM App interact with the end user, using simple business rules resulting in a practical business process, as presented in Fig. 14.
6. The description of how, during this process, the team used state-of-the-art tools for the data management and models building. For instance, the possibility to carry out the dimensionality analysis with RapidMiner ®, as presented in Fig. 12, could speed up the use of analytics while maintaining the quality of the solution for diagnosis.

The paper is then organized as follows: Section 2 presents the background of the research, which includes the context and previous information that has been collected and analyzed on the specific topic of this paper. Section 3 describes the DT of shaft bearings for CBM purposes, using the DMM framework suggested in (Crespo Márquez, 2022). Sections 4, 5 and 6 describe the detection, diagnostic and prognostic models included in this solution, as implemented in real life in the organization. Section 7 explains how the organization currently interacts with the DT. Section 8 reviews compliance with the six recommended features for the DT presented in this Section. Finally, the reader can find the conclusions of this work in Section 9.

2. Research background

This Section refers to the context and prior information that has been gathered and analyzed about the specific topic being addressed in this work. The two Subsections provide an overview of previous studies and existing literature related to Digital Twins (DTs) for CBM and to the more significant features to be considered when designing a DT.

In the context of the fourth industrial revolution, we have seen the emerging popularity of the concept of DTs, which aim to replicate physical equipment and systems in the digital world through effective integration of data, models, and decision-support systems, promising a step change in productivity and sustainable performance.

2.1. Literature review: challenges in DTs for CBM purposes

This Sections explains the fact that present investigations are at a very incipient stage and must deal with strong challenges that are very important to this work.

Despite encouraging developments in digital solutions, several challenges remain to be addressed before the potential opportunities they present can be realized effectively for maintenance. The challenges include a general lack of awareness of which techniques and technologies are suitable to tackle specific maintenance management problems. In the case of the DT, this is also generated by the fact that the definition of a DT and its characteristics are not yet fully established and there is a lack of conceptual basis (Schleich et al., 2017). Other authors affirm that one of the biggest challenges is focused on the availability and quality of the data, proposing alternatives for the creation of synthetic data through simulation (Animah and Shafiee, 2018).

The use of DT for CBM purposes is recently increasing. From a generic perspective, Yang el al. (Yang et al., 2021) present the differences of DT-based CBM compared to traditional CBM applications, focusing on changes and challenges that DTs brings to fault diagnosis, fault prognosis, and maintenance decisions. The work divides the changes into three aspects: i) a new CBM framework, ii) the data for CBM modeling and iii) the visualization tools. Same work divides the challenges in i) establishment of DTs model of complex system with multiple attributes, ii) CBM based on virtual and real data fusion, and iii) the verification of CBM by DTs. Another very interesting contribution in this sense is presented by Abbas et al. (Abbas et al., 2021)concerning the use of DT applications for prognostics and health management of subsea systems. They find interesting opportunities for i) real time degradation control, ii) dynamic maintenance optimization, iii) remote operations and unmanned facilities, iii) effective synchronization of operation and maintenance, and iv) cost reduction and lower emissions. However, in their application they find relevant challenges in i)the mitigation of the effect of sensor drift, ii) lack of models and analytical algorithms capturing the effects of maintenance actions, iii) lack of online models for real-time diagnostic and prognostics (currently most of the existing diagnostic and prognostic analysis are designed to run offline), and iv) lack of algorithms for multiple twins' synchronization and iv)validation on experimental prototypes and equipment.

Another relevant contribution presenting a systematic literature review of DTs for predictive maintenance can be found in van Dinter et al. (van Dinter et al., 2022). They identify several interesting aspects of 42 primary studies in this field like the objectives, application domains, platforms, representation types, etc. In this case, the aim is to contribute to a Software Engineering approach for developing predictive maintenance using DTs. This analysis observed a low adaptation level of industrial DT platforms with key challenges related to computational burden, the complexity of data, models, and assets, and the lack of reference architectures and standards.

Concerning rolling bearings, up to now there are still few investigations on rolling bearings under the concept of DT technology. In their recent review, Peng et al. (Peng et al., 2022) review the brief history of DT technology and describe the core technologies of rolling bearings for DT construction, including detection, modeling, and Prognostics Health Management (PHM) techniques. They conclude that the present investigations are far from enough and must deal with strong challenges related to the real-time online detection and multi-physics coupling models using fast algorithms. They also identify the importance of incorporating the knowledge of statistics and artificial intelligence algorithms in modelling, to take the place of conventional physic and mathematic models, to improve scalability and flexibility of DTs. Concerning modeling, recent contributions present innovative and varied hybrid approaches, for instance combining mechanism models together with sensor data models like in (Shi et al., 2023).

To summarize, the recurrent challenges identified in the literature with respect to DT technology to be used for CBM purposes include: 1) data acquisition and preprocessing to accurately reflects the operating conditions of the asset; 2) model complexity, variety and accuracy that can be affected by factors such as sensor placement, data quality, and model assumptions; 3) fast computing resources for organizations with limited computing infrastructure; 4) integration with existing maintenance and monitoring systems due to existing compatibility and interoperability problems; 5) shortage of skilled personnel to interact with this technology; and 6) cybersecurity to protect a large amount of sensitive data, and for real-time data processing reliance.

2.2. Literature review: relevant digital twins features

A recent review of DT applications (Errandonea et al., 2020), identified that their most common application was maintenance, followed by Prognostic Health Management (PHM) and lifecycle optimization, while the sectors where they are most applied are manufacturing, energy industry and aerospace.

Some authors sustain that in order to be considered a DT, a model must have some specific characteristics (Schleich et al., 2017) such as: scalability (ability to analyze different scales of information); interoperability (ability to convert, match and establish equivalence between representation models); expansibility (ability to integrate models); and fidelity (ability to conform to the physical model). However, the features or characteristics that a DT model should possess to be widely used in industry remains an open question in the literature.

Some contributions aim to narrow this research gap by proposing an initial synthesis of DT recommended features (Durão et al., 2018). In Durão et al. (2018) research, they found that the most frequent attributes of DTs in the context of Industry 4.0 are real-time data, integration, and fidelity. These are crucial attributes for connecting the Product Model and the real conditions of a product.

Real-time data is used for the optimization of products and production processes (Zhang et al., 2017) and it is important for knowing the status of the product and to focus on the management and optimization of processes through monitoring and data analytics (Konstantinov et al., 2017).

Integration is the most important value creation in the DT world (Uhlemann et al., 2017). A real- world object is represented by different models. The integration of the different models is essential for creating valuable data (Canedo, 2016).

Regarding fidelity, the DT allows the description of different operations in the real world. It is the fidelity of the model that provides the closeness to the physical product (Schleich et al., 2017).

But besides the definition of the DT as a very realistic model of the current state of the process, this model must have the possibility to show its own behavior in interaction with the environment in the real world. Therefore, in the framework where DTs are considered, their interaction with the real world is another important feature for our digital and intelligent maintenance management (Rosen et al., 2015). In a survey concerning the difficulties to implement DTs in industry (Durão et al., 2018) the main obstacles found are: robust integration of data and real-time control of the assets. The fact is, therefore, that most of the companies use the DT model as a conventional simulation model. Fig. 1 can serve as a summary visual representation of the literature review conducted on desirable characteristics in DT applications as outlined in the preceding paragraphs.

3. An approach to DTs for CBM Apps

This section is dedicated to present the different functional elements that can be found in a CBM App, as well as to introduce the selected framework for the representation of the DT CBM App.

3.1. The functional elements in a digital twin CBM App

A DT CBM App is an intelligent system for maintenance and asset management. It works together with Extraction, Transformation and Loading (ETL) Apps to generate ad hoc data repositories with a certain data structure model (See Fig. 2). Apps will ensure that data extracted from selected data sources, is transformed according to information requirements, formats, etc. These data repositories are expected to be placed in databases that are available in the cloud, using IaaS (Infrastructure as a Service) and PaaS (Platform as a Service) Tools. Then, it is just a question of intelligence, i.e., selecting the specific data for each

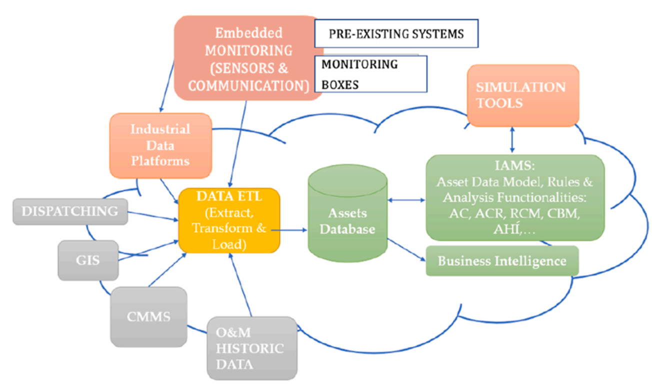

Fig. 2. Functional elements of an Intelligent Asset Management Platform (Marquez et al., 2020).

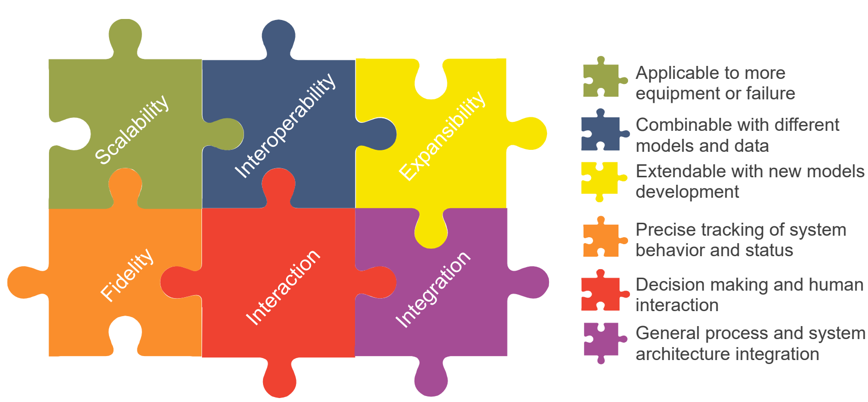

Fig. 1. Six DT recommended features.

occasion and purpose, and the proper business rules for each specific decision-making (Marquez et al., 2020).

