UNISINOS

Programa de Pós-Graduação em

Computação Aplicada

Doutorado Acadêmico

Marcos Leandro Hoffmann Souza

Prognosis & Health Management System (PHMS): a Machine Learning Framework to support decision-making in predictive maintenance in a production system

São Leopoldo, 2023

Marcos Leandro Hoffmann Souza

PROGNOSIS & HEALTH MANAGEMENT SYSTEM (PHMS): A MACHINE LEARNING FRAMEWORK TO SUPPORT DECISION-MAKING IN PREDICTIVE MAINTENANCE IN A PRODUCTION SYSTEM

Thesis presented as a partial requirement to obtain the Doctor's degree from the Applied Computing Graduate Program of the Universidade do Vale do Rio dos Sinos - UNISINOS

Advisor:
Prof. Dr. Cristiano André da Costa

Co-advisor:
Prof. Dr. Gabriel de Oliveira Ramos

São Leopoldo
2023

S729p Souza, Marcos Leandro Hoffmann.
Prognosis & Health Management System (PHMS) : a machine learning framework to support decision-making in predictive maintenance in a production system / by Marcos Leandro Hoffmann Souza. – 2023.
140 f. : il. ; 30 cm.

Thesis (Doctorate) — Universidade do Vale do Rio dos Sinos, Applied Computing Graduate Program, São Leopoldo, RS, 2023.
Advisor: Dr. Cristiano André da Costa.
Co-advisor: Dr. Gabriel de Oliveira Ramos.

1. Decision-making. 2. Industry 4.0. 3. Internet of things. 4. Deep learning. 5. Machine learning. 6. Reliability. 7. Remaining useful life. 8. Predictive maintenance. I. Title.

UDC: 658.58:004.85

Cataloging in Publication (CIP):
Librarian Alessandro Dietrich - CRB 10/2338

À memória do meu pai,Júlio Mauro Huff de Souza, que deixou muita saudade,e que eu gostaria muito que estivesse presente neste momento.

AGRADECIMENTOS

Como já dizia Anitelli: “Sonho parece verdade quando a gente esquece de acordar”. Hoje, vivo uma realidade que parece um sonho, mas foi preciso muito esforço, determinação, paciência, perseverança, ousadia e maleabilidade para chegar até aqui, e nada disso eu teria conseguido sozinho. Minha terna gratidão a todos os que colaboraram para que este sonho pudesse ser concretizado. Agradeço primeiramente a Deus, pela minha saúde e disposição para seguir até o fim dessa caminhada. Aos meus pais Nadir Luiza e Júlio Mauro (In memoriam), que de uma forma simples e correta me ensinaram a valorizar os estudos e me incentivaram, acreditando na minha capacidade. Aos meus irmãos Carlos, Eliane e Joceli, que sempre me incentivaram a ir em busca dos meus sonhos. À minhas filhas Ana Caroline e Helena, pela compreensão nos momentos de ausência e pelo amor destas minhas princesas que incentivou nos momentos de tensão. Agradeço aos sorrisos da Helena quando abria a porta do quarto me olhava e sorria, isso redobrava minhas energias. A meus netos Rafael e Maria Luiza, pela compreensão nos momentos de ausência. À minha esposa, amiga e companheira Lia, que dedicou grande compreensão, amor, carinho e incentivo cuidando de tudo enquanto me dedicava na realização deste sonho. À família de minha esposa, Sirlene, Leo, Luciano, Lucídio e Leia pela compreensão dos momentos de ausência. Ao Prof. Dr. Cristiano André da Costa, por sua amizade, confiança, apoio e orientação no desenvolvimento do trabalho. Agradeço também ao Prof. Dr. Gabriel de Oliveira Ramos pela coorientação e todos os ensinamentos no desenvolvimento do trabalho. Agradeço aos professores do PPGCA, principalmente ao Prof. Dr. Rodrigo Da Rosa Righi pelo acolhimento no programa. Agradeço aos colegas de trabalho: Marcus Leite, Victor, Danilo e Marcia por todo apoio principalmente da reta final deste trabalho. A todos os amigos que de uma forma ou de outra contribuíram para a realização deste sonho. Agradeço a Unisinos que me proporcionou uma formação completa e pela excelência dos professores que tive, tanto na graduação quando na pós-graduação.

RESUMO

A busca pela utilização efetiva dos ativos de produção tem sido constante, principalmente em indústrias com mecanização em evolução. Desta forma, a gestão da manutenção ganha visibilidade por ser responsável por garantir a disponibilidade dos ativos. A manutenção preditiva (MP) é uma das principais estratégias de gestão da manutenção. Permite a detecção precoce de falhas, evitando paradas não programadas e custos desnecessários. À medida que as tecnologias avançaram, a manutenção preditiva contribui para que gestão e prognóstico de saúde seja aperfeiçoada e fornece os meios para reconhecer padrões, entender anomalias e estimar a vida útil restante do equipamento. Paralelamente, tecnologias como internet das coisas, aprendizagem de máquina e computação em nuvem permitem a digitalização de ativos, proporcionando uma manufatura inteligente. No entanto, este cenário torna a MP uma tarefa complexa e cara quando aplicada a sistemas com equipamentos interligados em série. Por um lado, os dados são abundantemente gerados, coletados e armazenados. Por outro lado, há dificuldade em converter os dados em informações úteis para suportar MP e gestão e prognóstico de saúde. Diante das lacunas referentes a MP e confiabilidade, sugerimos nesta tese o Prognosis and Health Management System (PHMS) que é suportado por um framework analítico que utiliza um conjunto de técnicas e modelos aprendizagem de máquina. Para avaliar a proposição, realizamos um estudo de caso com dados reais da indústria de processo. No desenvolvimento do framework usamos aprendizagem de máquina semi-supervisionado com Autoencoder (AE) para construção do limiar operacional e identificação de anomalias. Para a etapa de identificação de variáveis aplicamos o XGBoost e o método SHAP. Na sequência, testamos diferentes arquiteturas de aprendizagem profunda para previsão de vida útil restante do sistema. Na previsão vida útil restante apresentamos diferentes arquiteturas de aprendizagem profunda. Neste sentido, destacamos a arquitetura de aprendizagem profunda N-BEATS como uma alternativa importante, em comparação com arquiteturas tradicionais como Redes Neurais Recorrentes (RNN). Por meio do framework aplicado ao estudo de caso, foi possível identificar uma anomalia, o comportamento das variáveis mais relevantes para a falha, e prever a vida útil restante do equipamento com R^2 superior a 90% com o N-BEATS. Desta forma, de acordo com os resultados apresentados, as equipes de operação e manutenção podem realizar ações preventivas evitando paradas não programadas do sistema produtivo. Neste sentido, o desenvolvimento do framework contribui para a adoção de tecnologias emergentes em processos reais. Além dos benefícios apresentados, destacamos o desenvolvimento de estudos de MP em dados reais desconhecidos do ambiente acadêmico. Chamamos atenção para este ponto, pois a ampla maioria dos estudos de confiabilidade são realizados sobre dados amplamente conhecidos e tratados.

Palavras-chave: Tomada de decisão. Indústria 4.0. Internet das Coisas. Aprendizagem de máquina. Confiabilidade. Vida útil restante.

ABSTRACT

The search for the effective use of production assets has been constant, mainly in industries with evolving mechanization. In this way, maintenance management gains visibility as it ensures asset availability. Predictive maintenance (PDM) is one of the main maintenance management strategies. Allows early detection of failures, avoiding unscheduled downtime and unnecessary costs. As technologies have advanced, predictive maintenance improves Prognosis and Health Management (PHM). It provides the means to recognize patterns, understand anomalies and estimate the equipment's remaining useful life (RUL). At the same time, technologies such as the internet of things (IoT), machine learning (ML), and cloud computing enable the digitization of assets, providing intelligent manufacturing. However, this scenario makes PDM a complex and expensive task when applied to systems with equipment connected in series. On the one hand, data is abundantly generated, collected, and stored. On the other hand, it is difficult to convert data into useful information to support PDM and PHM. Given the gaps related to PDM and reliability, we suggest the Prognosis and Health Management System (PHMS) in this thesis, which is supported by an analytical framework that uses a set of techniques and ML. First, we performed a case study to evaluate the proposition with real data from the process industry. In developing the framework, we used semi-supervised ML with Autoencoder (AE) to build the operational threshold and identify anomalies. For the Feature Identification step, we applied XGBoost and the SHAP method. Next, we test different deep learning architectures to predict the RUL of the system. In the RUL prediction, we present different deep learning architectures. In this sense, we highlight the N-BEATS deep learning architecture as an essential alternative to traditional architectures such as Recurrent Neural Networks (RNN). Through the framework applied to the case study, it was possible to identify an anomaly and the behavior of the most relevant variables for the failure and predict the RUL of the equipment with R^2 greater than 90% with N-BEATS. In this way, according to the results presented, the operation and maintenance teams can carry out preventive actions, avoiding unscheduled stops of the production system. In this sense, the development of the framework contributes to the adoption of emerging technologies in real processes. In addition to the benefits presented, we highlight the development of PDM studies on real data unknown in the academic environment. We draw attention to this point, as most reliability studies are based on widely known and treated data.

Keywords: Decision-making. Industry 4.0. Internet of Things. Deep learning. Machine Learning. Reliability. Remaining Useful Life.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

API	Application Programming Interface
AE	Autoencoder
AI	Artificial Intelligence
BP	Back-propagation
BS	British Standard
CBM	Maintenance-based conditions
CM	Condition Monitoring
CNN	Convolutional Neural Network
CPS	Cyber-Physical System
DCS	Distributed Control System
DL	Deep Learning
ELU	Exponential Linear Unit
FE	Feature Extraction
FI	Feature Identification
GRU	Gated Recurrent Unit
I4.0	Industry 4.0
ICS	Industrial Control Systems
IoT	Internet of Things
ISO	International Organization for Standardization
LSTM	Long Short-Term Memory
LOF	Local Outlier Factor
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multilayer Perceptron
MQ	Main Question
MSE	Mean Squared Error
MTBF	Mean Time Between Failures
MTTR	Mean Time to Repair
N-BEATS	Neural Basis Expansion Analysis For Interpretable TS Forecasting
PDM	Predictive Maintenance
PHM	Prognostics and Health Management
PHMS	Prognosis and Health Management System

ReLU	Rectified Linear Unit
RF	Random Forest
RFR	Random Forest Regressor
RNN	Recurrent Neural Networks
RMSE	Root Mean Square Error
RL	Reconstruction Loss
RUL	Remaining Useful Life
SCADA	Control and Data Acquisition
SHAP	SHapley Additive exPlanations
SM	Styrene Monomer
SVM	Support Vector Machine
LSM	Smart Local Moving
SQ	Sub-questions
SQL	Structured Query Language

CONTENTS

1 INTRODUCTION

Technology has supported and catalyzed advances in different industries and processes in recent years (WANG et al., 2018). In this context, in production systems, it is essential to adapt to the needs and requirements of different markets, such as a high product mix, low life cycle, and constant introduction of new competitors (GÄRTNER, 2018). Therefore, production models need to differ from linear models, where processes are like silos without interaction and communication between operations. For this, a new paradigm of production system emerges as a solution, which has as its primordial requirement the use of technology for the integration of devices and digital systems (Kagermann; WAHLSTER; HELBIG, 2013; TERRISSA et al., 2016).

The German government first declared the Fourth Industrial Revolution (a.k.a *Industry 4.0 (I4.0)*) at the Hannover Fair in 2011, viewed as a high-tech strategy for 2020, and notably based on the development of cyber systems (HENNING LUKAS WOLF-DIETER, 2011). Modern industrial systems are complex systems that integrate physical, software, and network components into so-called cyber-physical systems (CPS) (LIU et al., 2018). Under this perspective, a production system consists of several interconnected types of equipment with different functions. In process industries comprising a wide variety of manufacturing systems, from petrochemical facilities to glass, food, and pharmaceutical manufacturing, the adoption of technology has been constant, mainly to increase the availability of production plants (PERNO; HVAM; HAUG, 2022). Thus, continuous advances in several areas have increased expectations about the performance of these systems in terms of reliability and responsiveness to support decision-making (SÉNÉCHAL; TRENTESAUX, 2019; SIAFARA et al., 2017).

One of the Predictive maintenance (PDM) main objectives is to determine how long a system will run by analyzing data from its component equipment (FILZ et al., 2021; MÁRQUEZ et al., 2020). However, implementing PDM at the system level is a complex and costly task when assets are interconnected. However, the results tend to offset the adoption efforts (LARRINAGA et al., 2018). Furthermore, technological advances driven by I4.0, such as the IoT, Artificial Intelligence (AI), and cloud computing, enable the digitalization of assets creating smart manufacturing. Thus, they contribute to PDM being at the center of attention of researchers and the business area. Thus, through technologies, PDM becomes viable in the industry in complex environments (PERNO; HVAM; HAUG, 2022; CHRISTOU et al., 2022).

In the smart manufacturing context, using technologies provides several gains, especially for assets with Condition Monitoring (CM). In this regard, Condition-based Maintenance (CBM) has evolved through equipment data. Condition-based monitoring consists of monitoring a machine condition parameter so that a deviation indicates a developing fault and may be related to a specific variable, enabling the system to trigger a warning or alarm (CALLE et al., 2019). As technologies have advanced, CBM has evolved into Prognosis Health Management (PHM), which provides robust capabilities for dynamic pattern identification and enables understanding and estimating the Remaining Useful Life (RUL) of equipment and systems (KONG et al., 2020; CALLE et al., 2019). Therefore, the enabling technologies play an important role in CBM and PHM adoption, allowing for early fault detection and avoiding losses for organizations (CAIAZZO et al., 2022; PERNO; HVAM; HAUG, 2022; SCHWARTZ et al., 2022; ROSATI et al., 2022).

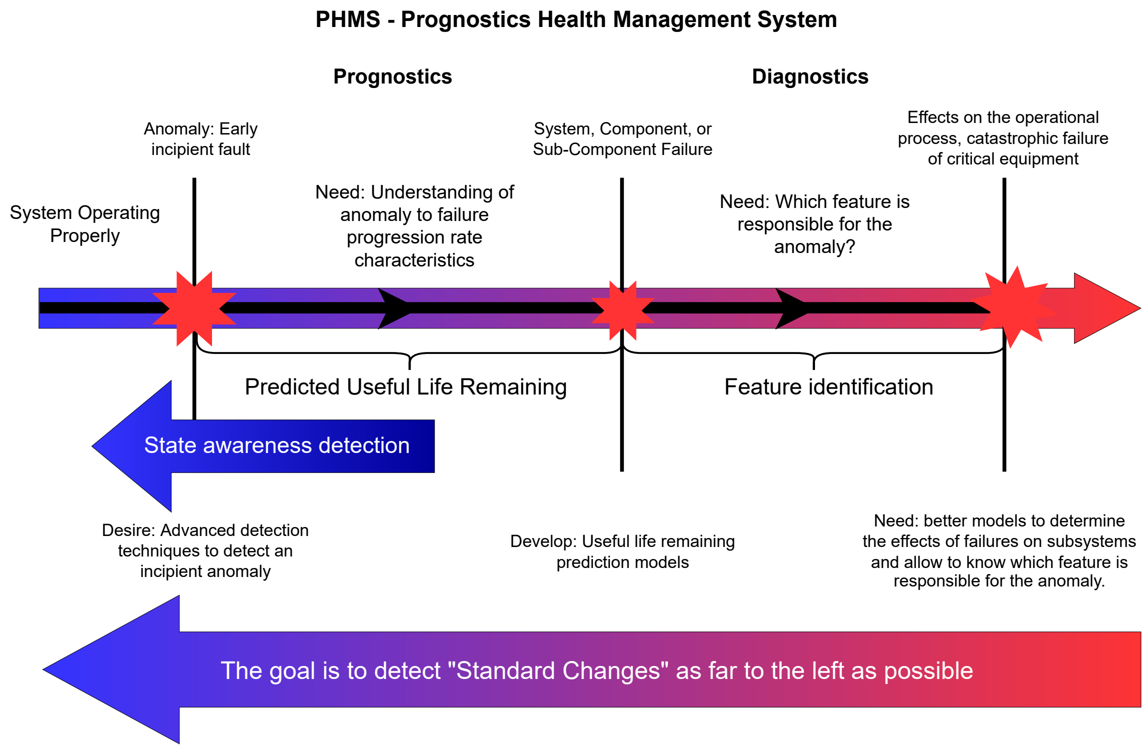
Early anomaly detection is the starting point for understanding a potential failure. When an abnormal condition is detected in the PHM approach, a possible action is to estimate the time until the equipment or system does not meet the functional requirements. In this sense, the interval between anomaly detection and failure consists of the equipment's RUL. Thus, a clear understanding of the system is essential for fault diagnosis and prognosis, and this is currently made possible through the advancement of enabling technologies (ROSATI et al., 2022; CHEN et al., 2021; YANG; ZHENG; QI, 2020; KRAUS; FEUERRIEGEL, 2019).

Approaches for anomaly detection and maintenance estimation are typically used to identify failure in a system component or subcomponent (ZOU et al., 2022). In essence, prognosis provides the predictive element complementing the diagnostic capabilities to identify and quantify the potential defect (SCHWARTZ et al., 2022). From an operator and maintainer's perspective, the provision of the warning time for useful life or failure is what separates prognosis from diagnosis. Therefore, prognostics is the PHM component that estimates the time until a failure occurs, enabling operators and maintainers to take preventive actions (CAIAZZO et al., 2022; SCHWARTZ et al., 2022). According to Figure 1, detection and correction actions should be as close to the anomaly as possible. Thus, managing the relationships between diagnostics and prognostics is of utmost importance to avoid unplanned downtime and consequent unnecessary costs.

Studies addressing PHM point to the need to improve the forecasts' reliability. Regarding accuracy, it is crucial since RUL is still a developing subject (FERREIRA; GONÇALVES, 2022; ROSATI et al., 2022). Therefore, anomaly detection is critical, particularly in an environment in which the data is noisy, such as the industrial environment (YANG et al., 2022; BENKER et al., 2021). However, getting consistent samples to train ML models for new types of equipment and peripherals takes a lot of work. As such, this field is highly relevant and challenging, especially for estimating RUL in a real industrial environment (BENKER et al., 2021; MA; MAO, 2021).

According to Figure 1, the cornerstone of a successful fault detection and prognosis approach lies in Feature Identification (FI) (OMRI et al., 2020). When developing a CBM and PHM approach for complex systems, converting raw sensor data is critical since it is how to generate information (CALLE et al., 2019). The application of PHM gains greater importance when applied to systems since recent studies focus particularly on equipment (CHRISTOU et al., 2022). In this sense, when studies address the term system, they refer to equipment components and peripherals, such as bearings, fasteners, and electrical systems. Thus, studies addressing systems' PHM, considering different equipment and peripherals, still represent a

Figure 1 – PHMS - Prognosis and Health Management System and anomaly progression timeline.



Source: Prepared by the author.

gap to be explored (CHRISTOU et al., 2022).

1.1 Motivation

The reliability of systems supported by PDM has motivated not only research whose objective is to highlight the importance of the topic for decision-makers. Moreover, with the computational power and abundant data generation resulting from technological advances, the subject of reliability has promoted studies where sophisticated computational models are used to improve the prediction of equipment failures (CHEN et al., 2020; SAHAL; BRESLIN; ALI, 2020).

However, it is possible to see through the current literature that addresses PDM to have focused on equipment-centric decisions that do not encompass the entire production system (LAZAROVA-MOLNAR; MOHAMED, 2019; NAPOLEONE; MACCHI; POZZETTI, 2020; VOGL; WEISS; HELU, 2019). Thus, the prioritization of maintenance and operation actions end up not being effective since knowing the conditions of the equipment is fundamental for the strategies of the maintenance team (SELLITTO, 2005). However, if the equipment is not a bottleneck, directing efforts to this system may not offer gains for the entire organization

(CHEN et al., 2020; SAHAL; BRESLIN; ALI, 2020). This way, using the latest manufacturing technologies in synergy with computational power opens opportunities to improve the business's strategic levels.

Knowing when a failure in equipment or system will occur contributes to better decisions in this context. Generally, when a specific action needs to be performed in an emergency, the available alternatives are reduced without prior planning, and the cost increases. Figure 1 presents the central point at which decision-making is of fundamental strategic importance for the business. In this way, the PHMS, as shown in Figure 1, is an alternative approach to increase system reliability through more efficient asset management.

1.2 Research Question

The motivation was confirmed through a systematic review related to systems' reliability as a support for decision-making. Thus, some open challenges and opportunities were listed, particularly as already presented, the possibility of using PDM of systems strategically to support better decisions that contribute to increasing the availability of the systems. From the observed opportunities and challenges, the following research question was listed:

How to develop a prognosis and health management for a production system for support decision-making using data from equipment and operation for anomaly detection and failure prediction in Industry 4.0?

The research question, as well as the proposal's theme, are related to:

- Possibility to increase system reliability by applying the PHMS;
- Evaluate the benefits of I4.0, such as IoT and ML, to improve the performance of industrial operations considering operation and maintenance;
- Provide support for decision-making operation and maintenance teams with data analysis.

1.3 Scientific Contributions

The main scientific contributions of this work are:

1. Propose a taxonomy with reliability constructs for decision-making in I4.0;
2. Develop an ML pipeline for treatment, analyze and predict failures in a process industry;
3. Create an operational threshold considering data from different equipment with peripheral systems and operational features applying Deep Learning (DL);
4. Develop a framework for PHMS that aims to identify an anomaly, the possible root cause, and the prediction of the RUL;

5. Provide the scientific community with a contextualized dataset from a real environment for future research.

In addition to the scientific contributions, this work was able to point out some technological contributions, such as: use of a dataset without publications related to the problem of predicting the RUL, evaluate different DL architectures for RUL prediction in real and noisy environments; adoption of interpretability of FI models in the field of PDM.

1.4 Objectives

The general objective of this study is to present an intelligent maintenance approach. To do this, we take as a backdrop Figure 1 and propose the PHMS as a pipeline. This provides a robust method that can enrich the decision-making process in maintenance and operations.

1.4.1 Specific Objectives

Specific objectives include:

1. A detailed review of the state of the art, challenges, and applications related to decision-making in the context of systems reliability in I4.0;
2. Develop a case study with real data from a petrochemical plant to evaluate the proposed framework;
3. Present a FI model considering the features of the production system and highlighting the most relevant to generate abnormal conditions over time in a process industry;
4. Test the adherence of different DL architectures in time series with noisy and non-stationary data to predict the RUL of the system;

The proposed objectives aim to offer decision-makers a robust asset or operations management approach. In line with the presented gaps and opportunities, the objectives also provide consistent advances for academia in terms of PDM. Thus, the following subsection presents the outline of this thesis.

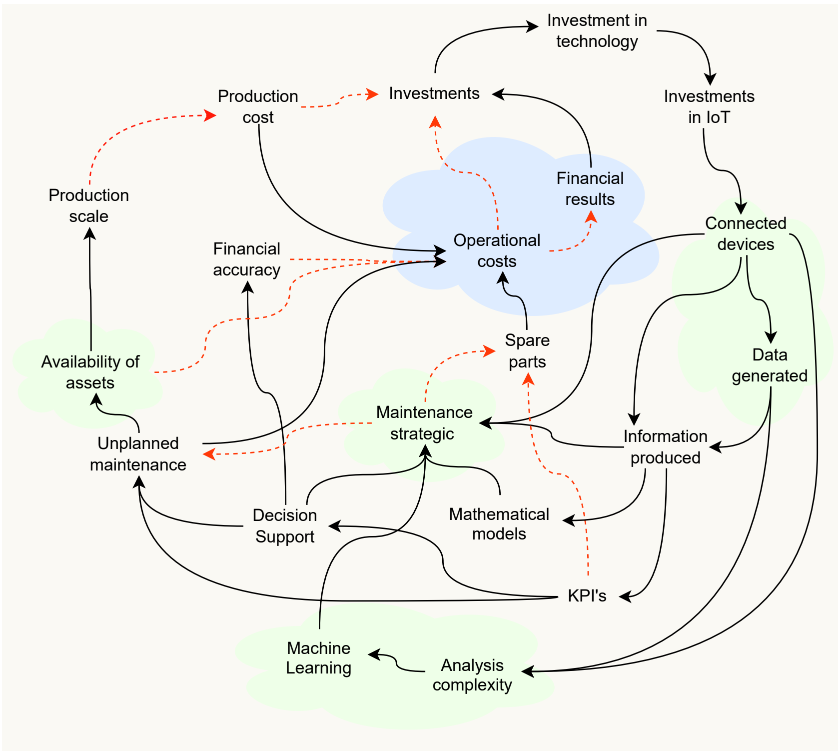
1.5 Document Organization

Having completed the introductory part, this work is organized into seven chapters. Chapter 2 presents the theoretical foundation and the background of the contents used to formulate this thesis proposal, information about I4.0, reliability, maintenance management, and IoT. Chapter 3 highlights the related works based on the previously published Souza et al. (2020) systematic review. This Chapter justifies the intention and hypotheses for carrying out this work. Chapter 4 describes the general structure of the proposed framework for PHMS. Chapter 5 details the methodological procedures to meet the objectives of this thesis. Finally, Chapter 6 presents and discusses the results obtained based on the previously published research Souza et al. (2021). Finally, the conclusion in Chapter 7, along with future work.

2 BACKGROUND

At first, to identify the main concepts to be studied and their relationships, a systemic map displayed in Figure 2 was created. With the systemic diagram, it is possible to visualize the benefits of adopting technologies to production systems, especially the IoT. Systemic diagrams facilitate understanding of the causal relationships between the topics addressed in the research and highlight the gains obtained with the technologies. When reading the systemic structure, we have the relationships represented by colors and meaning. For example, solid black lines represent directly proportional effects. The red dotted lines represent an inversely proportional relationship (SOUZA; RODRIGUES; MORANDI, 2018). In this way, we can say that the more "Investments in IoT", the more "connected devices" and the more "connected devices", the more "generated data"; the more "generated data", the greater the "analysis complexity" and thus, the more "Machine Learning" and the more "Machine Learning", the better "Maintenance strategies" can be used. Following the reasoning, the better the "Maintenance strategy", the less "Unplanned maintenance" and the less "Spare parts" and with that, the less "Operating costs". Through this analysis, the most relevant themes are evident, which justifies attention in this section.

Figure 2 – Systemic background scope map: Green clouds represent the focus of this research.



Source: Prepared by the author.

In the systemic map, the concepts are related and converge to a focal point of financial increment for the organization. Financial gains are not part of the scope of this thesis and are highlighted in blue in Figure 2. However, for financial gains to become a reality, we position the IoT technology driven by I4.0 as a fundamental link concerning the approaches proposed in this research. In this way, we started detailing these concepts, outlining the reliability obtained through the maintenance processes and, following with the challenges and technologies applied to the I4.0, which are fundamental themes for the elaboration of this thesis.

2.1 Reliability

According to Ziegel (2004), the first paper published to address the term reliability was in 1963 in the IEEE Journal. In this way, Kirkmant (1963) defines reliability as the probability that a technological system will operate correctly for a given period in a given environment. However, in this study, we use reliability definitions according to standards such as ISO[1] 8402 and BS[2] 4778, as described in the sequence.

Reliability is *the ability of an item to perform a necessary function, under environmental and operational conditions over a defined period (ISO 8402).*

- The expression "item" in this case indicates equipment, subsystem, or system that can be an entity;

- An essential role can be a single role or a set of functions required to provide a particular service.

Countless efforts have been made since the 1970s in the field of reliability to evaluate equipment and systems, as well as to meet the requirements for which they were designed (BURDICK et al., 1977; ZIEGEL, 2004).

Preliminary reliability studies typically classify systems based on mutually exclusive states *working* or *failed* (KIRKMANT, 1963). However, the need for improvement in reliability studies has evolved along with the requirements of products and processes (BLACK; MEJABI, 1995; Kagermann; WAHLSTER; HELBIG, 2013; LIN; CHANG, 2012; NGUYEN et al., 2016).

With the concept of reliability presented and positioned within this research, the next step is to understand how this concept relates to equipment briefly.

2.1.1 Equipment Reliability

The equipment's reliability corresponds to the mission of meeting the requirements, and the performance expected by the manufacturing systems (BIANCHINI; PELLEGRINI; ROSSI, 2019; USTUNDAG; CEVIKCAN, 2018). The equipment is subject to continuous degradation, depending on the operations in which it performs.

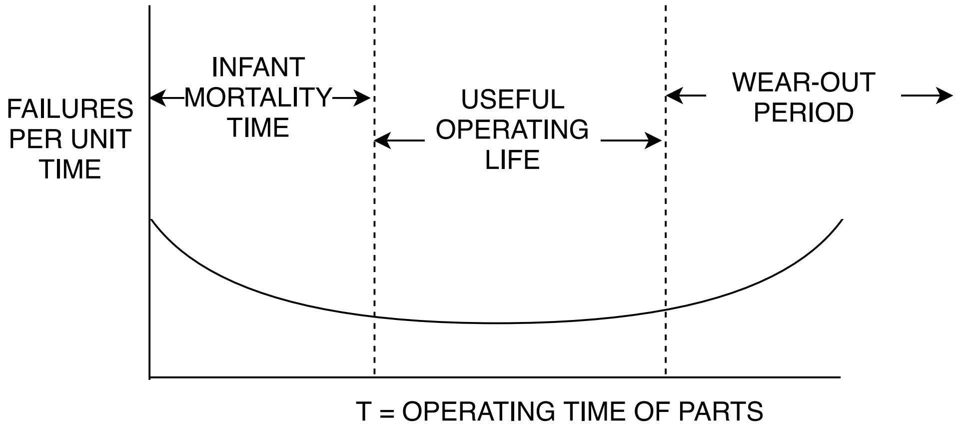
[1]International Organization for Standardization
[2]British Standard

Traditional equipment monitoring models are focused on analyzing availability where the times between failure and the times between repairs are recorded and provide information for calculating availability. It is possible to describe the Availability Equation 2.1 employing the Mean Time Between Failures (MTBF) and the Mean Time to Repair (MTTR) (GOPALAKRISHNAN et al., 2019; ZIEGEL, 2004; SALUNKHE; FUMERO, 2017).

$$A_{av} = \frac{MTBF}{MTBF + MTTR} \tag{2.1}$$

With the data used to calculate availability, it is possible to estimate the probability distribution that best corresponds to the failure mode of the equipment (HASHEMIAN; BEAN, 2011; LAZAROVA-MOLNAR; MOHAMED, 2019). The principal probability distributions used are; exponential, Weibull, gamma, and log-normal (SAGE; ROUSE, 1999). With the reliability distribution that best represents the failure mode, it is possible to analyze the equipment's health. The bathtub curve, Figure 3, has been used to determine the best maintenance strategy based on the time of operation (KIRKMANT, 1963; ZIEGEL, 2004).

Figure 3 – The bathtub curve.



Source: Prepared by the author.

In the bathtub curve, the authors propose three phases, as well as the maintenance strategy, suggested for the equipment (CANITO et al., 2017; KIRKMANT, 1963; MACCHI et al., 2012):

- **Infant Mortality**: in this region, premature failures occur, such as failures or errors in the manufacturing, installation, or operation processes. In this situation, the best strategy for the period is corrective maintenance, analyzing and resolving the causes of premature failures;

- **Useful Life**: in this phase, the equipment has a constant failure rate, originated by random failures, as an insufficient safety factor or errors in the operation process. In this case, the maintenance strategy indicated would be PDM;

- **Wear-out**: in this region, the failure rate has a gradual increase as time progresses, caused by usage time, fatigue, or wear. At this moment, the indicated strategy would be preventive maintenance.

In the survey conducted by (HASHEMIAN; BEAN, 2011), he pointed out that few industries benefit from PDM techniques (HASHEMIAN; BEAN, 2011). Instead, the engineers use PHM to detect anomalies or defects in the equipment and its systems (NEMETH et al., 2018). This maintenance is usually done in an invasive manner, being visually and physically on the machine where, in some cases, it is necessary to shut down the system (CANITO et al., 2017).

Although maintenance based on periodic reviews is the most applied and used method, these techniques are an unproductive and unreliable approach, according to recent research classification (GAO et al., 2015). This reliability management model is less efficient and expensive due to unplanned downtime, and possible damage to equipment (BIANCHINI; PELLEGRINI; ROSSI, 2019). Also, in some cases, the need for a large number of spare parts in stock, which would not be necessary in many cases, ends up increasing costs for the company (KINNUNEN et al., 2018; KOLINSKA; KOLINSKI, 2018; LIN; CHANG, 2012).

The evolution of technology collaborated with PDM, providing advanced and less invasive techniques with passive monitoring (CANIZO et al., 2017). The use of wireless devices and Supervisory Control and Data Acquisition (SCADA) provides companies with new ways to collect information about the performance of their industrial assets (NEMETH et al., 2018). These systems can obtain more data in a more straightforward and less complicated manner (KIANGALA; WANG, 2018; NEMETH et al., 2018; MCA; MCA; MANDA, 2018). Therefore, the volumes and variety of data available are more significant. However, a problem that arises is identifying which data is the most important and which is not. Consequently, the challenge is to know how to obtain valuable information from this data and generate knowledge to support decision-making processes (KIANGALA; WANG, 2018; ZHONG et al., 2017).

As part of new technologies, Data Analytic (DA) algorithms have started to extract information from historical data. Such algorithms allow us to identify industrial equipment's behavior over time and estimate possible future failures based on the extracted information (ZHONG et al., 2017). This technological approach used in the industries has been gaining importance and speed due to the movement provided by I4.0 (BOUSDEKIS; MENTZAS, 2017). It has its bases centered on information technology and which, in turn, provides the collection, storage, and analysis of a large amount and a variety of data (GAO et al., 2015; Kagermann; WAHLSTER; HELBIG, 2013; LAZAROVA-MOLNAR; MOHAMED, 2019; ZHONG et al., 2017).

Like the human body, production systems are composed of several operations. In this analogy, systems composed of different equipment must be operating within the best conditions for the production to deliver the expected results (BIANCHINI; PELLEGRINI; ROSSI, 2019; KIANGALA; WANG, 2018). That is, they need the system's reliability to be the best possible. Therefore, in the next topic, we place this concept within the scope of this research.

2.1.2 Systems Reliability

Currently, changes in the context of productive systems have been persistent. The diversity of products with distinct characteristics has been a requirement of different markets (CHEN et al., 2020). Regardless of the sector and the activity, the production systems' features are similar, where high productivity, low cost, and demand with high variability are competitive factors (NGUYEN et al., 2016).

Researchers have suggested different maintenance strategies to save system reliability and meet production requirements along with the manufacturing change context. A system typically comprises several subsystems and components that are interconnected so that the system can perform a set of necessary functions (ZIEGEL, 2004).

It is necessary to analyze the system as a whole, design, manufacture, and application to meet markets' requirements, such as demands for high reliability and durability. Therefore, it is required to maximize reliability throughout the product's life cycle(BOUSDEKIS; MENTZAS, 2017; CHEN et al., 2020). Thus, companies' competitiveness is deeply related to the reliability of the manufacturing system (LEE, 2020).

To some extent, the trouble-free operation capability of the manufacturing system determines the level of inherent reliability formed in the manufacturing process (BOUSDEKIS; MENTZAS, 2017). The traditional reliability modeling of manufacturing systems tends to follow the classic methods of the reliability block diagram, Fault Tree Analysis, Petri Nets, and so on (LEE, 2020). These methods provide a comprehensive analysis of the manufacturing system. However, they end up being complicated or inconvenient, considering that it depends on the prior analysis of the subsystem and equipment reliability individually (LEE, 2020; MEJÍA; PEREIRA, 2020; ZIEGEL, 2004)

The introduction of technology to efficiently check the equipment's condition is currently one of the most relevant maintenance tools, contributing to predictive models' improvements. Detecting failures or threats on a device before they occur and suggesting repairs to reduce the likelihood of failure is an essential contribution to productive systems. According to this scenario, PDM, which uses monitoring and data analysis techniques to predict when a failure may occur in equipment, gains even more relevance in maintenance management (KIANGALA; WANG, 2018; TERRISSA et al., 2016).

2.1.3 PDM - Predictive Maintenance

PDM has gained prominence in the scientific field in several multidisciplinary research groups. Combining data collection, infrastructure, storage, and AI disciplines has allowed the development of consistent lines of research. This section is intended to highlight some of the most important topics for understanding and adopting the PDM (LAMONACA et al., 2018; RODRÍGUEZ-MAZAHUA et al., 2016).

Data generated and collected from various sensors in industrial environments provide alternative opportunities for system or equipment lifecycle prediction solutions. The notion that a PDM can provide an intervention schedule based on equipment performance or condition becomes crucial for the industry's future. Data from different stages of the production system is one of the critical requirements for effective PDM execution. As a result, this can save maintenance costs, and downtime, as well as improve productivity and quality (KIANGALA; WANG, 2018; YAN et al., 2017).

One of the fundamental concepts related to PDM is PHM, which provides for managing the life cycle of industrial systems and employs four maintenance techniques: corrective, preventive at fixed intervals, failure detection, and CBM. The activity is monitored using sensors and mathematical methods that make it possible to assess the RUL of the equipment. PDM is one of the key technologies that is being implemented in I4.0 environments to enable smart manufacturing (KWON; DO; KIM, 2020; NEMETH et al., 2018; RUIZ-SARMIENTO et al., 2018).

2.1.4 PHM - Prognostics and Health Management

Prognostics and Health Management is a field of study that entails forecasting a system's future performance and health and then implementing measures to maximize that performance and prolong the system's life (BENKER et al., 2021; LEE et al., 2013). This may include monitoring the equipment for indicators of wear or deterioration and utilizing this data to estimate when maintenance or repairs will be required. PHM is applicable to a wide variety of systems, including mechanical, electrical, and biological systems (FERREIRA; GONÇALVES, 2022; BERRI; VEDOVA; MAININI, 2021; OMRI et al., 2020; TERRISSA et al., 2016).

PHM and PDM are related but distinct fields. PDM is a repair technique that utilizes data and analytics to forecast when equipment is likely to break so that maintenance can be planned proactively in advance of the failure. This may assist in decreasing equipment downtime and increasing its lifespan (WEN et al., 2021; XIA et al., 2018; LEE et al., 2014).

PHM is a broader strategy than anticipating equipment breakdowns. It entails projecting a system's future performance and health and then taking steps to maximize that performance and prolong the system's life. This may include monitoring the equipment for indicators of wear or deterioration and utilizing this data to estimate when maintenance or repairs will be required. It also covers decision-making and maintenance action optimization. Thus, PDM is a subset of PHM as a particular application of PHM that focuses on failure prediction, and maintenance schedule (FERREIRA; GONÇALVES, 2022; MA; MAO, 2021; BERRI; VEDOVA; MAININI, 2021; OMRI et al., 2020).

According to several researches, PHM can offer several benefits driven by I4.0, such as ML and IoT (FERREIRA; GONÇALVES, 2022; WEN et al., 2021; MA; MAO, 2021; BERRI; VEDOVA; MAININI, 2021; BENKER et al., 2021; CHEN et al., 2021; MAO et al., 2020; OMRI et al., 2020; VOGL; WEISS; HELU, 2019; XIA; XI, 2019; FERREIRO et al., 2016; LEE et al., 2014; LEE et al., 2013). These benefits including:

1. Increased uptime: By predicting when equipment is likely to fail, PHM can help to schedule maintenance proactively and reduce downtime.

2. Extended equipment life: By monitoring equipment for signs of wear or degradation, PHM can help to identify issues early and take action to extend the life of the equipment.

3. Cost savings: PHM can help lower maintenance and repair costs by reducing downtime and extending equipment life.

4. Improved safety: By identifying potential equipment failures before they occur, PHM can help to prevent accidents and improve overall safety.

5. Better decision-making: By providing accurate and actionable information about the health and performance of equipment, PHM can help to inform decisions about maintenance and repair.

6. Better optimization of maintenance actions: By providing information about the health state of the equipment and the Autoencoder (AE), PHM can help to optimize the timing and scope of maintenance actions, avoiding unnecessary or early maintenance.

7. Predictive Maintenance: By providing information about the state of the equipment, PHM can support PDM to schedule maintenance proactively before the failure occurs, reducing downtime and costs.

In this sense, PHM, in the context of I4.0, has started to receive special attention to increasing the competitiveness of the business.

2.2 Industry 4.0

I4.0 refers to the current trend of automation and data exchange in manufacturing technologies including IoT, AI, and cloud computing. PDM is a strategy that uses data from sensors and other monitoring devices to predict when equipment is likely to fail so that maintenance can be performed before failure occurs. In this way, I4.0 contributes significantly to improving maintenance techniques and helping to reduce downtime, improve efficiency and extend equipment life. (ROSATI et al., 2022; DALZOCHIO et al., 2020; ADU-AMANKWA et al., 2019; GÄRTNER, 2018).