IAMS Apps may also interact with additional tools such as simulation tools, providing extra analytical services, and they may add complementary data to the database records with results provided by these software elements. In addition to this asset knowledge discovering, creation and storing (Marquez et al., 2020), these IAMS Apps are provided by vendors together with business intelligence features or Apps (BI Apps). The BI App is designed for the interaction with the end user and extract database records to present the information according to the reporting needs and end user requirements, on demand or at the time needed by the business. A simple data flow of the process is presented in Fig. 1 (adapted from (Marquez et al., 2020).

3.2. *The selected framework for the DT CBM App representation*

A graphical tool that eases the conceptualization of the new ecosystem in Fig. 2 is the Input—Process—Output diagram presented in Fig. 3. This diagram details the digital maintenance management framework once the new processes detailed in the previous Section are incorporated. In Fig. 3, input is raw data from different business systems to be transformed or converted. Systems can be, for instance, ERP systems, dispatching systems, GIS systems, DCS, etc. Wherever relevant asset information is stored in the business (Crespo Márquez et al., 2020; Crespo et al., 2018). The process is then divided into different building blocks (similarly to what is done in the standard ISO 14224:2016), each of the blocks representing a system: ETL systems, Database Systems, IAM Systems, AI systems and BI systems. Each one of these systems will have a certain function (or group of functions) to ensure an effective and efficient digital maintenance management. The different systems will handle a very precise and predetermined data model to generate an output. These outputs can be for different purposes: to identify risk in assets, to assess that risk, to mitigate the risk, etc. Whatever is needed to control risk for the business in assets normal operations. Each one of these outputs is related and assigned to a different IAM App.

DTs (DT Apps) and Business Intelligence Apps (BI Apps) are very important supporting Apps within this framework. They may interact with the IAMS Apps to allow the introduction of powerful data analytics and visualization tools. Fig. 3 is a simple schematic of a complex digital maintenance management system.

4. Case study. Train axle bearings CBM digital twin

In the broadest understanding, CBM Apps include detection, diagnostics, and prediction of failure modes that can be interpreted to provide maintenance decision-making (Guillén et al., 2016). The case study in this paper is about the design of a DT in a CBM App elaborated to detect, classify, and predict train axle bearing failures using bearings monitored variables, in this case each bearing was only monitored capturing its temperature. A train axle bearing temperature depends on a set of factors, which when the train is running at the uninterrupted regime, consisting of the type and dimensions of bearings, the antifrictional and hydrodynamic properties of the lubricant, the spaces between the bearing rollers and rings, the static and dynamical loads of the bearing, the train running speed, the duration of travel without stops, the ambient air temperature, and the road curves (Lunys et al., 2015) (see Fig. 4).

Each train axle, in the train model of this case study, is equipped with four axle bearings, two inner and two outers. The temperature of each bearing is controlled by a PT-100 type temperature probe. Each PT-100 probe conforms to the requirements of standard EN 60751:2008 (industrial platinum resistance thermometers and platinum temperature sensors) and allows continuous monitoring of bearing temperature by the train control monitoring systems (TCMS). The TCMS will perform data capture of bearing temperatures with sampling frequency every minute, with the associated variables of date, time, outdoor temperature, and train speed.

The DT to be analyzed in this paper, departs from the fact that the theoretical physical model to calculate axle bearing temperatures could be replaced by a data-driven bearing temperature model as in (Crespo Márquez et al., 2020). The data-driven model inputs and outputs are presented in Fig. 5.

Notice that, to estimate an axle bearing temperature, the remaining axle bearings temperatures plus the ambient temperature are the only inputs considered. This is the capital principle, and very innovative approach, that is used to build all DT required predictive analytics. Artificial neural network algorithms were used to build the modes for temperatures and the subsequent prediction rules offered the following metrics (see Crespo et al., 2020, Table 12): Precision (100 %), Negative predictive value (22,87 %), Sensitivity (13.81 %), Specificity (100 %) Accuracy (31,35 %). With precision resulting the critical factor to accept the model to be put into operation (as presented in Fig. 7).

The estimation of the bearing temperatures with the desired precision, besides allowing anomalies detection, opens the door to further

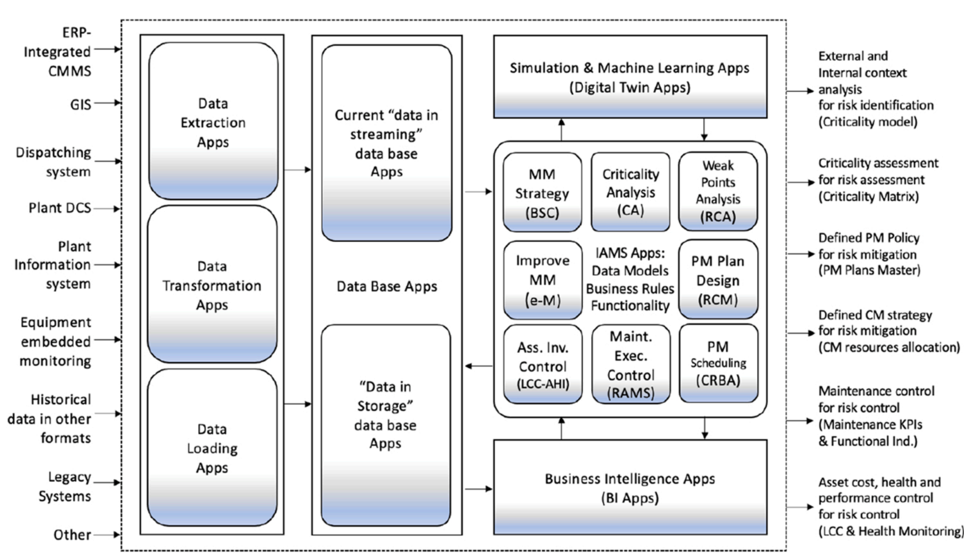

Fig. 3. Input—Process—Output diagram of the framework in (Crespo Márquez, 2022).

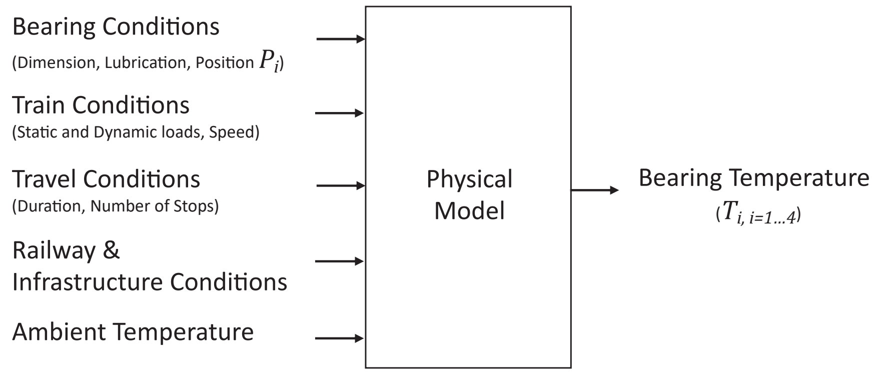

Fig. 4. Factors (physical model inputs) conditioning a train axle bearing temperature.

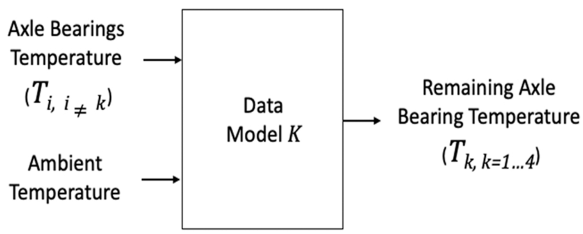

Fig. 5. Crespo et al. (2020) approach to predict axle bearing temperatures.

modeling development to classify the potential failure modes that could be generating the excess temperatures in a certain bearing. Once the failure mode is identified, reliability models are used to determine the remaining useful life (RUL) of the component at a given moment, leading to a subsequent maintenance decision.

In this paper the comprehensive functional description of the DT CBM App, as per the original design in collaboration with the Smart Maintenance Department of the company TALGO, is presented in Fig. 6. In Fig. 6, all algorithms, and models to simulate and classify bearings behavior are placed within the DT area, while business rules are implemented in the Apps located within the CBM processes apps area. This Figure is built using the framework in Fig. 2. In the subsequent Sections 5, 6 and 7, tool design details are presented to the reader, then the paper will describe how the DT in the designed App presents the recommended features that were referred in Section 3.1. (see Table 7 Section 9).

5. DT Analytics for anomalies detection

Concerning analytics for anomalies detection, the CBM App is using an ANN predictive model, that was selected because of a good ratio correlation coefficient vs total time, assuming a minimum required

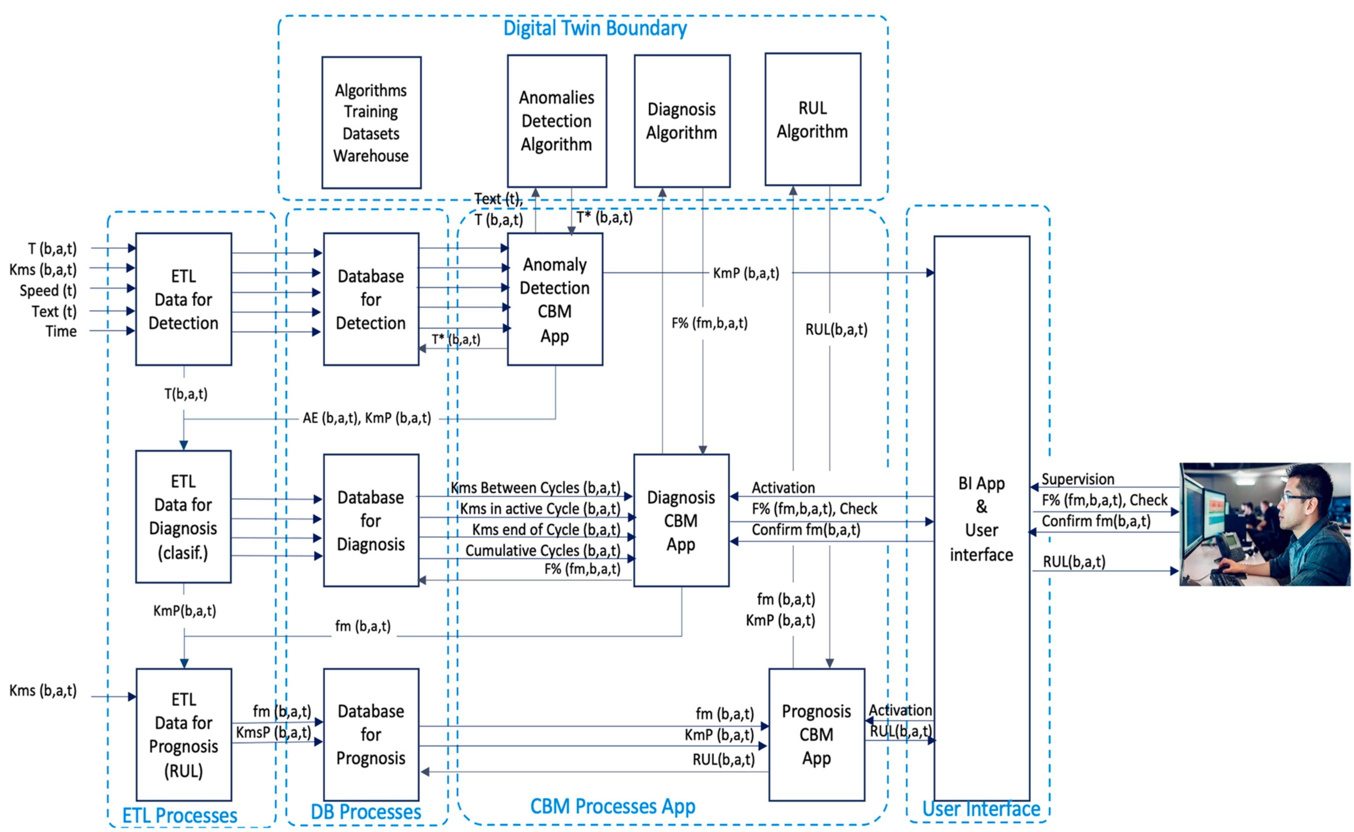

Fig. 6. Details of the DT CBM App structure using the framework in Fig. 2.

correlation factor of 0.96, compared to other models such as Generalized Linear Models (GLM), Decision Trees (DT), Random Forest (RF), Gradient Boosted Trees (GBT) and Support Vector Machines (SVM) that were compared in the study. This model was published in a previous paper by Crespo et al. (Crespo Márquez, de la Fuente Carmona et al., 2020), where their authors demonstrated that the suggested universal model per bearing position performed well (>0.98 correlation coefficient) for all four bearing positions. This made to overcome the non-ergodicity of the assets possible and permitted to develop only four bearing temperature prediction algorithms to cover all the fleet of trains bearings. The anomalies detection rule designed could identify damaged bearings with 100 % precision, at any speed of the train, based on a 10 ∘C Absolute Error (AE) threshold for the predicted temperature of the bearing. A threshold in train speed was introduced in the rule just for scoring data sets reduction, and the expected subsequent accuracy of the rule's improvement. However, accuracy improvement was found not to be very significant for all cases. No relevant results were obtained regarding the accuracy and sensitivity of the algorithm related to the increase of the train speed threshold up to 90 km/h. The motivation for the selection of the threshold, in addition to the one mentioned above, serve to rule out data related to train starts and stops, where bearing temperatures have transient values characteristic of these modes of operation.

To illustrate the difference in AE data distribution when the bearing is in good conditions versus when it comes to a degraded state, Fig. 7 represents the temperature prediction AE distribution in periods of good (green) vs. degraded (blue) conditions, with train speeds $TS_t \geq 90$ km/k. Notice that when the bearing is in good conditions (green) the absolute error (AE) of its temperature prediction will never be higher that 10ºC, while when the bearing is in degraded condition (blue curve) the AE can reach much higher values with significant probability. Therefore, there is a clear possibility to predict abnormal bearing temperature behavior with maximum precision and specificity, when selecting the Absolute Error descriptor with a threshold of 10 ºC in the rule, and regardless the speed of the train.

6. DT Analytics for failure mode classification

6.1. ETL Process and data base generation

The train axle bearing FM classification model is the second model contained in the DT of the CBM App in this paper case study. This modelling effort, to identify a certain bearing failure mode, required further ETL processes and different modeling tools. The most significant challenge was the decision (of the Smart Maintenance Department

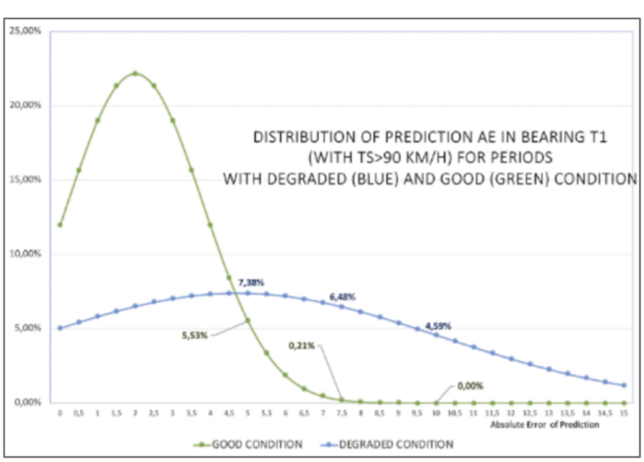

Fig. 7. Distribution of EA for good (green) and degraded condition (blue) periods, for a train speed $TS_t \geq 90$ km/h. Taken from (Crespo Márquez, de la Fuente Carmona et al., 2020).

together with the Maintenance Engineering Department of the company), to approach this problem modeling temperature cycles instead of temperature points. This is a popular method (Healey et al., 2021) to study fatigue data analysis of mechanical components. In these cases, it is common to reduce a variable stress spectrum into a simpler, equivalent set of stresses. Methods that extract successively smaller cycles from a sequence are used to simplify the calculation of the fatigue life of a component from these simpler cycles (Healey et al., 2021). In this case study this adds the benefit of the data reduction, as it will be detailed later.

The data used in the study come from a fleet of 16 identical trains, some with more or fewer cars but with identical and interchangeable bearings among all of them. Each train has provided an average of 800,000 temperature records for each of its bearings over the course of a year. Regarding cleanliness, only complete series of erroneous values have been eliminated, since there are very few errors, there has not been any considerable condition that has affected its bias. The size of the sample has been considerably large, in the order of Gigabytes, and the tool used to manage it has been RapidMiner.

In the following paragraphs it is shown how temperature cycle analysis works, the steps to follow and the key variables for this analysis.

The analysis departs with two steps dealing with data extraction as follows:

1. The input data received for each bearing of the train, which is the following:

- *Date and time:* This data is provided every minute.
- *Train Speed* (*TS*)**:** This data is provided every minute.
- *Ambient Temperature* (*Ta*)**:** This is the train outside temperature.
- *Bearing temperature* **(***Ti***):** Each axle has 4 bearings temperature measurements, named: T0, T1, T2 and T3.
- *Bearing Temperature prediction* ($\widehat{Ti}$)**:** This the prediction of each bearing temperature ($\widehat{T}0$, $\widehat{T}1$, $\widehat{T}2$ and $\widehat{T}3$) as obtained with baseline predictive analytics.
- *Kilometers traveled* (*Km*)**:** These are the total kilometers the train has traveled at a given time.

2. Data received from the CBM Anomalies Detection App:

- *Absolute Error* (*AE*)**:** This is the difference between real and predicted temperatures Ti-$\widehat{Ti}$ per bearing *i*. According to developed detection analytics the anomaly is detected, and *a positive* is registered, when $AETi > 10°C$.
- *Time / Km of the first registered positive:* For each bearing, the appearance of a positive ($AETi > 10°C$) will trigger the utilization of the algorithm for failure mode classification.

After these extraction steps, the analysis continues with the data transformation process, in three more steps, as follows:

3. Determination of the following variables calculated from the extracted ones (Fig. 8):

- *Accumulated absolute error* (*Acc AE*)**:** This variable accumulates the AE when a positive is registered, since the first positive.
- *Accumulated kilometers since the first positive***:** This is the total number of kms the train run since the first positive was registered.
- *Accumulated kilometers in positive***:** This is the total number of the kilometers the train run in positive, since the first positive.

4. Determination of the temperature cycles. This is done using an algorithm (see Annex 1 for the pseudo code) that is developed, which considers that when there is a short distance between two positives points (between which are interspersed negatives), these two points

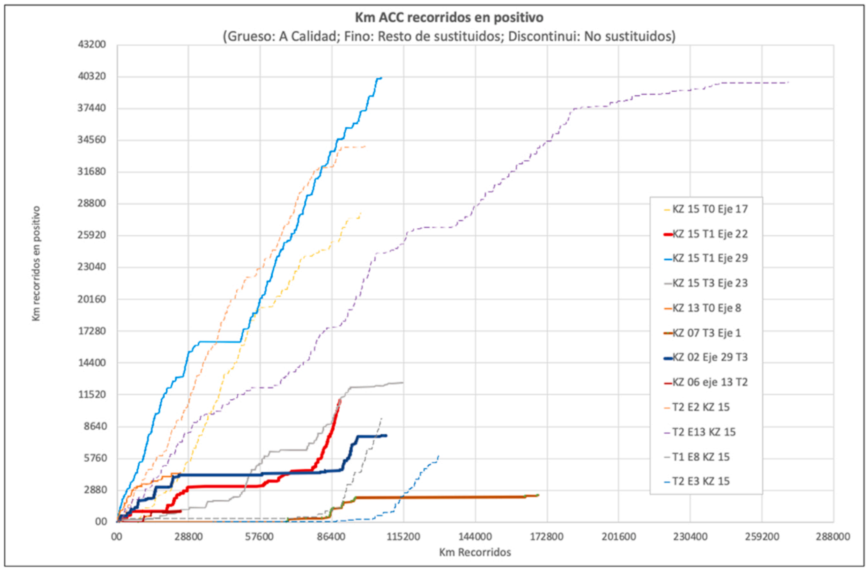

Fig. 8. Sample data regarding Kms traveled in positive for different bearings.

belong to the same bearing high temperature cycle. In this way, it is possible to simplify the data for the analysis of the bearing deterioration, under the assumption that this deterioration will be associated with the set of overtemperature cycles that are identified.