Maintenance techniques associated with reliability need constant evolution to continue with the advancement of technologies embedded in products and manufacturing equipment (SONY; NAIK, 2020). Coupled with advancing technologies, quality in manufacturing is a prerequisite for continuously producing reliable products, and proactive product reliability assurance is always a crucial routine for production. In this scenario, I4.0 provides a favorable environment for the evolution of reliability models (HE et al., 2019; LI et al., 2019).

The goal of I4.0 is to integrate factories with modern technologies such as CPS, IoT, and Internet of Services, to enhance agility and efficiency of manufacturing systems and to account for changing business environments (Kagermann; WAHLSTER; HELBIG, 2013; RUIZ-SARMIENTO et al., 2018; GÄRTNER, 2018). CPS's background provided technologies such as Digital Twin, which simulate the production system and monitor the health of equipment and products (CHENG et al., 2018; ROSEN et al., 2015).

Thus, the technological advances promoted by I4.0 provide continuous monitoring in real-time of several assets, generating and sending alerts based on predictive and prescriptive techniques, that it is possible to use ML models (NING et al., 2020; SCHEER, 2019; GÄRTNER, 2018). However, monitoring conditions in real-time only promotes a certain level of reliability, in which unpredictable and unexplained failures still occur. Unforeseen failures tend to have a significant impact, and in some cases, harm the organization as a whole (HE et al., 2018).

Several authors have focused their reliability studies, using advanced analytical techniques to I4.0 data (MOSAVI; LOPEZ; VARKONYI-KOCZY, 2018). However, these studies are, in general, limited to descriptive analysis without considering the possible impacts in financial results and quality service with suppliers or customers. As such, they do not address the potential of integrating I4.0 technologies with customer needs, equipment health, and suppliers (NAPOLEONE; MACCHI; POZZETTI, 2020; PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018; SANDENGEN et al., 2016; TERRISSA et al., 2016).

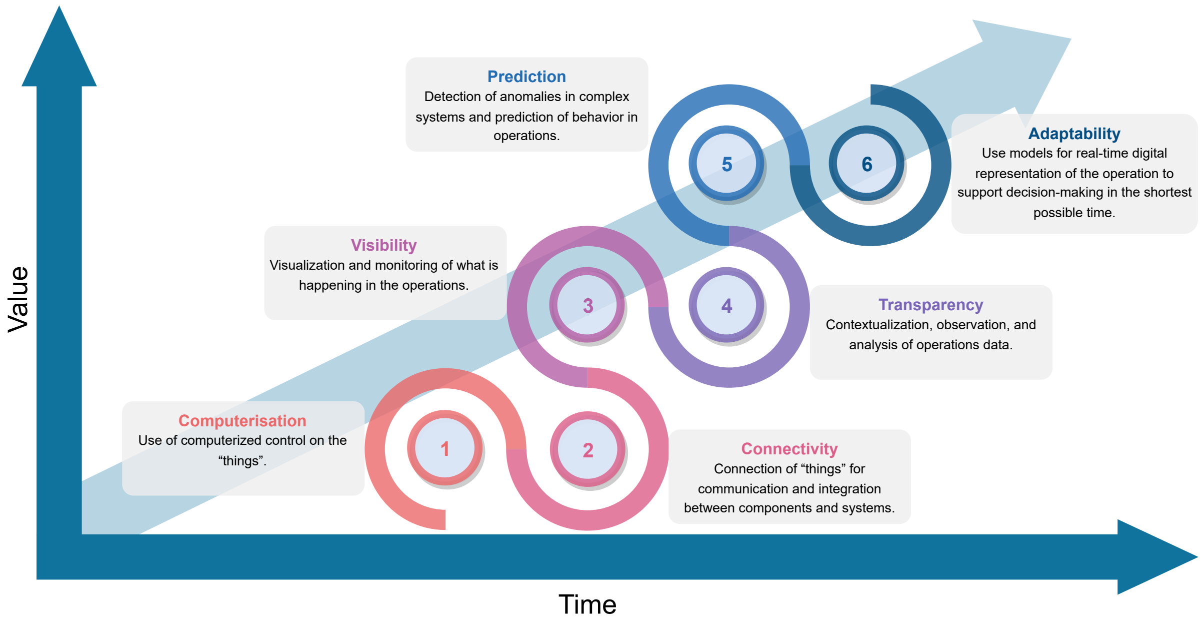
2.2.1 Internet of Things

IoT refers to the concept that any device, component, asset, or another item that can connect to a network and send data can become a data source that can be used in some way. Miniaturization and falling prices of sensors that measure things like temperature, pressure, humidity, and vibration have contributed to this growing trend (ROSATI et al., 2022; LAMONACA et al., 2018).

As sensors and devices cannot have a lot of storage space or processing power, they need energy efficiency. Most IoT-related approaches, such as cloud computing, need data distribution and storage strategies, even with cutting-edge methods such as fog and edge computing in some instances. In this sense, IoT consistently contributes to maintenance strategies gaining the spotlight (ROSATI et al., 2022; LAMONACA et al., 2018). However, as the data supply increases, the models become more complicated because the treatment and training steps consume a lot of computational resources and analysts' knowledge (SHCHERBAKOV; SAI, 2022; CHO et al., 2018).

As shown in Figure 4, implementing technologies such as IoT follows a journey that is not always linear. However, from I4.0 onwards, the data generated by the equipment begins to create value for the business. Through connectivity, as shown in Figure 4, in stage 3, visibility, data collected from sensors begins to be transformed into information providing value to the organization. The gains for the organization with IoT are usually perceived by improving the availability of operations, mainly with the reduction of operating costs with downtime.

Figure 4 – Stages in the Industry 4.0 development path, through IoT and the benefits.



Source: Prepared by the author, adapted from Gärtner (2018).

IoT is often used to implement PDM, as IoT devices collect real-time data from the system or equipment. Using ML, the collected data can be analyzed to identify patterns or anomalies that could indicate the possibility of future failures. Based on this information, those responsible for maintenance can take steps to correct problems before they occur, extending the life of the system or equipment and preventing downtime. In addition, IoT can also be used to monitor maintenance performed on a system or equipment, allowing maintenance personnel to track progress and ensure jobs are being performed efficiently and effectively (GUNGOR; ROSING; AKSANLI, 2022; SHCHERBAKOV; SAI, 2022; XIA et al., 2020; MÁRQUEZ et al., 2020; VOGL; WEISS; HELU, 2019).

In the context of I4.0, IoT plays an important role, as it contributes to system maintainability. Concerning industrial processes, maintainability measures the ease with which a system or component can be modified to correct errors, adapt to new needs, or improve. This includes making it easier to find and fix system problems and making changes and updates to improve or enhance the system. Maintainability is an important quality criterion for systems, as a poorly maintained system can be difficult and expensive to maintain and upgrade, leading to performance issues and system downtime (SHCHERBAKOV; SAI, 2022; VOGL; WEISS; HELU, 2019).

2.2.2 Machine Learning

In the context of I4.0, ML is used to analyze and make sense of the large amounts of data generated by the various connected devices and systems. This data is used to optimize and improve industrial processes, such as PDM, quality control, and energy management (CHEN et al., 2020; ZHOU; LIU; ZHOU, 2016).

One of the key benefits of ML in I4.0 is the ability to detect patterns and anomalies in the data that would be difficult or impossible for humans to detect. This can be used to improve the efficiency and productivity of industrial systems, as well as to detect and prevent problems before they occur (JIMENEZ et al., 2020; CHEN et al., 2020).

There are several specific ways ML is being used in I4.0, such as:

1. Predictive maintenance: ML models can predict when equipment is likely to fail, allowing maintenance to be scheduled before a breakdown occurs and reducing downtime and maintenance costs.

2. Quality control: ML models can be used to monitor and control the quality of products being produced in real-time, identifying and correcting issues before they reach the end customer.

3. Optimization of industrial processes: ML models can be used to optimize the parameters of industrial processes, such as energy consumption, to improve efficiency and reduce costs.

4. Autonomous systems: ML is a key technology in the development of autonomous systems, such as self-driving cars and drones, which are becoming increasingly relevant in the industrial context.

As shown in Figure 4, the development of I4.0, mainly in stages 1 and 2, makes it possible to connect devices through IoT and store a larger volume of data. With this, the evolution to stages 4 and 5, where ML gains relevance, is made possible. In this way, the application of ML contributes significantly to decision support (MOSAVI; LOPEZ; VARKONYI-KOCZY, 2018; GÄRTNER, 2018).

ML can be a powerful tool to help with PHM in several ways, as following:

1. Data analysis: analyze large amounts of sensor data from equipment, looking for patterns and trends that indicate potential issues.

2. Failure prediction: predict when equipment is likely to fail based on sensor and historical failure data.

3. RUL prediction: predict the RUL of equipment based on sensor data and degradation models.

4. Anomaly detection: detect abnormal equipment behavior, indicating potential issues.

5. Decision-making: support decision-making by providing recommendations for maintenance and repair actions based on the equipment's health and performance.

6. Optimization: optimize maintenance actions by providing information about the RUL and the health state of the equipment.

7. Online monitoring: monitor the equipment in real-time and detect any abnormal behavior, providing early warning of potential failures.

8. Self-learning: learn from historical data and improve over time, providing better predictions and better decision-making.

Adopting ML techniques in PHM can provide a more accurate and automated system for predicting equipment failure and optimizing maintenance actions, thus reducing downtime and costs and improving safety.

2.2.2.1 Machine Learning to anomaly detection

Anomaly detection plays a crucial role in I4.0 as it allows monitoring and analyzing the data from sensors and identifying abnormal behavior or patterns. This can help improve the efficiency and performance of industrial processes and reduce the likelihood of equipment failure or other issues. For example, in manufacturing, anomaly detection can be used to monitor the performance of machines and detect any potential problems before they lead to breakdowns, which can minimize downtime and reduce maintenance costs (YU; KIM; MECHEFSKE, 2021; JIMENEZ et al., 2020).

Notably, the choice of an ML model for anomaly detection depends on the type of data generated in the process, the nature of the anomaly, and the application's requirements. As the main ML models, we can highlight the PCA and the AE. The choice between using PCA or an AE for anomaly detection depends on the specific characteristics of the data and the task at hand (JIMENEZ et al., 2020; YANG; ZHENG; QI, 2020).

PCA can be a good option for linear data as it is computationally efficient and easy to interpret. However, if the data is non-linear or high-dimensional, PCA may not be able to capture the relevant features, and an AE may be a better choice. AE can also be useful for handling missing or noisy data (FERREIRA; GONÇALVES, 2022; JIMENEZ et al., 2020).

Regarding anomaly detection, the AE approach can be effective as it can learn a compressed representation of the normal data and then detect anomalies based on reconstruction loss. However, it is sensitive to the quality of the training data, and it may not be suitable for detecting certain types of anomalies, such as rare events or novel classes. Other techniques, such as one-class Support Vector Machine (SVM), Isolation Forest, and Local Outlier Factor (LOF), can be an option to detect the anomalies in such cases (DALZOCHIO et al., 2020; DIEZ-OLIVAN et al., 2019). These methods have their trade-offs, and it is important to test multiple algorithms to find the best one for a specific dataset and problem (YU; KIM; MECHEFSKE, 2021; KRAUS; FEUERRIEGEL, 2019; VERMA; CHANDRA; DWIVEDI, 2016).

2.2.2.2 Autoencoder for anomaly detection

An AE is a type of neural network that is used for unsupervised learning. It aims to learn a compressed representation (encoding) of the input data and reconstruct the original input from that encoding. The network typically comprises two parts: an encoder that maps the input to a lower-dimensional encoding and a decoder that maps the encoding back to the original input space. The training aims to minimize the difference between the original input and the reconstructed loss. Autoencoders have been used for many tasks, such as anomaly detection, dimensionality reduction, and generative modeling (FERREIRA; GONÇALVES, 2022; SCHWARTZ et al., 2022; KONG et al., 2020).

For anomaly detection, the basic concept is to train the AE on a dataset that contains only "normal" examples and then use it to detect anomalies in new, unseen data. The intuition behind this approach is that the AE has learned a compressed representation of the "normal" data during training and will have difficulty reconstructing examples significantly different from this "normal" data. One way to use the AE for anomaly detection is to calculate the reconstruction loss, which is the difference between the input and the reconstructed loss. If the reconstruction loss is high for a given input, the input is likely an anomaly. Another way is using the encoded output to compute the anomaly score (GOODFELLOW; BENGIO; COURVILLE, 2016; LECUN; BENGIO; HINTON, 2015; SOCHER et al., 2012).

In the PDM field, data from machines can be analyzed to detect patterns or anomalies that indicate an imminent failure, allowing maintenance to be scheduled before failure occurs. In this way, anomaly detection provides the ability to identify and respond to abnormal behavior, leading to more efficient, reliable, and cost-effective industrial processes. However, due to the complexity of the models and the analyzed system, identifying which process variable or equipment parameter is responsible for an anomaly is of paramount importance in the context of PHM (VOGL; WEISS; HELU, 2019).

2.2.2.3 Feature identification for anomaly detection

In PHM, feature importance is used to identify which sensor data or operating conditions are most strongly associated with the health and performance of a system. For example, in aircraft engine health management, feature importance can be used to identify which sensor data, such as vibration or temperature readings, are most strongly associated with the likelihood of engine failure (ALFEO; CIMINO; VAGLINI, 2022).

SHAP (SHapley Additive exPlanations) and Random Forest (RF) are two different methods for determining feature importance in ML. SHAP is a model-agnostic method that can be used to explain the output of any ML model. It assigns each feature an importance value for a particular prediction by calculating the average marginal contribution of that feature to all possible coalitions of features. The method is based on Shapley values from cooperative game theory, which assigns the average marginal contribution to each player (feature) to all possible coalitions (predictions). This method has been proven to have several desirable properties, such as consistency concerning the feature importance measures obtained by different methods (YANG, 2021; ZHANG et al., 2020; AGIUS et al., 2020).

On the other hand, RF is a specific ML model often used for feature selection. It is an ensemble of decision trees, and the feature importance is calculated based on the average decrease in impurity over all trees in the forest. RF is a robust algorithm for classification and regression problems, and it is beneficial for high-dimensional datasets and datasets with many features.

In summary, SHAP is a model-agnostic method that can explain any ML model's output, while RF is a specific ML model often used for feature selection (CAMPBELL et al., 2022). In this sense, due to the complexity of identifying which feature is most important for the health of equipment and systems, we applied SHAP for FI in the framework proposed.

For the implementation of PHM, the RUL prediction step is of paramount importance. In this sense, with the abundance of data and powerful algorithms, the task of the RUL gains prominence in I4.0 (KRAUS; FEUERRIEGEL, 2019).

2.2.2.4 Machine Learning to RUL prediction

Nowadays, there are several techniques for RUL forecasting. However, DL has gained relevance with the increasing data generation and computational power. DL is a subset of ML that involves training multi-layered neural networks to learn complex patterns in data. These networks can automatically extract features from incoming data, making them suitable for tasks such as RUL prediction involving large amounts of data with complex patterns (ALFEO; CIMINO; VAGLINI, 2022; LI et al., 2020; ZHAO; WANG; CHU, 2019).

RUL prediction with DL is to use sensor data from the system as input to the model. This data could include temperature, vibration, pressure, and other relevant measurements. The model would then learn to identify patterns in this data that correspond to different stages of the system's life and use those patterns to predict when the system is likely to fail. DL models such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), and others can be used for this task, depending on the nature and structure of the data (YU; KIM; MECHEFSKE, 2021; JIMENEZ et al., 2020). With the evolution of DL models, several architectures have received attention due to the results. In this sense, neural basis expansion analysis for interpretable time series forecasting (N-BEATS) appears as an alternative to be tested due to the results obtained in forecasting time series (MAKRIDAKIS; SPILIOTIS; ASSIMAKOPOULOS, 2022; ZHANG; SUZUKI; SHIOYA, 2022; ORESHKIN et al., 2021; ORESHKIN et al., 2019).

N-BEATS is a neural network architecture for time series forecasting. It is a generalization of the traditional feedforward neural network, where the model learns to predict the next value in a time series based on a fixed number of past values. N-BEATS allows the model to learn the underlying patterns in the time series data and make predictions based on those patterns. The architecture comprises the "backstage" and the "forecast" model. The "backstage" model is a stack of fully connected layers that extract features from the input time series, while the "forecast" model is a stack of fully connected layers that use those features to make predictions. The model is trained end-to-end to minimize the prediction error on the training data (ZHANG; SUZUKI; SHIOYA, 2022; ORESHKIN et al., 2021; ORESHKIN et al., 2019).

2.3 Final Remarks

As presented in this chapter, PDM collaborates so that PHM increases system availability. In this sense, the reliability and maintainability of systems can be considerably improved through PHM. However, as the connectivity of devices advances and the generation of data from different sources advances, the complexity of analysis tends to increase in proportions that are often unequal. In this sense, while I4.0 enables greater data generation, ML models also evolve and enable better decisions to be taken.

In this sense, to meet this research's general and specific objectives, we shed light on data analysis, anomaly detection, and RUL prediction. With these applications combined, it is possible to develop robust applications. Given this scenario, we position the PHMS, that is, the application of the PHM in a systemic way, considering several equipments in a production system and applying combined techniques for anomaly detection, anomaly identification, and RUL prediction.

3 RELATED WORK

This chapter groups together the main works related to supporting the proposed solution and collaborating with developing a robust methodology. We divided the chapter into four sections. The first and second sections have a broader level of detail, as it presents the step-by-step development of the survey methodology concerning the state of employment of systems reliability for decision-making. These sections also show related works to support and justify the relevance of this research. The third section presents the results and discussions obtained from the related works. On the other hand, Section four addresses the opportunities and motivations presented by the related works.

As an initial step in developing comprehensive research that contemplates related works and exposes opportunities and motivations, we have defined the main research question preceded by four other sub-questions. As highlighted at the beginning of this chapter, we detail each step employed in the next sections. As a starting point, we begin with the research question.

3.1 Research Question

Having defined the challenges and the desired scientific results, we now formulate the main question (MQ) and the corresponding sub-questions (SQ) that guide this review. Table 1 presents these questions.

Table 1 – Research Questions

Identifier	**Issue**
MQ	What types of interactions are used by reliability in Industry 4.0 to Decision-Making?
SQ1	How are the researches on reliability to support decision-making in the context of Industry 4.0?
SQ2	What are the technical methods applied in reliability to support decision-making?
SQ3	How would be a taxonomy using the terms found for reliability applications in the Industry 4.0?
SQ4	How has reliability helped the value chain management in the context of Industry 4.0?

Source: Prepared by the author.

The **MQ** was formulated to report how reliability has been applied to I4.0 to support decision-making. **SQ1** lists the primary means of disseminating research following reliability to support decision-making in the I4.0. **SQ2** understands which main methods and models are applied in reliability studies to support decision-making. **SQ3** searches the main terms found, for the creation of standardization and presentation of the taxonomy proposal. **SQ4** analyzes the kind of processes integration the reliability has support in the decision-making for the I4.0.

3.2 Search Strategy

The search string was focused on decision-making using reliability applied in I4.0, based on previous readings, we identified the need to carry out a literature review and evaluate the types of decision-making applying concepts of reliability making use of the approach of I4.0.

The string was constructed considering that reliability is associated, most of the time, with the several commonly used maintenance strategies in the industry. Figure 5 shows the string used in the search.

Figure 5 – Search String

((("decision-making") ***and*** (reliability ***or*** availability) ***and*** ("industry 4.0" ***or*** "fourth industrial revolution" ***or*** "smart factory" ***or*** "intelligent factory" ***or*** "digital factory") ***and*** (prescriptive ***or*** predictive ***or*** diagnostic ***or*** descriptive)))

Source: Prepared by the author.

3.2.1 Selected studies

We applied the string in Figure 5 until 2019 on Google Scholar. We filtered out papers published earlier than 2011 and also patents and quotations, resulting in 4040 research papers. Our screening intended to evaluate the search string to return a more significant number of studies, for responding to the MQ and SQ.

Out of this analysis, we cataloged papers following the criteria listed in Table 2.

Table 2 – Criteria used to filter out research out of the scope of this paper

Section	**Description**
Criterion 1	Papers published earlier than 2011
Criterion 2	Papers without abstract
Criterion 3	Books, technical reports, dissertations and theses
Criterion 4	Studies less than 4 pages long which are not in English
Criterion 5	Publications that do not use the search terms Industry 4.0, intelligent factory, smart factory, Fourth Industrial Revolution and reliability in the abstract or keywords
Criterion 6	Publications that do not address Industry 4.0, intelligent factory, smart factory, Fourth Industrial Revolution and reliability as to decision-making

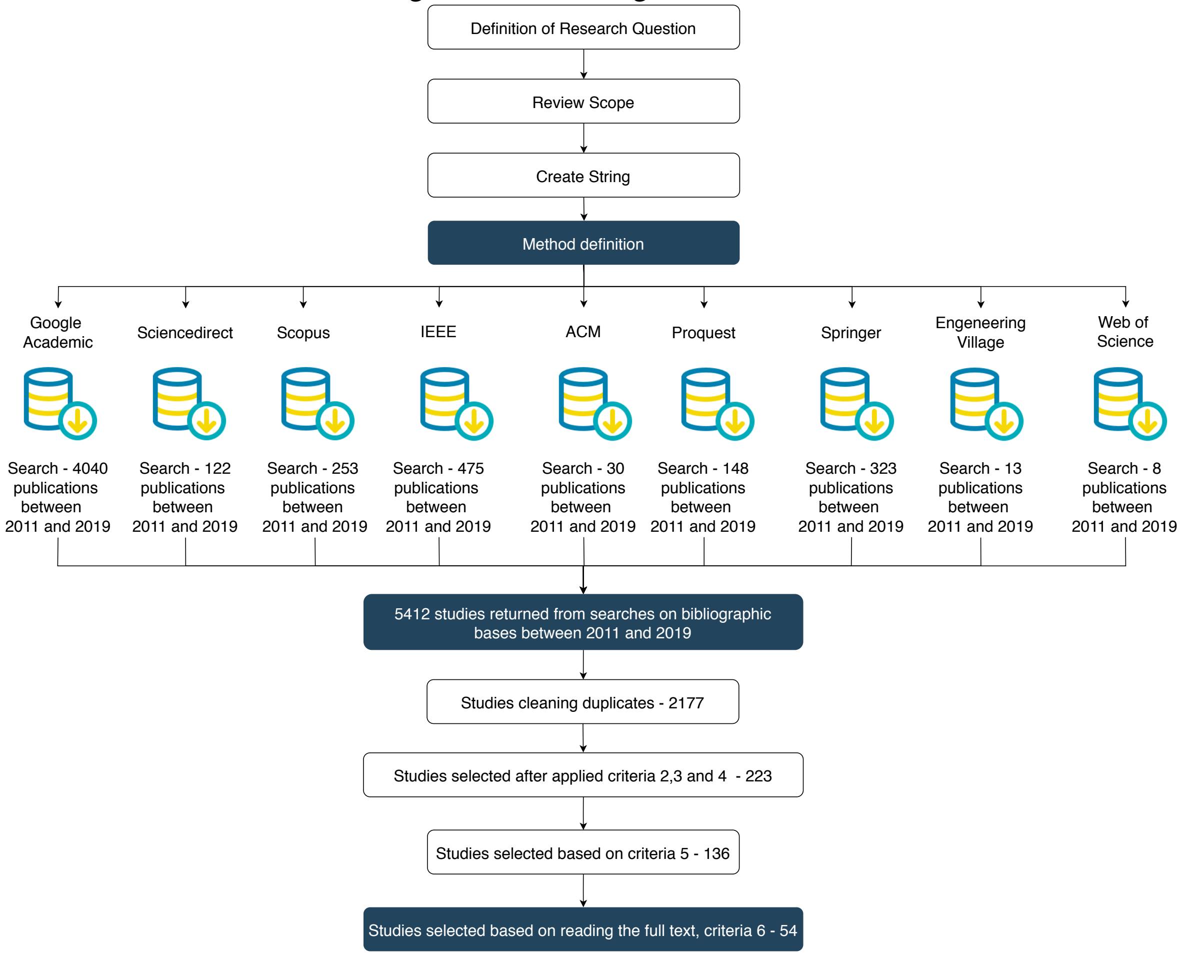
Source: Prepared by the author.

After the cataloging performed in Google Scholar, we performed the same procedure on the bases of ACM, IEEE, ScienceDirect, Scopus, Springer, Engineering Village, and Web of Science. The resulting search strategy is shown in Figure 6.

Enhance the selection strategy of articles that were part of the study is fundamental. According to criterion 1, the date for searching began in 2011, when the term was coined

Zhou, Liu e Zhou (2016), resulting in 5412 articles. After that, we removed duplicate titles, after which 2177 papers remained. We then applied criteria 2, 3, and 4, resulting in only 223 studies. Then, criterion 5 was applied, where 136 articles remained. We then carefully read these articles while applying criterion 6. Some studies presented the concepts of reliability but not applied to I4.0. The articles' validity to compose the literature review should consider I4.0 concepts in reliability to support decision-making.

Figure 6 – Screening of research



Source: Prepared by the author.

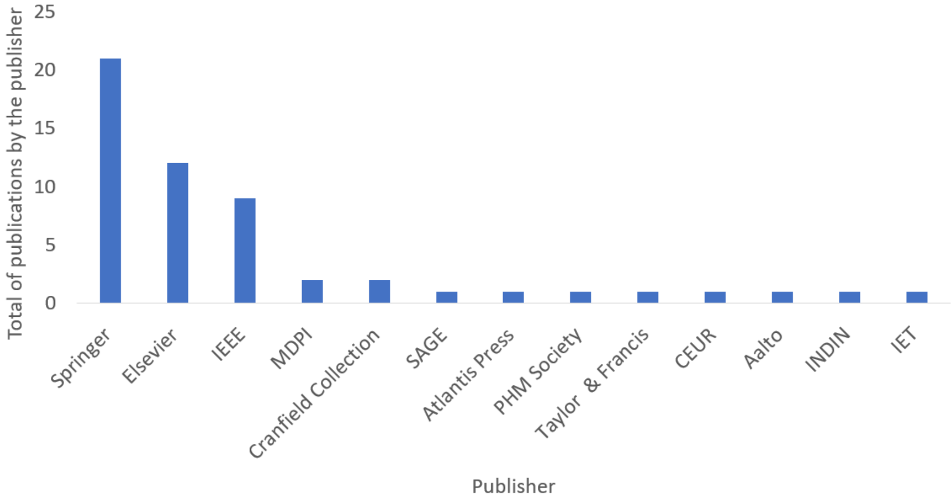
In the Appendix E and F, are shown the resulting selected articles, which list the year of each publication and the publisher. The separation between journal and conference aims to highlight differences in the methodological approaches and objectives of the studies according to the type of publication.

In Appendix E, we present the selected articles which list the year of each publication, the publisher, and journal where the papers were published.

The resulting selected articles are shown in Appendix F, which lists the year of each publication, the publisher, and the conference where the papers were published.

In the following section, we present the results and discussion in response to scientific challenges and research questions based on the studies selected in the literature review.

Figure 7 – Total aggregate distribution of publications by Publisher.



Source: Prepared by the author.

3.3 Results and Discussion

3.3.1 How are the researches on reliability to support decision-making in the context of I4.0?

In this subsection, we discuss and respond to how research on reliability support decision-making in the context of I4.0 are being reported. In Appendix E and F we summarize the answer to the research question. The tables show the types of publications, publishers, and names of the articles that compose the study corpus. In Figure 7, the number of publications is distributed by the publisher, where we can observe that the publishers Springer, Elsevier and, IEEE have the highest representativeness of publications.

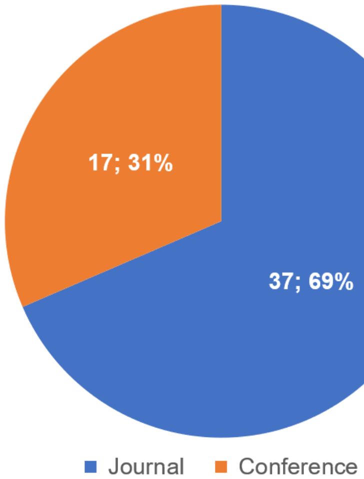
Although I4.0 has the fundamentals in different technologies, publishers such as the ACM still do not provide a representative amount of publications when associated with decision-making through reliability studies. However, it is possible to observe in Figure 7, that the bar graph presents a multidisciplinary characteristic, that is, the publications are in several areas of knowledge. Meanwhile, Figure 8 shows the ratio of paper by type.

Also show in Figure 8, the pie chart with journals and conference distributions. Moreover, we can observe that journals have more representability in publications selected to compose the survey corpus.

In Figure 9, we show the annual evolution of publications, with an emphasis on 2016 and rapid growth where the trend line confirms it. One factor that can be highlighted is the growing expansion of IoT and the advanced models of ML in I4.0 as responsible.

Accompanying the rapid growth of publications from 2016, we show that articles published in the journal had a significant increase in 2018, according to Figure 10. Such evolution collaborates with the relevance of the theme in the academic and scientific environment. However, understanding what types of publications the scientific academia has

Figure 8 – Distribution of publications by Type



Source: Prepared by the author.

contributed is necessary. In this matter, Figure 11 presents an extract with a methodological division used in the studies that compose the present survey.

As we can detect in Figure 11, publications applying case studies have occurred more frequently. The greater representativeness of this type of research indicates that the reliability studies for decision-making in the context of I4.0 have been directing efforts to practical applications with organizations. Such evolution was, until mid-2016 studies, focused on trends and architectures, which can be noticed in Figure 11. These trends are in line with what is proposed by Kagermann, Wahlster e Helbig (2013) in the study entitled, Recommendations for I4.0 Implementation, which suggests that enablers technologies should be used for problem-solving and decision-making.

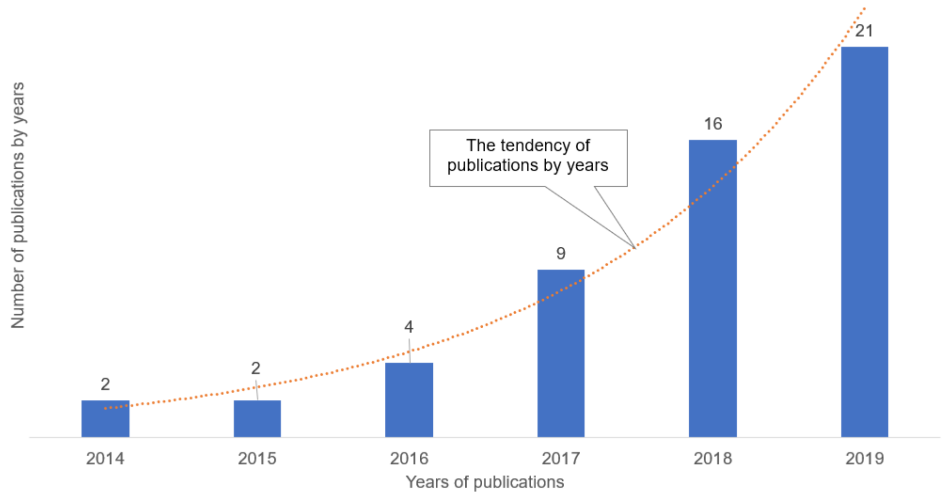
To summarize and collaborate with the answer to the research question **SQ1**, we produced Appendix G with the division between the corpus articles and the individual approach employed in conducting the research that was intended to be solved.

After answering the research question, **How are the researches on reliability to support decision-making in the context of I4.0?**. The next step will be using the studies that compose the survey and answer the second question that guides the present study.

3.3.2 What are the technical methods applied in reliability to support decision-making?

To answer this question, we used different contexts to evaluate the articles. First, we analyzed information related to the main methods and techniques applied in reliability of systems studies to support decision-making. Second, we examined the methods and techniques applied according to the purposes of the studies. Third, we discuss the techniques and methods used according to the main focus of studies.

Figure 9 – Distribution of publications by year with exponential tendency.



Source: Prepared by the author.

3.3.2.1 The main methods and techniques

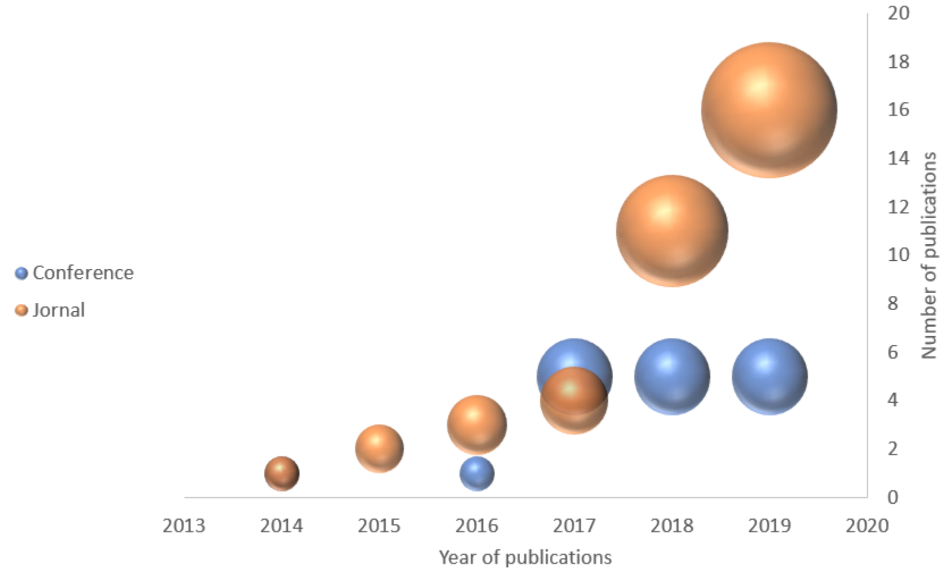
Initially, we performed a classification of each technique applied in the articles selected for the corpus. The subdivision was carried according to the primary technique related to the main focus of the study (ISMAIL; TRUONG; KASTNER, 2019; KRUMEICH et al., 2014; LIU et al., 2018; QIAO et al., 2012; REHMAN et al., 2016; SAHAL; BRESLIN; ALI, 2020; SAWANT; SHAH, 2013). Thus, we present in Figure 12 the proportion of techniques applied in the articles: Communication, Ingestion, Analysis and Storage.

- We catalogued as Communication the technique which the articles address sensor applications, integration, micro-services, CPS propositions, IoT on the shop floor, better equipment information capture techniques, multi-agent proportion, and wireless architectures (SALAZAR et al., 2019; LEE; ZHANG; NG, 2017; LEE et al., 2014; MOURTZIS; VLACHOU, 2018; ZHENG et al., 2018).

- Ingestion is the technique related to the ability to validate, clean, transform, noise reduction, and integration of data from equipment and systems (ISMAIL; TRUONG; KASTNER, 2019; MOURTZIS; VLACHOU, 2018; O'DONOVAN et al., 2015; REHMAN et al., 2016; SAWANT; SHAH, 2013).

- The main methods that we grouped in this category are: Feature Extraction (FE) to evaluate sensors in IoT systems, computer simulation to evaluate failure signals in systems and test models Schreiber et al. (2019), Syafrudin et al. (2018), Thoppil, Vasu e Rao (2019), Wang et al. (2017), and also, the method Fault Tree, employed to identify variables that are most representative for use in the decision model (FUMAGALLI et al., 2016).

Figure 10 – Distribution of publications by Type of year



Source: Prepared by the author.

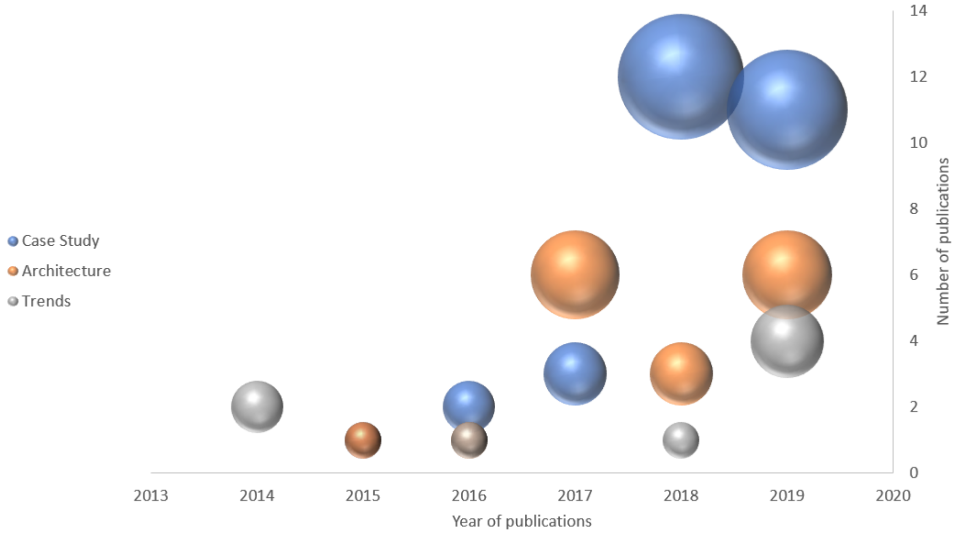
- The analysis is associated with data modeling using statistical methods and algorithms (ISMAIL; TRUONG; KASTNER, 2019; SAWANT; SHAH, 2013). In this group, we consider the use of ML to analyze Big Data sources for decision-making. As an example, the use of the RF method for adaptive maintenance with the purpose of PDM (CANIZO et al., 2017). In this same way, Decision Tree was used with a Simulation method, to simulate the equipment failure (SEZER et al., 2018). Clustering methods were also applied for PDM to building models with a hybrid approach to improve machine availability (CAO et al., 2019; CHO et al., 2018; SIAFARA et al., 2017).

- In the analysis technique, mathematics methods, such as Petri Network, Weibull distribution, and computer simulation, were applied in PDM studies with a focus on reducing costs of maintenance (HE et al., 2018; KŁOS; PATALAS-MALISZEWSKA, 2019; MYERS; TICKEM; EVANS, 2016). With similar objectives in lowering costs, a probabilistic economic method was used to analyze the quality of PDM associated with rework (KLEIN; BERGMANN, 2018; RIMPAULT; BALAZINSKI; CHATELAIN, 2018; TSAO et al., 2020).

- With the storage for this aggregation of techniques, we considered the articles in which the technical focus was linked to the data storage process (ISMAIL; TRUONG; KASTNER, 2019; SAWANT; SHAH, 2013). Thus, research focused on both platform development, information security, and CPS with cloud computing has been classified as a storage technique.

Studies, where the technical focus is on data storage, have different application purposes. Taking as an example, a platform for Big Data in an industrial environment using RHadoop, to

Figure 11 – Methodological application of the study by year



Source: Prepared by the author.

increase the quantity and quality of data and build better PDM studies (KU, 2018). However, as well as the quantity and quality of data, the concern with information security is an essential factor (TAN et al., 2017). Graph-theory was applied in the study in which the purpose was to analyze information security in a smart factory (HÄCKEL et al., 2019).

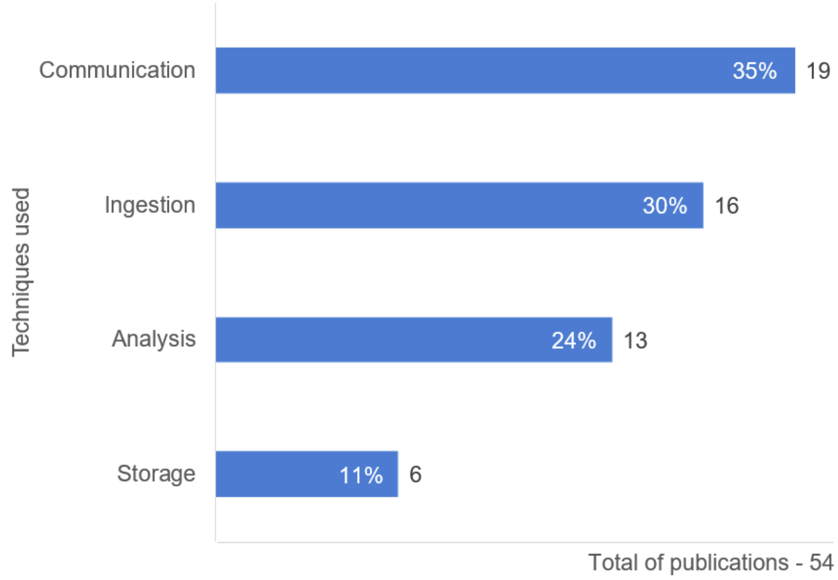
When the storage technique is related to architectures and trends, the main approaches are related to CPS (BALOGH et al., 2018; LEE et al., 2014; TERRISSA et al., 2016). Some authors comment that the use of CPS provides a significant competitive advantage by facilitating the capture and management of data (BALOGH et al., 2018; TERRISSA et al., 2016). In this way, the analysis that deal with the type of maintenance can present a better performance. As more data is collected and made available, better models can be built, and more confident decisions are possible (TAN et al., 2017).

3.3.2.2 Methods and techniques applied according to the purposes of the studies

In the second context considered to answer **SQ2**, we classified the techniques and methods linked to the purposes of the analyzed articles. In **SQ1**, we categorized according to the purpose, being Trends, Architecture, and Case Study. With these definitions, we were able to analyze according to the techniques and methods used, as seen on the Appendix H.