In Fig. 8 there is an illustration of one of the new variables, calculated from the extracted ones. This variable in Fig. 8: *Accumulated kilometers in positive*, is then used as an input in the model for failure mode classification. In the Figure, the variable is plotted for a total of 12 different bearings since the km of their first positive to the total accumulated kms at their failure.

Fig. 9 illustrates what is mentioned in the previous paragraph, detailing the number of kilometers that elapse between recorded positives and counting the number of cycles depending on the maximum distance selected between positives of the same cycle (this count appears below the graph).

5. Obtention of new following variables as per the cycle analysis performed:

- *Kilometers at the beginning of the cycle*: These are the kilometers that the bearing traveled, from the first positive, until a new cycle started.
- *Kilometers at the end of the cycle:* These are the kilometers that the bearing has traveled, from the first positive, until the end of the cycle.
- *Cycle Kms:* Kilometers that the train travels in a cycle (the cycle ends when the next positive is farther away from the previous one, than the limit in km established in each case).
- *Kilometers traveled between cycles*: These are the kilometers traveled between the end of one cycle and the beginning of the next one.
- *Cumulative cycles:* Cumulative number of cycles since the first positive.
- *Percentage of kilometers in active cycle*: Percentage of kilometers that the bearing accumulates in a cycle since the appearance of the first cycle.
- *Total kilometers in active cycle*: Accumulation of kilometers that the bearing run within cycles.
- *Accumulated kilometers between cycles*: This is the sum of the kilometers that a bearing traveled between cycles, up to the last cycle.
- *Average of the kilometers between cycles*: In this section we have the average of the kilometers traveled between cycles. Making this average gradually as we go from cycle to cycle.

Fig. 10 shows the results obtained for different cycle variables (detailed below), when cycles are selected with different maximum distances between positives (cycles have been calculated with these

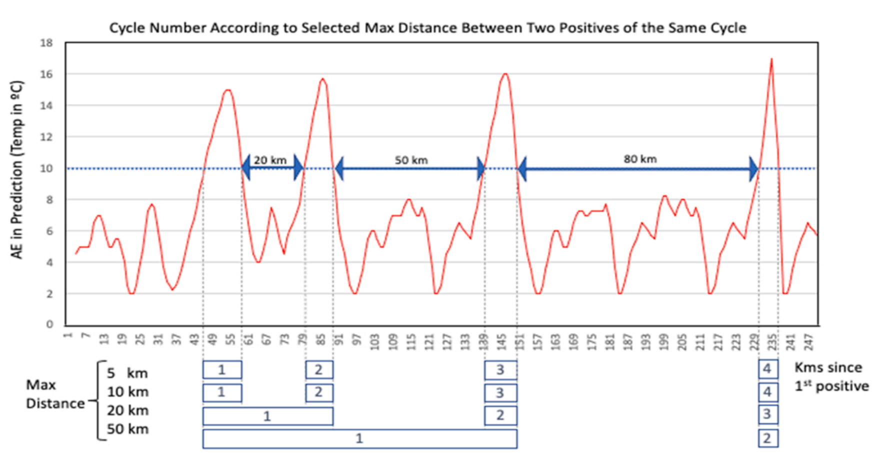

Fig. 9. Cycle count by varying the maximum distance between positives of the same cycle.

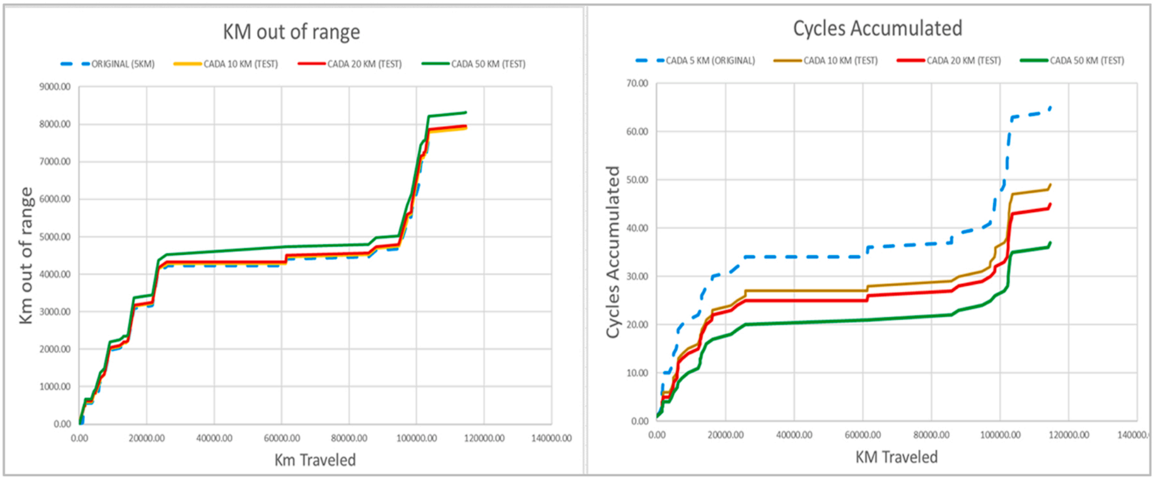

Fig. 10. Sample of values obtained for cycle variables, when varying the maximum distance selected between positives of the same cycle.

maximum distances between their positives set at 5, 10, 20 and 50 km).

Although the main aim of the transformation process is to approximate the physical degradation model in a simpler way, it is observed that the amount of data to be considered and stored for the bearing analysis is also significantly reduced. The reduction achieved in the data to be stored per bearing studied is presented in Table 1.

Finally, after these transformation steps, a specific-to-the-problem data model is generated and loaded into a new data base as explained in step 6:

6. Specific data base generation and load

An extract of the final tables obtained, of data per bearing, is presented in Table 2. The details of the data are also included in Annex 1 of this document. In this table the inner rows of each bearing records have been hidden to illustrate the number of records for the first stored cycles (e.g., 62–1 =61 records for bearing Train KZ15, T0 Axis1, or 201–62 =149 records for bearing Train KZ15 T1 Axis22).

6.2. The ML failure mode sorting algorithm

Once the required data base is ready for model generation training and production, the process continues with the algorithm design, testing and validation.

The algorithm attempts to separate bearings with internal deterioration from those with overtemperature caused by external causes, mainly due to the train axle guidance system problems. To that end, it is necessary to know the final diagnosis of all the bearings observed to have suffered overtemperature cycles. It is essential to have data on whether the bearing was replaced or not, and if once it was replaced, whether the analysis performed by the quality department found it with internal deterioration or not.

Bearings in the train that were not replaced, but which had overtemperature cycles recorded, are obviously classified as "non-deteriorating" bearings. Basically, most of these bearings went back to normal temperature conditions when the train guidance problems were solved.

Table 1
Reduction of the number of data records to be captured per bearing when applying the cycle algorithm.

Bearing Samples	REDUCTION of DATA POINTS for a Maximum distance between positives of a cycle of				
	1 km	5 km	10 km	20 km	50 km
KZ02 T3 AXLE 29	89.346	65	55	51	42
KZ15 T2 AXLE 1	78.318	416	281	207	152

All these records helped to better train the classification algorithm. In this way, it is possible to generate an ML algorithm that, receiving the data of the bearing temperature cycles over time, classifies it with a higher or lower probability of being a bearing with incipient deterioration vs. a bearing suffering of guidance problems.

The algorithm selected for this classification functionality can be chosen among different possibilities: according to its ROC curve (see Fig. 11), classification error, gain, execution time, training time, etc. For this case, the selected algorithm has been Deep Learning.

Concerning the feature selection, the software allows to carry out a trade-off between complexity of the algorithm (number of features considered as input) and the classification error (see in Fig. 12 the final 4 features set selected, corresponding with 2, 3, 4 & 5 features in the set, best sets evaluated by the software). This is an algorithm dimensionality analysis adding substantial value at this step of the design process. In the plot of Fig. 12 each point represents a different features set, i.e., a subset of the original variables, related to cycles, previously defined. Notice that the set selected by the software, of a complexity of 4 (i.e., a dimension of 4 variables), achieves a lower error rate that the original selected set of 5 (that was also including the duration of the cycle, as feature). So, the model is less complex (one feature less) and still more accurate than the original feature space (square in the graph). Using less features also means that models can be trained faster. The feature set which has been used to build the final model is shown. The weights assigned to the features by the software explain the relative importance of each variable in the dataset used to validate the model; they do not correspond to any internal weights used to configure the model algorithm.

Error rate in Fig. 12 is training error rate. This performance, and therefore the feature set optimization, is calculated on a 40 % hold-out set which has not been used for any of the performed model optimizations. This hold-out set is then used as input for a multi-hold-out-set validation where the performance for 7 disjoint subsets is calculated. The largest and the highest performance are removed and the average of the remaining 5 performances is reported here. Although this validation is not as thorough as a full cross-validation, this approach strikes a good balance between runtime and model validation quality (Kotu and Deshpande, 2019). Tables 3 & 4 show algorithm performance metrics and confusion matrix, respectively.

The company decided to use this algorithm in the App initially, according to good levels in precision when classifying the internal bearing failure mode, regardless the fact that the level of false negatives was high. To improve general algorithm accuracy, more bearings failure must be classified, and their data added to the database for training. A process that has been already implemented in the company.

Table 2
Example of an extract with data from several bearings, showing the number of data lines per bearing (assuming 5 km as max distance between positives of a cycle).