In case studies, the vast majority of reviews are grouped in Ingestion and Analysis according to the Appendix H. In the technique Ingestion, algorithms, such as RF and Deep Belief Network (DBN) are used for FE and data preprocessing (CANIZO et al., 2017; LI et al., 2019; SEZER et al., 2018; SYAFRUDIN et al., 2018; WANG et al., 2017). Traditional methods such as computer simulation are also utilized for Ingestion in which the main objective is to validate the data for reliability analysis (SCHREIBER et al., 2019).

Figure 12 – Articles per technique applied for the studies in reliability to support decision-making.



Source: Prepared by the author.

For analysis, the methods applied are the most diverse. Weibull distribution and Decision Tree are used for cost-oriented PDM (HE et al., 2018; SEZER et al., 2018). Decision Tree is also used for real-time PDM using Big Data environment (CANIZO et al., 2017). Fractal Analysis was proposed for monitoring the machine maintenance in the manufacturing (RIMPAULT; BALAZINSKI; CHATELAIN, 2018).

The main objectives of communication in case studies are associated with the use of sensors to build CPS environments (PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018; SHIHUNDLA; MPOFU; ADENUGA, 2019). When the applied technique is linked to Storage, the methods are centered on the construction of Big Data and risk assessment of the environment (KU, 2018; TAN et al., 2017).

In articles in which the objective was classified as architecture, the predominant methods are linked to communication techniques. The proposed architectures use different approaches such as microservice, blockchain, and multi-agent systems to build architecture for Big Data (SALAZAR et al., 2019; LEE; ZHANG; NG, 2017; MOHAMED; AL-JAROODI, 2019; PALAU; DHADA; PARLIKAD, 2019). For the development of CPS, some studies focus on the use of sensors to capture data with the equipment (BOUSDEKIS; MENTZAS, 2017; SANDENGEN et al., 2016; ZHENG et al., 2018).

In the articles in which we classified as trends, the prevalence centered on the main methods in proposing tables or scripts addressing the technologies of I4.0 for data collection and preparation (LAZAROVA-MOLNAR; MOHAMED, 2019; LEE et al., 2014). In these situations, the results presented are oriented towards theoretical and conceptual models (LAZAROVA-MOLNAR; MOHAMED, 2019; LEE et al., 2014; NEMETH et al., 2018).

3.3.2.3 Techniques and methods used according to the main focus

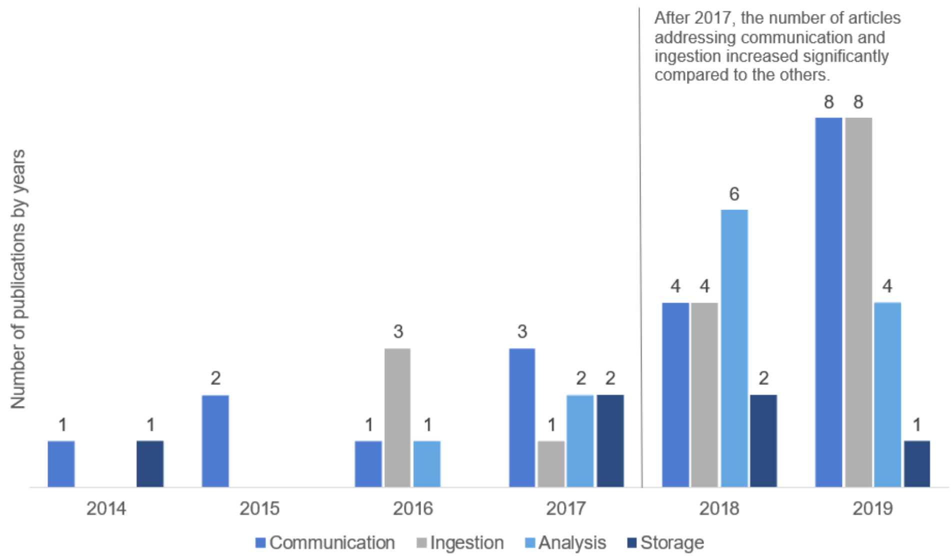
Finally, we discuss the third context. For this purpose, we present Appendix I with the main applied techniques pointed out, as well as the predominant focus related to reliability.

As the main focus shown in Appendix I, we call attention to advanced approaches linked to maintenance strategies. As an example, diagnostics and prognoses that aim to provide a structure to predict the degradation and maintenance of machines and devices (XIA et al., 2018; XIA; XI, 2019). Another important point to highlight is the proactive maintenance because different from the preventive and PDM, it provides actions that aim at the causes of the failure, not just the symptoms (CANITO et al., 2017).

According to the results found and presented in Appendix I, the technique that prevails in reliability studies for decision-making is linked to communication. When we analyzed the main techniques and methods together with the main focus of the studies, we realized that the efforts are centered on the collecting and processing data, bearing in mind that "PDM", "Diagnosis and prognosis" and "Operation based on conditions" are the main objectives of studies involving communication in reliability found in this research.

With **SQ2**, we realized that the vast majority of reliability studies focus on two main purposes when addressing the concepts of I4.0. We found articles in which methods and techniques are focused on the development of environments for data collection and preparation (CANIZO et al., 2017; KU, 2018; WANG et al., 2017). On the other hand, some studies have focused on the development of computational models to support decision-making (CHO et al., 2018; LI et al., 2019; LEE et al., 2015).

Figure 13 – Evolution of the techniques used over the years.



Source: Prepared by the author.

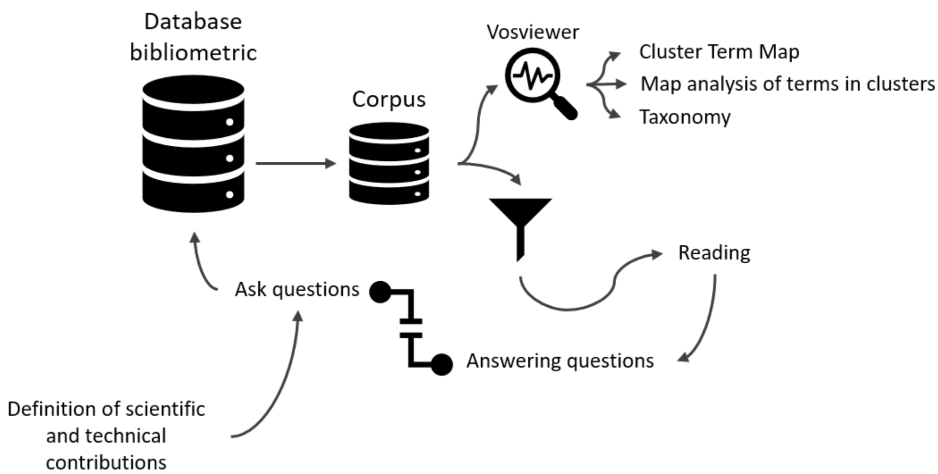
When we analyzed the proportion of each technique in the studies that were part of this survey, we realized that storage has the lowest representativeness. This is more evident when compared to other technologies, such as communication and ingestion, according to Figure 13. One possible explanation is related to the researchers interest in proposing architectures to ease the collection of data generated by equipment and systems.

We realized that part of the interest in this topic is related to the diversity of manufacturers, which have their middleware and communication protocols (CANITO et al., 2017; PÉREZ-LARA et al., 2018; ZHENG et al., 2018). As a result, some systems end up not communicating natively with other solutions (LEE; ZHANG; NG, 2017). In this way, they end up limiting communication between systems produced by different manufacturers (SALAZAR et al., 2019). And with that, they generate a movement that has raised interest in proposing architectures to get around communication problems.

3.3.3 How would be a taxonomy using the terms found for reliability applications in the I4.0?

To respond to the third sub-question of search, we performed separation of terms and proposed a taxonomy for reliability in the I4.0. Figure 14, shows the adopted methodological sequence. The taxonomy construction process begins with technical and scientific contributions, that were presented in the opening chapters of this article. This step is crucial because it will guide the research questions. With the research questions defined and the articles selected, we started to separate, discuss, and analyze the results. For this step, we use the VOSviewer tool as a support and to assist in visualizing the main terms covered in the selected articles.

Figure 14 – Methodology to generate the taxonomy

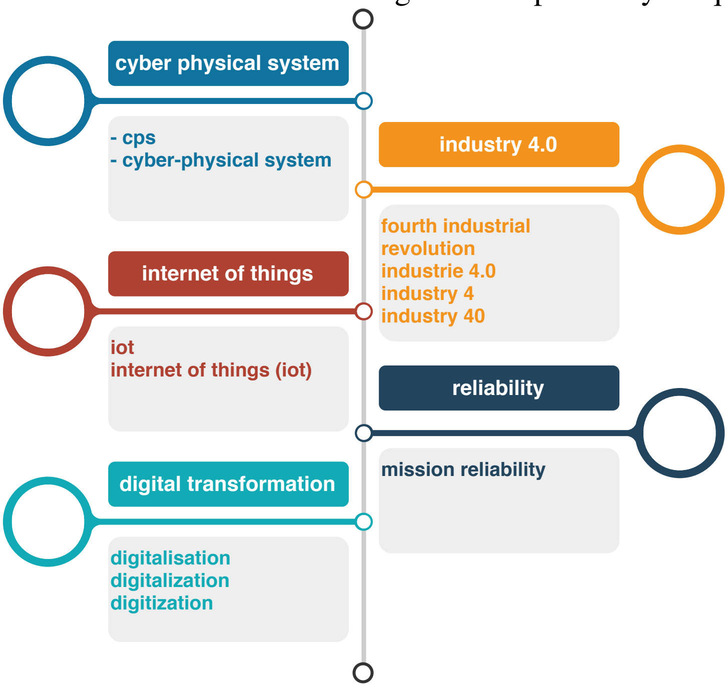


Source: Prepared by the author.

To establish logical reasons in the taxonomic definition process, we adopted three criteria to create the taxonomy for I4.0, focusing on reliability for decision-making.

Firts criterion: In the first step, a map and cluster were generated using VOSviewer, which applies the smart local moving algorithm and optimization (KLAVANS; BOYACK, 2017). For the VOSviwer map and cluster construction process, we imported the bibliometric data from the corpus and utilized the filters related by keyword and co-occurrences (NAUKKARINEN; BRAGGE, 2016). Co-occurrence analysis consists of analyzing the relationships between words and terms using the natural language algorithm (NAUKKARINEN; BRAGGE, 2016). To avoid redundancy, we applied a grouping of similar terms according to Figure 15. We adopted this action to prevent synonymous words and terms, writing, or even meaning from being plotted on the map and cluster separately.

Figure 15 – Co-occurrence terms configuration replaced by unique term.



Source: Prepared by the author.

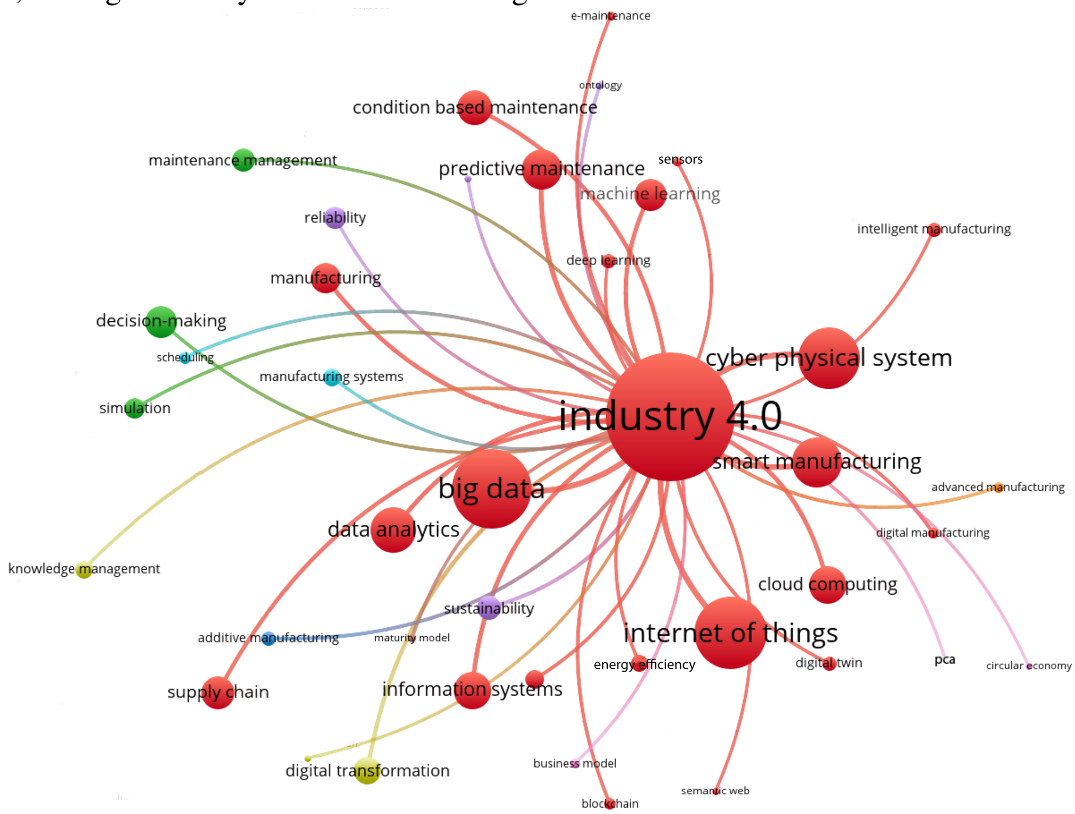
Second criterion: In the second criterion, the objective is to verify all terms hierarchically related to the term I4.0, focusing on reliability. Through VOSviewer, it was possible to build a network of relationships in Figure 16 and Heat Map Figure 17 of the main terms of the articles selected in the literature review. Still, in this criterion, we aim to highlight how the terms decision-making and reliability are related to I4.0 and synonyms.

In Figures 16 and 17, we can see that the term I4.0 is far from the term decision-making as well as reliability. On the other hand, we note that enabling technologies, as well as I4.0 synonyms, are close in the network formed by terms found in the studies.

From criterion 2, it was possible to highlight the main terms found for I4.0, focusing on reliability in Figure 18. With criterion 2 completed, the next step, criterion 3, is to identify which terms are closest and related to decision-making and reliability.

Third criterion: For each relevant term presented in Figure 18, a representation was made to generate the taxonomic tree with the most relevant links. Thus, it was possible to verify the

Figure 16 – The relevance of clusters that form a network between relationships with the main terms, linking reliability and decision-making in I4.0.



Source: Prepared by the author.

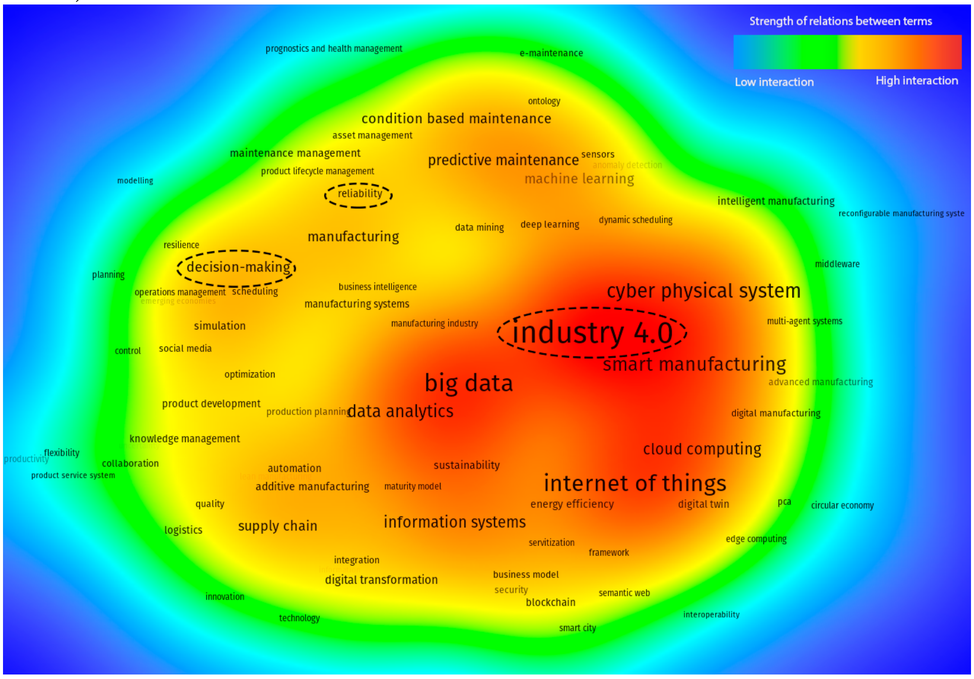
existence of direct relations and select the terms that belong to the same cluster as Figure 19 and Figure 20.

Figure 19 presents the primary connections with the respective terms: manufacturing, supply chain, maintenance management, manufacturing systems, quality, planning, and information systems. When we look at the term decision-making, it is clear that essential terms that guide the main actions of companies are slightly close. However, when compare with enabling technologies in I4.0, we note that there is still a gap between decision types and techniques.

The reliability cluster in Figure 20 presented the mapping of the central relationships and the following terms: PDM, condition-based maintenance, prognostics, and health management, manufacturing, and emerging economies. However, when the term reliability is analyzed, greater interaction with I4.0 and enabling technologies is noticed. According to the analyzed studies and the built clusters, we evidenced a strong connection and contribution of I4.0 in reliability researches. As noted in Figure 20, the terms Big Data and IoT are the ones that are most closely related to reliability. When it comes to decision-making, reliability acts as a link to I4.0. Thus, Figures 19 and 20 confirm the importance of reliability in decision-making processes.

To summarize criterion 2 of the taxonomy construction, we show the clusters resulting from

Figure 17 – A heat map with the strength of the connections between the main terms of decision-making, and reliability with I4.0: where the intense red shows a high amount of relation with the other terms, and the blue one a low connection with the additional terms.



Source: Prepared by the author.

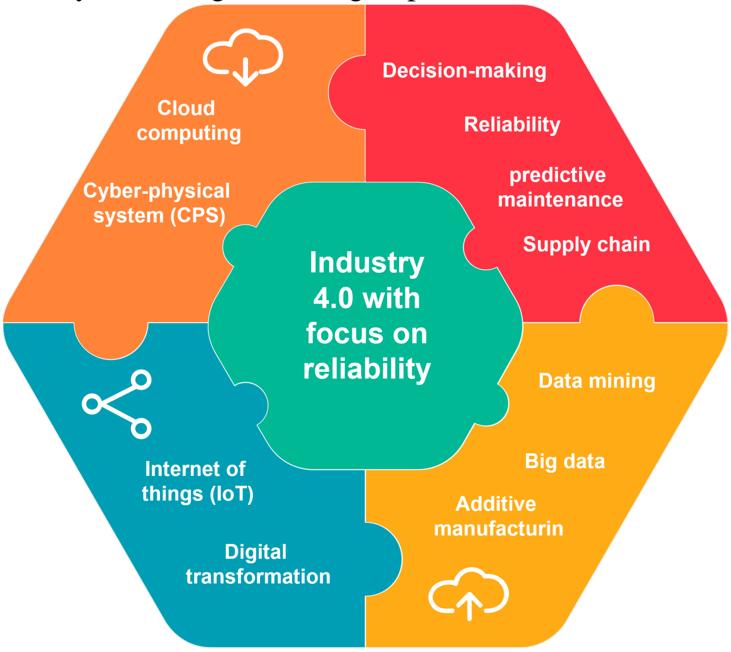
mapping with critical terms in Table 3. Behind the elaboration of the clusters, it was possible to analyze the types of decision-making, as well as the relations with the reliability. We were able to identify the existing relationships that underpinned the construction of the taxonomy and thus answer the second research question according to this study.

After defining the criteria, we elaborate on the taxonomy in Figure 21, with the main terms associated with reliability in I4.0 to support decision-making. We develop taxonomy based on the clusters generated in the second criterion. Throughout the analysis, we noted that reliability studies in I4.0 can be divided into applications with local and global purposes (PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018; SANDENGEN et al., 2016). In Figure 21 it is possible to perceive the global goals in the region highlighted in blue and the local goals in green. In the construction of taxonomy, the different decision levels were considered, where it is feasible to prioritize actions aimed at the overall gain of the organization and the network where the company is part. This approach is the main contribution of the developed taxonomy.

Conceptually I4.0 is divided into three types of integration: vertical, horizontal, and end-to-end (Kagermann; WAHLSTER; HELBIG, 2013). Several authors define end-to-end integration as I4.0, the main contribution (Kagermann; WAHLSTER; HELBIG, 2013; GÄRTNER, 2018; WANG et al., 2016). However, for the gains to be realized, the vertical and horizontal integration must be present and actively operating, using the technologies that provide the application of a smart factory (UHLMANN; FRAZZON, 2018).

Vertical integration in I4.0, when applied in reliability studies, is related to the traditional

Figure 18 – Key terms to generate a group for I4.0 with focus on reliability.



Source: Prepared by the author.

types of maintenance practiced. The goal of vertical integration is to combine the subsystems with the production and process into the company (CHO et al., 2018; GRACEL, 2018; UHLMANN; FRAZZON, 2018). The leading technologies and approaches are similar to those already employed in maintenance reliability studies. As a significant difference, we present Figure 21 highlighting the objectives of vertical integration over traditional models of reliability-centered maintenance (NEMETH et al., 2018; SANDENGEN et al., 2016).

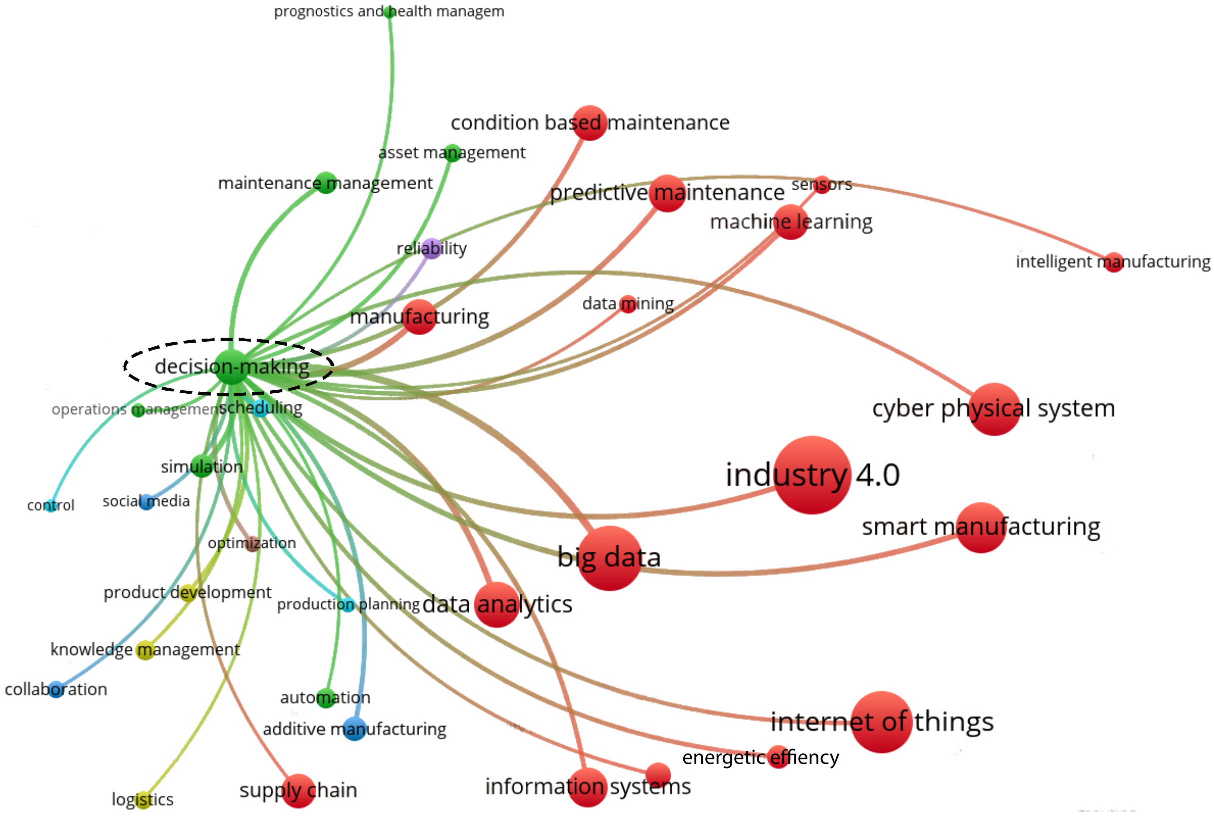
Horizontal integration is intended to provide value to the entire network where the organization is a part (SANDENGEN et al., 2016). Reliability, therefore, plays a crucial role in ensuring reliable and flexible production. Unreliable processes can affect the entire supply chain increasing costs, reducing competitiveness, and cooperation between companies,

Table 3 – Key terms and related clusters mapped

Key terms	**Clusters**
Big Data	CPS, smart manufacturing, cloud computing, IoT, sustainability, energy efficiency, Digital Twin and digital manufacturing
Reliability	predictive maintenance, condition based maintenance, asset management, e-maintenance, prognostics and health management
Data Analytics	artificial intelligence, data mining, deep learning and machine learning
Decision-making	maintenance management, product life cycle management, supply chain, innovation, flexibility, digital transformation, product development, additive manufacturing, production planning, simulation, optimization, quality, innovation and social media

Source: Prepared by the author.

Figure 19 – Clusters formed with the central relationships between types of decision-making in I4.0, focusing on links with decision-making.

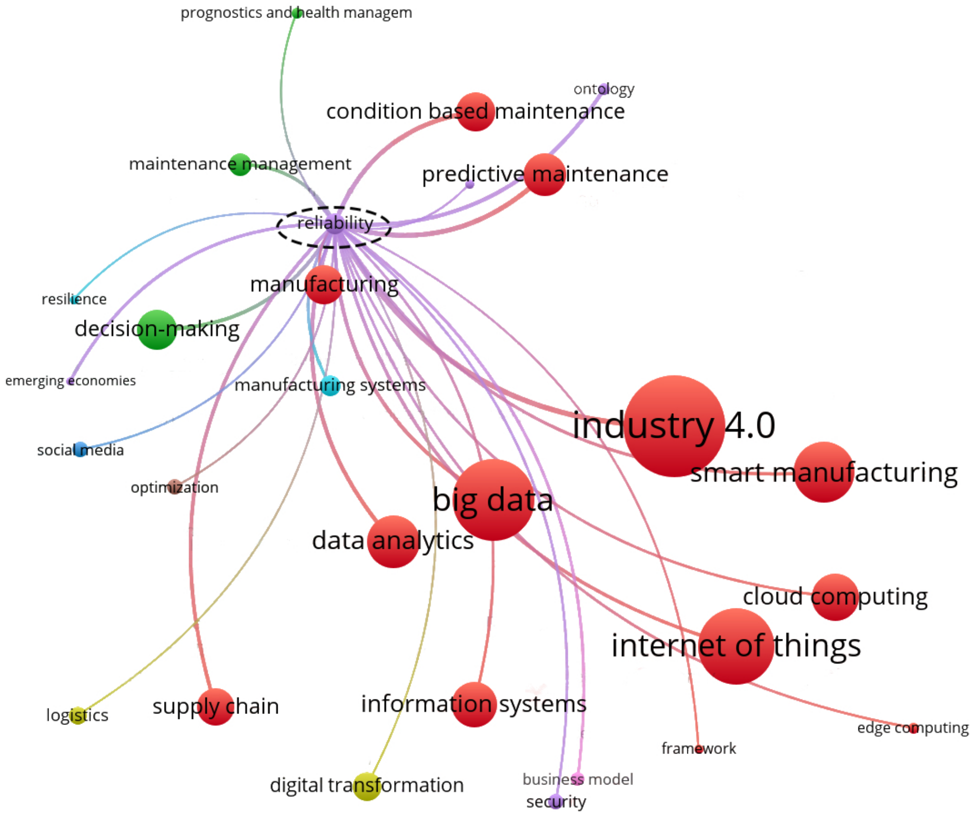


Source: Prepared by the author.

negatively impacting the benefits from I4.0 (LEE et al., 2015; TAN et al., 2017). Intelligent supply chain management, on the other hand, provides key performance indicators by analyzing historical data, including different sources such as financial data and market demand, thereby predicting and quantifying critical indicators based on various factors (MADHIKERMI et al., 2018). Finally, past health and condition information can feed back into machinery and equipment designs to continually redesign the life cycle. These actions offer consumers and users the possibility to enjoy higher productivity while minimizing reliability concerns. And for companies, they provide quick response and flexibility in satisfying customer demand (CANITO et al., 2017; MADHIKERMI et al., 2018; PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018; SANDENGEN et al., 2016; TERRISSA et al., 2016). These concepts were fundamental to the development of taxonomy and discussions about the application of reliability for decision-making.

After the construction of the taxonomy, it was possible to analyze the different types of decision-making that reliability studies provide in the context of I4.0. The taxonomy developed is fundamental to answer the fourth guiding question of this research because we highlighted the essential concepts linked to the value chain. Thus, we can answer the following question: How has reliability helped the value chain management in the context of I4.0?

Figure 20 – Clusters formed with the central relationships between reliability and I4.0.



Source: Prepared by the author.

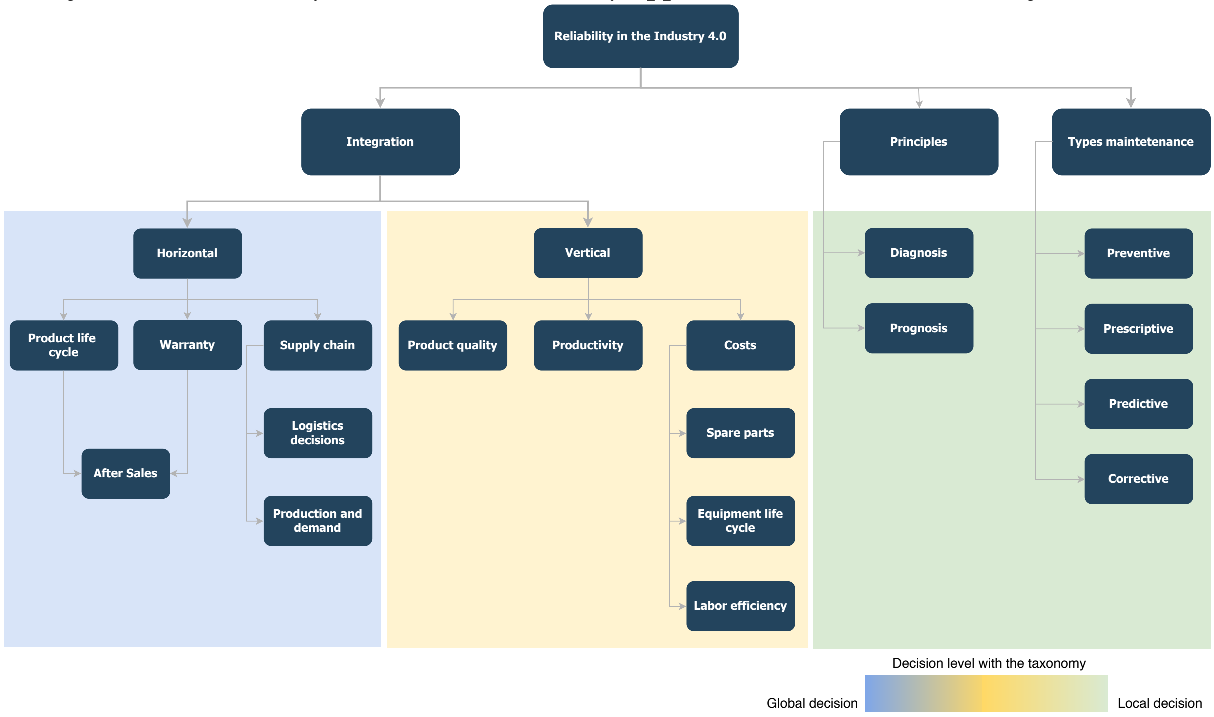
3.3.4 How has reliability helped the value chain management in the context of I4.0?

Intelligent production systems consist of three approaches to decision support, where we can separate them vertically, horizontally, and from end-to-end (SCAPOLO et al., 2014; UHLMANN; FRAZZON, 2018). Being a vertical approach, where there is an integration between the company's internal operations, horizontal external processes are used with inputs for decision support. End-to-end integration is where business cooperation takes place. The first two approaches are considered the main ones in I4.0 (Kagermann; WAHLSTER; HELBIG, 2013; GÄRTNER, 2018).

Figure 22 presents, in a simplified way, how horizontal and vertical integration interact. Distinct colors represent vertical integration, and the group of colors represent horizontal integration process in the organization (Kagermann; WAHLSTER; HELBIG, 2013). Vertical integration is the process of limited interaction between boundaries. As for horizontal integration, processes flow across the organization, allowing production, development, and post-production information to be used as feedback for process improvements (GRACEL, 2018).

During the analysis of the selected studies, we noticed the occurrence of approaches where the purpose was the focus on local equipment decisions. This type of study is linked to traditional reliability search. The so-called conventional and reliability articles concentrate on

Figure 21 – Taxonomy created with reliability application for decision-making in the I4.0



Source: Prepared by the author.

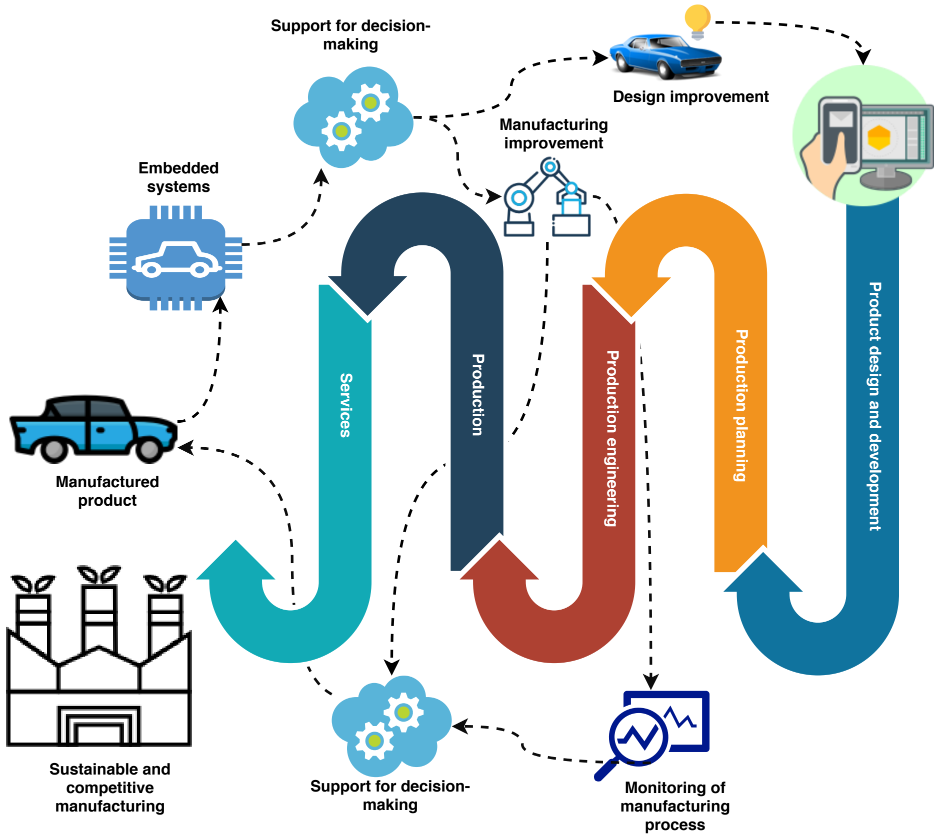
analyzing the health of the machine through predictive or preventive maintenance and, in some cases, evolving to prescriptive models. In the studies of Bousdekis e Mentzas (2017), Ferreiro et al. (2016) the concepts of Big Data were employed. However, the main objectives of the research were the health of the equipment. In these cases, the benefits made possible by applying Big Data were not realized when decisions were centered solely on the machine.

In Appendix J, we gather the studies we classified as focusing on equipment, and this summary is interesting when compared to Appendix G because it shows that the proportion of case studies is higher compared to the other research proposals.

A significant point that articles classified as focus on equipment bring is the proposal of applying Big Data technologies for information collection (GODREAU et al., 2019; KLEIN; BERGMANN, 2018; SYAFRUDIN et al., 2018). Because collecting and maintaining information extracted from different sources is a critical factor for success in I4.0 (LAZAROVA-MOLNAR; MOHAMED, 2019). Thus, as shown in Figure 23, decisions such as: **when will it fail**, as in the studies of Myers, Tickem e Evans (2016), Rimpault, Balazinski e Chatelain (2018), is useful when the intention is to analyze the specific conditions of equipment. In addition, integration models with intelligent processes Preuveneers, Joosen e Ilie-Zudor (2018), Xu et al. (2019) are helpful to support decisions of the type when revising the assets. However, when the goal is to understand what happened to the equipment Bianchini, Pellegrini e Rossi (2019), Sezer et al. (2018), Thoppil, Vasu e Rao (2019), it is essential to support managers.

Thus, decision-making based on the principles and types of taxonomy maintenance focuses

Figure 22 – Typical process flow along the value chain: the colors individually represent vertical integration, and the union of colors represent the horizontal integration.



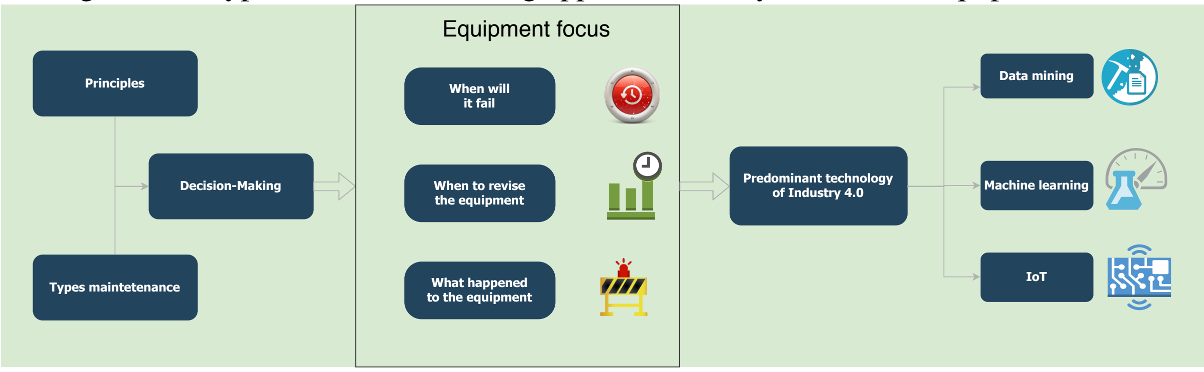
Source: Prepared by the author.

on decisions at local levels within organizations. Figure 23 presents decisions that are predominant to equipment. In this type of decision-making, the dominant technologies of I4.0 are applied to maximize equipment performance (CANITO et al., 2017; FERREIRO et al., 2016; HE et al., 2018; KIANGALA; WANG, 2018; LI et al., 2019; TERRISSA et al., 2016). Equipment-focused decisions are primarily intended to ensure equipment availability (CAO et al., 2019). However, great locations do not always provide a great overall result.

During the construction of the taxonomy, it was possible to realize that the expected decisions characterize the types of integration that are conceptualized in the context of I4.0 (PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018; SANDENGEN et al., 2016). Vertical integration is characterized by arrangements where the goal goes beyond equipment issues. Figure 24, we graphically simplify how the main types of decisions occur in the main resolution employed in vertical processes. In this case, we can cite decisions such as how much to produce and what level of inventory must be maintained to preserve the organization's gain (KIANGALA; WANG, 2018).

Still, in this context, the definition of labor resources is an essential factor for the business's

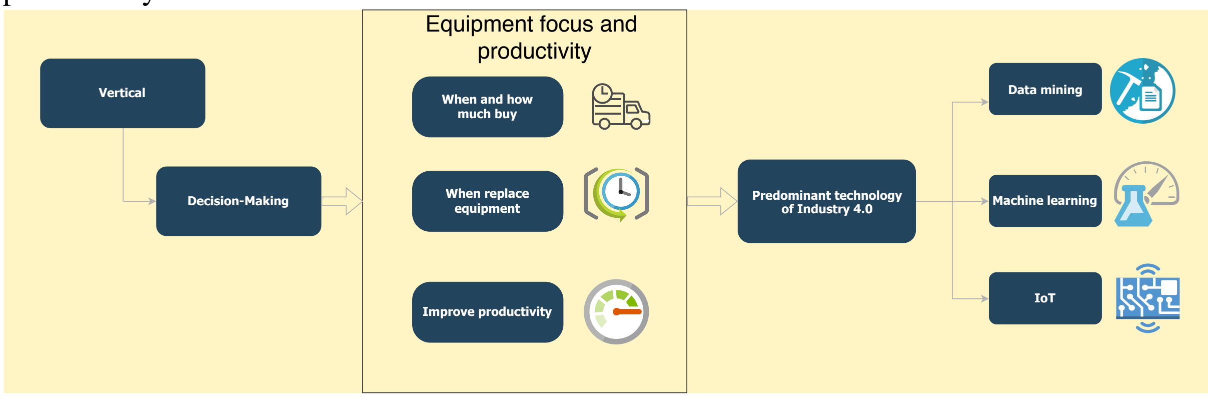
Figure 23 – Types of decision-making applied reliability in I4.0 with equipment focus.