Register Count	Bearing	Replaced	Damaged	Kms Enf of Cycle	Kms between Cycles	Cycle Kms	Cummulative cycles	Kms in active cycle
2	KZ15T0A17	1	1	345.6	2619.9	342.6	1	342.6
…	…	…	…	…	…	…	…	…
62	KZ15T0A17	1	1	66704.7	58.5	31.7	61	10667.0
63	KZ15T1A22	1	1	25.1	8.0	22.0	1	22.0
…	…	…	…	…	…	…	…	…
201	KZ15T1A22	1	1	92086.1	41.1	154.1	139	9694.0
202	KZ15T1A29	1	1	1211.8	13.4	1209.9	1	1209.9
…	…	…	…	…	…	…	…	…
459	KZ15T1A29	1	1	105226.1	870.3	18.3	258	38379.0
…	…	…	…	…	…	…	…	…

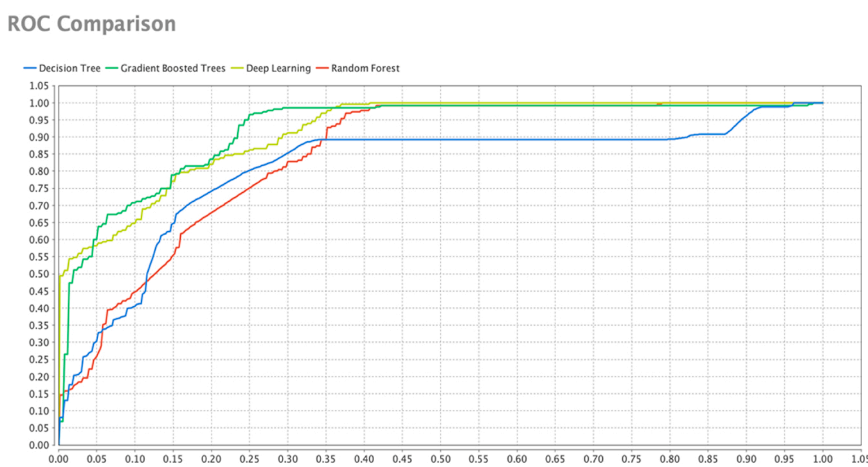

Fig. 11. ROC curves of the different techniques (RapidMiner ®).

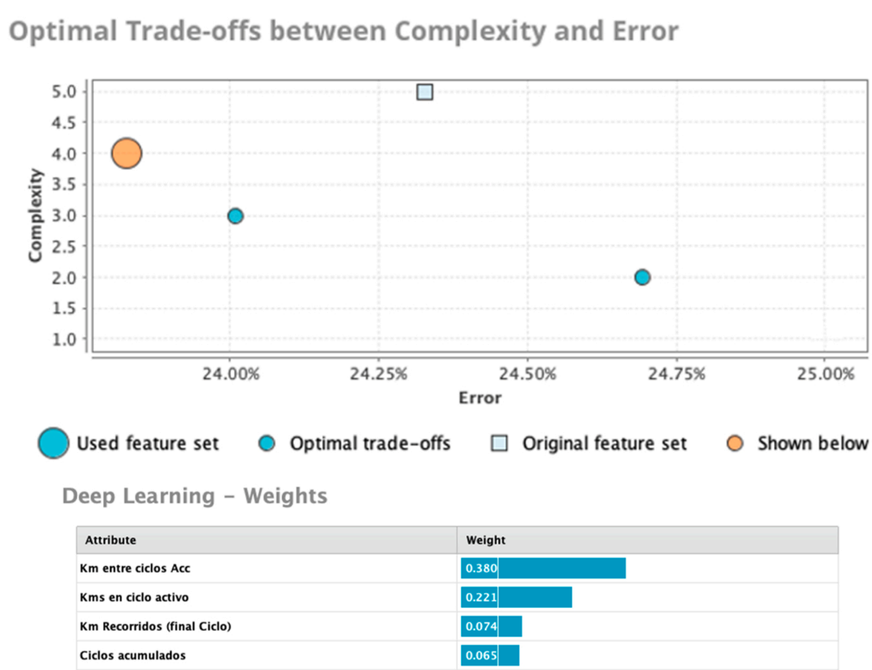

Fig. 12. Trade-offs between model dimensionality (complexity) and error, including final features selection and their weights (RapidMiner ® output).

7. DT Analytics for prognostics. The RUL determination

7.1. The RUL concept and definition considered for this case study

Failure prognostics is defined (ISO 13381–1:2004) as "the Estimation of the Time to Failure (ETTF) and the risk of existence or later appearance of one or more failure modes". However, in most of the literature related to prognostics, the terminology Remaining Useful Life (RUL) is used, instead of ETTF (Medjaher et al., 2012). The concept of the RUL has been widely used in operational research, reliability, and statistics

Table 3
Classification algorithm performance metrics.

Criterion	Value	STD
Accuracy	76.3 %	±0.2 %
Classification error	23.7 %	±0.2 %
AUC	91.6 %	±0.2 %
Precision	100 %	±0.0 %
Recall	5 %	±1.0 %
F Measure	9.5 %	±1.9 %
Sensitivity	5 %	±1.0 %
Specificity	100 %	±0.0 %

Table 4
Sorting algorithm confusion matrix (range 1: Guidance FM; range2: Internal FM).

	True range 1	True range 2	Class Precision
Predicted range 1	785	248	**75.99 %**
Predicted range 2	0	13	**100.00 %**
Class Recall	**100.00%**	**4.98 %**	

literature with important applications in other fields such as material science, biostatistics, and econometrics. Clearly the definition of the useful life depends on the context and operational characteristics (Si et al., 2011).

Most of the prognostic activities are related to estimating the RUL of physical assets based on their current health and operating conditions that are expected for that asset in the future (Medjaher et al., 2012). Furthermore, estimation of RUL has been extensively investigated in recent years as a key aspect of Prognostics and Health Management (Hu et al., 2018). The research on how to best estimate the RUL has gained popularity. However, due to its complicated relationship with observable health information, there is no such best approach which can be used universally to achieve the best estimate (Si et al., 2011).

If the reader searches the literature related to the term RUL, a multitude of possible definitions for this term can be found, although they all coincide in one part. They define the RUL as the useful life that the asset has left from the current time (in which a symptom is detected, a condition monitored out of thresholds, etc.) until the end of its useful life, appearance of the failure or requires maintenance intervention (Okoh et al., 2014). The Remaining Useful Life is typically random and unknown, although it depends on the current age of the asset, future operating conditions and the observed condition monitoring or information on the asset's health.

Concerning the estimation of the RUL, the existing approaches fall into three main categories (Jardine et al., 2006): statistical approaches, artificial intelligence (AI) approaches and model-based approaches.

- The first category corresponds to statistical approaches or degradation models. Estimating the RUL is achieved by evaluating the conditional lifetime distribution given that a system has survived up to a specific time. Life characteristics of a population of identical systems and lifetime data are required. However, such data are scarce or even nonexistent. With the advances in CM technologies, degradation data can be obtained from routine CM as feasible and low-cost alternatives to estimate the RUL. These data are usually correlated with the underlying physical degradation process. Properly modeled, degradation data can be used to predict unexpected failures and accurately estimate the lifetime of gradually degraded systems. Weibull analysis, PHM models or PIM models can be used to stablished a predetermined level of failure probability that can be used to estimate the RUL (Goode et al., 2000).
- AI approaches must face the problem of the curse of dimensionality and the problem of long-term dependencies when building the ML algorithms. To that end, some popular algorithms are: self-organizing neural networks (S-ONN), dynamic wavelet neural networks (DWNN), recurrent neural network (RNN), Echo State Networks (ESN) and Long Short-Term Memory (LSTM) networks for RUL estimation. (Ramezani et al., 2020; Zhang et al., 2019).
- Model-based approaches to prognosis require specific mechanistic knowledge and theory relevant to the monitored machine (like modeling for instance crack dynamics or other physics of failure). Mechanistic models are based on mathematical description of mechanical, chemical, biological, etc. phenomenon or process (Cempel et al., 1997) and (Qiu et al., 2002). A different way of applying model-based approaches to prognosis is to derive the explicit relationship between the condition variables and the lifetimes (current lifetime and failure lifetime) via mechanistic modelling. This has been done with bearings using vibration signals in (Cempel et al., 1997) and (Qiu et al., 2002).

7.2. RUL calculation and a procedure for its determination

In this paper a statistical approach is followed to estimate the RUL (of any bearing of a train), once a positive (or anomaly detected for a failure mode) appears in a train axle bearing. A positive (according to the Procedure for the Design and Implementation of CBM Plans in the company) is defined as the occurrence of an absolute error (AE) of prediction greater than 10ºC between the actual bearing temperature and that predicted by the ANN designed for detection, when the train is running at more than 90 km/h (i.e., $AE \geq 10\ ^\circ\mathrm{C}$, $TS \geq 90\ km/h$) and for more than one minute.

RUL is now defined as a random variable that, estimated from the appearance of the first positive, offers a good prediction of the life of the element until its replacement due to over temperature or noise. This replacement is nowadays performed after the activation of the safety alert in the train monitoring and control system (TCMS) and/or because of a certain inspection (probably during a weekly train inspection in the workshop). The safety alert is triggered when the temperature difference between the four bearings of the same axle is higher than 25 °C — ($T_{max} - T_{min}) \geq 25\ ^\circ\mathrm{C}$ — and this condition is maintained for more than 1 min. Company's objective through the analysis included in this part of the paper is to foresee the recommended time of bearing replacement, after its first positive, even without prior inspection, according to statistical estimates.

The calculation will be applicable to any bearing regardless of its position in the train. The data used comes from the record of bearing replacements that could be traced, and linked to the recorded equipment monitoring data, and the predictions of bearing failures made with artificial intelligence algorithms. In general, the RUL is the estimated time of equipment operation until failure (point F) and it is estimated for any point (A in Fig. 13) located from point P (from which it is already possible to verify the existence of a potential failure) on the equipment. Therefore, the following always occurs: *RUL≤ P-F Interval* (See Fig. 13).

To calculate the RUL of a bearing at point A, it will first be necessary to model the random variable "PF interval", i.e. the interval (in time, km,

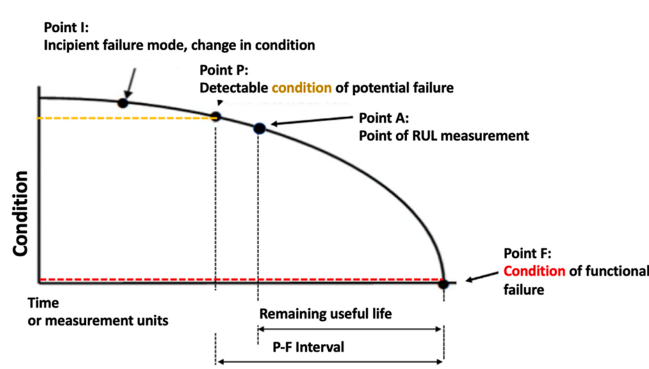

Fig. 13. P-F Curve and P-F time interval. Estimated time to failure (RUL).

or representative unit of measurement) that elapses between the first positive (point P, agreed in the CBM procedure for: $AE \geq 10$. C, $TS \geq 90$ km/h) and the possible replacement due to overtemperature and/or noise of a bearing. The point F considered takes place, in general, after the activation of the safety alert in the train monitoring and control system (TCMS), this condition is not of functional loss of the bearing, but of operation in conditions of lower safety level. Then it is possible to define, for this case study:

$$RUL = RUL \quad AF \quad interval = (PF \quad Interval - PA \quad Interval).$$

The calculation of the RUL is non-deterministic and is therefore affected by a certain level of uncertainty (depending on the uncertainty of the calculation model and the quality and quantity of the existing data).

The determination of the RUL will be made from the estimation of the distribution function of the PF interval, using a statistical technique such as the Weibull analysis. The application of the Weibull method will be carried out considering uncensored data of the random variable, i.e., samples of times from the first positive to the alarm and subsequent replacement of the bearing. It is considered that the bearing degrades from the first positive (point P), and that this degradation has a positive correlation with the number of positives recorded. It is also considered that the degradation is irreversible. In this procedure, the RUL is expressed in km (useful life in km), $F(t, \beta, \eta) = Weibull(t, 4.14, 105652.23)$.

From the first positive (point P) it is then possible to determine how many kilometers the train can run, without requiring bearing replacement due to over-temperature or excessive noise, with a certain probability. For example, it is possible for the train to run 30,000 km after the first positive with a 99.46 % probability that an alarm will not appear on the train. Similarly, it is possible for the train to run 50,000 km after the first positive with a 95.59 % probability that no alert will appear on its TCMS. It is then decided to opt for 30,000 km as the PF Interval for the subsequent remaining useful life calculations, an interval that is going to be achieved with a 99.46 % probability without the need for bearing replacement, which usually happens after the TCMS alert on the train.

There are simple goodness-of-fit tests such as the Kolmogorov-Smirnov test. The test is done for what is called a significance level (α) given the sample size. For example, for the case at hand, looking in the Kolmogorov-Smirnov Test tables at a significance level of $\alpha = 0.20$, the critical value of the difference D for a sample size $N = 4$ is 0.494; this means that in 20 % of random samples of size 4, the maximum absolute deviation between the cumulative distribution of the sample [i.e. SN(x) = k/N, where k is the number of observations less than or equal to x] and the cumulative distribution of the population (theoretical function calculated above) will be at least 0.494. Calculated the differences in our case we obtain a maximum deviation of Dmax= 0.21907 with Dmax= $0.21907 \leq$ D_(α = 0.2) = 0.494. Therefore, we should not reject the hypothesis made, in other words, we can work with the function selected in principle.

Once the goodness-of-fit test has been performed, it is possible to perform an additional specific experiment to get a true dimension of the uncertainty that currently exists given a small sample size for a certain failure mode. What is done in this experiment is the following: 1) Consider the WEIBULL (t;4,14; 105,652.23) distribution as the "true a priori" of the PF interval; 2) Obtain a large sample (100 observations) of PF intervals generated as pseudo-random numbers from that distribution. 3)Generate a series of experiments (10 have been selected in principle) by randomly selecting 4 of these numbers each time. 4) Calculate the best approximation of a Weibull for each of the 10 experiments. 5) Obtain the variables of interest of this analysis that have been initially selected: Kms at R= 0.9946; R (30000) and R (15000). 6) Conclude on the uncertainty of the decision making, currently having a sample of only four observations.

In Table 5 it is possible to appreciate the variations of the measures that could be considered for a very low sample size, in the selected variables of interest, for example:

Table 5
Results of experiments with Easyfit Software. ® Mathwave. Data analysis and simulations.

Experiment	Weibull best fit		Kms with R= 0,9946	R (30000) In %	R (15000) In %
	β	η			
1	1.16	120770	1343	81.97	91.49
2	1.45	88153	2399	81.06	93.08
3	1.59	107840	4049	87.70	95,749
4	9.46	96370	55509	99.99	100
5	4.30	107340	31892	99.58	99.97
6	1.93	105830	7084	91.59	97.72
7	5.44	108080	41412	99.90	99.99
8	1.45	85994	2351	80.48	92.35
9	4.93	80484	27943	99.23	99.97
10	4.16	113360	32352	99.60	99.98
Max	9.46	120770	55509	99.99	100
Min	1.16	80484	1343	80.48	91.49
Mean	3.59	101422	20633	92.11	97.03
STD	2.63	13059	19616	8.60	3.56

1. The Km to be traveled with a reliability of not having TCMS alerts of 99.46 % that had been considered in a value of 30,000 in this work, could come to be estimated in 1343 km in the case of the first experiment (lower limit) or in 55,509 km in experiment 4 (upper limit). The average value of this variable in the experiments would be 22,633.4 km.
2. The reliability of the period selected as the PF interval in this work, R (30,000), could be estimated at 80.48 % in the case of the eighth experiment (lower limit) or 99.99 % in experiment 4 (upper limit). The average value of this variable in the experiments would be 92.11 km.
3. The reliability of 50 % of the period selected as the PF interval in this work, R(15,000), could be estimated at 91.49 % in the case of the first experiment (lower limit) or 100 % in experiment 4 (upper limit). The average value of this variable in the experiments would be 97.03 %.
4. Another interesting observation is that the ratio between the upper and lower limits of the estimates made. For example, this ratio is $R30000_{MIN-MAX} = 1,24$ for the R(30,000) limits, while it is of lower value for the reliability at 15,000 km, with $R15000_{MIN-MAX} = 1,09$.

All the above indicates the existing limitations in the accuracy of the calculations, which will disappear as the size of the observations obtained increases, for future bearing replacements.

8. The interaction with the digital twin

The functionality of the DT allows the evaluation of the failure mode risk level and the subsequent control actions, this will allow the maintenance staff to schedule convenient maintenance activities. Interaction with the DT must be done using simple business rules resulting in a practical business process. Any new event detected by the DT leading to a new state of the asset concerning a failure mode will be a call for maintenance action. For the correct interpretation of the method of interaction with the DT, Table 6 describes the necessary concepts to be reviewed (taken from an original work in (Martínez-Galán Fernández et al., 2022).

This method of interacting with the DT takes into consideration four proposed levels or states of FM risk: low, medium, high and failure; and two different types of events: monitoring and preventive maintenance events. The limits of the 4 risk levels (See Fig. 14), are determined by the company maintenance organization as follows (this could be agreed differently):

Table 6
Key concepts in DT interaction with maintenance techs.

Concept	Types
Event Recordable, scheduled, or supervening time, at which the risk level of the affected failure modes must be reanalyzed.	• *Monitoring Event:* Events taking place because of the CBM App (and its DT algorithms). They can be detection events, diagnostics events, or prognostics events. • *Preventive Maintenance Events:* Maintenance programmed or unforeseen events. They can be for example inspections or any PM activity.
State Qualitative level of risk at a given time. Each event causes a possible change in the level of risk.	• *Fault*: State after the failure has occurred. State in immediate replacement or repair of the item is required. • *High Risk*: State of operation closest to failure. Short-term activities are scheduled to reduce the level of risk. • *Medium Risk*: State in which an anomaly has been detected but with some security it is possible to continue operating under normal conditions. Medium-term activities are planned to confirm the risk and analyze how it evolves. • *Low Risk*: Normal operating state of the item
Failure Mode Failure modes involved that can be fully or partially managed by CBM. Monitoring solutions and maintenance tasks are applied at failure mode level.	• Primary failure mode (PFM)* • Secondary failure mode (SFM)*: initiated by a PFM

* Terminology adopted from ISO 13381, (ISO, 2015).
Source:Adapted from Martínez-Galán Fernández et al. (2022).

- Low Risk: Normal operation with no positive in the detection algorithm, or with a positive detected but less than 15.000 km ago (this number of kilometers is selected in consonance with the resulting RULs for both FMs in Section 6).
- Medium Risk: A failure mode enters this level when it is classified with > 70 % probability by the classification algorithm (Section 5) or when is running 15.000 kms. since the first detection positive (Section 4) and no definitive failure mode classification is made (Section 5).
- High Risk: A failure mode enters this level when it is confirmed by a technician inspection, that is released when entering in Medium Risk. A failure mode leaves this level when there is a bearing replacement or reconditioned.
- Fault: This would assume the bearing functional loss and will never be reached under normal operations of the DT App.

It is considered that both, monitoring events and PM events, may lead to a change in the risk level of one or more failure modes of the asset. This is because these events release a new assessment of the FMs. Then, when a new failure mode state is reached, a certain maintenance action is accomplished (for instance, an algorithm for detection is launched, an inspection is done, a replacement takes place, etc.).

To describe these concepts in a graphical manner, a CBM sequence affecting two failure modes is pictured in Fig. 14. In this case, Monitoring Events and PM Events may change the each one of the FMs risk level (FM1: Internal degradation and FM2: External guidance failure). Let's assume that the error in bearing temperature prediction provides valuable information to be used as a symptom of both FMs. Temperature.

In the example of Fig. 14 we represent a monitoring solution that provides information that can be linked with FM1 and FM2 risk level control. Event 1 appears when a first threshold on the prediction error is exceeded). At this time, the maintenance technician will not observe an increase in the risk level of any failure modes. However, the maintenance technician will supervise the action of launching the classification algorithm and commence counting kilometers since the first positive was recorded.

In this case, Event 2 takes place after a certain number of kms after event 1, for instance 15,000 kms, and non-concluding result offered by the classification algorithm. Then, it is agreed that both failure modes (FM1 and FM2) will be escalated to the Medium-risk level. Moreover, it is agreed that since the classification algorithm does not distinguish FM1 from FM2, an inspection event (Event 3 in Fig. 14) will be scheduled. Event 3 is therefore a PM activity, an inspection task that will confirm the existence of FM1 or FM2, turning back the non-existing FM to a low-risk state (green color in Fig. 14). At the same, the maintenance technician could decide, for instance and always in accordance to his/her inspection outcome, to escalate FM1 risk level to high (red) and to program the substitution of the bearing (Event 4 in Fig. 14). This interaction was agreed and described to guide maintenance actions and to make sure of proper systems supervision.