Source: Prepared by the author.

profitability. Soon having the ability to estimate the need for this type of support becomes a significant competitive factor. In the studies, Lee, Zhang e Ng (2017), Mourtzis e Vlachou (2018), Schreiber et al. (2019), the search for better use of human capital is made possible through the use of technologies that come from intelligent factories. However, in the environment of vertical decisions, a recurring concern is linked to production in the right quantity, just in time, and with the expected quality. In the studies analyzed, we noticed that this concern is presented in research related to reliability and I4.0. Information improvements make it possible to deliver quality in products and processes (LEE et al., 2014; TAN et al., 2017). These are examples of applying Big Data and analytics that provide value to the organization (TERRISSA et al., 2016).

Figure 24 – Types of decision-making applied reliability in I4.0 with equipment and productivity focus.



Source: Prepared by the author.

In the Appendix K, we summarized the articles in which we selected with the approaches used in the context of vertical decisions. Appendix K shows that the prevalence of articles with the case study approach was higher in comparison with other methods. This result is relevant because it shows applied research where the type of decision goes beyond the boundaries between machine and process. Such a purpose is expected in I4.0 applications. Another point that articles characterized as applied to verticalization addresses is that the vast majority still have productivity and optimization.

According to the studies summarized in the vertical decision type, it is essential to highlight the diversity of decision types that have been found in the literature. Decisions not only involving human resources, such as operators and maintainers, but also decisions connected to the production environment. As an example, research involving ecological issues Siafara et al. (2017), where equipment operating under normal conditions will be environmentally viable. A vital consideration to reflect is preparing the environment for I4.0 approaches to be applied. That is, an ecosystem must be assembled and operational so that the collected data can be used to generate relevant and consistent insights.

With a significant data infrastructure utilizing CPS connected to IoT devices, data-centric decision-making produces better decisions. Another significant benefit is the velocity with which critical business decisions are made. Depending on the severity of events, data-driving decisions have better effectiveness. According to this context, studies analyzing sensor applications and management employed in equipment are essential for smart factories (FORDAL; RØDSETH; SCHJØLBERG, 2019).

In a framework of CPS, it is possible to analyze employing ML to be used to estimate equipment life (KU, 2018; LEE et al., 2014). With models based on massive amounts of data, greater assertiveness in decisions is achieved (CANIZO et al., 2017; CHO et al., 2018). With possibilities where uncertainties are minimized, the costs involved tend to be reduced (PALAU; DHADA; PARLIKAD, 2019; VOGL; WEISS; HELU, 2019). An example of the benefit of this type of arrangement is with early equipment replacements or unforeseen breakdowns (NEMETH et al., 2018). This type of decision is made possible by connected systems providing data for more certain choices (FUMAGALLI et al., 2016). Benefits like this prevent equipment from being considered disposable, where parts replacement is performed outside the economic life of the machine (ADU-AMANKWA et al., 2019; HE et al., 2018; SÉNÉCHAL; TRENTESAUX, 2019; WANG et al., 2017).

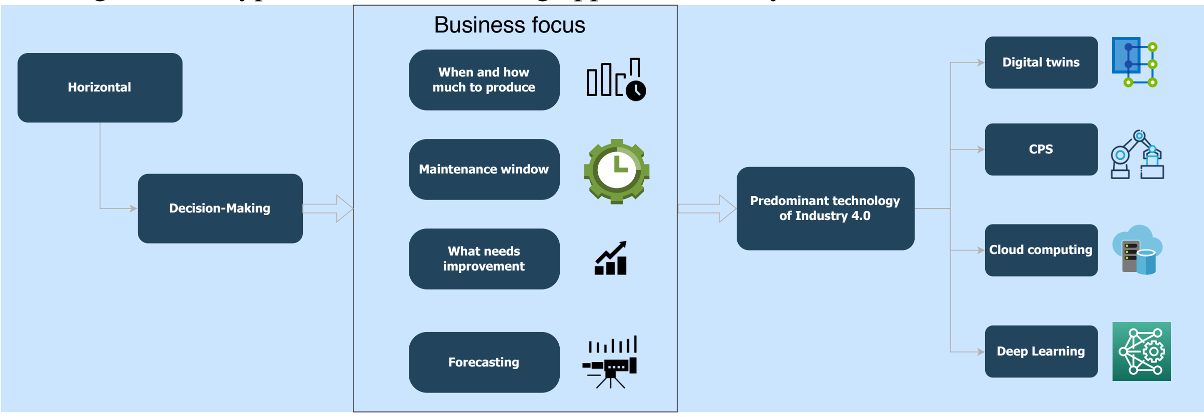
Despite the possibility of important decisions, vertical integration is limited to organizational boundaries, and a company is usually embedded in an industry, or productive sector (SCHEER, 2019; UHLMANN; FRAZZON, 2018). Thus, managers need to make conclusions beyond the company border. Furthermore, as evidenced by taxonomy, this kind of arrangement is defined as decisions at the horizontal level (GÄRTNER, 2018).

In Figure 25, decisions arising from horizontal integration, applying I4.0 technologies, are intended to use business-focused reliability (LEE et al., 2015). In other words, decisions from a horizontal perspective search results where global objectives are reached, according to Figure 20. It can be observed that reliability is even addressed in social networks, supply chain management, and enterprise resource planning (TAN et al., 2017). Thus, decisions about the lens of horizontal integration have an impact on the whole organization (BALOGH et al., 2018; LEE; ZHANG; NG, 2017; LEE et al., 2014; PREUVENEERS; JOOSEN;

ILIE-ZUDOR, 2018; WAN et al., 2018).

Among the leading technologies of I4.0, as shown Figure 25, there is the Digital Twin, which associated with production strategies that enable a flexible reconfiguration of manufacturing strategies (STRAKA et al., 2018). The facility with which alternatives combined to production, processes, and market are examples of the benefits of technologies associated with I4.0 in business focus (CHENG et al., 2018; ROSEN et al., 2015).

Figure 25 – Types of decision-making applied reliability in I4.0 with business focus.



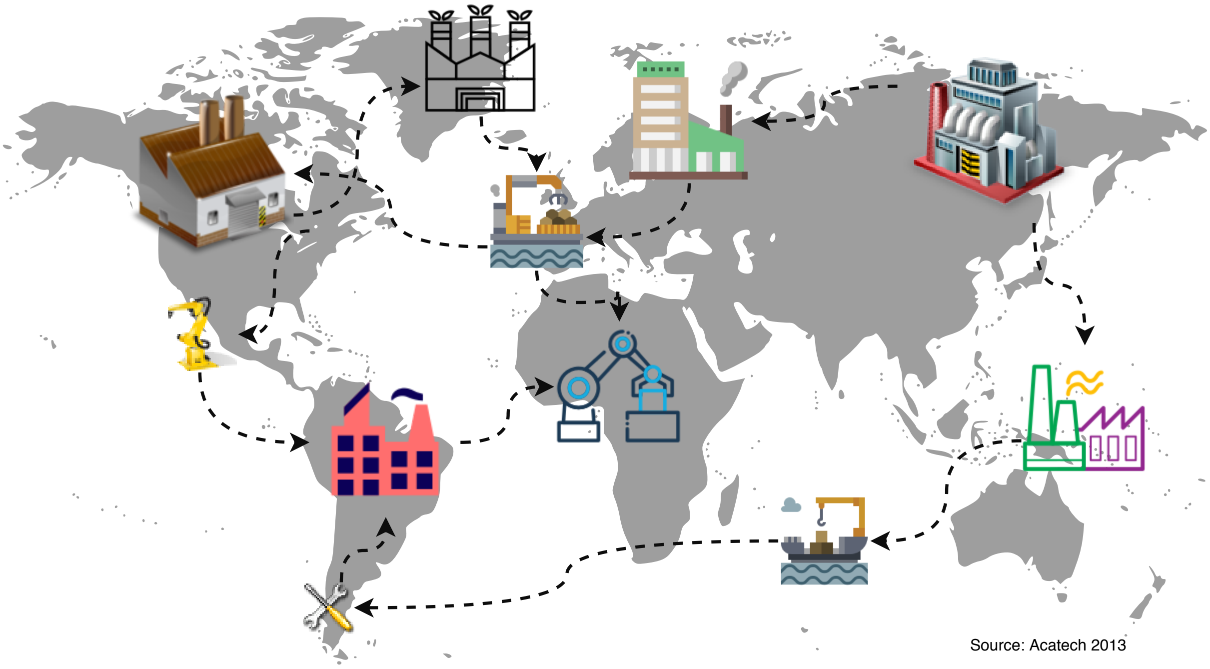
Source: Prepared by the author.

In the analysis of the selected articles, we extracted in Appendix L the studies positioned where the objectives go beyond the processes of the organization. It should be noted that the amount of case study is still limited compared to the other approaches that have been proposed in the construction of taxonomy. In selected articles, where the case study was applied, the main focus is on the product. As an example, studying the impacts related to PDM on product quality, the purpose of the decision is beyond the organization (TSAO et al., 2020) boundary. Likewise, identifying barriers and understanding the after-sales impact when applying PDM is essential information for decision-makers (SALAZAR et al., 2019; MADHIKERMI et al., 2018).

As shown in Figure 26, there is an interconnection between companies where one organization cannot be independent of another (Kagermann; WAHLSTER; HELBIG, 2013). This type of integration is the core of I4.0 (GÄRTNER, 2018). Because there is interconnection between the companies, it is possible to quickly adjust product inventory levels when there is a problem with suppliers or customers (ZHOU; LIU; ZHOU, 2016). However, if there are no interconnected processes, there is a delay between the occurrence of the events and the corrective actions (GRACEL, 2018; SCAPOLO et al., 2014). Similarly, the ability to make internal adjustments to the company based on demand is an essential competitive advantage (NGUYEN et al., 2016; XIA; XI, 2019). Reliability studies are, therefore, of great importance to the I4.0 (SCHEER, 2019). It should be noted that in the event of a failure in the organization's internal equipment, the harm generated may damage the supply chain and may even cause customer dissatisfaction (MEHDIYEV et al., 2017; PÉREZ-LARA et al., 2018).

A relevant fact that we highlighted is related to the number of studies in which the target is to propose or develop architecture for data collection (LEE et al., 2015; SANDENGEN et al., 2016). Another important technology to be considered with reliability is Digital Twin, especially in product designs, since the manufacturing can be accelerated, due to the possibility of emulating real systems in virtual environments (KUEHN, 2018; LIM; ZHENG; CHEN, 2020; MADNI; MADNI; LUCERO, 2019).

Figure 26 – Interconnection and cooperation of companies in a worldwide business network, the core of I4.0.



Source: Adapted by the author.

Reliability has been in the researchers' agendas, as we can present in the analysis of the selected articles. However, in our research, it was possible to note that the main efforts are focused on integrating and collecting data. The data collection, storage, and processing is a critical stage. We noticed as a significant part of the studies analyzed that the main focus was the development of Big Data management architecture (MADHIKERMI et al., 2018; ROSSIT; TOHMÉ; FRUTOS, 2019). These findings confirm the importance of the data-driven stage for decision-making. In this case, it reaffirms the importance of data management in the context of I4.0 to generate gains for the organization and value chain (KLINGENBERG; BORGES; ANTUNES, 2021; LEE et al., 2015; PÉREZ-LARA et al., 2018; SANDENGEN et al., 2016). The next step is to understand which challenges and future directions that I4.0 is immersed in, concerning reliability for decision-making.

3.4 Opportunities and Motivations

It is now possible to generate a large amount of data from machines and processes, as the number of sensors and IoT devices are becoming more widespread. However, there is still a contrast between device opportunities and reality. Technologies have improved at higher rates than the need for machine replacement. Another adversity is that while new machines and equipment are already produced with several sensors, old machines are difficult to connect due to obsolete controllers and different manufacturers (BALOGH et al., 2018; ROCHA; BARRETO; SEMAN, 2019).

Equipment manufacturers have realized the advantages of adopting IoT systems, and many, especially the largest, have already migrated to connected systems. Either way, small and midsize businesses need a viable strategy to realize the benefits of using IoT systems to be competitive (BALOGH et al., 2018; CANIZO et al., 2017; CHO et al., 2018; RIMPAULT; BALAZINSKI; CHATELAIN, 2018; TERRISSA et al., 2016).

As a relevant factor for the business to be competitive, asset management should be an essential item on the decision-makers agenda (SCHEER, 2019). Given that, not providing a customer as a result of unplanned downtime may expose the client to the competitor. In this scenario, reliability proves to be fundamental for the organization since asset management, considering reliability, helps improve results in general (GRACEL, 2018).

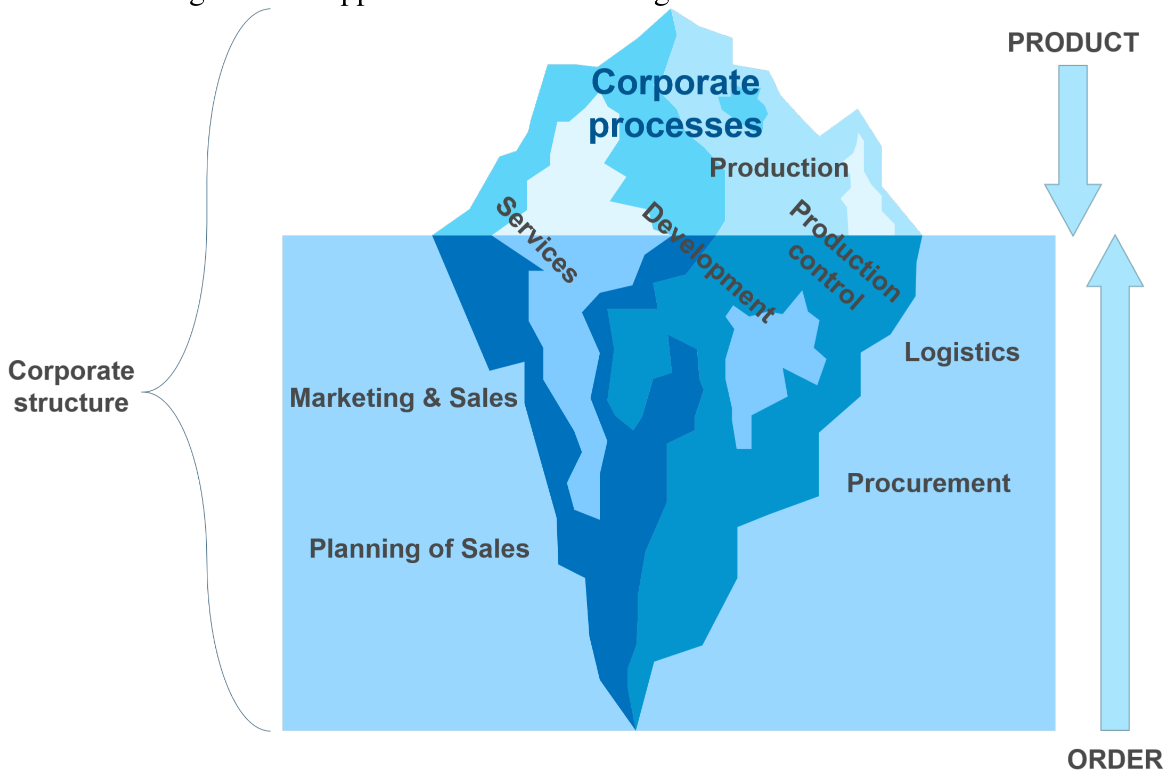
However, employing reliability taking into account only equipment and peripherals, limits the achievement of maximized results for the business. Thus, it is essential to analyze issues such as inventory level, material and sales prices, logistics issues, and the market as a whole (GRACEL, 2018; SANDENGEN et al., 2016; SCAPOLO et al., 2014). In this context, we present Figure 27 as an iceberg of possibility and opportunity for I4.0 (SCHEER, 2019; GÄRTNER, 2018).

Figure 27 summarizes the areas of the corporation and also symbolizes the order flow, upward and downward, the stream of production. It is observed in Figure 27 that the production process is visible, that is, having a higher perspective in reliability studies (BALOGH et al., 2018; FERREIRO et al., 2016; FORDAL; RØDSETH; SCHJØLBERG, 2019; KIANGALA; WANG, 2018; KŁOS; PATALAS-MALISZEWSKA, 2019; KU, 2018; LEE et al., 2015; MOURTZIS; VLACHOU, 2018; NEMETH et al., 2018; PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018; RUIZ-SARMIENTO et al., 2018; SYAFRUDIN et al., 2018; TSAO et al., 2020; WANG et al., 2017; ZHENG et al., 2018). In addition to production control, services, and development, there are already studies that address this issue associated with reliability, even if to a limited extent (MADHIKERMI et al., 2018; PÉREZ-LARA et al., 2018).

The vast majority of authors considered and had excellent prospects for applying decision-making reliability (ESMAEILIAN; BEHDAD; WANG, 2016; Kagermann; WAHLSTER; HELBIG, 2013; SCHEER, 2019; GÄRTNER, 2018). Decisions go beyond vertical integration and find the entire heat chain and corporate structure, as shown in Figure 27. Examples of opportunities include:

Production Flexibility: With digital factory and horizontal integration, custom production is feasible as it will be possible to identify the most important consumer needs. With smaller

Figure 27 – Opportunities and Challenges for the business in I4.0.



Source: Prepared by the author.

batch production, the variety of products can be significant. In this scenario, the reliability of production equipment should receive special attention, as equipment fault tolerance must be minimal. Another significant impact on customized production is related to logistics processes, as the number of suppliers tends to increase, and the company's dependence on specific suppliers tends to decrease. In this way, the production chain will have a faster response in meeting the particular wishes of customers (RYCK; VERSTEYHE; SHARIATMADAR, 2020). Manufacturing flexibility assists in the rapid reconfiguration of the production system, where it can be automated through a Digital Twin to simulating different scenarios (CHENG et al., 2018; ROSEN et al., 2015).

Planning: Important questions refer to such decisions - when to produce, how much to produce, how much to buy, and how much to store. In all planning decisions, reliability has a high magnitude function, because when such choices are optimized, availability of equipment is considered. Failure to meet availability in many cases causes significant damage to the corporation.

Environmental: Monitoring equipment and minimizing unnecessary interventions, replacing still functional parts, is a way to save environmental resources. Another point to note is the possibility of optimizing materials with the production of defective items. In this way, I4.0 makes it possible to use resources efficiently Szalavetz (2017). Still, on this topic, the consumption of renewable resources, such as energy, can be reduced with the use of energy-oriented reliability to promote the application of sustainable manufacturing. Thus, integrating the energy attribute to dynamically schedule preventive maintenance intervals as performed in the study by Xia et al. (XIA et al., 2018).

A vision of computing in I4.0: from computing in the scenarios presented, it will play a key role, as enabling technologies of I4.0 are predominantly digitally connected. So to take advantage of the ecosystem providing integration of all processes, information technology will be crucial for organizations. Thus, the area of computing should have a cross-sectional combination with the entire corporation. As a result, the data generated in different departments will be used to build numerous models to support decision-making. These scenarios emerge as a future challenge for applied computing.

Reliability for decision-making: according to the results found, in some systems reliability studies, the concept is still linked only to the equipment. We can see this in the response of **SQ4**, where the vast majority of articles analyzed, addressed issues related to equipment health, productivity, and operating costs (BALOGH et al., 2018; XU et al., 2019). However, according to the future perspectives presented, reliability has a role that goes beyond and must be considered so that companies' strategic decisions are made efficiently (SANDENGEN et al., 2016).

When decisions covering the value chain are required, only sophisticated algorithms for measuring and determining equipment availability are no longer sufficient (BORGI et al., 2018; GAO et al., 2015). According to what was presented in response to **SQ2**, the data are critical factors for the use of sophisticated forecasting models (PÉREZ-LARA et al., 2018; ROSSIT; TOHMÉ; FRUTOS, 2019). Therefore, the performance and options provided by the models are dependent on the data that is used (LEE; ZHANG; NG, 2017; ROSSIT; TOHMÉ; FRUTOS, 2019).

Thus, the collection and conditioning of data are critical factors for the algorithms to deliver expected results (NEMETH et al., 2018). So data and modeling must be following the type of desirable decision. The need to include aspects related to the entire value chain in the reliability models are fundamental prerequisites for there to be a perception of gain in the use of technologies linked to I4.0.

This environment is conducive to the application of I4.0 technologies. However, decisions are still just the visible part of an iceberg of possibilities. Research, where reliability models consider not only the local environment but the existing connections in the systems, must be made so that the value provided by the I4.0 is perceived throughout the value chain.

3.5 Final Remarks

The related works to this research show that most research focuses on improvements in communication techniques and data ingestion. In addition, according to Figure 12, They address questions related to information collection. As for research related to analytics, the vast majority use already-known datasets in the quest to improve predictive models laterally.

In this sense, we perceive the need to move forward with ML applications for real and complex environments, given that investment in technology is only valid if there are benefits for the sponsor. Another finding we noticed through the literature review is the gap in applications that explore systems in the form of horizontal integration. In this way, with the proposed framework, we intend to offer an approach that makes it possible to benefit from I4.0 to increase the reliability of the systems applying the PHMS.

With a comprehensive and careful review of the related works, it was possible to confirm the academic motivations and shed light on specific gaps that this research aims to elucidate. For this purpose, in the next chapter, the proposed method adopted in this research will be introduced and detailed.

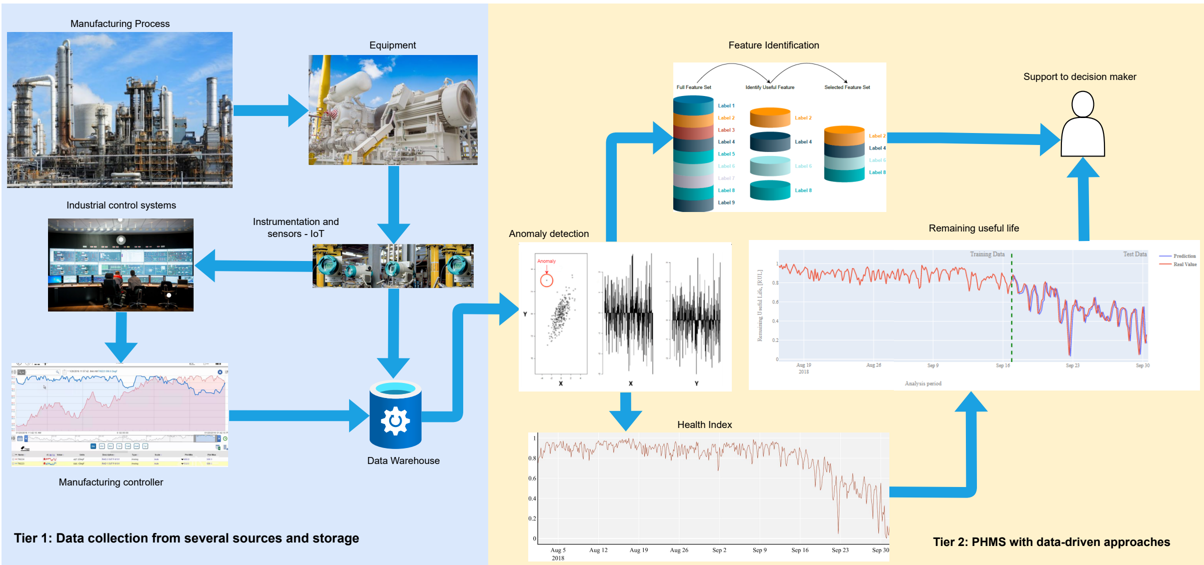
4 FRAMEWORK TO SUPPORT DECISION-MAKING IN THE I4.0 CONCEPT

We divided the framework development that supports the PHMS into three subtopics. The first subsection below discusses the AE model for anomaly detection. The second subsection presents the model development for FI of the main variables related to the failure. Finally, the third discusses the activities that make up the system's RUL prediction stages.

4.1 Framework overview

Figure 28 shows the general structure of the framework. The objective is to develop an approach in which anomalies are detected, and preventive actions are carried out. In addition, using the framework provides strategic decisions regarding asset availability. Thus, the main features of the proposed framework will be discussed in the following sections.

Figure 28 – Framework to support the decision-making in I4.0 concept.



Source: Prepared by the author.

The Figure 28 illustrates the proposed structure, which was divided into two tiers. The first tier is data collection from various sources and storage, which allows all relevant devices and data sources to be used for process control and monitoring and later stored. The second tier consists of anomaly detection, FI, and AE. Finally, the framework provides decision support, especially on the PHMS. The techniques and systems contained within the framework's tiers are detailed in the following sections.

4.2 Tier 1: Data collection from several sources and storage

CPS is an essential element among the enabling technologies for I4.0, as it integrates physical and virtual processes. As an example, the integration of operational equipment with cloud computing or networks. In this way, it is possible to capture and manipulate data from operations, communicating with physical devices, and receiving feedback from the executed actions on the entire system and vice versa.

The CPS module plays an important role and provides the basis for intelligent manufacturing. The CPS module's main contribution is to give the manufacturing system understanding and perception, i.e., perceiving its condition and the surrounding environment's state. In other words, to integrate operations as a system consisting of different operations and devices. Equipment, production process, and product constitute a single manufacturing system in which the production conditions are monitored by connecting the sensors and equipment that constitute the respective system.

In the context of the framework presented in Figure 28, the CPS provides the integration between the different systems and provides feedback on the process's conditions according to the actions performed. Thus, providing bases for maintenance and production teams to carry out specific interventions.

According to the concept presented, the CPS in this framework consists of a link between physical equipment and virtual systems. Thus, in the following subsection, we address the data and the source that make up the framework.

4.2.1 Instrumentation and sensors

In a smart factory, knowledge automation is manifested by multi-agent collaboration in a distributed environment. In the smart factory environment, multiple agents collaborate and act in combination to execute specific predefined actions. IoT devices are applied for monitoring and control purposes in the petrochemical industry and production management and control purposes. Online data systems in the petrochemical industry consist of sensing and measurement, industrial networks, and various measurement and analysis systems. To access the production data in real-time usually, the following infrastructure usually requires APIs (application programming interfaces) and UIP (user interface program) (MIN et al., 2019).

Automation systems are based on controllers and offer, in some cases, control possibilities ranging from simple to advanced constraints where optimal parameters are predefined. This means it is possible, under stable conditions, to come near to the optimal operation. Standard rules in this case, as in petrochemical and oil refining in general, are the capacities of heat exchangers, pumps, valves, and other process devices, whose connections and capabilities interact to form a connected system (LI, 2016).

4.2.2 Industrial control systems

Industrial Control Systems (ICS) are composed of a complex set of sensors, actuators, and control agents, such as control systems, including Supervisory Control and Data Acquisition

(SCADA) and Distributed Control System (DCS) (MCA; MCA; MANDA, 2018). ICS are widely employed in different critical operations, where it plays an essential role in monitoring and controlling physical and chemical products. These systems are responsible for receiving the signal coming from an instrument, such as flow control, and provide the operators with the possibility of making adjustments if necessary. This type of system also enables remote operation, i.e., based on a target value, the system automatically adjusts the output signals to search for the defined target (MOKHTARI et al., 2021; BALADOR; ERICSSON; BAKHSHI, 2017).

In the petrochemical industry, the DCS operation is critical. This system acts as a "brain" for production, providing operators with the possibility to intervene to keep output in optimal condition and ensure the system's safety. In conjunction with DCS in industrial environments, historians are generally used to converting data and directing it to the data warehouse to be available for future analysis (LI, 2016). Thus, the CPS can be considered a set of industrial equipment that is reproduced virtually, with communication and feedback(MOKHTARI et al., 2021; ISMAIL; TRUONG; KASTNER, 2019; BALADOR; ERICSSON; BAKHSHI, 2017).

Due to the inherent operational complexity of petrochemical processes, this industry is considered the forerunner of industrial automation. Thus, with the introduction of the fourth industrial revolution, this industry can be considered a propitious environment for the use of new technologies. Data mining can help operators identify hidden patterns due to the immense amount of variables that operators usually need to control (ISMAIL; TRUONG; KASTNER, 2019). The proposed framework provides an advantage because patterns shuffled among the monitored variables can be displayed and ranked according to the priority level regarding operational interventions using an analytical model (LI, 2016; MIN et al., 2019).

4.2.3 Data source

Together with the CPS, IoT is the primary driver of I4.0. Therefore, as the main element of Data Sources, we can consider the IoT as a direct role. It enables all common objects that integrate operations in the enterprise to perform independent functions and realize the connection between physical and virtual systems.

The function of the IoT in the Data source module is primarily data transmission. Data collected by sensors in the device is transmitted to local databases or cloud data centers via the IoT for real-time or subsequent use. Some plant equipment is equipped with supervisory control and a data acquisition system DCS to monitor its equipment.

In the anomaly detection module, sensor data from the DCS equipment can be fused to extract pertinent rules in-depth and, consequently, the system's health status and subsequently analyzed and diagnosed. Data from the entire operational area will be considered regarding the framework's source of data, focusing on the plant's reaction. The data come from different types of equipment and therefore have distinct characteristics. The data source block was subdivided into two parts, data coming from the DCS and data acquisition. The acquisition part is linked to the collection point, while the DCS consists of controlling the equipment and directing the data. The data is processed in the DCS and then directed to the company's operational database. The data is stored in relational databases and is available for future analysis. In the context of the main objectives that guide this research, the data source will be the raw material that feeds the proposed architecture's next block. Furthermore, thus, enable data-driven management considering the reliability of the system. The following Tier consists of utilizing equipment data and applying data mining to identifying anomalies that can put the plant's operational continuity at risk.

4.3 Tier 2: PHMS with data-driven approaches

This section will present the core of this research, contextualizing the PHMS model, which is not limited only to identifying anomalies but also provides insights for strategic actions. We have separated this section into sub-sections to facilitate understanding and shed light on the importance of each stage in the desired result.

4.3.1 Anomaly detection

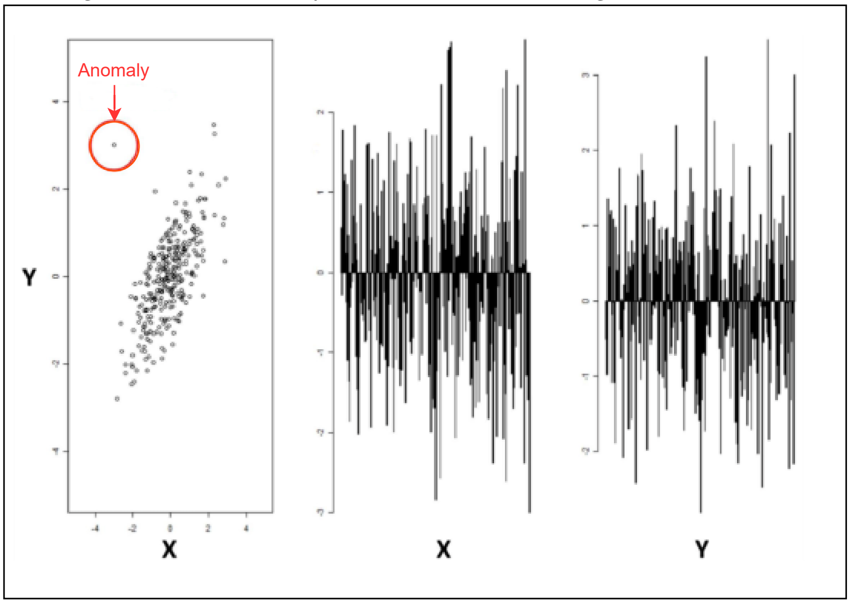
Connectivity through sensors connected with equipment allows for an abundance of data. The main challenge for manufacturing is using these large amounts of available data and extract useful information, making it possible to reduce costs, optimize capacity and maintain asset availability. In this panorama, the most recent ML and data analysis resources emerge as an excellent possibility to improve managers' decisions. For this step, an AE model was applied through a CNN type architecture. In Section 5.2.1.1, we discuss the implementation of the model in more detail.

In this scenario, anomaly detection (or outlier detection) has been applied to assist reliability engineering using advanced ML techniques. Anomaly detection identifies rare elements, events, or observations that arouse suspicion as they differ substantially from most data. Anomalous data can usually be linked to some problem or rare event, such as bank fraud, medical problems, structural defects, equipment malfunction. This connection is useful to identify which data points can be considered abnormal since identifying these occurrences is usually strategic from an operational point of view.

Thus, it guides us to one of this model's main objectives: to identify whether the data points generated by the different sensors are normal or anomalous? As in Figure 29, analysis through data visualization can provide relevant information to decision-makers in some simple cases. However, in production environments, the number of features is high, and analysis utilizes graphics as Figure 29 becomes a considerable challenge.

Any equipment, whether of the rotary type (pump, compressor, gas, or steam turbine, among

Figure 29 – Anomaly detection considering two variables.



Source: Prepared by the author.

others) or stationary (heat exchanger, distillation column, pressure, and level control valves), will eventually reach a stage in which health is deficient. This point may not necessarily lead to real failure, but conditions in which the equipment is not operating under ideal conditions. This signals that there may be a need for some intervention in maintenance to restore all operational potential or even some operational adjustment.

The most common way to perform condition monitoring is to examine each measurement of the equipment's sensor and assign a minimum and maximum acceptable value limit. If the measured value is within limits, the equipment or system is normal. On the other hand, if the measured value is outside the limits, it is considered unhealthy, and an alert is sent.

This strategy of imposing alert limits is known to send many false alarms, alarms for situations where the measures represent acceptable states for the equipment and systems. On the other hand, in some cases, alerts are missing, that is, problematic situations that should be alerted. The first problem wastes time, effort, and equipment availability, as an improper repair stop reduces operational availability. The second problem is more crucial, leading to real damage with associated impacts and production losses.

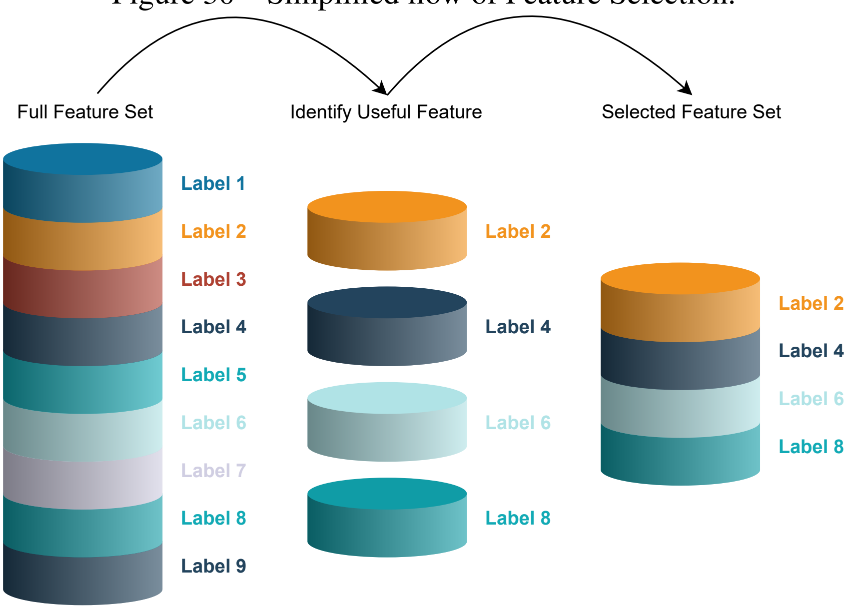
Both problems result from the same cause: the condition of an equipment or system cannot be reliably assessed based on the analysis of each measurement alone, as illustrated in Figure 29. Thus, we must consider a combination of the various measures to obtain a more accurate indication of the analyzed system's real condition.

In this research, the identification of anomalies contributes to decision-making by different areas. The environment in which it will be developed is a complex system consisting of different equipment types with different measurements. Therefore, traditional anomaly identification techniques usually do not show satisfactory results for this type of problem. Thus, to guide the maintenance and production teams' actions by identifying anomalies, the combination of advanced ML techniques is presented as a viable alternative (SAHAL; BRESLIN; ALI, 2020; BIANCHINI; PELLEGRINI; ROSSI, 2019). Among the possible techniques, the selection of variables has been successfully applied in several studies (VOGL; WEISS; HELU, 2019; ZHANG et al., 2018; XU et al., 2019; WANG et al., 2017). That said, in the next section, we outline the application of variable selection in the context of this research.

4.3.2 Feature Identification

The anomaly identification step helps to identify the existence of a measure outside an established standard. Undoubtedly, the identification of anomalies is relevant to highlight the need to take some action (FERNANDES et al., 2019; HU et al., 2018). However, in complex systems, identifying which measure or measures are responsible for a particular anomaly becomes difficult (SAHAL; BRESLIN; ALI, 2020). In this way, FI techniques have contributed to reducing the complexity of determining which variables are responsible for a specific abnormal condition (LIU et al., 2021). Figure 30 presents a simplification of the flow of identification and selection of the main variables to identify the system's unusual state.

Figure 30 – Simplified flow of Feature Selection.



Source: Prepared by the author.

The most relevant features are extracted in the first place by transforming data captured in different domains, for example, statistical domain, frequency, and time-frequency, to obtain representative and relevant information. FI is then applied to improve the relevance and reduce redundancy between features before feeding the ML model (MOKHTARI et al., 2021; MUTLU; ALTUNTAS, 2019). Thus, the performance of the developed model reduces the dependence only on the optimization of the algorithms. Typically, feature extraction and identification are time-consuming and dependent on domain knowledge (MOORTHY; GANDHI, 2020). In Section 5.2.1.3, we discussed the implementation of the model in more detail.

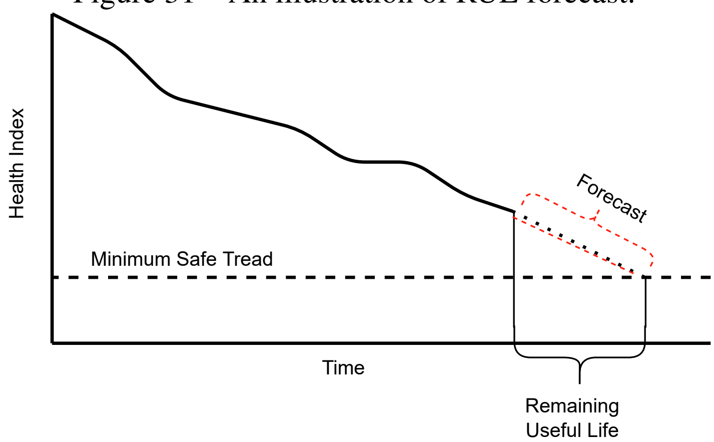
Traditional feature selection applications have focused on extracting latent variables that predict a particular action or event (TSAI; SUNG, 2020; CHEN et al., 2020). The framework proposed to apply the identification of variables to, in advance, identify which variables are responsible for an eventual anomaly in the process. Thus, with the variables ranked at the level of importance, a manager can perform operational and maintenance actions, and eventual failures can be avoided, thereby increasing systems' reliability.

After presenting the FI step, the following subsection presents the step of the AE of the system. This contributes to the maintenance team's direction and decision-making.

4.3.3 RUL - Remaining Useful Life

Maintenance actions have traditionally been based on information representing the equipment's current conditions, which researchers have called "based on a non-predictive condition" (LI, 2016; TERRISSA et al., 2016; WEN et al., 2021). The notable aspect of PDM is using methods and models to estimate additional conditions and RUL (AYDEMIR; ACAR, 2020; LI et al., 2019; SCHREIBER et al., 2019; GUO et al., 2017; TERRISSA et al., 2016). PDM differs from traditional CM maintenance in which it recommends maintenance actions based on information extracted through CM, and the goal is in the current condition, compared to establishing a prevision when using PDM (WEN et al., 2021; WU et al., 2021; LIU et al., 2021; CHEN et al., 2021; AYDEMIR; ACAR, 2020).

Figure 31 – An illustration of RUL forecast.



Source: Prepared by the author.

A RUL estimate of a failed component, the assignment of uncertainty thresholds to the trend curve that will provide the maintainer with the earliest and latest time (with increasing risk) to perform maintenance, and the associated risk factor when maintenance action is delayed are necessary to answer the maintainer's question. Figure 31 presents the RUL concept as a function of time and the system's Health Index (AYDEMIR; ACAR, 2020; WEN et al., 2021).

Several studies have focused on developing models to provide greater equipment availability. However, DL has provided excellent results among ML models, especially when the system's complexity is high (CHEN et al., 2021; LIU et al., 2021; XIA et al., 2020; GUO et al., 2017). In this sense, in Session 5.2.2, procedures used to carry out this critical step in the proposed framework will be presented in detail.