The reader may notice that whatever actions is agreed by the organization, to follow a given event, must consider the limits of failure mode RULs, established by the prognosis algorithms. This limits the reaction times when scheduling inspection manually or when moving the failure modes to a higher risk level automatically.

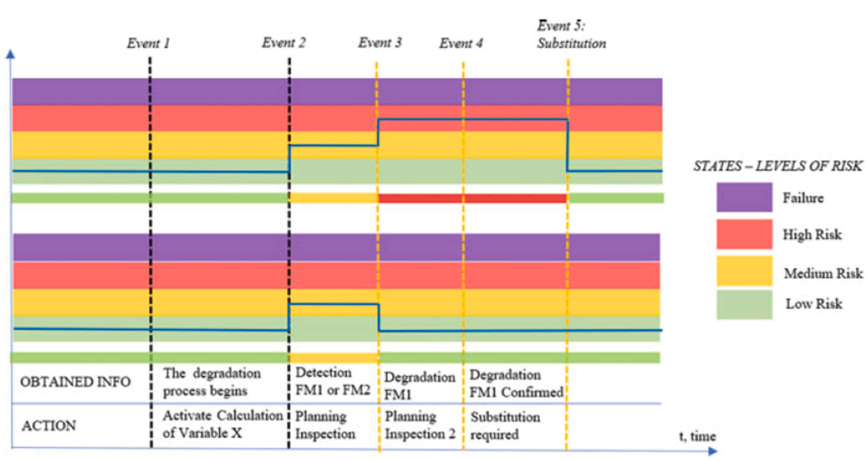

Fig. 14. Graphic representation of the CBM APP DT interaction with Maintenance technicians. Adapted from (Martínez-Galán Fernández et al., 2022).

9. The digital twin features

After reviewing the comprehensive description of the DT in Fig. 6, let's now discuss about whether the CBM App DT fulfills the comprehensive list of high level characteristics that could be found in the literature for DTs. This is an exercise that we must do see if we speak properly when we refer to these tools and algorithms that we embed in CBM apps as DTs, and if so, to see the potentials and the quality of the DT. Table 7 presents the fulfillment of desirable DT features in this paper case study.

10. Conclusions

This paper presents a design of a DT configuration for Condition Based Maintenance and the description of its functionality using the framework in (Crespo Márquez, 2022), overcoming many of the challenges identified for the implementation of this type of tool for maintenance. This configuration introduces three analytical models, with two of them being contributions of the paper specifically designed for failure mode diagnosis/classification and RUL estimations. In these models, machine learning techniques are conjugated with statistical techniques to gain flexibility when modelling in these initial stages of the DT for CBM development.

A very interesting point of the paper is the data transformation strategy to move from temperature data points to temperature cycles data points, allowing the classification of failure modes with a considerable reduction in the quantity of data required for reasonable diagnosis quality.

A selection of six features (scalability, interoperability, expansibility, fidelity, interaction, and integration) to consider when designing these types of DT solutions/systems is a recommendation of this paper, that is done after a careful review of the literature.

Considerable attention was devoted in this work to examine the interaction with the DT using simple and practical business rules.The inclusion of a use case from TALGO, a high-speed train manufacturer, presented as a case study in the paper, adds significant value to the research. During the development of this use case, an approach was adopted to describe the interaction between the DT CBM App and the end users, as illustrated in Fig. 14. This approach employs straightforward business rules to handle events that modify the asset's condition, resulting in a practical business process that effectively addresses the challenges identified in the aforementioned contributions.

In the case study, the reader can appreciate how, to deal with data related challenges, the team used state-of-the-art tools for the data management and models building. For instance, the possibility to carry out the dimensionality analysis with RapidMiner ®, as presented in Fig. 12, could speed up the use of analytics while maintaining the quality of the solution for diagnosis. Also, this tool allows the selection of the more balanced model in terms of fast computation and accuracy of results, to overcome potential limitations in computing infrastructure.

The paper wishes to contribute to bridging the existing gap in digital configurations for maintenance management, designed to benefit of DTs.

Potential limitations of these tools are associated to the existing data, models, and services/apps management. Notice that digital transformation requires an important effort in the maintenance organization to keep all these three layers updated. The need for additional skills in the organization is a clear issue to pay attention to.

Future lines of research are associated to the discovery of new configuration for different critical assets maintenance (horizontal expansion), but also with the incorporation of more advanced tool for integration, from the monitoring infrastructure to the high-level intelligent maintenance apps (vertical integration). Another important development can be using this framework not with maintenance operational services, as the one we are dealing with in this paper, but also for services/apps with a more strategic nature, such as those for dynamic criticality analysis, or those for long-term asset health analysis.

Table 7
CBM App DT features in the case study.

Feature	Case study description
Scalability.	The DT model has been scalable to all train bearings requiring only the development of models per axle bearing position, regardless the axle in the train or the train in the fleet.
Interoperability	Data used to train the three different types of models came from the same source and there is a procedure explaining how original data is converted and matches the different predictive analytics data models. Real time data is now used to generate an on-line output.
Expansibility.	There is a clear possibility to integrate new models. For instance, RUL models based on machine learning models have been introduced to replace statistical models in some applications with more consistent data.
Fidelity	The ML models for anomalies detection replace in this case, with high tested precision, the very complex physical models related to the calculation of the dynamic behavior of loads in the train per axle bearings in each railway point at a certain speed.
Interaction	This part has been found a very interesting feature. When modeling a given failure mode (FM) different risk levels or states are proposed: low, medium, high and fault. At the same time two different types of events may show up: monitoring and preventive maintenance events. It is considered that both monitoring events and PM events (with human intervention) may lead to a change in the risk level of one or more failure modes of the asset. This is because these events trigger a new risk assessment of the affected FMs. A given event may affect different failure modes and in different ways. It is also assumed that reaching a new failure mode state triggers a maintenance action (the release of an algorithm for detection or prediction, an inspection, a replacement, etc.). This human supervision of the model's performance and interaction with the DT resulted to be critical for the DT success.
Integration	The DT is to be integrated in the App in place, to control the trains fleet dynamic maintenance. Axle bearing DT must be incorporated into the comprehensive train CBM App. In this App, a total of 10 train critical systems are monitored to generate an on-line train risk assessment and to suggest an immediate action. Understanding the implications of each system risk, according to each system criticality, is critical to establish an effective dynamic maintenance strategy. In this case this DT has been integrated within Google cloud infrastructure/services (The reader is referred to https://cloud.google.com/customers/talgo for precise details about this integration).

Author Statement

We have carefully reviewed all Editor's and reviewers ́comments and we are resubmitting the paper for your revision. We hope to have address all your concerns.

Declaration of Competing Interest

The authors of this paper declare no conflicts of interest that could be inherent in their submissions.

Data availability

The data that has been used is confidential.

Acknowledgements

This paper has been written within the framework of the projects INMA "Asset Digitalization for INtelligent MAintenace" (Grant PY20 RE014 AICIA, founded by Junta de Andalucía PAIDI 2020, Andalucía FEDER 2014-2020) and Geminhi (Digital model for Intelligent Maintenance based on Hybrid prognostics models), (Grant US-1381456, founded by Junta de Andalucía, Andalucía FEDER 2014-2020). The authors would like to thank the TALGO company for the facilities provided for the design and implementation of these tools.

ANNEX 1. Algorithm generating overtemperature cycles

Template for the initial analysis and data organization.

```
For i
    If absolute error (i)=>10
        Absolute error only in positive(i) = relative error(i);
        Km travelled in positive(i) = Km travelled(i);
        Km accumulated in positive(i) = Km accumulated in positive(i-1) + Km travelled in positive(i);
        Accumulated absolute error(i) = Accumulated absolute error(i-1) + absolute error only in positive(i);
    Else
        Absolute error only in positive(i) = 0;
        Km travelled in positive(i) = 0;
        Accumulated Km in positive(i) = accumulated Km in positive(i-1);
        Accumulated absolute error(i) = accumulated absolute error(i-1);
    End if
    Accumulated Km since first positive(i) = accumulated Km since first positive(i-1) + Km travelled(i);
End_for
```

.

Cycles analysis

```
Import "template for cycles analysis. Beginning and basic calculations. Column (date/time)" = "Template for the initial analysis and data organization. Column (date/time)";
Import "template for cycles analysis. Beginning and basic calculations. Column (VEL)" = "Template for the initial analysis and data organization. Column (Velocity)";
Import "template for cycles analysis. Beginning and basic calculations. Column (Absolute error)" = "Template for the initial analysis and data organization. Column (Absolute error)";
Import "template for cycles analysis. Beginning and basic calculations. Column (Elapsed time)" = "Template for the initial analysis and data organization. Column (Elapsed time)";
Import "template for cycles analysis. Beginning and basic calculations. Column (Km/min)" = "Template for the initial analysis and data organization. Column (average speed(Km/min))";
For i
        Time between measures(i) = Elapsed time(i) *24+60+60;
        Km between measures(i) = (Km/min(i) * Time between measures(i)) / 60;
        Km travelled(i) = Km travelled(i-1) + Km between measures(i);
End_for
Import "template for cycles analysis. Cycle detection. Column (Km travelled)" = "template for cycles analysis. Beginning and basic calculations. Column (Km travelled)";
For i
        If [km travelled(i) – Km travelled(i-1)]<5
                Cycles detection(i) = 1;
        Else
        Cycles detection(i) = 0;
        End_if
        If cycles detection(i) = 0
                Km cycle start(i) = km travelled(i);
                Km cycle end(i) = km travelled(i);
                Km between cycles(i) = km cycle start(i+1) – km cycle end(i);
        Else
                Km cycle start(i) = km cycle start(i-1);
                Km cycle end(i) = km cycle end(i+1);
                Km between cycles(i) = km between cycles(i+1);
        End if
        Km of the cycle(i) = km cycle end(i) – km cycle start(i);
        If cycle detection(i-1) = 0
                If cycle detection(i) = 1
                        Accumulated cycles(i) = accumulated cycles(i-1) + 1;
                Else
                        Accumulated cycles(i) = accumulated cycles(i-1);
                End if
                Accumulated cycles(i) = accumulated cycles(i-1);
        End if
End_for
Import "template for cycles analysis. Cycles representation. Column (km cycle start)" = "template for cycles analysis. Cycles detection. Column (km cycle start)";
Import "template for cycles analysis. Cycles representation. Column (km cycle end)" = "template for cycles analysis. Cycles detection. Column (km cycle end)";
Import "template for cycles analysis. Cycles representation. Column (km between cycles)" = "template for cycles analysis. Cycles detection. Column (km between cycles)";
Import "template for cycles analysis. Cycles representation. Column (km of the cycle)" = "template for cycles analysis. Cycles detection. Column (km of the cycle)";
Import "template for cycles analysis. Cycles representation. Column (accumulated cycles)" = "template for cycles analysis. Cycles detection. Column (accumulated cycles)";
For i
        Cycles/10000 km(i) = accumulated cycles(i)/ [km cycle end(i) * 10000];
        Cycles/3000 km(i) = accumulated cycles(i)/ [km cycle end(i) * 3000];
        Km in active cycle(i) = km in active cycle(i-1) + km of the cycle(i);
        % of km in active cycle(i) = km in active cycle(i)/ [km cycle end(i) + km in first positive];
        Km between accumulated cycles(i) = Km between accumulated cycles(i-1) + km between cycles(i);
        Media of km between cycles(i) = Km between accumulated cycles(i)/accumulated cycles(i);
End_for
```

References

Abbas, M., Liu, Y., Cai, B. (2021). Digital Twin-based Prognostics and Health Management for Subsea systems: Concepts, Classification, Opportunities and Challenges. *Proceedings of the 31st European Safety and Reliability Conference, ESREL 2021*. ⟨https://doi.org/10.3850/978–981-18–2016-8_070-cd⟩.

Animah, I., Shafiee, M., 2018. Condition assessment, remaining useful life prediction and life extension decision making for offshore oil and gas assets (2018). J. Loss Prev. Process Ind. 53, 17–28. https://doi.org/10.1016/j.jlp.2017.04.030.

Canedo, A. (2016). Industrial IoT lifecycle via digital twins. *2016 International Conference on Hardware/Software Codesign and System Synthesis, CODES+ISSS 2016*. ⟨https://doi.org/10.1145/2968456.2974007⟩.

Cempel, C., Natke, H.G., Tabaszewski, M., 1997. A passive diagnostic experiment with ergodic properties. Mech. Syst. Signal Process. https://doi.org/10.1006/mssp.1996.0064.

Crespo, A., Sola, A., Moreu, P., Gómez, J.F., de la Fuente, A., Guillén, A., González-Prida, V., 2018. Criticality analysis for improving maintenance, felling and pruning cycles in power lines. IFAC-Pap. https://doi.org/10.1016/j.ifacol.2018.08.262.

Crespo Márquez, A., 2022. Driving the introduction of digital technologies to enhance the maintenance management process and framework. Springer Series in Reliability Engineering 25–30. https://doi.org/10.1007/978-3-030-97660-6_3.

Crespo Márquez, A., de la Fuente Carmona, A., Marcos, J.A., Navarro, J., 2020. Designing CBM plans, based on predictive analytics and big data tools, for train wheel bearings. Comput. Ind. 122 https://doi.org/10.1016/j.compind.2020.103292.

Crespo Márquez, A., Rosique, A.S., Moreu de León, P., Gómez Fernández, J.F., Diego, A. G., Fernández, E.C., 2020. Exploiting EAMS, GIS and dispatching systems data for criticality analysis. Value Based Intell. Asset Manag. https://doi.org/10.1007/978-3-030-20704-5_7.

Durão, L.F.C.S., Haag, S., Anderl, R., Schützer, K., Zancul, E., 2018. Digital twin requirements in the context of industry 4.0. IFIP Adv. Inf. Commun. Technol. https://doi.org/10.1007/978-3-030-01614-2_19.

Errandonea, I., Beltrán, S., Arrizabalaga, S., 2020. Digital Twin for maintenance: A literature review. Computers in Industry. https://doi.org/10.1016/j.compind.2020.103316160.

Goode, K.B., Moore, J., Roylance, B.J., 2000. Plant machinery working life prediction method utilizing reliability and condition-monitoring data. Proc. Inst. Mech. Eng., Part E: J. Process Mech. https://doi.org/10.1243/0954408001530146.

Guillén, A.J., Crespo, A., Gómez, J.F., Sanz, M.D., 2016. A framework for effective management of condition based maintenance programs in the context of industrial development of E-Maintenance strategies. Comput. Ind. 82 https://doi.org/10.1016/j.compind.2016.07.003.

Healey, R., Wang, J., Chiu, W.K., Chowdhury, N.M., Baker, A., Wallbrink, C., 2021. A review on aircraft spectra simplification techniques for composite structures. Compos. Part C: Open Access Vol. 5. https://doi.org/10.1016/j.jcomc.2021.100131.

Hu, C., Pei, H., Wang, Z., Si, X., Zhang, Z., 2018. A new remaining useful life estimation method for equipment subjected to intervention of imperfect maintenance activities. Chin. J. Aeronaut. https://doi.org/10.1016/j.cja.2018.01.009.

Jardine, A.K.S., Lin, D., Banjevic, D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Signal Process. 20 (7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012.

Konstantinov, S., Ahmad, M., Ananthanarayan, K., Harrison, R., 2017. The cyber-physical E-machine manufacturing system: virtual engineering for complete lifecycle support. Procedia CIRP. https://doi.org/10.1016/j.procir.2017.02.035.

Kotu, V., Deshpande, B., 2019. Data Science Concepts and Practice. Data Handling in Science and Technology.

Lunys, O., Dailydka, S., Bureika, G., 2015. Investigation on features and tendencies of axle-box heating. Transp. Probl. 10 (1) https://doi.org/10.21307/tp-2015-011.

Marquez, A.C., Fernandez, J.F.G., Fernández, P.M.G., Lopez, A.G., 2020. Maintenance management through intelligent asset management platforms (IAMP). Emerging factors, key impact areas and data models. Energies 13 (15). https://doi.org/10.3390/en13153762.

Martínez-Galán Fernández, P., Guillén López, A.J., Márquez, A.C., Gomez Fernández, J. F., Marcos, J.A., 2022. Dynamic Risk Assessment for CBM-based adaptation of maintenance planning. Reliab. Eng. Syst. Saf. 223 https://doi.org/10.1016/j.ress.2022.108359.

Medjaher, K., Tobon-Mejia, D.A., Zerhouni, N., 2012. Remaining useful life estimation of critical components with application to bearings. IEEE Trans. Reliab. https://doi.org/10.1109/TR.2012.2194175.

Okoh, C., Roy, R., Mehnen, J., Redding, L., 2014. Overview of Remaining Useful Life prediction techniques in Through-life Engineering Services. Procedia CIRP. https://doi.org/10.1016/j.procir.2014.02.006.

Peng, F., Zheng, L., Peng, Y., Fang, C., Meng, X., 2022. Digital Twin for rolling bearings: a review of current simulation and PHM techniques. Meas.: J. Int. Meas. Confed. Vol. 201 https://doi.org/10.1016/j.measurement.2022.111728.

Qiu, J., Seth, B.B., Liang, S.Y., Zhang, C., 2002. Damage mechanics approach for bearing lifetime prognostics. Mech. Syst. Signal Process. https://doi.org/10.1006/mssp.2002.1483.

Ramezani, S., Moini, A., Riahi, M., Marquez, A.C., 2020. A model to determining the remaining useful life of rotating equipment, based on a new approach to determining state of degradation. J. Cent. South Univ. 27, 2291–2310. https://doi.org/10.1007/s11771-020-4450-7.

Rosen, R., Von Wichert, G., Lo, G., Bettenhausen, K.D., 2015. About the importance of autonomy and digital twins for the future of manufacturing. IFAC-Pap. https://doi.org/10.1016/j.ifacol.2015.06.141.

Schleich, B., Anwer, N., Mathieu, L., Wartzack, S., 2017. Shaping the digital twin for design and production engineering. CIRP Ann. - Manuf. Technol. https://doi.org/10.1016/j.cirp.2017.04.040.

Shi, H., Song, Z., Bai, X., Hu, Y., Li, T., Zhang, K., 2023. A novel digital twin model for dynamical updating and real-time mapping of local defect extension in rolling bearings. Mech. Syst. Signal Process. 193 https://doi.org/10.1016/j.ymssp.2023.110255.

Si, X.S., Wang, W., Hu, C.H., Zhou, D.H., 2011. Remaining useful life estimation - a review on the statistical data driven approaches. Eur. J. Oper. Res. https://doi.org/10.1016/j.ejor.2010.11.018.

Uhlemann, T.H.J., Schock, C., Lehmann, C., Freiberger, S., Steinhilper, R., 2017. The digital twin: demonstrating the potential of real time data acquisition in production systems. Procedia Manuf. https://doi.org/10.1016/j.promfg.2017.04.043.

van Dinter, R., Tekinerdogan, B., Catal, C., 2022. Predictive maintenance using digital twins: a systematic literature review. Inf. Softw. Technol. Vol. 151 https://doi.org/10.1016/j.infsof.2022.107008.

Yang, C., Cai, B., Shao, X., Liu, Y., Liu, Y., Feng, Q., … Wang, H. (2021). Application of digital twins in condition-based maintenance. *Proceedings of the 31st European Safety and Reliability Conference, ESREL 2021*. ⟨https://doi.org/10.3850/978–981-18–2016-8_365-cd⟩.

Zhang, B., Zhang, S., Li, W., 2019. Bearing performance degradation assessment using long short-term memory recurrent network. Comput. Ind. https://doi.org/10.1016/j.compind.2018.12.016.

Zhang, H., Liu, Q., Chen, X., Zhang, D., Leng, J., 2017. A digital twin-based approach for designing and multi-objective optimization of hollow glass production line. IEEE Access. https://doi.org/10.1109/ACCESS.2017.2766453.