4.3.3.1 Machine Learning for RUL prediction

In the PHM context, answering the following question is one of the main goals of the PHM approach: "What is the RUL of a machine or component after a malfunction situation is detected, isolated, and identified"? In this regard, we propose an approach called PHMS, as presented in Figure 1, which aims to analyze the system's health.

This study sheds light on advanced DL techniques for estimating the system's RUL. Therefore, we compared different DL techniques applied in HI to estimate the system's RUL. Furthermore, we consider using the Neural Basis Expansion Analysis For Interpretable Time Series Forecasting (N-BEATS) architecture to estimate the system's RUL as a contribution of this study. This architecture has promising results compared to traditional methods for time series prediction (MAKRIDAKIS; SPILIOTIS; ASSIMAKOPOULOS, 2022; ORESHKIN et al., 2019). Figure 32 summarizes the architectures we used in the RUL prediction step.

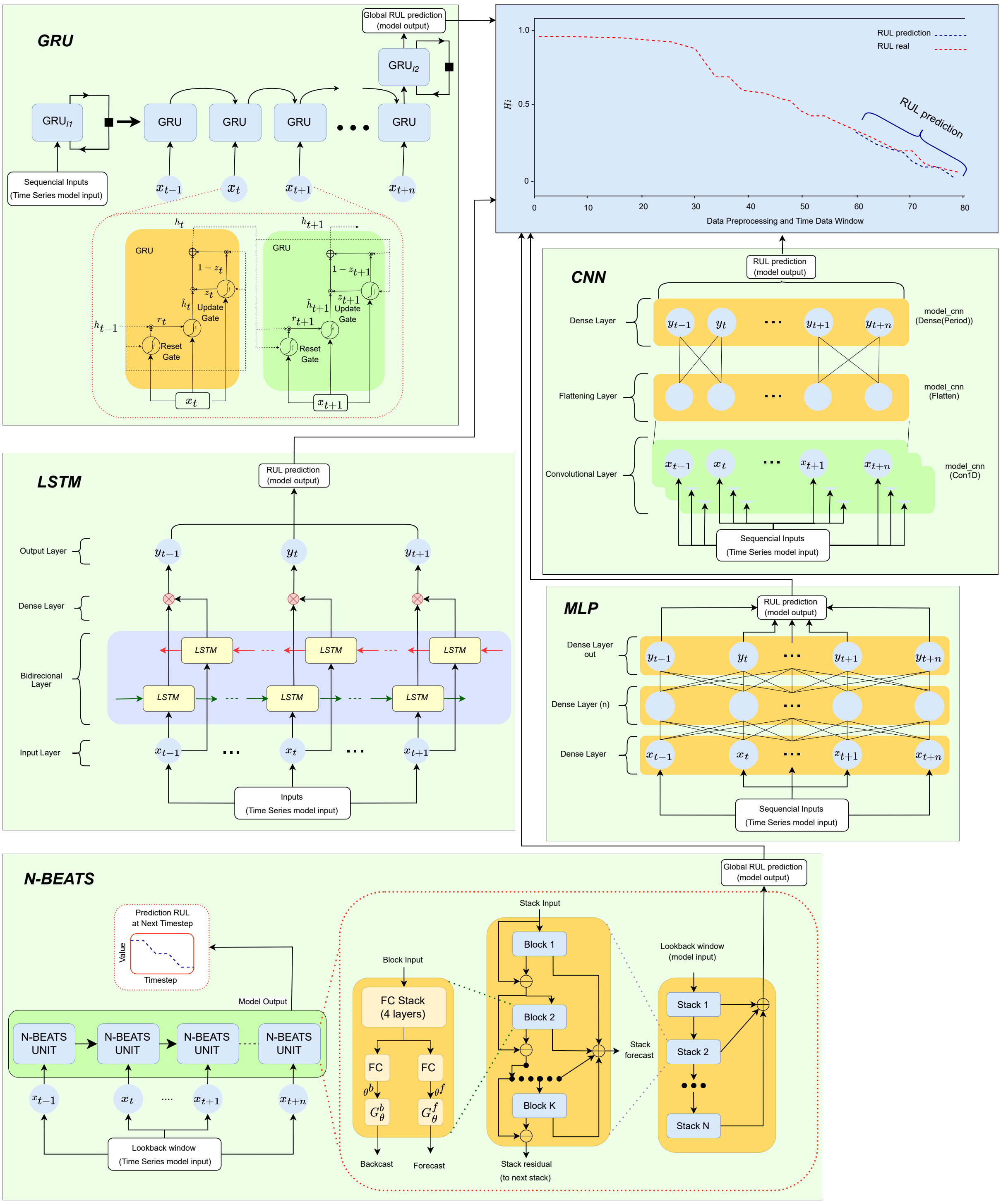
In Figure 32, we present the N-BEATS architecture, which is based on bi-directional residual links and a very deep stack of fully connected layers. Moreover, we bring LSTM and GRU architectures, which are RNN dealing with long-time dependencies in data sequences. We use the MLP, a feedforward neural network that can learn dependencies in a sequence. Finally, we present the CNN architecture, which can learn long-term dependencies on a complex data sequence. In Subsection 5.2.2.3, we present in more detail the architectures and the main differences.

4.4 Final Remarks

In this chapter, the steps for developing the framework were discussed. Each phase has importance for the final result. First, however, the issue related to data collection is highlighted. For example, when making an analogy with a car engine, we can say that the fuel is the data because if we don't use quality fuel, the engine won not work correctly. In this sense, there is no point in using sophisticated models on inconsistent data, "garbage in, garbage out".

Tier 2 deals with the modeling phases, where each step already allows the visualization of results. For example, anomaly detection signals that something may be out of behavior patterns. However, it is not always clear where the efforts of the operation and maintenance teams should

Figure 32 – Simplified DL architectures for RUL prediction. As input, standardized time series windows and the models' forecasts as output.



Source: Prepared by the author.

be concentrated, especially in a complex system such as a petrochemical plant. In this sense, the FI stage provides a north pointing possible causes of the anomaly detected and, thus, reducing the time of action to correct the deviation. Finally, knowing how long the system can operate before a failure is fundamental for allocating efforts and prioritizing actions. In this sense, the RUL forecast stage gains relevance. Given these scenarios, the PHMS supported by the framework is relevant to support decision-making in the context of intelligent manufacturing.

In the next chapter, it will be discussed how to operationalize the framework detailing each step.

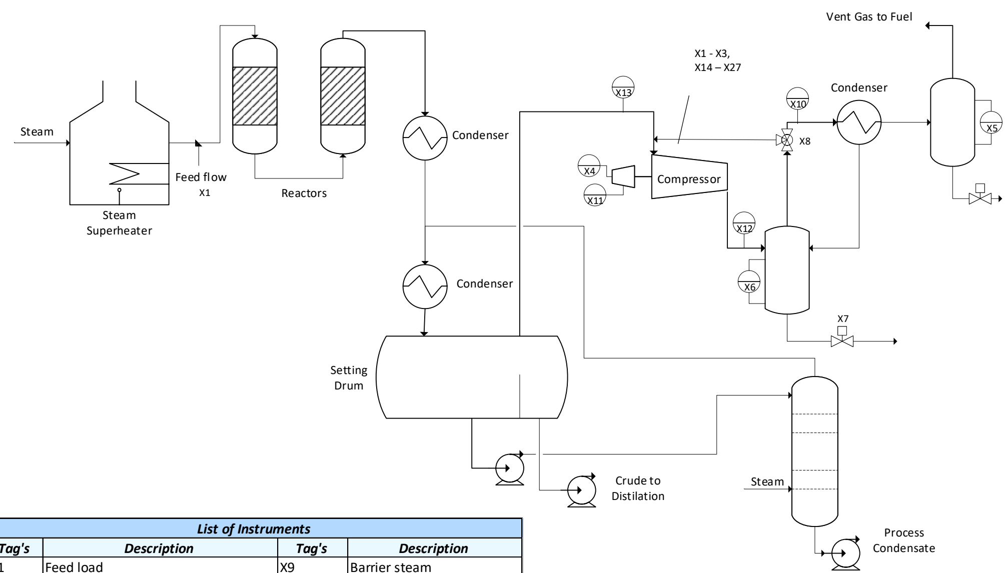
5 MATERIALS AND METHODS

According to the framework proposed in Chapter 4, this chapter presents the methodological procedures used to propose and validate the PHMS considering a real case study. Therefore, we have divided this chapter into two tiers. The first deals with data acquisition and the criteria that define the analysis periods considering the case study. The second tier describes data-driven approaches and models used in the research and the validation criteria.

5.1 Tier 1: Data collection

The Styrene Monomer (SM) is a process industry that is intensive in control parameters due to the criticality in which minor operational disturbances can cause damage to equipment and production out of specification. In this sense, the equipment that has significant importance is the compressor. For that matter, the possibility of analyzing all variables and detecting small changes that may cause the compressor to fail would significantly benefit the business. For example, in Figure 33, the main control variables are organized from X1 to X27. As the number of parameters and control variables is massive, there is a need to develop a model in which it is possible to identify the process variable that has tremendous significance in generating a compressor shutdown.

Figure 33 – Simplified diagram with the equipment and instruments of the reaction area of the petrochemical plant where we applied the study.



List of Instruments			
Tag's	*Description*	*Tag's*	*Description*
X1	Feed load	X9	Barrier steam
X2	Condensed for cooling	X10	Discharge pressure
X3	Injection of anti polymerization	X11	Speed control
X4	Compressor motor electrical current	X12	Output temperature control
X5	Sealing water	X13	Input temperature control
X6, X7	Discharge drum level control	X14 to X24	Vibration control
X8	Compressor recirculation	X25 to X27	Displacement control

Source: Prepared by the author.

The Chemical Industry is intensive in control and automation. For this, process data is transmitted in real-time to the supervisors through electronic devices installed on the equipment. The data from the assets, in the vast majority, is used to monitor the production process in a DCS. In the case were carried out the study, the data management architecture relies on the process monitoring software. It allows operators and managers to monitor in real-time the main process parameters. Another benefit of this system is that it supports the staff in investigating historical events since the data generated by the process is stored in a Structured Query Language (SQL) database and is available for analysis when necessary.

Table 4 – Process variables with units of measure.

Variables	**Process measurements**	**Unit**
X1	Feed load	t/h
X2	Condensed for cooling	kg/h
X3	Injection of anti polymerization	kg/h
X4	Compressor motor electrical current	A
X5	Sealing water	%
X6	Drum level control	kg/h
X7	Drum level control	%
X8	Compressor recirculation	mmHgA
X9	Barrier steam	Kgf/cm2g
X10	Discharge pressure	Kgf/cm2g
X11	Speed control	rmp
X12	Output temperature control	°C
X13	Input temperature control	°C
X14 - X27	Vibration sign	mm/s

Source: Prepared by the author.

As shown in Figure 33 different equipment generate several types of data, according to Table 4. Although the system usually provides data in real-time, in this study, the author defined the frequency of data collection in an average interval of 30 minutes for variables presented in Table 4. With this criterion to minimize possible noise in the indication of the instruments (MIN et al., 2019). Therefore, high dimensionality is frequently present in information from the instruments and sensors since it owns different measurement types.

The time interval selected was from Aug. 1 to Sep. 20, 2018, to train the model. This first interval includes a reaction system and compressor operation within the normal condition containing 2,377 data samples. The test set consists of twelve days before the compressor stops on Oct. 1, with 553 data samples. This approach aimed to use a period within normal limits for training the model, and with that, it identified a threshold allowing the recognition of the occurrence of anomalies. With the model built and trained, and the threshold defined, the second part of the dataset identified anomalies through the model.

The production of SM occurs through reactors with fixed-bed catalyst, overheated steam, and negative pressure to facilitate the reaction of ethylbenzene converted into SM. However, the reaction system does not convert EB entirely to SM because of the by-products (DIMIAN; BILDEA, 2019). The negative pressure is of paramount importance to minimize the generation of other products that are not SM. For it, the production uses a compressor to reduce the pressure of the reaction system (DIMIAN; BILDEA, 2019). The compressor also has a second important function regarding the gases removed from the reactors, the most significant percentage being hydrogen. They serve as fuel for the superheating furnace, which provides energy to promote the reaction. Figure 33 presents in a simplified way the SM reaction process.

According to what was presented, the compressor has paramount importance in SM production since the plant performance becomes compromised in possible compressor failure and the consequent stoppage. The generation of by-products will increase because the reaction will occur under positive pressure, in addition to the need for alternative fuel in overheating furnaces (DIMIAN; BILDEA, 2019).

The maintenance and production teams put efforts to keep the compressor in operation, considering that the impacts on the company's results are significant in case of a failure in the equipment. Therefore, an approach is necessary when it is possible to act in an anticipatory manner and avoid a compressor failure. Traditionally the production team's role is to control the variables linked to the production trend, such as flow, pressure, temperature, and manufacturing. On the other hand, the maintenance team monitors variables related to the equipment, such as; vibration, axial and radial displacements, and motor current. Considering that SM production occurs under high temperatures, and the generation of PS happens with SM heating, there is an increased risk of polymerization in the compressor (DARVISHI; RAHIMPOUR; RAEISSI, 2019). Traditionally the control of polymerization is carried out using vibration and displacement sensors. However, early detection of failures of this magnitude can prevent a system shutdown by increasing reliability.

After presenting the source of the data and selection criteria, the following section consists of applying data mining techniques so that the system's reliability is high and that the operational interventions occur with greater assertiveness.

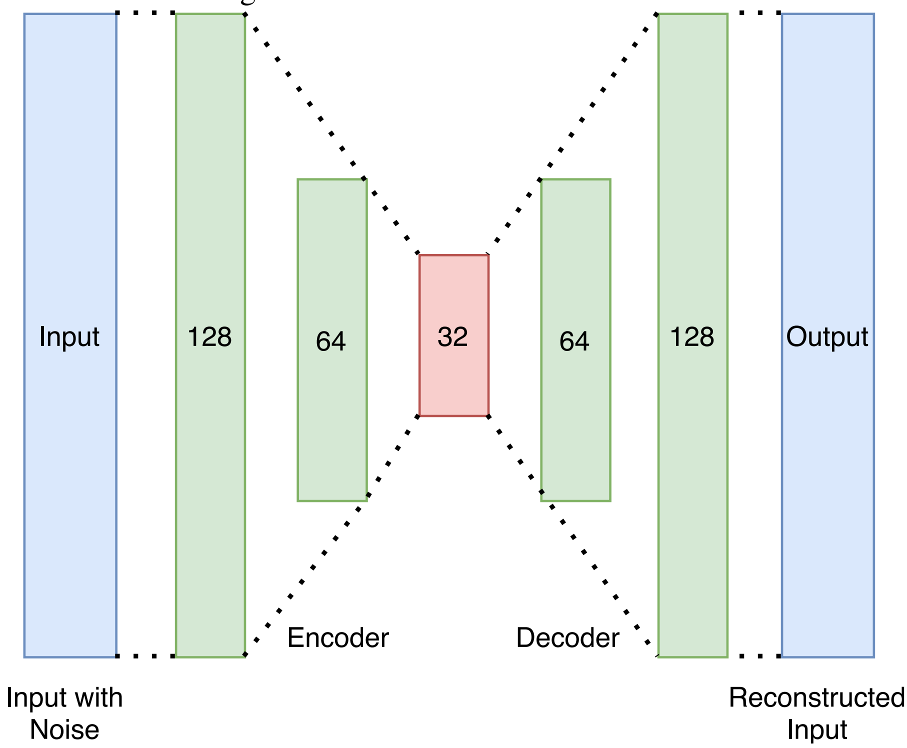
5.2 Tier 2: Data-driven approaches

For the models to perform better, data quality is an essential factor. Therefore, it is imperative to select the most critical variables according to the analysis. A FI approach that best represents the attributes is a fundamental step in modeling, focusing on the quality of the model input data (AREMU et al., 2020). A benefit of selecting variables is the possibility of reducing redundancy among features before feeding the ML model (WANG et al., 2018). As a result, it is possible to improve the prediction models' performance, reduce the overfitting, increase precision, spend less computational resources, and many further benefits (FERNANDES et al., 2019).

Among the powerful techniques for FI, the AE models are an efficient alternative. The study by Wang et al. (2018) introduced variants of AE models to assist in discovering features that provided better performance for the model. In the article by Kong et al. (2020), they applied a Deep Autoencoder model to select the most powerful features. The paper conducted by Li et al. (2020) proposed a method using CNN to identify the equipment's degradation.

AE is an unsupervised learning algorithm to extract characteristics from input data without initial label information. It mainly consists of two parts, including encoder and decoder. In Figure 34 a simplification architecture of AE is presented (LI et al., 2020; WANG et al., 2018; SHAO et al., 2017). The encoder can perform data compression, especially when dealing with high dimensional input, by mapping the input to a hidden layer.

Figure 34 – Autoencoder architecture.



Source: Prepared by the author.

In the first part there is an input $\mathbf{x} = \{x_1, x_2, ..., x_n\}$, which is converted into hidden layers, encoder $\mathbf{h} = \{h_1, h_2, ..., h_m\}$, through an activation function that can be symbolized as follows:

$$\mathbf{h} = \phi_{act}\left(\mathbf{W} + \mathbf{b}\right) \tag{5.1}$$

where ϕ_{act} denotes the activation function and W is the weight matrix, b is a bias vector, and the dimension h_m is usually smaller than x_n.

In the second part, decoder can reconstruct the input approach, where the hidden layer $\mathbf{h} = \{h_1, h_2, ..., h_m\}$ is converted into an output $\hat{\mathbf{x}} = \{\hat{x}_1, \hat{x}_2, ..., \hat{x}_n\}$ through an activation function, which can be expressed as follows:

$$\hat{\mathbf{x}} = \phi_{act}\left(\mathbf{W'h} + \mathbf{b'}\right) \tag{5.2}$$

where W' is the weight matrix and b' is a bias vector. The parameters of the AE are trained by minimizing the reconstruction error (loss function) between the input and output using Back-propagation (BP) algorithm (RUMELHART; HINTON; WILLIAMS, 1986). As an example of a loss function, we mention the Mean Squared Error (MSE), that can be symbolized as the Equation 5.3 (KONG et al., 2020):

$$L\left(\theta\right)_{AE} = \frac{1}{N}\sum_{i=1}^{N}(x - \hat{x})^2 \tag{5.3}$$

where i varies to N $\{i = 1, 2, ..., N\}$, N represents the number of samples, and θ is the parameter set which can be indicated as $\theta = \{W, b, W', b'\}$.

In the research carried out by Wang et al. (2018), they made a comparison highlighting the characteristics of CNN and AE models. The authors validate the benefits of using CNN algorithms in AE architectures (KONG et al., 2020; MAO et al., 2020; REN et al., 2018). The essential characteristics of the CNN need fewer parameters, which in some cases reduce the time and computational resources for training (KONG et al., 2020; LECUN; BENGIO; HINTON, 2015). This feature makes it possible to test the combination of different hyperparameters and thereby improve the accuracy of the models (RUMELHART; HINTON; WILLIAMS, 1986; LECUN; KAVUKCUOGLU; FARABET, 2010; WANG et al., 2018). Thus, when the data to be analyzed contains high dimensionality and noise, the use of CNN in AE architecture is appropriate (WANG et al., 2018).

Multi-layer CNNs are experts at learning complex, high-dimensional, nonlinear mappings of large amounts of data (LECUN et al., 1998; LI et al., 2020; KIM, 2014). Convolution models can also assist in removing noise from the input data (LI et al., 2020). Due to the ability to identify and maintain the most relevant characteristics, researchers have been widely applying them in several fields, such as image and speech recognition, among other applications (LECUN; BENGIO; HINTON, 2015; KIM, 2014; XIA et al., 2018; LECUN; KAVUKCUOGLU; FARABET, 2010; ZEILER; FERGUS, 2014).

CNN assumes local dependencies in input measurements, like local receptive fields in an image. In this study, the measures are dependent and structured (CHE et al., 2020). For example, an increase in the load flow, represented by variable X1, will impact all other measurements, as it is an integrated operating system (KANG; CATAL; TEKINERDOGAN, 2020). Therefore, using a CNN model is a viable solution for the present study.

In general, the purpose of CNN is to identify the new data's generic features, using a specific configuration of convolutional and grouping layers. The convolutional layer is named based on its operation to convince filters with raw input data to generate different resources. The grouping layer extracts the most representative local resources by applying a sliding window to the previous layer's features. A summary construction of CNN is presented as follows (LI et al., 2020; CHEN et al., 2020).

The input data with N in length is represented by $x = [x_1, x_2, ..., x_N]$. Each x_i can be one-dimensional, two-dimensional, or even higher dimensional. In this study, the relationship between process data and compressor condition monitoring (vibration and displacement signals) and compressor failure is considered. The input data, in this case, is one-dimensional. The convolution operation in each feature map is represented as:

$$\mathbf{a} = g\left(W^T \times x_{i:i+K_{L-1}} + b\right), i = 1, 2, 3, ..., N - K_L + 1 \tag{5.4}$$

which W is the filter with length K_L, b is the bias and g is the activation function. In this work as activation function we selected the Rectified Linear Unit (ReLU) i.e., $g(x) = max(0, x)$ (CHEN et al., 2020). With the output, it is possible to understand the most important characteristics learned concerning the input data. After layers of convolution, CNN uses a pooling to summarize the previous layer's output in the same map of features. This operation makes it possible to reduce the dimensions and consequently reduce the overfitting of the model. The features resulted after convolution layers are expressed as $a = [a_1, a_2, ..., a_n]$. Then, the function that defines a pooling is:

$$\mathbf{F} = \{max\{a_{i:i+P_L-1}\}, i = sl - s + 1, l = 1, 2, 3, ...\} \tag{5.5}$$

where P_L is the window size, and s is the stride of pooling function.

In ML, the models' interpretability is as crucial as the forecasting accuracy for most problems (BAPTISTA et al., 2018). As in health research, where understanding the model is of paramount importance, in industrial applications, the results of the model must be interpretable (QI, 2012; WANG et al., 2018). To be used in industrial applications, DL analytical solutions, despite their enormous potential, need to be understood by decision-makers. Otherwise, stakeholders can ignore the recommendations and decisions generated (WANG et al., 2018). DL models are generally considered a black box due to the inherent complexity, especially if the network becomes deeper (MAO et al., 2020; KONG et al., 2020).

It is difficult to interpret and explain the computation and reproduction procedures of the features. Thus, as suggested in the study by (WANG et al., 2018), to complement and facilitate to understand the abstract resources learned by the network, the fusion with other ML techniques can contribute to a more effective model.

Generally, there are three groups of FI algorithms: filter, wrapper, and embedded models (TSAI; SUNG, 2020). The wrapper's significant advantage and embedded compared to filter models is that they take into account the effects of the FI subset with predictor algorithm performance (HU et al., 2018; KASONGO; SUN, 2020).

This research proposes a model in which it provides subsidies for maintenance and production managers to support evaluation. That said, the model must also be interpretable and friendly to decision-makers and provide accuracy to address the problem. Otherwise, it may not be helpful for the purposes due to the limited understanding of the model's results (QI, 2012; WANG et al., 2018). Considering that a CNN model identifies anomalies and brings a challenge to understand the results, a complementary approach is necessary.

5.2.1 Anomaly detection

We subdivide into three subtopics the framework elaboration that supports the PHMS. The first subsection below discusses the AE model for anomaly detection. The second subsection presents the model development for FI of the main variables related to the failure. Finally, the third discusses the activities that make up the steps for forecasting the system's RUL.

5.2.1.1 CNN Autoencoder

Architecturally, in a simplified way, an AE is a feedforward neural network. Similar to single-layer perceptrons, AE resembles a multilayer perceptron (MLP), where it contains an input layer, an output layer, and one or more hidden layers. However, the output layer has the same number of neurons as the input layer to reconstruct the input data.

An essential step in building the model consists of pre-processing the data. Due to the difference in scale among values of features, we normalized the data before feeding the model. This action aims to avoid a possible model bias. For this activity, we used the Scikit-learning[1] pre-processing tool to scale the input variables the model. Therefore we used the *MinMaxScaler* function and put the data in a range of [0,1].

After pre-processing the data, the AE model is built using the hyperparameters shown in Figure 35. Moreover, in Figure 35, the encoder structure is in gray, the decoder structure is in yellow, and the convolution layers by dashed lines. Finally, for implementation, we apply the Python programming language with packages for developing ML models and Tensorflow and Keras frameworks to build the CNN AE Network.

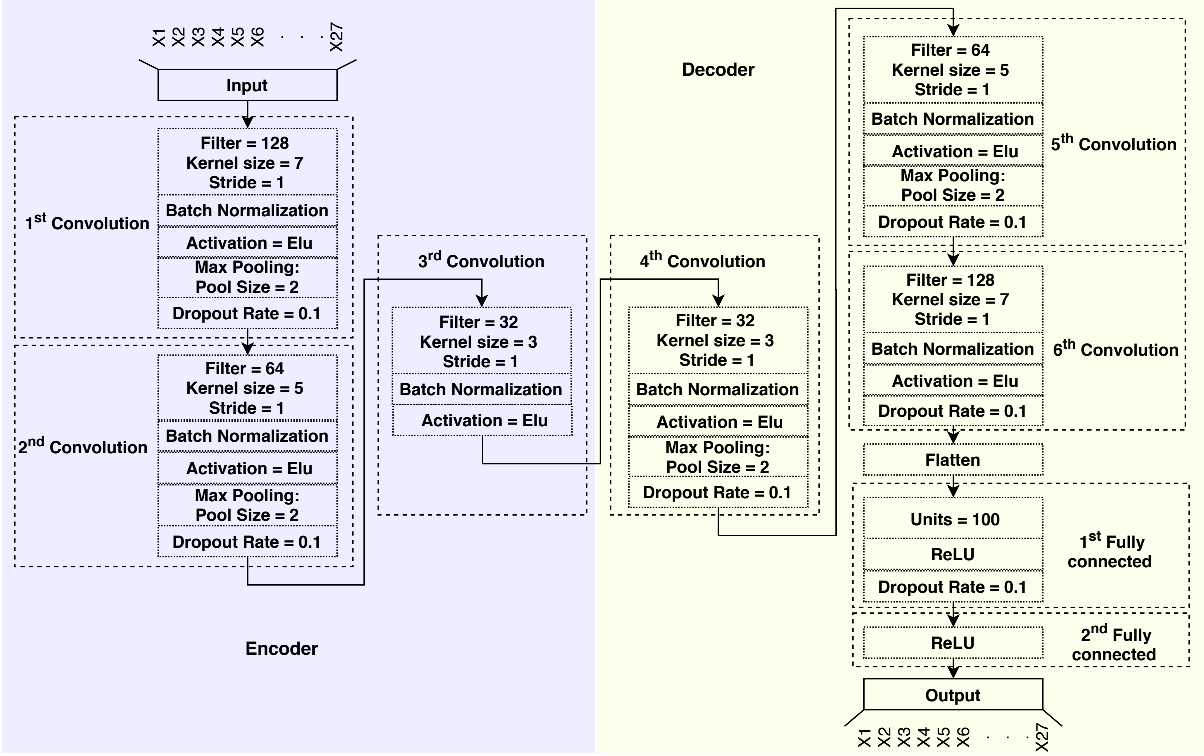
In the convolution layers, used the Exponential Linear Unit (ELU) (PEDAMONTI, 2018; CLEVERT; UNTERTHINER; HOCHREITER, 2016) activation function, Batch Size of 32, and 150 Epochs to train the model since it presented an enhanced performance in the analysis. In the same way, to avoid overfitting used the Callback[2] function of Keras and EarlyStopping metrics and 15% of the training data for validation and calculated the loss.

In identifying anomalies, the purpose is to use the AE network to compact the control and

[1] MinMaxScaler

[2] https://keras.io/api/callbacks/

Figure 35 – The proposed convolutional neural networks Autoencoder model for anomaly detection.



Source: Prepared by the author.

sensor variables' data readings. Thus, it is possible to obtain a reduced dimension that represents the interactions among the monitored variables. In the proposed model, the AE trains with data $\{x(1), x(2), ..., x(i)\}$ that describes the normal operating condition to compress and then reconstruct the input variables $\{\hat{x}(1), \hat{x}(2), ..., \hat{x}(i)\}$. Moreover, the error expressed by the Equation 5.6 becomes reduced, considering each data sample $x(i) \in \mathbb{R}^n$ is described by n different samples (HAROON, 2017).

$$MSE(x, \hat{x}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{5.6}$$

With dimensionality reduction, the network learns the interactions among several variables and must reconstruct the original ones in the output layer. The main idea is that as the monitored equipment presents some degradation, it might affect the interaction among the variables, such as temperature, pressure, vibration, flow rate, and liquid level. When this happens, there will be an increase in the network reconstruction error with the input features. By monitoring the reconstruction error, it's possible to estimate the monitored equipment condition, considering that the error tends to increase as the equipment degrades.

The proposed method used the error probability distribution of the input data reconstruction to identify whether a data point is normal or abnormal, higher than the threshold. The definition of the limit depends on the research problem. This study defines the standard limit based on a real data set from a petrochemical company.

The MSE distribution for reconstructing the region considered normal operating conditions can then limit value to consider it abnormal. From the Reconstruction Loss (RL), we can, for example, define a loss > 0.15 as an anomaly. The evaluation of the method to detect the equipment degradation then calculates the RL and compares the data points to the threshold. As a result, they are abnormal or normal.

This subsection aimed to present the methodological procedures applied to identify an anomaly in the system. In the following subsection, the objective is to offer the approaches in developing the model to determine which variables significantly contribute to the anomaly detected.

5.2.1.2 Model to FI

We used a supervised learning model to select the variables that contributed to the system failure. In supervised learning, we needed to present the model with a feature to be predicted, so we used the AE model's RL. In this step, we used the decision tree-based algorithms Random Forest Regressor (RFR)[3] and XGBoost[4]. We chose these algorithms due to their ability to handle high-dimensional data and the consistent results presented in different applications (LUNDBERG et al., 2020; DENG et al., 2021; MITCHELL; FRANK, 2017; BUITINCK et al., 2013; GEURTS; ERNST; WEHENKEL, 2006). As a metric for model evaluation, we defined that the values of R^2 should be greater than > 90% on test data.

For model training, we employed the exhaustive search technique of the GridSearchCV method. For defining the hyperparameters of the models, we used the GridSearchCV method available in the API Scikit-learn[5]. GridSearchCV involves testing all possible combinations of hyperparameters and selecting the one with the best performance in a search space (BUITINCK et al., 2013).

In order to present the model's interpretation, we show the results of the SHAP method applied in this study in Section 6.2.

5.2.1.3 FI with SHAP

In this step, we use the RL and threshold developed in the previous session and already published in Souza et al. (2021). However, this study differs regarding the resource identification phase. In order to do this, we used the SHAP method (SHapley Additive exPlanations) (LUNDBERG; ERION; LEE, 2018; SAYRES et al., 2019). The SHAP technique is a tool for interpreting ML models.

[3] Random Forest Regressor
[4] XGBoost
[5] GridSearchCV

Several methods have been proposed to explain the ML model predictions. The SHAP method has shown great results in explaining the results in various applications, including in health care, where interpreting the model's results is a fundamental requirement. To this end, the SHAP method uses game theory to calculate each feature's contribution to the model's result. The SHAP method determines the most relevant local characteristics of each feature in the model's result (CAMPBELL et al., 2022; AGIUS et al., 2020; ZHANG et al., 2020; SAYRES et al., 2019; LUNDBERG; ERION; LEE, 2018; YANG, 2021; ARIZA-GARZON et al., 2020).

There are three key ideas presented in the SHAP. The first is that the explanation of an ML model is the sum of the contributions of each feature. The second is that the contribution of each feature is the difference between the model prediction with and without the feature. The third is that the contribution of each feature is the difference between the model prediction with and without the feature, weighted by the feature's value (CAMPBELL et al., 2022; SAYRES et al., 2019; LUNDBERG; ERION; LEE, 2018; YANG, 2021). Therefore, we propose the following three steps for identifying the main variables related to the fault and consequent system downtime.

1. In terms of overall interpretability, the sum of the SHAP values can reveal each predictor's positive or negative impact on the outcome. Similar to the variable importance graphic, but with the additional ability to display whether or not each variable positively or negatively correlates with the outcome.

2. The second type is local interpretability and refers to the fact that each observation has unique SHAP values. We can articulate the reasoning behind a prediction and the roles played by the many predictors. The results of traditional variable significance methods are shown only globally and not separately for each instance. Because of local interpretability, we can isolate the effects of each element and compare them.

3. And third, unlike other approaches that rely on surrogate models, such as linear regression or logistic regression, SHAP values can be derived for any tree-based model. Accessible on a global scale, a SHAP value graphic can illustrate positive or negative associations between predictors and the outcome variable.

This study uses SHAP to explain tree-based models, such as random forests, decision trees, and gradient-boosted trees, which are popular nonlinear predictive models (LUNDBERG et al., 2020). Therefore, for us to use SHAP for FI, we first need to train an ML model. In this sense, the following section presents the model used.

5.2.2 RUL - Remaining Useful Life

PHM predict future behavior regarding the existing operating state and schedule maintenance activities required to preserve a system's health. In this sense, the following subsection presents the step to define the system's HI, which will be the basis for estimating the system's RUL.

5.2.2.1 Defining the System's Health Index

In this study, we normalized the RL output from AE and generated the time series representing the system's Health Index (HI). In order to do so, we applied Equation 5.7 (SCHWARTZ et al., 2022).

$$HI = 1 - \left[\frac{X_i - min(X_i)}{max(X_i) - min(X_i)}\right] \tag{5.7}$$

Where X_i relates to the i value of RL over time. That is, according to Equation 5.7, the system's HI is calculated from the RL, where the values are normalized between (0 and 1). Thus, the value closest to 1 indicates that the system is healthy, while the value close to 0 indicates that the system is more degraded with compromised health (SCHWARTZ et al., 2022; ZOU et al., 2022; YANG et al., 2022).

5.2.2.2 Data Preparation

Data preparation is an important step for the success of any ML project. We prepared the data for training and testing the models in this step. As input, the models will receive a previous data set, and, as output, they will produce a new predicted observation. To this end, we divided the HI time series into individual instances to train the model. As an example, we consider the following univariate data series: $[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]$.

The series can be divided into a collection of input/output patterns, with a one-time window acting as input and three-time steps working as output for a one-window prediction. The following sequence shows the division of the series into size 3 windows. The first input window is [1, 2, 3], and the corresponding output window is [4, 5]. The second input window is [2, 3, 4], and the corresponding output window is [5, 6]. And so on, as shown in Table 5.

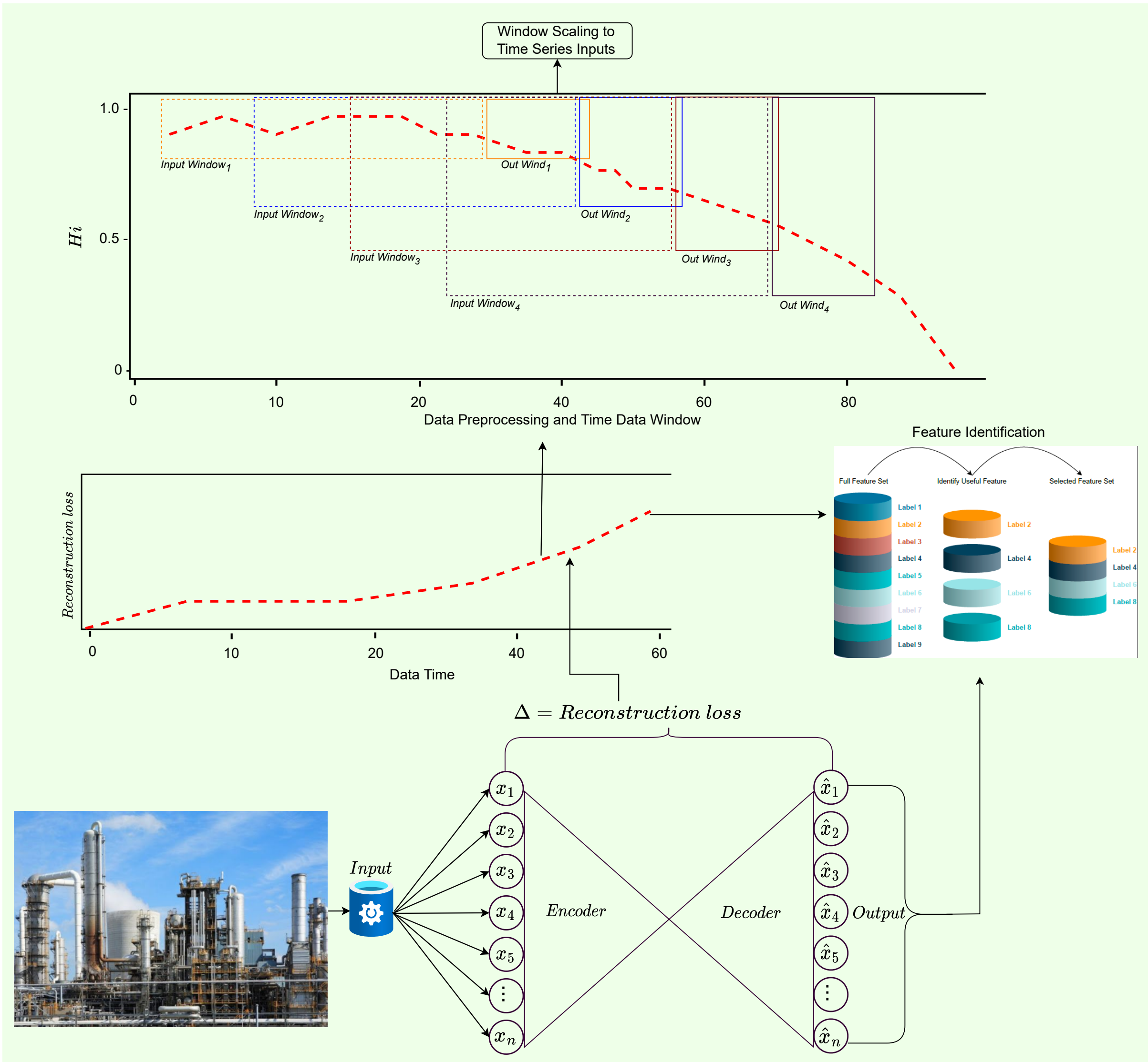
Table 5 – Input and output example for a multi-step time series forecast.

Input:	Output:
[1, 2, 3]	[4, 5]
[2, 3, 4]	[5, 6]
[3, 4, 5]	[6, 7]
⋮ ⋮ ⋮	⋮ ⋮

Source: Prepared by the author.

We grouped the HI measurements over a three-hour average because we considered there were no significant variations in the series over this interval. The training windows consider intervals of 1.5 days (36 hours), while the testing window considers one day (24 hours). Figure 36 presents in simplified form the steps involved in data preparation. Considering the collection, the AE model derivating the RL, which are inputs for the FI and HI steps, proceeds to divide the time windows to feed the models for predicting the system's RUL

Figure 36 – Data preparation for RUL prediction and FI.



Source: Prepared by the author.

In the next subsection, we present the Benchmarking step of the DL models for estimating the system's RUL.

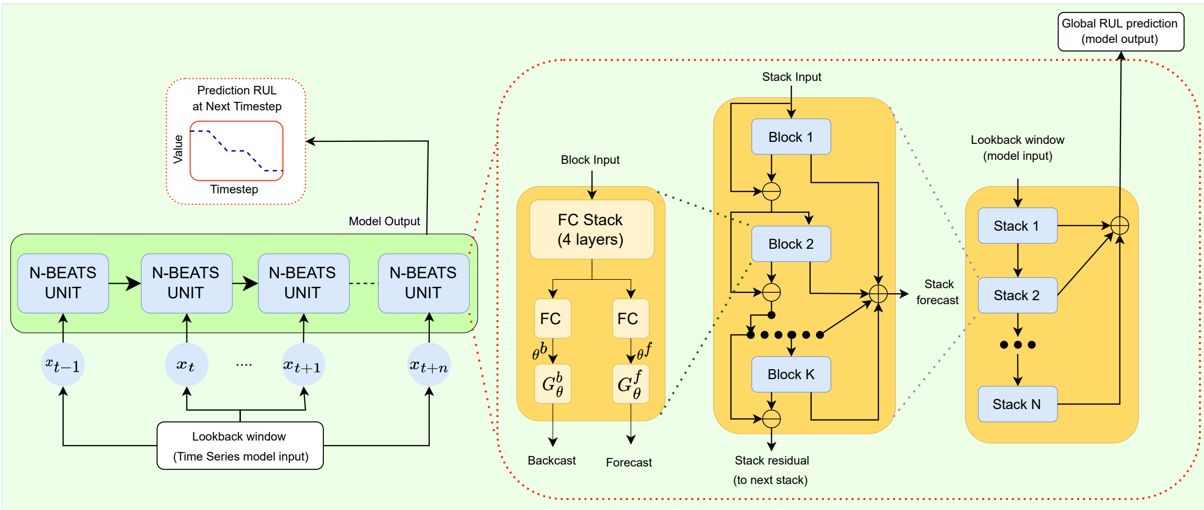
5.2.2.3 Building the models

In the following subsection, we address the main differences between the architectures presented in Section 4.3.3.1 concerning the purpose of this research.

5.2.2.4 N-BEATS architecture

Each N-BEATS block is mathematically described as a series of fully connected layers with a prediction/backcast bifurcation at the end. A block "removes the signal that can approximate well". Then, the block focuses on the remaining error the previous blocks failed to correct. Each block makes a partial prediction based on the local time series. The stack gathers incomplete predictions into its blocks and delivers them to the next stack. Stacks consider a lookback window to find non-local trends in time series. Finally, the partial forecasts are combined into a model-level global prediction, as shown in Figure 37 (ORESHKIN et al., 2019).

Figure 37 – Neural Basis Expansion Analysis For Interpretable time series Forecasting Architecture.



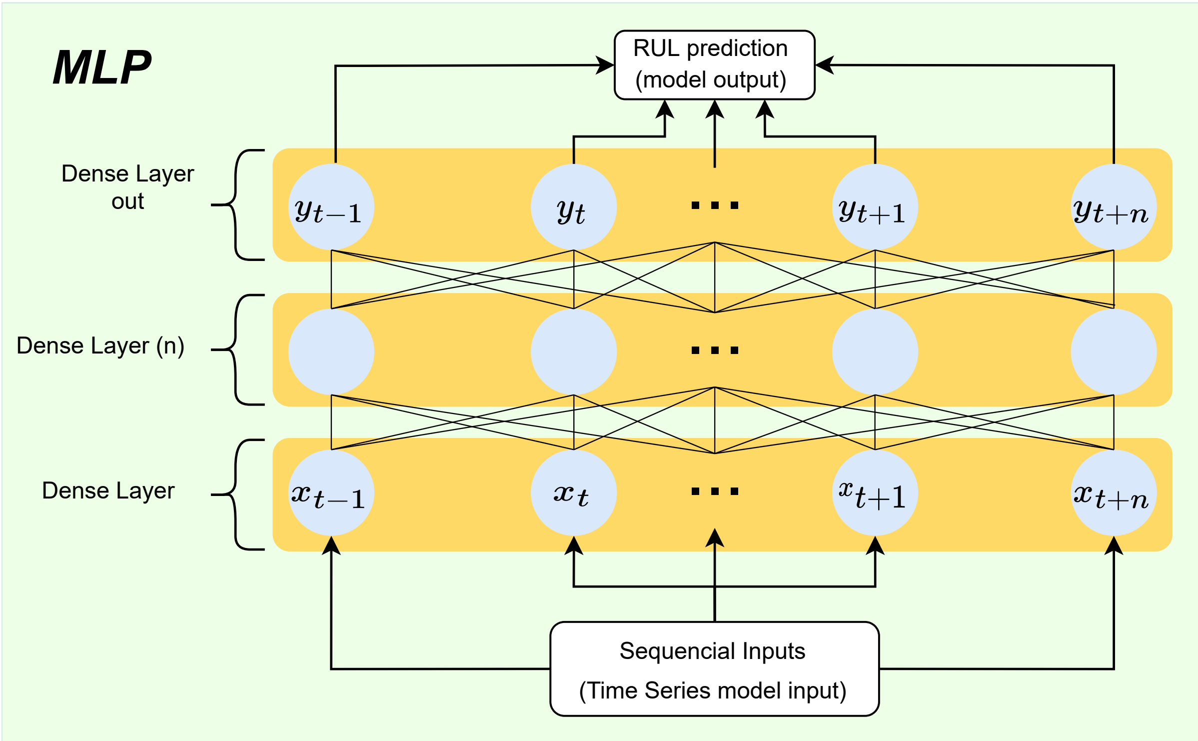
Source: Prepared by the author adapted from (ORESHKIN et al., 2019).

Given the results obtained in different time serie applications, N-BEATS has drawn researchers' attention in several fields, such as: sewage treatment prediction (ZHANG; SUZUKI; SHIOYA, 2022); atmospheric drag on spacecraft dynamics (STEVENSON et al., 2022); stock market prediction (SINGHAL; MATHEW, 2022); and electricity demand (ORESHKIN et al., 2021). Given these scenarios, we bring this architecture to compare the RUL's prediction results with traditional models

5.2.2.5 MLP architecture

The MLP is a simpler neural network since it approximates a mapping function from input to output variables. This ability is important for time series. In this regard, it can handle the noise often present in time series. Figure 38 presents the MLP architecture in simplified form.

Figure 38 – MLP architecture.



Source: Prepared by the author.

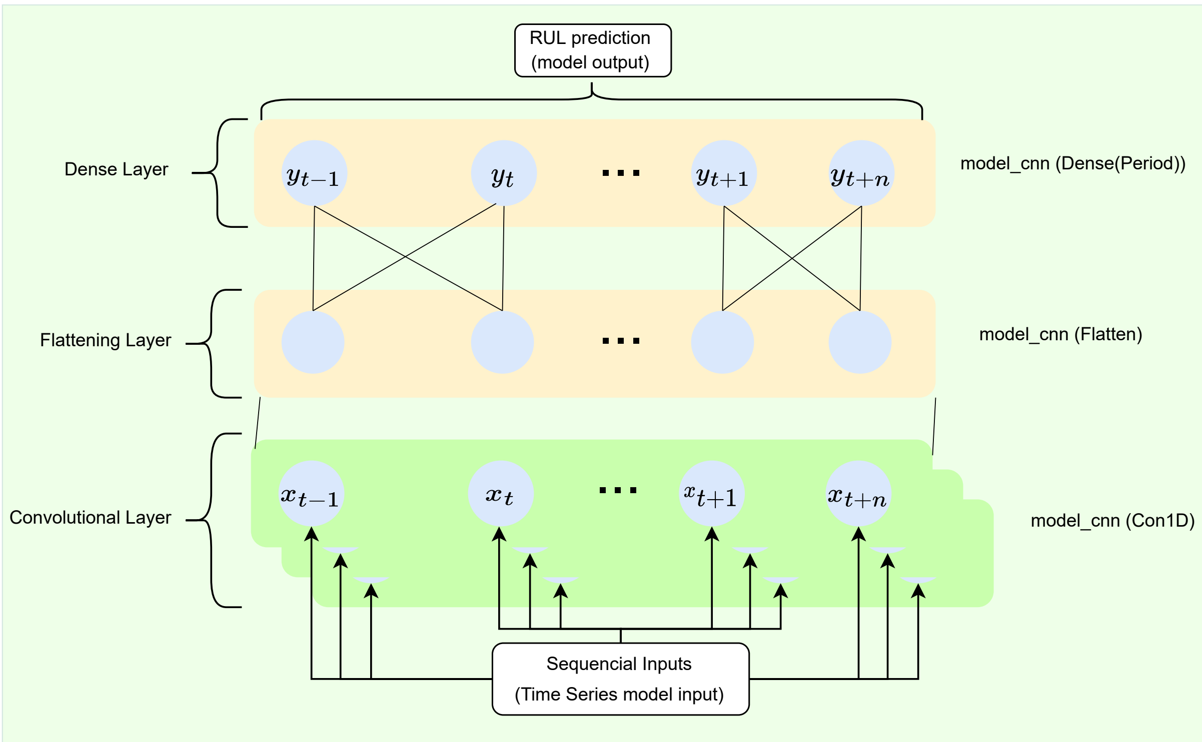
The MLPs are resistant to noise in the input data and the mapping function. Moreover, they allow learning and prediction even with missing values. Another important feature is that they make no assumptions regarding the mapping function and can learn linear and non-linear connections. These characteristics are important in time series processing and provide robust applications, especially in prediction (GOODFELLOW; BENGIO; COURVILLE, 2016).

For quality predictions to be made, the MLP needs a meaningful input mapping to learn the most important time series characteristics. On the other hand, MLP networks can have any number of inputs and outputs in the mapping function. Thus, they can be applied to multivariate time series predictions and still perform multi-period predictions (GOODFELLOW; BENGIO; COURVILLE, 2016; SUTSKEVER; VINYALS; LE, 2014; GRAVES; SCHMIDHUBER, 2005). Therefore, MLP networks can be effective for time series prediction, especially regarding this study's objective.

5.2.2.6 CNN architecture

In this study, we adopted the CNN comprising two convolutional layers to extract spatial characteristics and the fully connected neural network to obtain a regression model (LECUN et al., 1998). CNN were intended to process image data effectively. They have provided great results in demanding computer vision challenges, including image categorization and object localization, image captioning, and several applications in the computer vision field. They can learn and map characteristics from raw data for TS prediction. A CNN model can filter and refine a time series as a one-dimensional image (YANG et al., 2015; LECUN et al., 1998; LECUN; BENGIO et al., 1995). Figure 39 presents an abstraction of the CNN architecture.

Figure 39 – Simplified architecture of a CNN.



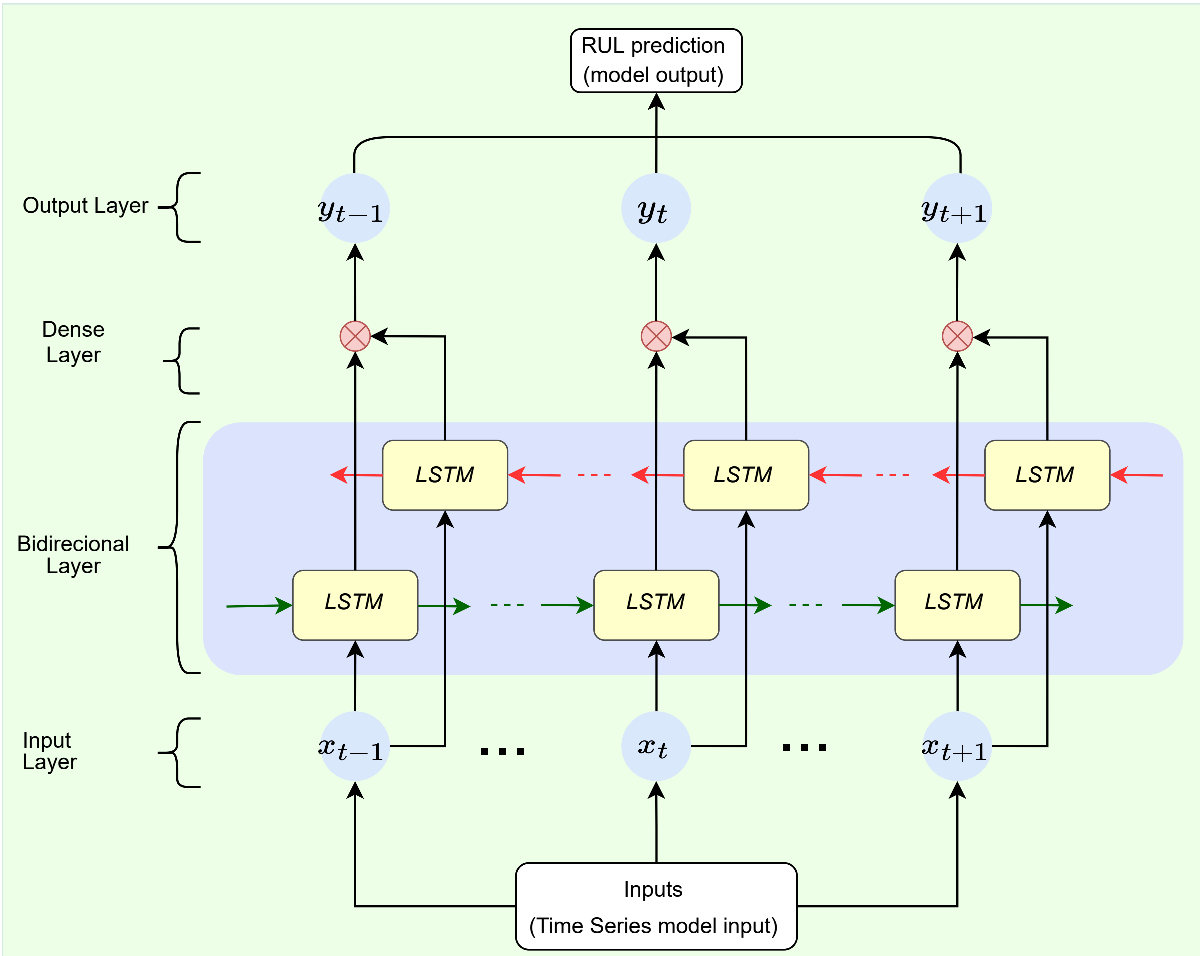
Source: Prepared by the author.

The CNNs obtain the advantages of Multilayer Perceptrons (MP) for time series prediction, including multivariate series and learning complex functional connections, with the benefit that the model does not learn directly with a time lag. Instead, the model can learn from an input sequence the most important representation for prediction (YANG et al., 2015; LECUN et al., 1998; LECUN; BENGIO et al., 1995). The CNN has shown consistent results when applied to identifying anomalies, such as bearing failure (KONG et al., 2020) and estimating RUL in industrial components (ZHANG et al., 2021).

5.2.2.7 LSTM architecture

LSTM is a type of RNN, a special neural network designed for sequence problems. Thus, given a standard feedforward MLP network, an RNN can be understood as adding loops to the architecture. For example, each neuron can pass its signal forward (laterally) in a given layer and transfer it to the next layer. The network's output can feedback as an input to the network with the next input vector, and so on. Recurrent connections add state or memory to the network and allow it to learn and take advantage of the ordered nature of time series (HOCHREITER; SCHMIDHUBER, 1997). Figure 40 presents an abstraction of the LSTM architecture.

Figure 40 – Simplified architecture of LSTM.



Source: Prepared by the author.

LSTM can handle time series where feedforward networks do not perform satisfactorily using fixed-size time windows. In addition to the general benefits of using neural networks for sequence prediction, temporal dependence of the data benefits RNNs. In other words, the network receives one observation of the sequence at a time as an input, which contributes to learning from previous relevant observations and, thus, the predictions of the analyzed sequence. Because of this ability to learn long-term correlations in a sequence, LSTM networks eliminate the need for a specific time window and can accurately model complex multivariate sequences (HOCHREITER; SCHMIDHUBER, 1997).

However, LSTM does not adequately capture the non-stationary information of a time series. Another LSTM limitation is dealing with autoregression with important information within a small time series window. In this case, an MLP network may be more efficient than LSTM (MA; MAO, 2021; SUTSKEVER; VINYALS; LE, 2014). The LSTM's limitation in dealing with nonstationary data may not even be the best option for predicting RUL (MA; MAO, 2021). However, since LSTM has shown consistent results in time series prediction, especially in classification problems, such as anomaly detection (KONG et al., 2020), RUL estimation in industrial components (ZHANG et al., 2021), and engine failure prediction (MA; MAO, 2021), we bring in LSTM for RUL prediction and analyze the result by benchmarking the models.

5.2.2.8 GRU architecture

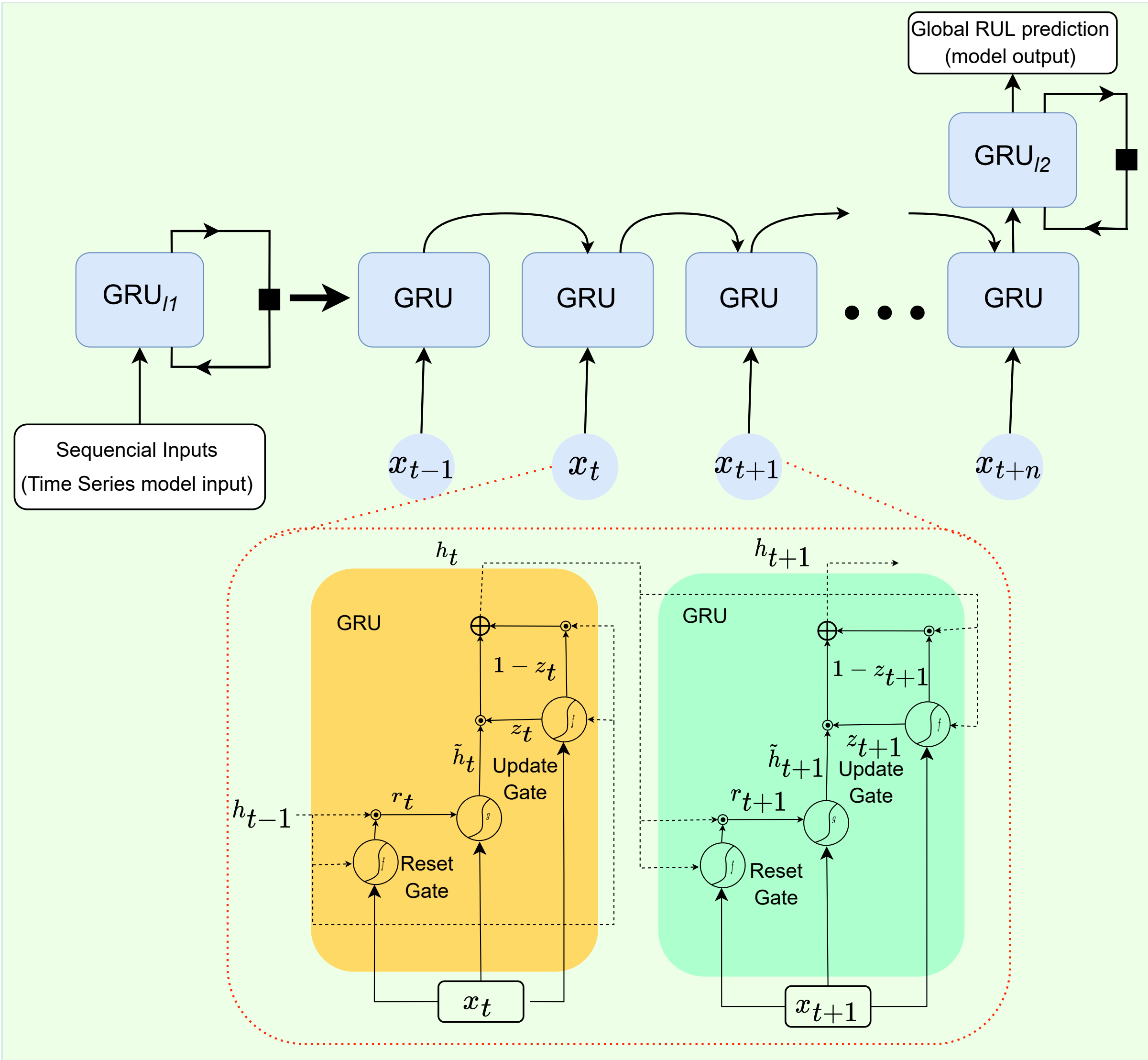
The GRU was proposed by (CHUNG et al., 2014) to allow each recurrent unit to record dependencies on multiple time scales adaptively. The GRU, similarly to the LSTM unit, includes gated units that control the information flow within the unit. However, it requires separate memory cells.

The most remarkable similarity shared by LSTM and GRU is the additive component of their memory cell upgrade. The classic RNN changes the activation or contents of a unit with a new value calculated from the current input and the previous hidden state. However, over time, the LSTM and GRU units retain the most relevant information. Figure 41 presents an abstraction of the GRU architecture.

Although they are RNNs, LSTM and GRU have subtle differences. Regulated disclosure of memory contents is a feature of the LSTM unit that GRU does not have. The output port on the LSTM unit controls the amount of memory content visible to or used by other units on the network. On the other hand, GRU exposes all its material without any restriction. Another distinction is the placement of the entry gate or the reset gate in the GRU case. When updating the candidate activation, the GRU controls the information flow from the previous activation using a recurrent memory update gate. Applications of RNNs, such as asset integrity monitoring, have shown promising results (WANG et al., 2017; ZHANG et al., 2021). In this regard, we use GRU to predict RUL and analyze the result by benchmarking the models

After the presentation of the neural network models, next, we present the methodology for evaluating the models.

Figure 41 – Simplified architecture of GRU.



Source: Prepared by the author.

5.2.2.9 RUL model performance metrics

We used different evaluation methods to assess the performance of the DL model predictions. Therefore, we used the Mean Absolute Error (MAE) Equation 5.8, MSE Equation 5.9, Mean Absolute Percentage Error (MAPE) Equation 5.10, and the Root Mean Squared Error (RMSE) Equation 5.11.

The MAE is defined as the mean of the absolute value of the errors, Equation 5.8.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|RUL_i - RUL_{\hat{y}_i}| \tag{5.8}$$

The MSE is defined as the mean squared errors, Equation 5.9.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(RUL_{y_i} - RUL_{\hat{y}_i})^2 \tag{5.9}$$

The MAE is more robust toward outliers, while the MSE assigns a higher penalty to larger errors. The MSE is a more popular error measure than the MAE because the MSE assigns a larger penalty to larger errors. It is also easier to interpret, as the MSE is the variance of the errors (CHEN, 2021).

The MAPE is defined as the average of the absolute value of the errors divided by the actual value, Equation 5.10.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|RUL_{y_i} - RUL_{\hat{y}_i}|}{RUL_{y_i}} \tag{5.10}$$

The RMSE is defined as the square root of the MSE Equation 5.11.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(RUL_{y_i} - RUL_{\hat{y}_i})^2} \tag{5.11}$$

We applied the Coefficient of Determination R^2 to evaluate the quality of the predictions. The R^2 is a measure of how well the predicted values fit the observed values. The R^2 is a value between 0 and 1 for regression data, where 0 means the model does not explain the variability of the data around its mean, and 1 means the model explains all the variability of the data around its mean. The R^2 is defined by Equation 5.12 (CHEN, 2021).

$$R^2_{RUL} = 1 - \frac{\sum_{i=1}^{N}(RUL_{y_i} - RUL_{\hat{y}_i})^2}{\sum_{i=1}^{N}(RUL_{y_i} - RUL_{\bar{y}_i})^2} \tag{5.12}$$

where i varies for N $\{i = 1, 2, ..., N\}$, where N represents the number of samples and RUL_{y_i} is the actual value and $RUL_{\hat{y}_i}$ is the predicted value of RUL, respectively. The total number of RUL true targets in the respective test set N.

The next section addresses the results obtained through the methodological procedures applied in a real case study.

5.3 Final Remarks

This chapter presented the methodological procedures for implementing the framework, evolving from the PHM, focused on specific equipment, to the PHMS. Initially, we approach the stages of data collection and treatment. To contextualize this step, the environment in which the case study will be carried out was used. Next, ML activities were treated. Initially, the details and importance of the AE for anomaly detection were explained, followed by the FI with the model validation criteria, and finalized with the RUL prediction stage. For RUL prediction, different DL architectures were explored and what draws attention is N-BEATS, which has shown consistent results for series predictions with noise and non-stationary.

For validation of the framework and methodological procedures, the next chapter will present the case study putting into practice the steps explained in this chapter.

6 RESULTS

This chapter presents the results of the framework through a case study applied in the petrochemical industry. This sector was chosen due to the digitalization maturity and to be intense in industrial automation and the researcher's knowledge in this operation, in addition to the relevance of this industry in revenue and complexity operation. In the petrochemical industry, the impacts of unplanned downtime are often severe for the organization, as financial losses tend to increase beyond damaged equipment.

For the development of the chapter, the problem to be treated is initially presented. Next, we present the development stages guided by the proposed framework and methodology. Finally, we conclude with a discussion of the results.

6.1 Case study

I4.0 provides massive data collection. As a benefit, it can provide the current state of the machines or processes through Big Data Analytics (AYDEMIR; ACAR, 2020; SAHAL; BRESLIN; ALI, 2020; RAUCH; LINDER; DALLASEGA, 2020). This approach provides support for the decision of business managers in the intelligent manufacturing process (ATZORI; IERA; MORABITO, 2010; DIEZ-OLIVAN et al., 2019; LU, 2017; ROY et al., 2016; RUBMANN et al., 2015; YOKOYAMA, 2015; SAHAL; BRESLIN; ALI, 2020; AYDEMIR; ACAR, 2020).

Process industries, such as petrochemicals, control parameters are usually liquid level, pressure, temperature, speed of pumps and compressors (TEWARI; DWIVEDI, 2019; MIN et al., 2019). In many cases, the control parameters monitor the quality of the product, the facilities' safety, and the people involved in these activities (CHENG; YAO; WU, 2013; MUTLU; ALTUNTAS, 2019). However, to maintain stability and continuity of operations, the equipment must be fully operational, with high standards of reliability (NAKAYAMA; SPÍNOLA; SILVA, 2020). Since the risks of an eventual accident due to equipment failure can damage the facilities and the community surrounding the plant (WANG et al., 2018; PANDARAKONE; MIZUNO; NAKAMURA, 2019).

In unplanned interruptions with process equipment failure, the impact ends up being caused by production costs, which reduces the company's competitiveness (AYDEMIR; ACAR, 2020). To assist in this approach, technicians and operators usually using control and signaling sensors to monitor indications of failures, such as vibration, acceleration signals, and temperature of abnormal equipment (MIN et al., 2019; ZHANG et al., 2015; WANG et al., 2018; LEE et al., 2013). In this way, manufacturing processes with these characteristics are responsible for driving Big Data in the context of I4.0. Therefore, it becomes an ideal environment for Big Data Analytics to support decisions and collaboration to guide managers towards the best judgment.

Several reliability studies have been carried out in the Petrochemical Industry to monitor equipment and detect anomalies using sensors installed at strategic points (MUTLU; ALTUNTAS, 2019; ZHANG et al., 2015; CHENG; YAO; WU, 2013; MULUBRHAN; MOKHTAR; MUHAMMAD, 2014; JIA et al., 2018; CHEN et al., 2017). This possibility has become an essential mode of increasing equipment availability (AYDEMIR; ACAR, 2020). However, when monitoring is performed only on specific equipment, the effects of operating conditions may not be perceived or evaluated.

Since the Petrochemical Industry is intensive in data generated through connected devices for the most diverse controls, the following arises: why not use process variables to help identify future anomalies? The process variables have *fingerprints*, which are known parameters for operation within the control points. However, small changes in isolated variables sometimes are not simple tasks for operators to notice, and even a minor disturbance might cause equipment failure. On the other hand, the maintenance sector monitoring the equipment in isolation may not be enough to predict possible failures effectively (AYDEMIR; ACAR, 2020; WANG et al., 2018; WANG; ZHENG; ZHANG, 2020).

Therefore, we believe that asset management that includes operational and monitoring critical equipment data benefits the organization. With this, the maintenance sector achieves better monitor and control assets, and production increases equipment availability. In this way, a model is needed to qualify which feature is more relevant to identify a possible failure Wang et al. (2018), Wang, Zheng e Zhang (2020) and, therefore, provide insights for preventive decision-making (ZHAO; WANG; CHU, 2019; SOUZA et al., 2020; WANG; ZHENG; ZHANG, 2020).

For this reason, this chapter aims to apply the PHMS framework to support decision-making in a petrochemical company and presents the following main contributions:

1. Apply a DL model for anomaly detection on a production system;

2. Propose an FI model considering process and equipment data to identify the possible root cause;

3. Compare the main DL architectures for RUL prediction in a noisy and non-stationary time series;

4. Performs system RUL prediction given an abnormal condition.

The following sections of this chapter detail each step to realizing the case study.

6.2 Development and results

One of the principal equipment to optimize an SM plant's efficiency is the compressor, which is responsible for reducing the system pressure and reusing the gas generated in the reaction. This way, it becomes essential to use a model that can anticipate an adjustment in preventive maintenance to avoid a system stop.

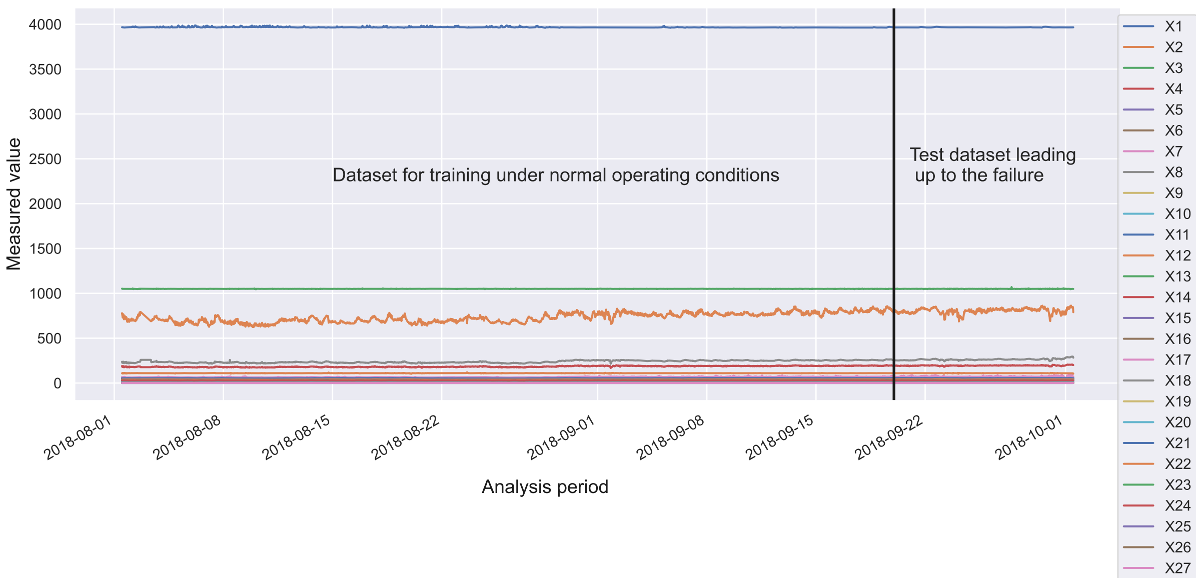
As discussed in Chapter 5, the data used in the model is from the reaction process. For the proposed model, all process variables may contain relevant insights to prevent system failure. As an example, we can mention the variable $X1$ on Table 4 that represents the feed load of the plant. This measure is essential since there is an impact on the system pressure in case of sudden variations, and therefore the compressor will have an overload. Another critical variable consists of the temperature parameters since the SM polymerization becomes facilitated in high-temperature conditions. The polymerization inside the compressor is usually crucial, considering cleaning the system and the expressive financial losses.

Detecting minor variations in process parameters is usually a complex activity for operators and engineers due to the number of variables and instruments in the operational process. However, with the AE, it is possible to identify small changes in the control parameters and warn of possible equipment anomalies. Accordingly, in the following subsection, it will be applied to identify a possible abnormal condition.

6.2.1 Autoencoder model to anomaly detection

To build the model, we need to separate the data in training and testing. For this activity, those responsible for the process provided production and maintenance reports to identify time windows in which the compressor operated under normal conditions. Also, to test, 12 days until the compressor stops, as shown in Figure 42. In Appendix A, we present the measured values in different ranges to facilitate the visualization of each feature. In Appendix B, we present a table with summary statistics with all Features.

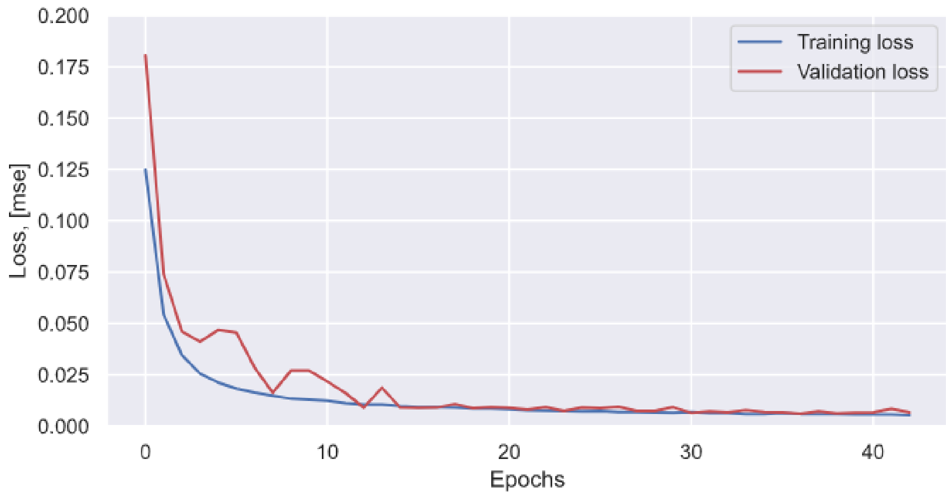
Figure 42 – Features included in the study with the periods used under normal conditions to train and pre-failure to test the model.



Source: Prepared by the author.

With the data sets defined, we applied the model shown in Figure 35, and to track training accuracy, we used 15% of the training data set for validation after each epoch. The complete model and the used data set are available in the GitHub repository[1]. In Figure 43, we presented the adjustment of the model in training and testing. In Appendix D, we presented a small summary of the hyperparameters tested. As we used the EarlyStopping parameter, the model stopped in epoch 53, although we defined it as the initial parameter with 150 epochs for training. This approach contributes to avoiding overfitting the model.

Figure 43 – Train and validation loss.



Source: Prepared by the author.

With the training set, it is possible to trace the calculated loss distribution and verify the appropriate threshold to identify an anomaly, as shown in Figure 44. It becomes evident that a value above the threshold can be considered statistically as an anomaly.

From the loss distribution Figure 44, we can consider a threshold of > 0.15 as a possible anomaly. With this, we calculate the loss in the test set to verify when it crosses the anomaly threshold.

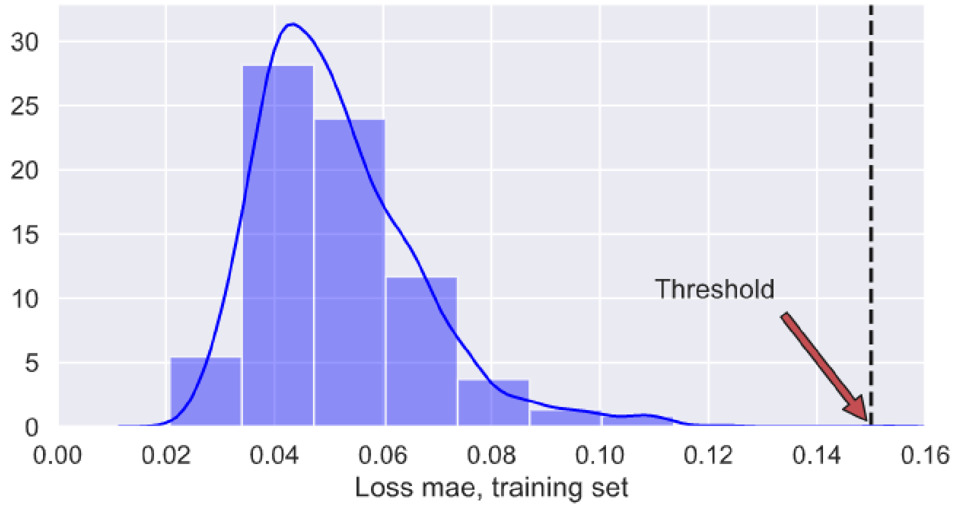
Next, we estimated the training set metrics and merged the test data into a unique data set. We could visualize the AE model's result with loss distribution and threshold and verify the period until the compressor stopped. In Figure 45, it is possible to observe that there were signs of an anomaly in September, and there is a tendency towards the threshold, which remains until the failure occurs on Oct. 1.

According to the proposed method, we used a new data set to validate the model and the limit considered for an anomaly. In this sense, we extracted a new set of data with 38 days with different compressor operating conditions. In the time interval, the compressor normally presented minor disturbances and a more serious fault, which generated a shutdown.

By applying the model to the new data set, it was possible to detect the anomalies in which the compressor operated and, finally, notice the equipment's shutdown. We observed an

[1] Repository

Figure 44 – Distribution of reconstruction loss for *healthy equipment*.



Source: Prepared by the author.

anomaly after Sep. 15, shown in Figure 46, and it demonstrated the equipment operating close to the threshold. This statement becomes evident when, in early October, there was a shutdown. Thus, we considered that the model, and the threshold, are relevant to support the operational and maintenance areas in managing this vital equipment.

After identifying the existence of an anomaly, it is necessary to determine which variables are responsible for the compressor failure. For this purpose, we employed a new step in which consists of the feature identification accountable for the abnormal operating conditions.

6.2.2 Model to Feature identification

Distributing the RL input data across the AE model makes it possible to analyze which features contribute most to the anomaly. This step is a relevant benefit in the PHMS proposal, as it contributes to the effective implementation of previous actions. To do so, we used the 27 variables that make up the system as characteristics. And for the label, we used the RL displayed in Figure 44 of the AE and already published in the study by Souza et al. (2021).

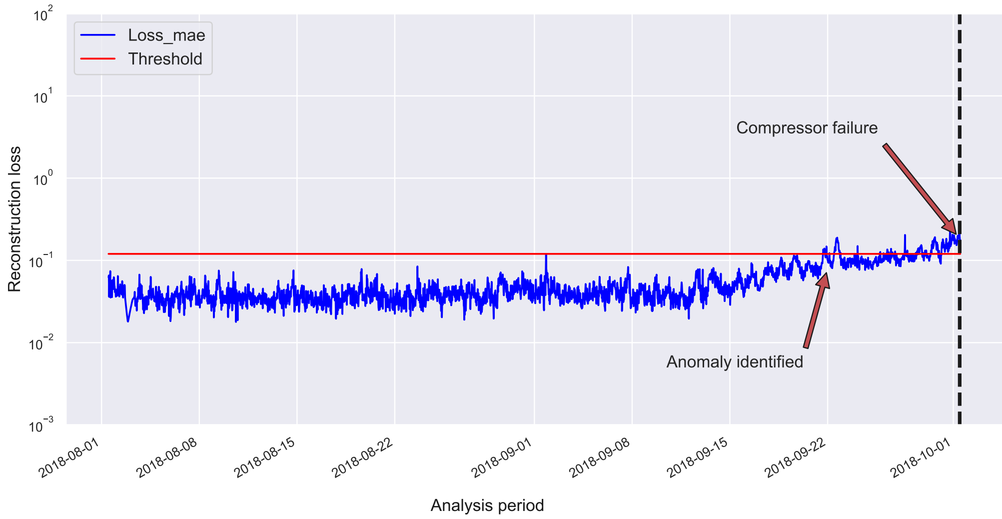
For defining the hyperparameters of the algorithms, we adopted the exhaustive search technique as a session. In Table 6, we display an extract of the main hyperparameters obtained for the models.

After running the models with the hyperparameters from Table 6, we obtained the following results:

- XGBoost: $R^2 = 96.80\%$.
- Random Forest: $R^2 = 95.24\%$

According to the in-session validation criteria, we consider the model valid, as we obtained

Figure 45 – The Autoencoder model results until compressor failure.



Source: Prepared by the author.

$R^2 = 96.80\%$ in test data. Thus, according to the adopted criteria, we consider XGBoost valid. Then, we applied the SHAP method to explain the XGBoost results to identify which Features are more relevant.

6.2.2.1 SHAP to Feature identification

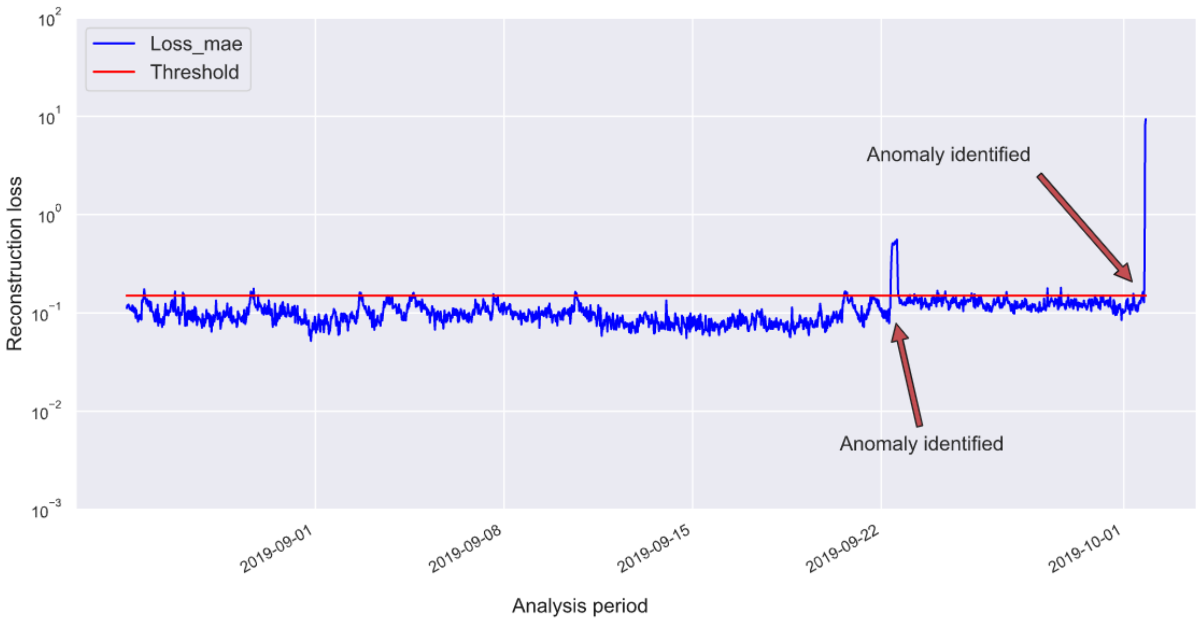
Understanding the predictions of sophisticated models remains a significant difficulty, and interpretations at the Feature level individually are usually of greater interest. Feature interpretability receives greater relevance when it comes to black box models. Therefore, in this research, using SHAP, we analyzed the behavior of the features in predicting the anomaly. We used the XGBoost model, which showed the best performance. In Figure 47, we present the global importance and local explanation summary of the features for training the model.

In Figure 47, the bar plot represents the average magnitude of the SHAP value. The most important features are listed in descending order according to their significance for model performance. The Beeswarm graphic [2] to the right of Figure 47 displays a dense summary of information on how key features of the dataset affected the model output. A single point represents each supplied instance. The SHAP value of the feature determines the X position of the point, and the points accumulate along the sequence of each feature to show the density. In the left plot in Figure 47, it shows that $X13$ is on average the most important feature. Whereas the right image of Figure 47 shows that for an identified anomaly, smaller values (blue) contribute less to the occurrence of the abnormal condition.

In particular, the Beeswarm plot to the right of Figure 47 provides the following information:

[2]Beeswarm

Figure 46 – Results of the model with new data set.



Source: Prepared by the author.

- **Feature Importance:** The Features are sorted from largest to smallest. The feature $X13$ is the most important, followed by $X15$ and $X8$.

- **Impact:** The horizontal position reveals whether the value positively or negatively influences the forecast.

- **Original value:** For each observation, the color indicates whether the value of that variable is high (in red) or low (in blue).

- **Correlation:** Having a high-temperature value (feature "$X13$") has a high impact on the prediction of the identified anomaly. Red represents a "high" effect, and the X axis represents a "positive" effect.

In order to analyze the dependency between features for local forecasts, we present Figure 48. In **"a"** in Figure 48, we notice that when features $X13$ and $X16$ are low (blue) the risk for anomalies is reduced. Whereas the image **"c"** in Figure 48 shows a strong relationship between $X8$ and $X16$. That is, upon a certain condition of the recirculation of the $X8$ system, the vibration $X16$ has a strong tendency for the anomaly to occur. This type of information allows identifying conditions and relationships between features that can cause instability in the system.

Finally, we analyze the contribution of each feature to the prediction of a single dataset instance. Considering instance 138 as an example, the force graphic in Figure 49 provides the following information:

- **Output value** $f(x)$**:** is the prediction for observation 138 of the dataset.

Table 6 – Best set of main hyperparameters using brute force according to the algorithm used.

Hyperparameters	XGBoost	Random Forest
n_estimators	2000	200
max_depth	None	80
learning_rate	0.01	Na
colsample_bylevel	0.4	Na
colsample_bytree	0.8	Na
min_samples_split	Na	Na
max_features	Na	sqrt
min_samples_leaf	Na	1

Source: Prepared by the author.

- **Base value:** is the predicted value if we knew no feature for the current value. In other words, it is the average prediction of the validation dataset. In this case, the mean value of the test is 0.07613.
- **Red/blue:** features that increase the prediction (to the right) are shown in red, and those that move the prediction down are in blue.

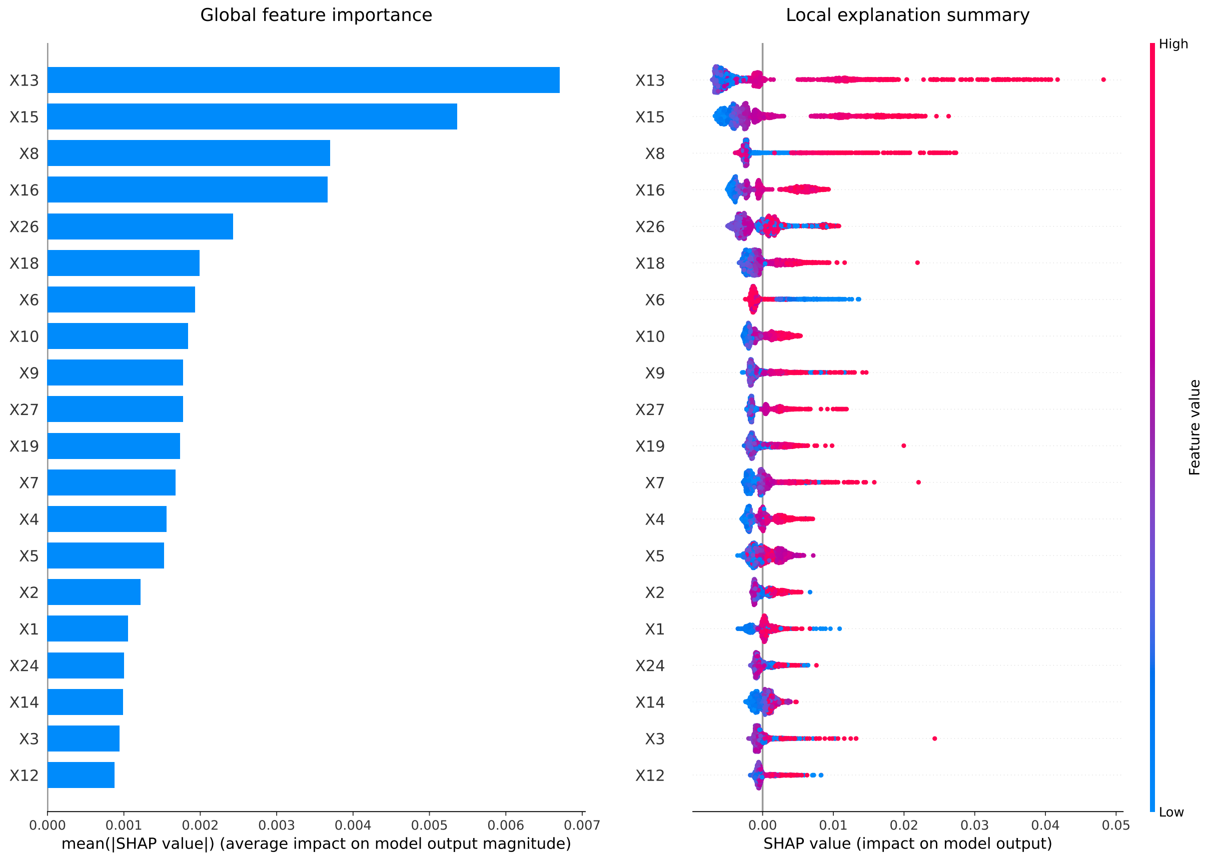
Thus, we can observe that the feature $X13$ is the most important for predicting instance 138, followed by $X15$ and $X8$. Whereas the failure that occurred on October 1, 2018, occurred due to the polymerization of the compressor discharge piping. According to the model results, the feature $X13$ is a signal that the compressor input temperature rise indicates polymer formation. Another relevant characteristic identified by the model was that $X8$, according to the fault investigation, had the compressor recirculation opening to compensate for the pressure in the compressor discharge piping. Thus, the model could identify the relationship between the variables and the abnormal condition. We point out that with SHAP, it was possible to analyze the contribution of each feature to the system failure locally. With the results presented, preventive actions could be taken to avoid unplanned system downtime

We performed the FI step in the PHMS context in this subsection. In the next subsection, we address the step dealing with RUL prediction.

6.2.3 Prediction RUL for the system

The RUL prediction is an important task in the PHM context. However, fully understanding the dynamics of a complex system is difficult. In this sense, we performed the RUL prediction for a system with several assets and different functions, raising the analysis's complexity. In this direction, we initially need the system's HI for RUL prediction. In order to do so, we followed

Figure 47 – Impact of each feature on the predictive model. On the left are average SHAP values for each feature and on the right is the local explanation of each feature.



Source: Prepared by the author.

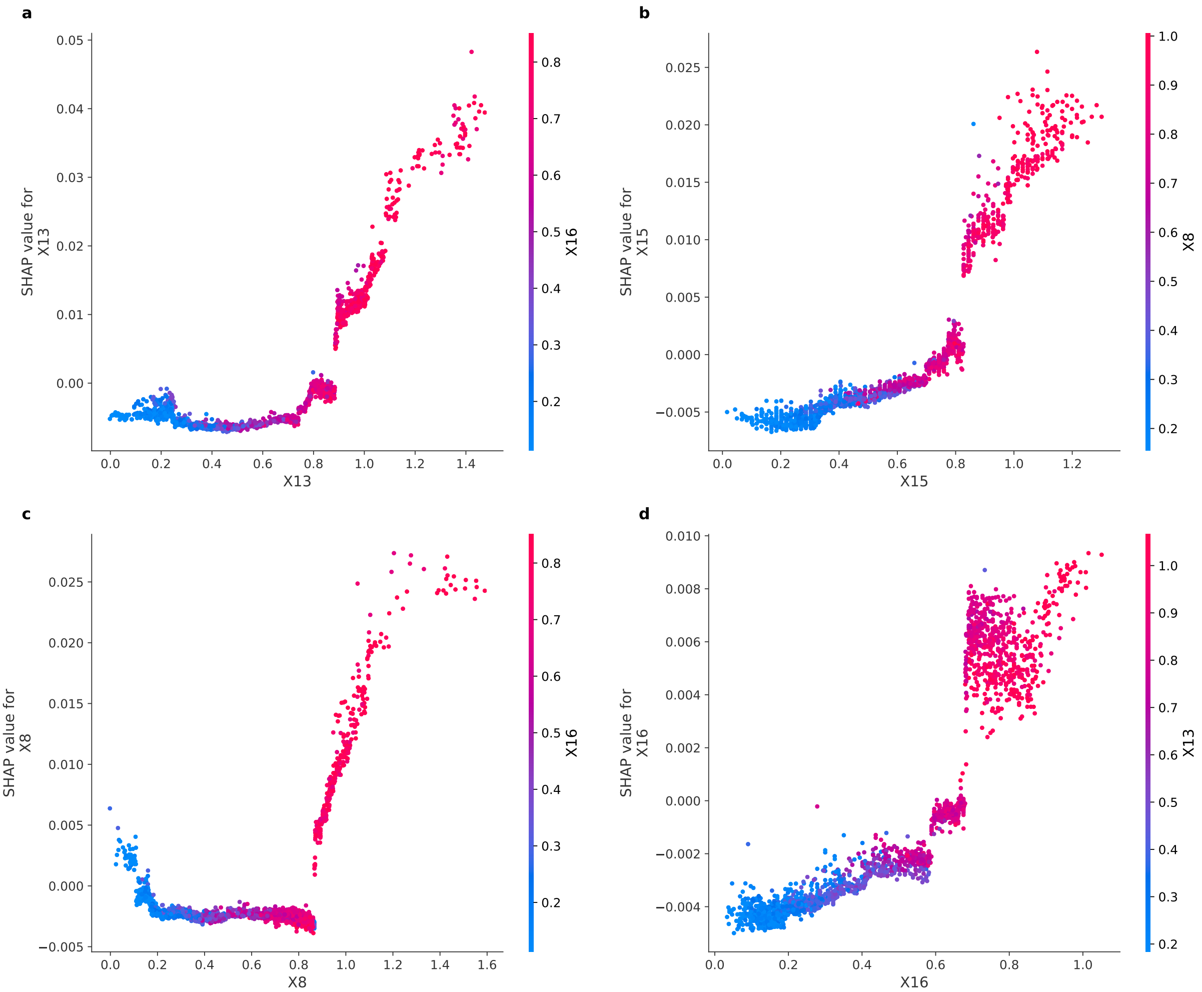
the steps presented in section 5.2.2.3. As a first step, we applied Equation 5.7 to normalize the RL of the AE. Figure 50 presents the resulting system's HI characteristic curve.

The analysis period comprised the time frame from 01/08/2018 to 01/10/2018 according to axis X of Figure 50. This interval contemplates stable operation until the failure in early October. We performed the following dataset division for training and testing. For training, we considered 60% of the sample from the beginning to 09/12/2018, and from this date on, we used it for testing the prediction models.

6.2.3.1 Model performance benchmark for RUL prediction.

Regarding the models' performance, we present a summary of the results in visual form in Figure 51 and a comparison of the metrics in Figure 52 and Table 7. We can notice in Figure 51 that the model showing the highest compliance to the test data is N-BEATS, and we confirmed it in Table 7 with the results for all models. This result can be explained by the fact that the model can capture the temporal dynamics of the data even under non-stationary conditions. On the other hand, the LSTM and GRU models underperformed since they failed to capture the temporal dynamics in non-stationary data. The results obtained with RNNs corroborate studies

Figure 48 – Combination of model local explanations. **a**, shows the relationship between Input Temperature Control $X13$ and Vibration Control $X16$. **b**, Vibration Control $X15$ and Compressor Recirculation $X8$. **c**, Compressor Recirculation $X8$ and Vibration Control $X16$. **d**, Vibration Control $X16$ and Input Temperature Control $X13$.



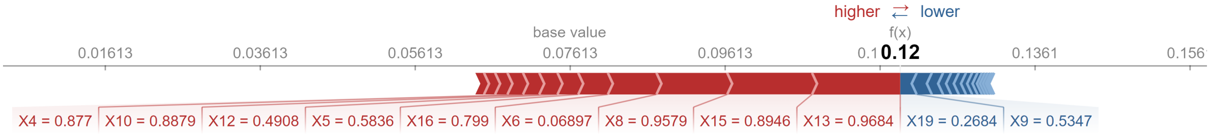
Source: Prepared by the author.

pointing out this type of RNN limitation (MA; MAO, 2021).

The results displayed in Table 7 and Figure 51 draw attention to the MLP model's performance because, despite being a simpler architecture, it was superior compared to more complex networks. We can infer that the ability to approximate the mapping function from input to output variables contributed to dealing with the existing noise in the HI time series. Another important MLP characteristic that contributes to the model's performance is the absence of temporal structure. In other words, time steps are modeled as input resources, meaning the network has no explicit manipulation or understanding of the temporal structure or order between observations.

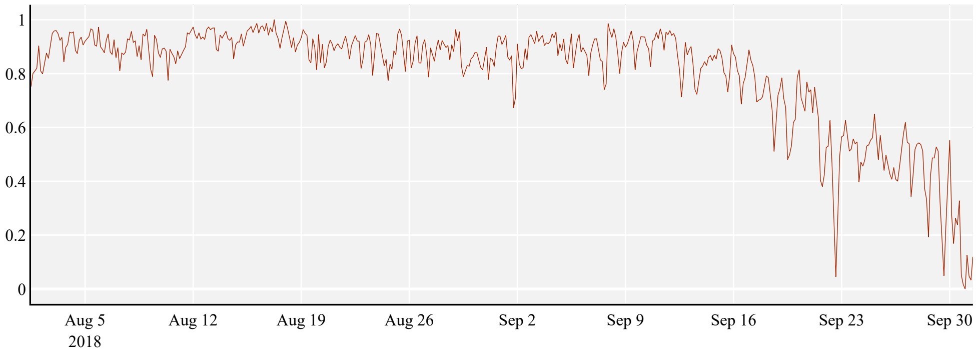
Our results also highlight CNN's ability to handle the time series with HI characteristics. This is justified because, like MLP, they handle multivariate input and output very well, do not depend on temporal relationships, and learn arbitrary functional connections with higher

Figure 49 – Local SHAP value for a specific dataset observation.



Source: Prepared by the author.

Figure 50 – The characteristic curve of the health index of the system in which we carried out the study.



Source: Prepared by the author.

complexity (YANG et al., 2015).

6.3 Final Remarks

According to the results presented in this case study, we can infer that it is difficult to deal with complex problems with a single approach. Therefore, for anomaly identification, the unsupervised models are consistent alternatives since we do not always know the behavior to be predicted. On the other hand, supervised models are more suitable for RUL prediction problems. Thus, hybrid ML modeling can be a relevant alternative in complex environments, such as industrial settings. Thus, as we present a framework approach, we can robustly handle the complexity of using data for decision support.

In this study, we can organize the framework into different steps. In the anomaly identification step, we applied AE, which makes it possible to reconstruct input data and, through the latent representation of the data, detect anomalies hidden in the high dimensionality of the features. In the feature detection stage, we used SHAP, which provides an analysis that goes beyond visualizing which features are most relevant to the model result, as it provides means of understanding the combination of feature states, as an example, the relationships we present in Figure 48. As a final step, we used different DL models, which enabled the prediction of the system's RUL using the latent representation input data generated by AE.

Using the framework proposed here, we sought to evolve regarding traditional PHM by

Table 7 – Benchmark the forecast models for the RUL

Metrics	**N-Beats**	MLP	CNN	LSTM	GRU
MSE	**0.63**	1.03	1.14	1.68	1.32
MAE	**6.01**	7.40	7.67	9.06	8.11
RMSE	**7.93**	10.15	10.68	12.96	11.49
MAPE	**11.32**	17.17	17.76	23.19	20.46
R^2	**91.36**	87.01	87.16	82.35	85.55

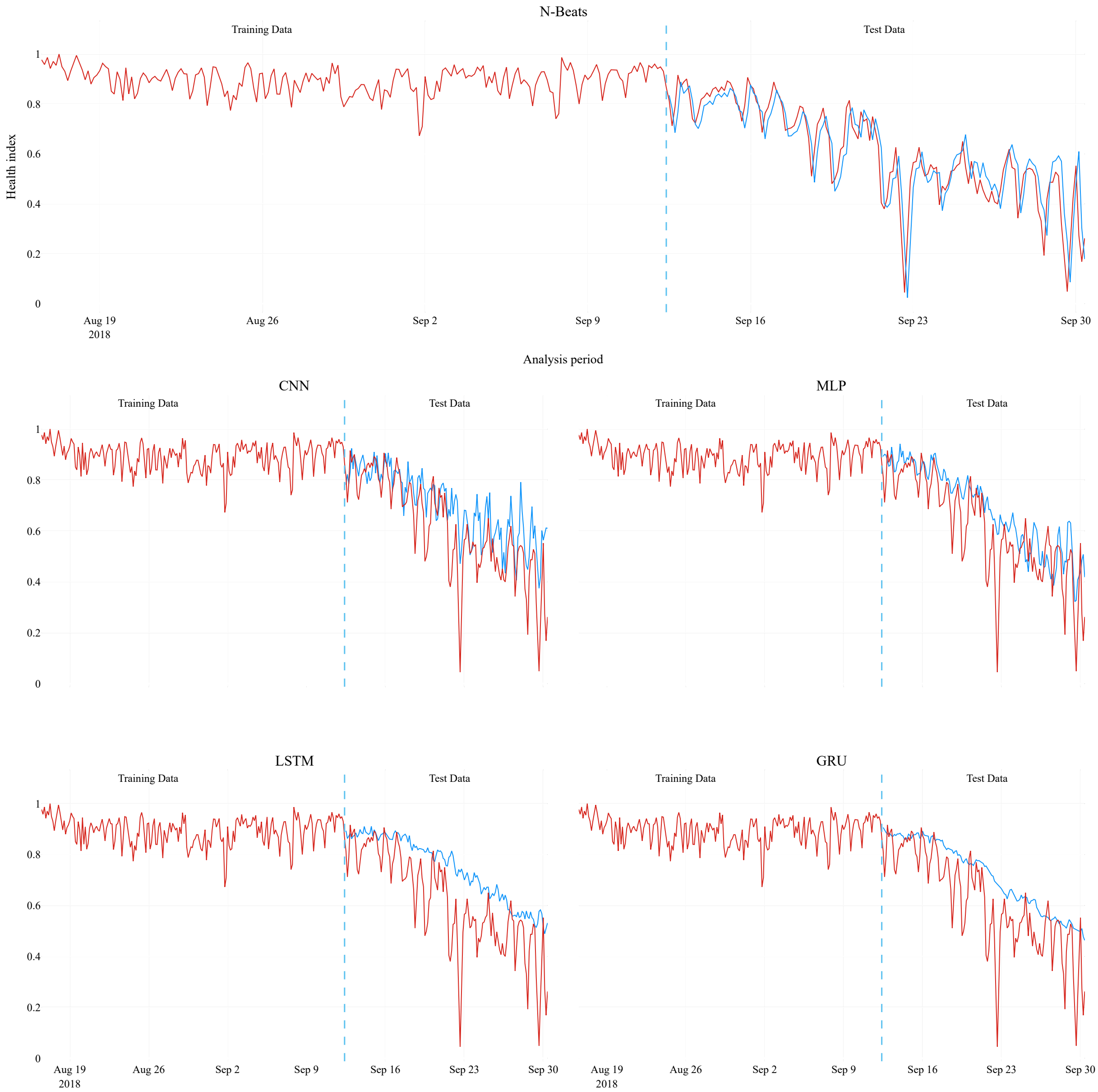
Source: Prepared by the author.

providing a methodology that allows using different ML techniques. With this, it is possible to progress toward analyzing not only a specific piece of equipment and its components but a system comprising several pieces of equipment, components, and subsystems. In this sense, we position the use of our proposition which we call PHMS. In our judgment, the implementation of PHMS in industrial systems is made possible through a pipeline combining advanced ML techniques.

Figure 53 in **"a"** demonstrates a prevention and reaction strategy. On the other hand, in **"b"**, it comprises a comprehensive analysis of system data and outputs with insights for decision-makers. In the smart manufacturing context, PHMS enables maintenance and operation actions to transition from preventive and reactive to predictive, preventive, prescriptive, and CBM strategies. The benefits that can derive from PHMS are aligned with smart manufacturing, which includes reducing costs, increasing system availability, reducing energy consumption, improving production yield, and avoiding damage to the environment due to accidents from critical system failures.

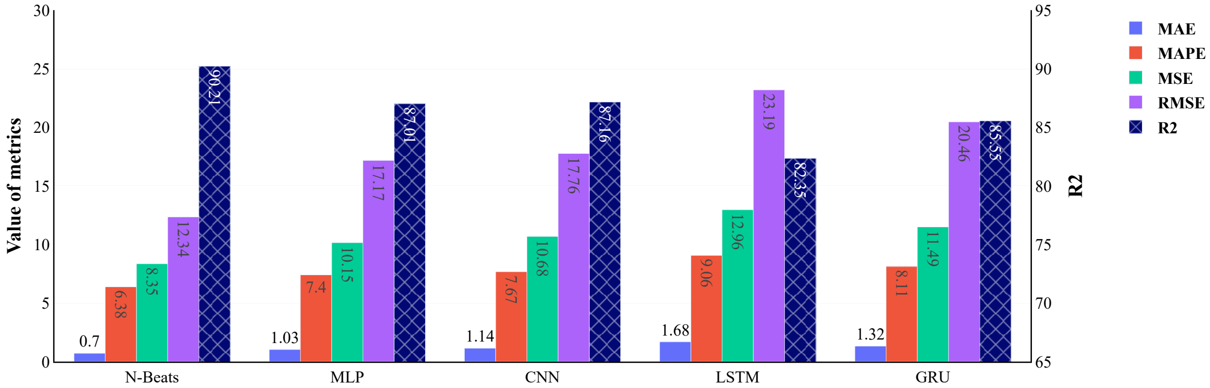
Although we have found the applied approach consistent and valid under real-world conditions, we have limited ourselves to a specific case study. It is a restriction regarding the application in other environments. In this regard, we consider replicating the research in similar systems and different contexts, such as capital markets, fraud detection in financial systems, and in the health of living beings, such as animals and humans. Despite the peculiarities, the different contexts have the complexity of the systems in common, just like the industrial processes.

Figure 51 – Results of models for predicting the RUL in test data compared to real data. The y axes represent the HI of the system and the x axes the analysis period. The light blue dashed line shows the separation of test and training windows.



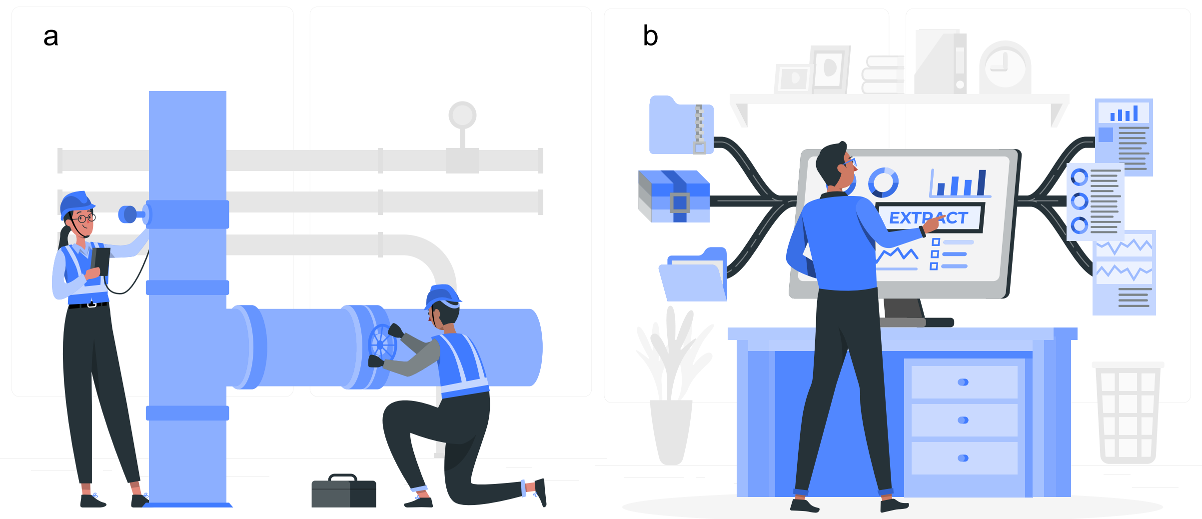
Source: Prepared by the author.

Figure 52 – Performance of the models according to the metrics used in test set. Values are in percent and the y axis on the right considers only R^2 because of the scale.



Source: Prepared by the author.

Figure 53 – In **"a"** preventive and reactive maintenance through local analysis. In **"b"**, PHMS is a data-driven pipeline for decisions in an intelligent manufacturing system.



Source: Prepared by the author adapted from Storyset.

7 CONCLUSION

This study was guided by the research question: "How to develop a PDM model in a production system to support decision-making in the context of I4.0?". In this sense, we introduce the concept of PHMS to support decision-making in PDM and increase reliability in the context of I4.0. This concept evolves concerning the traditional PHM as it includes a production system. In this sense, within the scope of this research, a production system is represented by various equipment and peripherals in addition to operating variables. The PHMS is justified because, to model complex systems, the traditional maintenance and reliability methods have limitations, as presented in the chapter 3 and previously published in (SOUZA et al., 2020). Furthermore, chapter 3 shows that most recent research dealing with reliability focuses on local improvements. In this way, PHMS allows investigations and corrective activities to be performed quickly and efficiently on systems, as presented in Figure 1 in the introduction section.

To answer the research question and meet the main specific objectives, we developed an ML framework to support the PHMS. Then, we performed a case study with real data to evaluate our proposition. With this, we highlight some essential contributions to the field of academia and business where part of the results have already been published in (SOUZA et al., 2021). As an initial step, we applied the AE to the system variables to map the latent representation of the features and generate a threshold of normality and the RL. Then, it was possible to detect anomalies in the system, perform the FI, generate the HI, and predict the RUL for the system.

Also, as an essential contribution, a taxonomy was created with the application of reliability for decision-making in I4.0 published in (SOUZA et al., 2020). The taxonomy presented the application types, principles, and integration model considering reliability in the context of I4.0. We noticed that in vertical integration, the gains are more focused on internal and local results. On the other hand, in horizontal integration, the benefits go beyond the boundaries of organizations. In this sense, with the evolution of the digitization of operations driven by I4.0, the use of data for decision support is a reality. However, when applied to complex systems, there are limitations and a lack of convergence by which ML techniques to apply. In this direction, we deem relevant the proposition of the framework that supports the PHMS developed in this research.

PHMS relies on using sensors and other monitoring devices to continually collect data on the performance and condition of system equipment. This data is then analyzed using the framework with algorithms and ML techniques to identify patterns and trends that could indicate a possible failure. Based on this analysis, tasks can be scheduled in advance, allowing maintenance and operations teams to resolve potential issues before they occur. As a result, PHMS can effectively streamline maintenance and operation activities and reduce costs, allowing teams to focus on proactive actions rather than reactive ones. It can also help improve equipment and system reliability and extend equipment life.

We position PHMS as a type of strategy that uses data from different sources and, through analysis, identify when the system is likely to fail, and with that, actions are taken in advance. In this way, PHMS aims to avoid equipment failures and unplanned stops, which can interrupt operations, reduce productivity and increase costs.

This research presents some considerable contributions regarding the ML techniques proposed in the framework. In the first review, we noted the benefits of SHAP for troubleshooting and health management of industrial equipment, as demonstrated in human health research. The adoption of SHAP has provided advances concerning traditional FI methods, where the averages of the characteristics are considered, as in (SOUZA et al., 2021). Furthermore, we note that RNNs may not be the best option in short-term non-stationary time series. In this scenario, we emphasize that simpler models, such as MLP, or complex ones, such as CNN, present better results. A relevant relationship between MLP and CNN is that in shorter time windows, the absence of time dependence contributed to better results than RNNs. In addition, we highlight the promising result obtained with N-BEATS, since, in all validation metrics, N-BEATS was superior. Thus, it is an effective option for predicting RUL in complex and noisy environments, such as industrial processes.

7.1 Contributions

With the framework, it was possible to create an operational threshold to alert the existence of an event outside the operational standards. The possibility of developing a maintenance alert based on system data contributes significantly. In systems such as petrochemical, it is often complex to define operation limits. Another benefit of the search is to detect the possible root cause for anomalies since the time to act can be reduced. Concerning FI, the application of SHAP to understand the relationship of features over time helped to understand the behavior of the type of failure, as shown in the case study. Also noteworthy is the result obtained with N-BEATS for forecasting a more complex time series. The contributions mentioned have already been partially published in important journals. Next, the publications derived from this thesis are presented

7.2 Publications

As partial contributions throughout the research, articles were produced for publication in journals listed below:

- Published articles:
  - SOUZA, Marcos Leandro Hoffmann et al. A survey on decision-making based on system reliability in the context of Industry 4.0. Journal of Manufacturing Systems, v. 56, p. 133–156, 2020.

  - SOUZA, Marcos Leandro Hoffmann et al. A feature identification method to explain anomalies in condition monitoring. Computers in Industry, Elsevier, v. 133, 2021.

- Articles under review:

  - SOUZA, Marcos Leandro Hoffmann et al. Proposition of a data-oriented pipeline for PHMS in a production system using ML. Computers in Industry, Elsevier.

7.3 Limitations and future work

Despite the relevant results presented, the approach presented in the research has limitations. As the main limitation, it can be highlighted that the data used in the study belong to the time domain and the frequency domain. Thus, the proposed solution was not designed to deal with discrete data. Another limitation is linked to the failure mode since, in electrical systems, failures are usually random, making early identification and signaling as an anomaly difficult. Finally, the amount of data available to evaluate different operating conditions. Furthermore, we put little effort into testing DL hyperparameters for RUL prediction.

In continuity, this research intends to present results contributing to the academic area and decision-makers in sectors similar to those discussed here. As a benefit, one can cite the RUL forecast model and optimize the results, where the primary data are related to the market in which the research is concentrated. Also, in future works, we consider evaluating the information from the operators and the field maintenance team and investigating the benefits of the model, mainly in the FI stage. In addition, the model can evolve and consider simulations of critical operating conditions and follow the model's performance. We also suggest including Concept drift to monitor the model's condition for changes in the system over time.

REFERENCES

ADU-AMANKWA, Kwaku et al. A predictive maintenance cost model for CNC SMEs in the era of industry 4.0. **International Journal of Advanced Manufacturing Technology**, Springer London, v. 104, n. 9-12, p. 3567–3587, 10 2019. ISSN 14333015. Disponível em: <http://link.springer.com/10.1007/s00170-019-04094-2>.

AGIUS, Rudi et al. Machine learning can identify newly diagnosed patients with CLL at high risk of infection. **Nature Communications**, Springer US, v. 11, n. 1, 2020. ISSN 20411723. Disponível em: <http://dx.doi.org/10.1038/s41467-019-14225-8>.

ALFEO, Antonio L.; CIMINO, Mario G.C.A.; VAGLINI, Gigliola. Degradation stage classification via interpretable feature learning. **Journal of Manufacturing Systems**, Elsevier Ltd, v. 62, n. May 2021, p. 972–983, 2022. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2021.05.003>.

AREMU, Oluseun Omotola et al. A relative entropy based feature selection framework for asset data in predictive maintenance. **Computers and Industrial Engineering**, Elsevier, v. 145, n. July 2019, p. 106536, 2020. ISSN 03608352. Disponível em: <https://doi.org/10.1016/j.cie.2020.106536>.

ARIZA-GARZON, Miller Janny et al. Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending. **IEEE Access**, v. 8, p. 64873–64890, 2020. ISSN 21693536.

ATZORI, Luigi; IERA, Antonio; MORABITO, Giacomo. The Internet of Things: A survey. **Computer Networks**, Elsevier B.V., v. 54, n. 15, p. 2787–2805, 2010. ISSN 13891286. Disponível em: <http://dx.doi.org/10.1016/j.comnet.2010.05.010>.

AYDEMIR, Gurkan; ACAR, Burak. Anomaly monitoring improves remaining useful life estimation of industrial machinery. **Journal of Manufacturing Systems**, Elsevier, v. 56, n. February, p. 463–469, 2020. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2020.06.014>.

BALADOR, Ali; ERICSSON, Niclas; BAKHSHI, Zeinab. Communication middleware technologies for industrial distributed control systems: A literature review. **IEEE International Conference on Emerging Technologies and Factory Automation, ETFA**, p. 1–6, 2017. ISSN 19460759.

BALOGH, Z. et al. Reference Architecture for a Collaborative Predictive Platform for Smart Maintenance in Manufacturing. **INES 2018 - IEEE 22nd International Conference on Intelligent Engineering Systems, Proceedings**, p. 000299–000304, 2018. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8523969/>.

BAPTISTA, Marcia et al. Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. **Computers and Industrial Engineering**, Elsevier, v. 115, n. October 2017, p. 41–53, 2018. ISSN 03608352. Disponível em: <https://doi.org/10.1016/j.cie.2017.10.033>.

BENKER, Maximilian et al. Utilizing uncertainty information in remaining useful life estimation via Bayesian neural networks and Hamiltonian Monte Carlo. **Journal of Manufacturing Systems**, Elsevier Ltd, v. 61, n. November 2020, p. 799–807, 2021. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2020.11.005>.

BERRI, Pier Carlo; VEDOVA, Matteo D.L. Dalla; MAININI, Laura. Computational framework for real-time diagnostics and prognostics of aircraft actuation systems. **Computers in Industry**, Elsevier, v. 132, 2021. ISSN 01663615. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166361521001305>.

BIANCHINI, Augusto; PELLEGRINI, Marco; ROSSI, Jessica. Maintenance scheduling optimization for industrial centrifugal pumps. **International Journal of System Assurance Engineering and Management**, v. 10, n. 4, p. 848–860, 2019. ISSN 09764348. Disponível em: <http://link.springer.com/10.1007/s13198-019-00819-4>.

BLACK, John J.; MEJABI, O. O. Simulation of complex manufacturing equipment reliability using object oriented methods. **Reliability Engineering and System Safety**, v. 48, n. 1, p. 11–18, 1995. ISSN 09518320.

BORGI, Tawfik et al. Big Data for Operational Efficiency of Transport and Logistics: A Review. **6th IEEE International Conference on Advanced Logistics and Transport, ICALT 2017 - Proceedings**, p. 184–188, 2018. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8547029/>.

BOUSDEKIS, Alexandros; MENTZAS, Gregoris. Condition-Based Predictive Maintenance in the Frame of Industry 4.0. In: **IFIP Advances in Information and Communication Technology**. [s.n.], 2017. v. 513, p. 399–406. ISBN 9783319669229. Disponível em: <http://link.springer.com/10.1007/978-3-319-66923-6_47>.

BUITINCK, Lars et al. API design for machine learning software: experiences from the scikit-learn project. **arXiv preprint arXiv:1309.0238**, p. 1–15, 2013. Disponível em: <http://arxiv.org/abs/1309.0238>.

BURDICK, G. R. et al. Phased Mission Analysis: A Review of New Developments and An Application. **IEEE Transactions on Reliability**, R-26, n. 1, p. 43–49, 1977. ISSN 15581721.

CAIAZZO, Bianca et al. Towards Zero Defect Manufacturing paradigm: A review of the state-of-the-art methods and open challenges. **Computers in Industry**, Elsevier, v. 134, p. 103548, 2022. ISSN 01663615. Disponível em: <https://doi.org/10.1016/j.compind.2021.103548>.

CALLE, Kerman López de et al. Dynamic condition monitoring method based on dimensionality reduction techniques for data-limited industrial environments. **Computers in Industry**, Elsevier, v. 112, 2019. ISSN 01663615. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166361519301174>.

CAMPBELL, Thomas W. et al. Exact Shapley values for local and model-true explanations of decision tree ensembles. **Machine Learning with Applications**, Elsevier Ltd., v. 9, n. May, p. 100345, 2022. ISSN 26668270. Disponível em: <https://doi.org/10.1016/j.mlwa.2022.100345>.

CANITO, Alda et al. An architecture for proactive maintenance in the machinery industry. **Advances in Intelligent Systems and Computing**, v. 615, p. 254–262, 2017. ISSN 21945357. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-319-61118-1_31>.

CANIZO, Mikel et al. Real-time predictive maintenance for wind turbines using Big Data frameworks. **2017 IEEE International Conference on Prognostics and Health Management, ICPHM 2017**, p. 70–77, 2017. Disponível em: <https://ieeexplore.ieee.org/abstract/document/7998308/>.

CAO, Qiushi et al. An ontology-based approach for failure classification in predictive maintenance using fuzzy c-means and SWRL rules. **Procedia Computer Science**, v. 159, p. 630–639, 2019. ISSN 18770509.

CHE, Changchang et al. Domain adaptive deep belief network for rolling bearing fault diagnosis. **Computers and Industrial Engineering**, Elsevier, v. 143, n. June 2019, p. 106427, 2020. ISSN 03608352. Disponível em: <https://doi.org/10.1016/j.cie.2020.106427>.

CHEN, Chih Wen et al. Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results. **Expert Systems**, v. 37, n. 5, p. 1–10, 2020. ISSN 14680394.

CHEN, Li-Pang. **Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python**. [S.l.]: O'Reilly Media, 2021. 272–273 p. ISSN 0040-1706.

CHEN, Zhiqiang et al. Deep neural networks-based rolling bearing fault diagnosis. **Microelectronics Reliability**, Elsevier Ltd, v. 75, p. 327–333, 2017. ISSN 00262714. Disponível em: <http://dx.doi.org/10.1016/j.microrel.2017.03.006>.

CHEN, Zhenghua et al. Machine Remaining Useful Life Prediction via an Attention-Based Deep Learning Approach. **IEEE Transactions on Industrial Electronics**, IEEE, v. 68, n. 3, p. 2521–2531, 2021. ISSN 15579948.

CHENG, Ching Wu; YAO, Hong Qing; WU, Tsung Chih. Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry. **Journal of Loss Prevention in the Process Industries**, Elsevier Ltd, v. 26, n. 6, p. 1269–1278, 2013. ISSN 09504230. Disponível em: <http://dx.doi.org/10.1016/j.jlp.2013.07.002>.

CHENG, Ying et al. Cyber-physical integration for moving digital factories forward towards smart manufacturing: a survey. **International Journal of Advanced Manufacturing Technology**, Springer London, v. 97, n. 1-4, p. 1209–1221, 7 2018. ISSN 14333015. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045842061&doi=10.1007%2Fs00170-018-2001-2&partnerID=40&md5=5a172289f8818e45b16ebfbb37afa27dhttp://link.springer.com/10.1007/s00170-018-2001-2>.

CHO, Sangje et al. A hybrid machine learning approach for predictive maintenance in smart factories of the future. In: **IFIP Advances in Information and Communication Technology**. [s.n.], 2018. v. 536, p. 311–317. ISBN 9783319997063. Disponível em: <http://link.springer.com/10.1007/978-3-319-99707-0_39>.

CHRISTOU, Ioannis T. et al. End-to-end industrial IoT platform for Quality 4.0 applications. **Computers in Industry**, Elsevier, v. 137, 2022. ISSN 01663615. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166361521001986>.

CHUNG, Junyoung et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. **arXiv preprint arXiv:1412.3555**, p. 1–9, 2014. Disponível em: <http://arxiv.org/abs/1412.3555>.

CLEVERT, Djork Arné; UNTERTHINER, Thomas; HOCHREITER, Sepp. Fast and accurate deep network learning by exponential linear units (ELUs). **4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings**, p. 1–14, 2016.

DALZOCHIO, Jovani et al. Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. **Computers in Industry**, Elsevier, v. 123, 2020. ISSN 01663615. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166361520305327>.

DARVISHI, Ali; RAHIMPOUR, Mohammad Reza; RAEISSI, Sona. A theoretical and experimental study for screening inhibitors for styrene polymerization. **Processes**, v. 7, n. 10, p. 1–18, 2019. ISSN 22279717.

DENG, Wenkai et al. A Data Mining Based System for Transaction Fraud Detection. **2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021**, p. 542–545, 2021.

DIEZ-OLIVAN, Alberto et al. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. **Information Fusion**, v. 50, p. 92–111, 2019. ISSN 15662535. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1566253518304706>.

DIMIAN, Alexandre C.; BILDEA, Costin Sorin. Energy Efficient Styrene Process: Design and Plantwide Control. **Industrial and Engineering Chemistry Research**, v. 58, n. 12, p. 4890–4905, 2019. ISSN 15205045.

ESMAEILIAN, Behzad; BEHDAD, Sara; WANG, Ben. The evolution and future of manufacturing: A review. **Journal of Manufacturing Systems**, The Society of Manufacturing Engineers, v. 39, p. 79–100, 2016. ISSN 02786125. Disponível em: <http://dx.doi.org/10.1016/j.jmsy.2016.03.001>.

FERNANDES, Marta et al. Data analysis and feature selection for predictive maintenance: A case-study in the metallurgic industry. **International Journal of Information Management**, v. 46, p. 252–262, 2019. ISSN 02684012. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0268401218304699>.

FERREIRA, Carlos; GONÇALVES, Gil. Remaining Useful Life prediction and challenges: A literature review on the use of Machine Learning Methods. **Journal of Manufacturing Systems**, Elsevier Ltd, v. 63, n. May, p. 550–562, 2022. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2022.05.010>.

FERREIRO, Susana et al. Industry 4.0 : Predictive Intelligent Maintenance for Production Equipment. **European Conference of the Prognostics and Health Management Society**, n. June 2016, p. 1–8, 2016. Disponível em: <http://papers.phmsociety.org/index.php/phme/article/view/1667>.

FILZ, Marc André et al. Data-driven failure mode and effect analysis (FMEA) to enhance maintenance planning. **Computers in Industry**, Elsevier, v. 129, 2021. ISSN 01663615.

FORDAL, Jon Martin; RØDSETH, Harald; SCHJØLBERG, Per. Initiating industrie 4.0 by implementing sensor management – Improving operational availability. **Lecture Notes in Electrical Engineering**, Springer Verlag, v. 484, p. 200–207, 2019. ISSN 18761119. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85059079009&doi=10.1007%2F978-981-13-2375-1_26&partnerID=40&md5=37a8fee5d8e03b1c97629821b2df7564>.

FUMAGALLI, Luca et al. A Smart Maintenance tool for a safe Electric Arc Furnace. **IFAC-PapersOnLine**, v. 49, n. 31, p. 19–24, 2016. ISSN 24058963. Disponível em: <http://www.sciencedirect.com/science/article/pii/S2405896316328269>.

GAO, R. et al. Cloud-enabled prognosis for manufacturing. **CIRP Annals - Manufacturing Technology**, v. 64, n. 2, p. 749–772, 2015. ISSN 17260604.

GÄRTNER, Bertolt. Industry 4.0 maturity index. **Assembly**, v. 61, n. 12, p. 32–35, 2018. ISSN 10508171. Disponível em: <https://www.ptc.com/-/media/Files/PDFs/IoT/acatech_STUDIE_Maturity_Index_eng_WEB.PDF%0Ahttp://www.acatech.de/fileadmin/user_upload/Baumstruktur_nach_Website/Acatech/root/de/Publikationen/Projektberichte/acatech_STUDIE_Maturity_Index_eng_WEB.pdf%0Awww.acat>.

GEURTS, Pierre; ERNST, Damien; WEHENKEL, Louis. Extremely randomized trees. **Machine Learning**, v. 63, n. 1, p. 3–42, 2006. ISSN 08856125.

GODREAU, Victor et al. Continuous improvement of HSM process by data mining. **Journal of Intelligent Manufacturing**, Springer New York LLC, v. 30, n. 7, p. 2781–2788, 2019. ISSN 15728145. Disponível em: <http://link.springer.com/10.1007/s10845-018-1426-7>.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. MIT press, 2016. 1–3 p. Disponível em: <http://www.deeplearningbook.org/front_matter.pdf>.

GOPALAKRISHNAN, Maheshwaran et al. Machine criticality assessment for productivity improvement: Smart maintenance decision support. **International Journal of Productivity and Performance Management**, v. 68, n. 5, p. 858–878, 2019. ISSN 17410401. Disponível em: <https://www.emeraldinsight.com/doi/abs/10.1108/IJPPM-03-2018-0091>.

GRACEL, Piotr Łebkowski Jarosław. Concept of Industry 4.0-Related Manufacturing Technology Maturity Model (ManuTech Maturity Model–MTMM). **Decision Making in Manufacturing and Services**, v. 12, 2018. Disponível em: <http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.baztech-e2e2199a-8f2d-47d9-8764-a944dcb44fc9>.

GRAVES, Alex; SCHMIDHUBER, Jürgen. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. **Neural Networks**, v. 18, n. 5-6, p. 602–610, 2005. ISSN 08936080.

GUNGOR, Onat; ROSING, Tajana; AKSANLI, Baris. STEWART: STacking Ensemble for White-Box AdversaRial Attacks Towards more resilient data-driven predictive maintenance. **Computers in Industry**, Elsevier, v. 140, 2022. ISSN 01663615. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166361522000574>.

GUO, Liang et al. A recurrent neural network based health indicator for remaining useful life prediction of bearings. **Neurocomputing**, Elsevier B.V., v. 240, n. 16, p. 98–109, 2017. ISSN 18728286. Disponível em: <http://dx.doi.org/10.1016/j.neucom.2017.02.045>.

HÄCKEL, Björn et al. Assessing IT availability risks in smart factory networks. **Business Research**, Springer, v. 12, n. 2, p. 523–558, 2019. ISSN 21982627.

HAROON, Danish. **Python Machine Learning Solutions**. [s.n.], 2017. ISBN 9781484228227. Disponível em: <https://www.safaribooksonline.com/library/view/python-machine-learning/9781787127692/>.

HASHEMIAN, H. M.; BEAN, Wendell C. State-of-the-art predictive maintenance techniques. **IEEE Transactions on Instrumentation and Measurement**, IEEE, v. 60, n. 10, p. 3480–3492, 2011. ISSN 00189456.

HE, Yihai et al. Cost-oriented predictive maintenance based on mission reliability state for cyber manufacturing systems. **Advances in Mechanical Engineering**, SAGE Publications Inc., v. 10, n. 1, 2018. ISSN 16878140. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85041566509&doi=10.1177%2F1687814017751467&partnerID=40&md5=cb57dc5abfc2c2af58cdf71aaa887843>.

____. Reliability-oriented design of integrated model of preventive maintenance and quality control policy with time-between-events control chart. **Computers and Industrial Engineering**, v. 129, p. 228–238, 2019. ISSN 03608352. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0360835219300518>.

HENNING LUKAS WOLF-DIETER, Wahlster Wofgang Kagermann. **Industrie 4.0: Mitdem Internet der Dinge auf demWegzur 4.industriellen Revolution**. 2011. 13 p. Disponível em: <http://www.wolfgang-wahlster.de/wordpress/wp-content/uploads/Industrie%257B/_%257D4%257B/_%257D0%257B/_%257DMit%257B/_%257Ddem%257B/_%257DInternet%257B/_%257Dder%257B/_%257DDinge%257B/_%257Dauf%257B/_%257Ddem%257B/_%257DWeg%257B/_%257Dzur%257B/_%257Dviert>.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 11 1997. ISSN 08997667.

HU, Liang et al. Feature selection considering two types of feature relevancy and feature interdependency. **Expert Systems with Applications**, Elsevier Ltd, v. 93, p. 423–434, 2018. ISSN 09574174. Disponível em: <https://doi.org/10.1016/j.eswa.2017.10.016>.

ISMAIL, Ahmed; TRUONG, Hong Linh; KASTNER, Wolfgang. Manufacturing process data analysis pipelines: a requirements analysis and survey. **Journal of Big Data**, SpringerOpen, v. 6, n. 1, 2019. ISSN 21961115. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85059746810&doi=10.1186%2Fs40537-018-0162-3&partnerID=40&md5=0ceb5cfbfc4d51b0186152664d1e128d>.

JIA, Feng et al. A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. **Neurocomputing**, Elsevier B.V., v. 272, p. 619–628, 2018. ISSN 18728286. Disponível em: <https://doi.org/10.1016/j.neucom.2017.07.032>.

JIMENEZ, Juan José Montero et al. Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. **Journal of Manufacturing Systems**, Elsevier, v. 56, n. March, p. 539–557, 2020. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2020.07.008>.

Kagermann; WAHLSTER, Wolfgang; HELBIG, Johannes. Recommendations for implementing the strategic initiative INDUSTRIE 4.0. **Final report of the Industrie 4.0 WG**, n. April, p. 82, 2013. ISSN 2405-8963.

KANG, Ziqiu; CATAL, Cagatay; TEKINERDOGAN, Bedir. Machine learning applications in production lines: A systematic literature review. **Computers and Industrial Engineering**, Elsevier Ltd, v. 149, n. April, p. 106773, 2020. ISSN 03608352. Disponível em: <https://doi.org/10.1016/j.cie.2020.106773>.

KASONGO, Sydney Mambwe; SUN, Yanxia. A deep learning method with wrapper based feature extraction for wireless intrusion detection system. **Computers and Security**, Elsevier Ltd, v. 92, 2020. ISSN 01674048.

KIANGALA, Kahiomba Sonia; WANG, Zenghui. Initiating predictive maintenance for a conveyor motor in a bottling plant using industry 4.0 concepts. **International Journal of Advanced Manufacturing Technology**, v. 97, n. 9-12, p. 3251–3271, 8 2018. ISSN 14333015. Disponível em: <http://link.springer.com/10.1007/s00170-018-2093-8>.

KIM, Yoon. Convolutional neural networks for sentence classification. **EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference**, p. 1746–1751, 2014.

KINNUNEN, Sini Kaisu et al. Internet of things in asset management: Insights from industrial professionals and academia. **International Journal of Service Science, Management, Engineering, and Technology**, IGI Global, v. 9, n. 2, p. 104–119, 2018. ISSN 19479603. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85052116110&doi=10.4018%2FIJSSMET.2018040105&partnerID=40&md5=e9ec2dc3db39846e84ec159515ec807d>.

KIRKMANT, Robert A. Failure Concepts in Reliability Theory. **IEEE Transactions on Reliability**, R-12, n. 4, p. 1–10, 1963. ISSN 15581721.

KLAVANS, Richard; BOYACK, Kevin W. Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge? **Journal of the Association for Information Science and Technology**, v. 68, n. 4, p. 984–998, 2017. ISSN 23301643.

KLEIN, Patrick; BERGMANN, Ralph. Data generation with a physical model to support machine learning research for predictive maintenance. **CEUR Workshop Proceedings**, v. 2191, p. 179–190, 2018. ISSN 16130073. Disponível em: <http://ceur-ws.org/Vol-2191/paper22.pdf>.

KLINGENBERG, Cristina Orsolin; BORGES, Marco Antônio Viana; ANTUNES, José Antônio Valle. Industry 4.0 as a data-driven paradigm: a systematic literature review on technologies. **Journal of Manufacturing Technology Management**, v. 32, n. 3, p. 570–592, 2021. ISSN 1741038X.

KŁOS, Sławomir; PATALAS-MALISZEWSKA, Justyna. The use of the simulation method in analysing the performance of a predictive maintenance system. In: **Advances in Intelligent Systems and Computing**. Springer, Cham, 2019. v. 801, p. 42–49. ISBN 9783319996073. Disponível em: <http://link.springer.com/10.1007/978-3-319-99608-0_5>.

KOLINSKA, Karolina; KOLINSKI, Adam. Analysis of Spare Parts in Terms of Their Availability Management for the Production Processes Needs. **Business Logistics in Modern Management**, v. 18, p. 191–204, 2018. Disponível em: <https://hrcak.srce.hr/ojs/index.php/plusm/article/view/7888>.

KONG, Xianguang et al. A multi-ensemble method based on deep auto-encoders for fault diagnosis of rolling bearings. **Measurement: Journal of the International Measurement Confederation**, Elsevier Ltd, v. 151, p. 107132, 2020. ISSN 02632241. Disponível em: <https://doi.org/10.1016/j.measurement.2019.107132>.

KRAUS, Mathias; FEUERRIEGEL, Stefan. Forecasting remaining useful life: Interpretable deep learning approach via variational Bayesian inferences. **Decision Support Systems**, v. 125, 2019. ISSN 01679236.

KRUMEICH, Julian et al. Advanced planning and control of manufacturing processes in steel industry through big data analytics: Case study and architecture proposal. **Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014**, IEEE, p. 16–24, 2014.

KU, Jin Hee. A Study on Prediction Model of Equipment Failure Through Analysis of Big Data Based on RHadoop. **Wireless Personal Communications**, v. 98, n. 4, p. 3163–3176, 2018. ISSN 1572834X. Disponível em: <http://link.springer.com/10.1007/s11277-017-4151-1>.

KUEHN, Wolfgang. Digital twins for decision making in complex production and logistic enterprises. **International Journal of Design and Nature and Ecodynamics**, v. 13, n. 3, p. 260–271, 2018. ISSN 17557445.

KWON, Hweeung; DO, Thai Ngan; KIM, Jiyong. Comprehensive Decision Framework Combining Price Prediction and Production-Planning Models for Strategic Operation of a Petrochemical Industry. **Industrial and Engineering Chemistry Research**, v. 59, n. 25, p. 11610–11620, 2020. ISSN 15205045.

LAMONACA, F. et al. Internet of Things for Structural Health Monitoring. **2018 Workshop on Metrology for Industry 4.0 and IoT, MetroInd 4.0 and IoT 2018 - Proceedings**, n. September, p. 95–100, 2018.

LARRINAGA, Felix et al. Implementation of a Reference Architecture for Cyber Physical Systems to support Condition Based Maintenance. **2018 5th International Conference on Control, Decision and Information Technologies, CoDIT 2018**, p. 773–778, 2018. Disponível em: <http://ebiltegia.mondragon.edu/xmlui/handle/20.500.11984/1174>.

LAZAROVA-MOLNAR, Sanja; MOHAMED, Nader. Reliability assessment in the context of industry 4.0: Data as a game changer. **Procedia Computer Science**, Elsevier B.V., v. 151, n. 2018, p. 691–698, 2019. ISSN 18770509. Disponível em: <https://doi.org/10.1016/j.procs.2019.04.092>.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. ISSN 14764687.

LECUN, Yann; BENGIO, Yoshua et al. Convolutional networks for images, speech, and time series. **The handbook of brain theory and neural networks**, Cambridge, MA USA, v. 3361, n. 1, p. 14, 1995. ISSN 17416485.

LECUN, Yann et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2323, 1998. ISSN 00189219.

LECUN, Yann; KAVUKCUOGLU, Koray; FARABET, Clément. Convolutional networks and applications in vision. **ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems**, p. 253–256, 2010.

LEE, C. K.M.; ZHANG, S. Z.; NG, K. K.H. Development of an industrial Internet of things suite for smart factory towards re-industrialization. **Advances in Manufacturing**, v. 5, n. 4, p. 335–343, 12 2017. ISSN 21953597. Disponível em: <http://link.springer.com/10.1007/s40436-017-0197-2>.

LEE, Hyunsoo. Development of real-time sketch-based on-the-spot process modeling and analysis system. **Journal of Manufacturing Systems**, Elsevier, v. 54, n. December 2019, p. 215–226, 2020. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2019.12.006>.

LEE, Jay et al. Industrial Big Data Analytics and Cyber-physical Systems for Future Maintenance & Service Innovation. **Procedia CIRP**, v. 38, p. 3–7, 2015. ISSN 22128271. Disponível em: <http://www.sciencedirect.com/science/article/pii/S2212827115008744>.

____. Recent advances and trends in predictive manufacturing systems in big data environment. **Manufacturing Letters**, Society of Manufacturing Engineers (SME), v. 1, n. 1, p. 38–41, 2013. ISSN 22138463. Disponível em: <http://dx.doi.org/10.1016/j.mfglet.2013.09.005>.

____. Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications. **Mechanical Systems and Signal Processing**, Elsevier, v. 42, n. 1-2, p. 314–334, 2014. ISSN 08883270. Disponível em: <http://dx.doi.org/10.1016/j.ymssp.2013.06.004>.

LI, Bao rui et al. Framework and case study of cognitive maintenance in Industry 4.0. **Frontiers of Information Technology and Electronic Engineering**, v. 20, n. 11, p. 1493–1504, 11 2019. ISSN 20959230. Disponível em: <http://link.springer.com/10.1631/FITEE.1900193>.

LI, Defang. Perspective for smart factory in petrochemical industry. **Computers and Chemical Engineering**, v. 91, p. 136–148, 2016. ISSN 00981354. Disponível em: <https://www.sciencedirect.com/science/article/pii/S009813541630059X>.

LI, Xiang et al. Deep learning-based adversarial multi-classifier optimization for cross-domain machinery fault diagnostics. **Journal of Manufacturing Systems**, v. 55, n. November 2019, p. 334–347, 2020. ISSN 02786125.

LIM, Kendrik Yan Hong; ZHENG, Pai; CHEN, Chun Hsien. A state-of-the-art survey of Digital Twin: techniques, engineering product lifecycle management and business innovation perspectives. **Journal of Intelligent Manufacturing**, Springer Science and Business Media LLC, v. 31, n. 6, p. 1313–1337, 2020. ISSN 15728145.

LIN, Yi Kuei; CHANG, Ping Chen. System reliability of a manufacturing network with reworking action and different failure rates. **International Journal of Production Research**, v. 50, n. 23, p. 6930–6944, 2012. ISSN 00207543. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/00207543.2011.638939>.

LIU, Chao et al. A systematic development method for cyber-physical machine tools. **Journal of Manufacturing Systems**, The Society of Manufacturing Engineers, v. 48, p. 13–24, 2018. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2018.02.001>.

LIU, He et al. A generalized cauchy method for remaining useful life prediction of wind turbine gearboxes. **Mechanical Systems and Signal Processing**, Elsevier Ltd, v. 153, p. 107471, 2021. ISSN 10961216. Disponível em: <https://doi.org/10.1016/j.ymssp.2020.107471>.

LU, Yang. Industry 4.0: A survey on technologies, applications and open research issues. **Journal of Industrial Information Integration**, Elsevier Inc., v. 6, p. 1–10, 2017. ISSN 2452414X.

LUNDBERG, Scott M. et al. From local explanations to global understanding with explainable AI for trees. **Nature Machine Intelligence**, Nature Publishing Group, v. 2, n. 1, p. 56–67, 2020. ISSN 25225839. Disponível em: <http://dx.doi.org/10.1038/s42256-019-0138-9>.

LUNDBERG, Scott M.; ERION, Gabriel G.; LEE, Su-In. Consistent Individualized Feature Attribution for Tree Ensembles. **arXiv preprint arXiv:1802.03888**, n. 2, 2018. Disponível em: <http://arxiv.org/abs/1802.03888>.

MA, Meng; MAO, Zhu. Deep-Convolution-Based LSTM Network for Remaining Useful Life Prediction. **IEEE Transactions on Industrial Informatics**, v. 17, n. 3, p. 1658–1667, 2021. ISSN 19410050.

MACCHI, Marco et al. Maintenance management of railway infrastructures based on reliability analysis. **Reliability Engineering and System Safety**, Elsevier, v. 104, p. 71–83, 2012. ISSN 09518320. Disponível em: <http://dx.doi.org/10.1016/j.ress.2012.03.017>.

MADHIKERMI, Manik et al. Key data quality pitfalls for condition based maintenance. **2017 2nd International Conference on System Reliability and Safety, ICSRS 2017**, v. 2018-Janua, p. 474–480, 2018.

MADNI, Azad M.; MADNI, Carla C.; LUCERO, Scott D. Leveraging digital twin technology in model-based systems engineering. **Systems**, v. 7, n. 1, p. 7, 2019. ISSN 20798954. Disponível em: <https://www.mdpi.com/2079-8954/7/1/7>.

MAKRIDAKIS, Spyros; SPILIOTIS, Evangelos; ASSIMAKOPOULOS, Vassilios. M5 accuracy competition: Results, findings, and conclusions. **International Journal of Forecasting**, v. 38, n. 4, p. 1346–1364, 2022. ISSN 01692070.

MAO, Wentao et al. Online detection of bearing incipient fault with semi-supervised architecture and deep feature representation. **Journal of Manufacturing Systems**, Elsevier, v. 55, n. February, p. 179–198, 2020. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2020.03.005>.

MÁRQUEZ, Adolfo Crespo et al. Designing CBM Plans, Based on Predictive Analytics and Big Data Tools, for Train Wheel Bearings. **Computers in Industry**, v. 122, 2020. ISSN 01663615.

MCA, VijayaRamaraju Poosapati; MCA, Vedavathi Katneni; MANDA, Vijay Killu. Super SCADA Systems: A Prototype for Next Gen SCADA System. **Iaetsdjaras.Org**, v. 5, n. 3, p. 107–115, 2018. Disponível em: <http://iaetsdjaras.org/gallery/18-march-559.pdf>.

MEHDIYEV, Nijat et al. iPRODICT - Intelligent process prediction based on big data analytics. **CEUR Workshop Proceedings**, v. 1985, p. 13–24, 2017. ISSN 16130073. Disponível em: <https://pdfs.semanticscholar.org/469e/584be90e70cd7b77d43c2f23b5f3c0787bd1.pdf>.

MEJÍA, Gonzalo; PEREIRA, Jordi. Multiobjective scheduling algorithm for flexible manufacturing systems with Petri nets. **Journal of Manufacturing Systems**, Elsevier, v. 54, n. December 2019, p. 272–284, 2020. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2020.01.003>.

MIN, Qingfei et al. Machine Learning based Digital Twin Framework for Production Optimization in Petrochemical Industry. **International Journal of Information Management**, Elsevier, v. 49, n. May, p. 502–519, 2019. ISSN 02684012. Disponível em: <https://doi.org/10.1016/j.ijinfomgt.2019.05.020>.

MITCHELL, Rory; FRANK, Eibe. Accelerating the XGBoost algorithm using GPU computing. **PeerJ Computer Science**, v. 2017, n. 7, 2017. ISSN 23765992.

MOHAMED, Nader; AL-JAROODI, Jameela. Applying blockchain in industry 4.0 applications. **2019 IEEE 9th Annual Computing and Communication Workshop and Conference, CCWC 2019**, p. 852–858, 2019. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8666558/>.

MOKHTARI, Sohrab et al. A machine learning approach for anomaly detection in industrial control systems based on measurement data. **Electronics (Switzerland)**, v. 10, n. 4, p. 1–13, 2021. ISSN 20799292.

MOORTHY, Usha; GANDHI, Usha Devi. Forest optimization algorithm-based feature selection using classifier ensemble. **Computational Intelligence**, v. 36, n. 4, p. 1445–1462, 2020. ISSN 14678640.

MOSAVI, Amir; LOPEZ, Alvaro; VARKONYI-KOCZY, Annamária R. Industrial applications of big data: State of the art survey. **Advances in Intelligent Systems and Computing**, v. 660, p. 225–232, 2018. ISSN 21945357. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-319-67459-9_29>.

MOURTZIS, Dimitris; VLACHOU, Ekaterini. A cloud-based cyber-physical system for adaptive shop-floor scheduling and condition-based maintenance. **Journal of Manufacturing Systems**, v. 47, p. 179–198, 2018. ISSN 02786125. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0278612518300700>.

MULUBRHAN, Freselam; MOKHTAR, Ainul Akmar; MUHAMMAD, Masdi. Integrating reliability analysis in life cycle cost estimation of heat exchanger and pump. **Advanced Materials Research**, v. 903, p. 408–413, 2014. ISSN 10226680.

MUTLU, Nazli G.; ALTUNTAS, Serkan. Assessment of occupational risks In Turkish manufacturing systems with data-driven models. **Journal of Manufacturing Systems**, Elsevier, v. 53, n. October, p. 169–182, 2019. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2019.09.008>.

MYERS, A.; TICKEM, D.; EVANS, J. People centred intelligent predict and prevent (PCIPP) a novel approach to remote condition monitoring. In: **IET Conference Publications**. [S.l.: s.n.], 2016. v. 2016, n. CP701, p. 9 (6 .)–9 (6 .). ISBN 9781785611384.

NAKAYAMA, Ruy Somei; SPÍNOLA, Mauro de Mesquita; SILVA, José Reinaldo. Towards I4.0: A comprehensive analysis of evolution from I3.0. **Computers and Industrial Engineering**, Elsevier, v. 144, n. April, p. 106453, 2020. ISSN 03608352. Disponível em: <https://doi.org/10.1016/j.cie.2020.106453>.

NAPOLEONE, Alessia; MACCHI, Marco; POZZETTI, Alessandro. A review on the characteristics of cyber-physical systems for the future smart factories. **Journal of Manufacturing Systems**, Elsevier, v. 54, n. August 2019, p. 305–335, 2020. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2020.01.007>.

NAUKKARINEN, Ossi; BRAGGE, Johanna. Aesthetics in the age of digital humanities. **Journal of Aesthetics and Culture**, v. 8, n. January, 2016. ISSN 20004214.

NEMETH, Tanja et al. PriMa-X: A reference model for realizing prescriptive maintenance and assessing its maturity enhanced by machine learning. **Procedia CIRP**, v. 72, p. 1039–1044, 2018. ISSN 22128271. Disponível em: <https://publik.tuwien.ac.at/files/publik_271447.pdf>.

NGUYEN, Dang Trinh et al. Fault diagnosis for the complex manufacturing system. **Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability**, v. 230, n. 2, p. 178–194, 2016. ISSN 17480078.

NING, Fangwei et al. Manufacturing cost estimation based on a deep-learning method. **Journal of Manufacturing Systems**, Elsevier, v. 54, n. December 2019, p. 186–195, 2020. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2019.12.005>.

OMRI, N. et al. Industrial data management strategy towards an SME-oriented PHM. **Journal of Manufacturing Systems**, Elsevier, v. 56, n. November 2019, p. 23–36, 2020. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2020.04.002>.

ORESHKIN, Boris N. et al. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. **arXiv preprint arXiv:1905.10437**, p. 1–31, 2019. ISSN 2331-8422. Disponível em: <http://arxiv.org/abs/1905.10437>.

____. N-BEATS neural network for mid-term electricity load forecasting. **Applied Energy**, Elsevier Ltd, v. 293, n. April, p. 116918, 2021. ISSN 03062619. Disponível em: <https://doi.org/10.1016/j.apenergy.2021.116918>.

O'DONOVAN, Peter et al. Big data in manufacturing: a systematic mapping study. **Journal of Big Data**, v. 2, n. 1, 2015. ISSN 21961115. Disponível em: <https://www.biomedcentral.com/openurl?doi=10.1186/s40537-015-0028-x>.

PALAU, Adrià Salvador; DHADA, Maharshi Harshadbhai; PARLIKAD, Ajith Kumar. Multi-agent system architectures for collaborative prognostics. **Journal of Intelligent Manufacturing**, Springer New York LLC, v. 30, n. 8, p. 2999–3013, 12 2019. ISSN 15728145.

PANDARAKONE, Shrinathan Esakimuthu; MIZUNO, Yukio; NAKAMURA, Hisahide. Algorithm and Artificial Intelligence Neural Network. **Energies**, v. 12, p. 2105, 2019.

PEDAMONTI, Dabal. Comparison of non-linear activation functions for deep neural networks on MNIST classification task. **arXiv**, n. 3, 2018. Disponível em: <http://arxiv.org/abs/1804.02763>.

PÉREZ-LARA, Magdiel et al. Organizational systems convergence with the industry 4.0 challenge. In: **Best Practices in Manufacturing Processes: Experiences from Latin America**. Cham: Springer International Publishing, 2018. p. 411–431. ISBN 9783319991900. Disponível em: <http://link.springer.com/10.1007/978-3-319-99190-0_19>.

PERNO, Matteo; HVAM, Lars; HAUG, Anders. Implementation of digital twins in the process industry: A systematic literature review of enablers and barriers. **Computers in Industry**, Elsevier, v. 134, 2022. ISSN 01663615. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166361521001652>.

PREUVENEERS, Davy; JOOSEN, Wouter; ILIE-ZUDOR, Elisabeth. Robust Digital Twin Compositions for Industry 4.0 Smart Manufacturing Systems. **Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOCW**, v. 2018-Octob, p. 69–78, 2018. ISSN 15417719. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8536107/>.

QI, Yanjun. Ensemble Machine Learning. **Ensemble Machine Learning**, p. 307–323, 2012.

QIAO, Zhihua et al. PVAm–PIP/PS composite membrane with high performance for CO₂/N₂ separation. **AIChE Journal**, v. 59, n. 4, p. 215–228, 2012. ISSN 12350621.

RAUCH, Erwin; LINDER, Christian; DALLASEGA, Patrick. Anthropocentric perspective of production before and within Industry 4.0. **Computers and Industrial Engineering**, v. 139, 1 2020. ISSN 03608352. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0360835219300233>.

REHMAN, Muhammad Habib Ur et al. Big data reduction framework for value creation in sustainable enterprises. **International Journal of Information Management**, Elsevier Ltd, v. 36, n. 6, p. 917–928, 2016. ISSN 02684012. Disponível em: <http://dx.doi.org/10.1016/j.ijinfomgt.2016.05.013>.

REN, Lei et al. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. **Journal of Manufacturing Systems**, Elsevier, v. 48, n. November 2017, p. 71–77, 2018. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2018.04.008>.

RIMPAULT, Xavier; BALAZINSKI, Marek; CHATELAIN, Jean François. Fractal analysis application outlook for improving process monitoring and machine maintenance in manufacturing 4.0. **Journal of Manufacturing and Materials Processing**, v. 2, n. 3, p. 62, 2018. ISSN 25044494.

ROCHA, Lucio A; BARRETO, Fernando; SEMAN, Laio O. **The Internet of Things in the Industrial Sector**. Springer International Publishing, 2019. 326 p. ISBN 978-3-030-24891-8. Disponível em: <http://dx.doi.org/10.1007/978-3-030-24892-5_9%0Ahttp://link.springer.com/10.1007/978-3-030-24892-5>.

RODRÍGUEZ-MAZAHUA, Lisbeth et al. A general perspective of Big Data: applications, tools, challenges and trends. **Journal of Supercomputing**, v. 72, n. 8, p. 3073–3113, 8 2016. ISSN 15730484. Disponível em: <http://link.springer.com/10.1007/s11227-015-1501-1>.

ROSATI, Riccardo et al. From knowledge-based to big data analytic model: a novel IoT and machine learning based decision support system for predictive maintenance in Industry 4.0. **Journal of Intelligent Manufacturing**, Springer US, 2022. ISSN 15728145. Disponível em: <https://doi.org/10.1007/s10845-022-01960-x>.

ROSEN, Roland et al. ScienceDirect About The for. **IFAC-PapersOnLine**, p. 567–572, 2015.

ROSSIT, Daniel Alejandro; TOHMÉ, Fernando; FRUTOS, Mariano. A data-driven scheduling approach to smart manufacturing. **Journal of Industrial Information Integration**, Departamento de Ingeniería, Universidad Nacional del Sur, Av. Alem 1253, Bahía Blanca, Buenos Aires CP 8000, Argentina, v. 15, p. 69–79, 2019. ISSN 2452414X. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2452414X18300475>.

ROY, R. et al. Continuous maintenance and the future – Foundations and technological challenges. **CIRP Annals - Manufacturing Technology**, CIRP, v. 65, n. 2, p. 667–688, 2016. ISSN 17260604.

RUBMANN, Michael et al. **Industry 4.0: World Economic Forum**. [S.l.], 2015. 1–20 p.

RUIZ-SARMIENTO, J. R. et al. Analysis of data from the industrial machinery ithin the hot rolling process for predictive maintenance. In: **Frontiers in Artificial Intelligence and Applications**. Machine Perception and Intelligent Robotics Group, System Engineering and Auto. Dept., University of Malaga, SpainInstitute of Biomedical Research of Malaga (IBIMA), SpainACERINOX Europa S.A.U., Spain: IOS Press, 2018. (Frontiers in Artificial Intelligence and Applications, v. 310), p. 122–133. ISBN 9781614999287. ISSN 09226389. Disponível em: <http://dx.doi.org/10.3233/978-1-61499-929-4-122>.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, 1986. ISSN 00280836.

RYCK, M. De; VERSTEYHE, M.; SHARIATMADAR, K. Resource management in decentralized industrial Automated Guided Vehicle systems. **Journal of Manufacturing Systems**, Elsevier, v. 54, n. June 2019, p. 204–214, 2020. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2019.11.003>.

SAGE, Andrew P; ROUSE, William B. **Handbook of systems engineering and management**. [S.l.]: John Wiley {\&} Sons, Inc., 1999. 378 p. ISSN 02786125. ISBN 9783540773405.

SAHAL, Radhya; BRESLIN, John G.; ALI, Muhammad Intizar. Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case. **Journal of Manufacturing Systems**, Elsevier, v. 54, n. December 2019, p. 138–151, 2020. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2019.11.004>.

SALAZAR, Luis Alberto Cruz et al. Cyber-physical production systems architecture based on multi-agent's design pattern—comparison of selected approaches mapping four agent patterns. **International Journal of Advanced Manufacturing Technology**, Springer London, v. 105, n. 9, p. 4005–4034, 2019. ISSN 14333015.

SALUNKHE, Omkar; FUMERO, Patricia Quijada. Demonstration of real-time criticality assessment using a test-bed. **Chalmers Master thesis**, 2017. Disponível em: <http://publications.lib.chalmers.se/records/fulltext/250242/250242.pdf>.

SANDENGEN, Odd Christian et al. High Performance Manufacturing - An Innovative Contribution towards Industry 4.0. **atlantis-press.com**, 2016. Disponível em: <https://www.atlantis-press.com/proceedings/iwama-16/25862212>.

SAWANT, Nitin; SHAH, Himanshu. Big Data Application Architecture Q & A. **Big Data Application Architecture Q & A**, 2013.

SAYRES, Rory et al. Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. In: **Ophthalmology**. Red Hook, NY, USA: Curran Associates Inc., 2019. (NIPS'17, 4), p. 552–564. ISBN 9781510860964. ISSN 15494713.

SCAPOLO, Fabiana et al. JRC Foresight Study: How will standards facilitate new production systems in the context of EU innovation and competitiveness in 2025 ? Final Report. **Core.Ac.Uk**, p. 144, 2014. Disponível em: <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC93699/jrc_27ap15_2rep_web.pdf>.

SCHEER, August-wilhelm. **Enterprise 4.0-From disruptive business model to the automation of business processes**. [S.l.]: Saarbrucken: AWSi Publish, 2019. ISBN 9783981991604.

SCHREIBER, Martin et al. Integrated production and maintenance planning in cyber-physical production systems. **Procedia CIRP**, Elsevier B.V., v. 79, p. 534–539, 2019. ISSN 22128271. Disponível em: <https://doi.org/10.1016/j.procir.2019.02.095>.

SCHWARTZ, Sébastien et al. An unsupervised approach for health index building and for similarity-based remaining useful life estimation. **Computers in Industry**, Elsevier, v. 141, 2022. ISSN 01663615. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166361522001130>.

SELLITTO, Miguel Afonso. Formulação estratégica da manutenção industrial com base na confiabilidade dos equipamentos. **Production**, v. 15, n. 1, p. 44–59, 2005. ISSN 1980-5411.

SÉNÉCHAL, Olivier; TRENTESAUX, Damien. A framework to help decision makers to be environmentally aware during the maintenance of cyber physical systems. **Environmental Impact Assessment Review**, v. 77, p. 11–22, 2019. ISSN 01959255. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0195925517304882>.

SEZER, Erim et al. An Industry 4.0-Enabled Low Cost Predictive Maintenance Approach for SMEs. **2018 IEEE International Conference on Engineering, Technology and Innovation, ICE/ITMC 2018 - Proceedings**, p. 8, 2018. Disponível em: <https://dspace.lib.cranfield.ac.uk/handle/1826/14010>.

SHAO, Haidong et al. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. **Mechanical Systems and Signal Processing**, Elsevier Ltd, v. 95, p. 187–204, 2017. ISSN 10961216. Disponível em: <http://dx.doi.org/10.1016/j.ymssp.2017.03.034>.

SHCHERBAKOV, Maxim; SAI, Cuong. A Hybrid Deep Learning Framework for Intelligent Predictive Maintenance of Cyber-physical Systems. **ACM Transactions on Cyber-Physical Systems**, v. 6, n. 2, 2022. ISSN 23789638.

SHIHUNDLA, Trevor Bogani; MPOFU, Khumbulani; ADENUGA, Olukorede Tijani. Integrating product-service systems into the manufacturing industry: Industry 4.0 perspectives. **Procedia CIRP**, Elsevier B.V., v. 83, p. 8–13, 2019. ISSN 22128271. Disponível em: <https://doi.org/10.1016/j.procir.2019.03.147>.

SIAFARA, Lydia C. et al. SAMBA: A self-aware health monitoring architecture for distributed industrial systems. **Proceedings IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society**, v. 2017-Janua, p. 3512–3517, 2017. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8216594/>.

SINGHAL, Vatsal; MATHEW, Jimson. Fusion of Wavelet Decomposition and N-BEATS for improved Stock Market Forecasting. **Research Square**, p. 0–21, 2022. Disponível em: <https://www.researchsquare.com/article/rs-2003731/v1?utm_source=researcher_app&utm_medium=referral&utm_campaign=RESR_MRKT_Researcher_inbound>.

SOCHER, Richard et al. Convolutional-recursive deep learning for 3D object classification. **Advances in Neural Information Processing Systems**, v. 1, n. i, p. 656–664, 2012. ISSN 10495258.

SONY, Michael; NAIK, Subhash. Industry 4.0 integration with socio-technical systems theory: A systematic review and proposed theoretical model. **Technology in Society**, Elsevier Ltd, v. 61, n. August 2019, p. 101248, 2020. ISSN 0160791X. Disponível em: <https://doi.org/10.1016/j.techsoc.2020.101248>.

SOUZA, Marcos Leandro Hoffmann et al. A survey on decision-making based on system reliability in the context of Industry 4.0. **Journal of Manufacturing Systems**, v. 56, p. 133–156, 2020. ISSN 02786125.

____. A feature identification method to explain anomalies in condition monitoring. **Computers in Industry**, Elsevier, v. 133, 2021. ISSN 01663615. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166361521001354>.

SOUZA, Marcos Leandro Hoffmann; RODRIGUES, Luis Henrique; MORANDI, Maria Isabel Wolf Motta. Design of a System Dynamics Model to Analyze the Styrene Demand in the Brazilian Market. **Systemic Practice and Action Research**, Systemic Practice and Action Research, v. 31, n. 1, p. 87–104, 2018. ISSN 1094429X.

STEVENSON, Emma et al. A deep learning approach to solar radio flux forecasting. **Acta Astronautica**, Elsevier Ltd, v. 193, n. August 2021, p. 595–606, 2022. ISSN 00945765.

STRAKA, M. et al. Design of large-scale logistics systems using computer simulation hierarchic structure. **International Journal of Simulation Modelling**, v. 17, n. 1, p. 105–118, 2018. ISSN 19968566.

SUTSKEVER, Ilya; VINYALS, Oriol; LE, Quoc V. Sequence to sequence learning with neural networks. **Advances in Neural Information Processing Systems**, v. 4, n. January, p. 3104–3112, 2014. ISSN 10495258.

SYAFRUDIN, Muhammad et al. Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. **Sensors (Switzerland)**, v. 18, n. 9, 2018. ISSN 14248220. Disponível em: <https://www.mdpi.com/1424-8220/18/9/2946>.

SZALAVETZ, Andrea. The environmental impact of advanced manufacturing technologies: Examples from Hungary. **Central European Business Review**, v. 6, n. 2, p. 18–29, 2017. ISSN 18054862. Disponível em: <https://www.ceeol.com/search/article-detail?id=562089>.

TAN, Julian S.K. et al. Quality Analytics in a Big Data supply chain: Commodity data analytics for quality engineering. **IEEE Region 10 Annual International Conference, Proceedings/TENCON**, p. 3455–3463, 2017. ISSN 21593450. Disponível em: <https://ieeexplore.ieee.org/abstract/document/7848697/>.

TERRISSA, Labib Sadek et al. A new approach of PHM as a service in cloud computing. **Colloquium in Information Science and Technology, CIST**, v. 0, p. 610–614, 2016. ISSN 23271884.

TEWARI, Saurabh; DWIVEDI, U. D. Ensemble-based big data analytics of lithofacies for automatic development of petroleum reservoirs. **Computers and Industrial Engineering**, Elsevier, v. 128, n. August 2018, p. 937–947, 2019. ISSN 03608352. Disponível em: <https://doi.org/10.1016/j.cie.2018.08.018>.

THOPPIL, Nikhil M.; VASU, V.; RAO, C. S.P. Failure Mode Identification and Prioritization Using FMECA: A Study on Computer Numerical Control Lathe for Predictive Maintenance. **Journal of Failure Analysis and Prevention**, Springer New York LLC, v. 19, n. 4, p. 1153–1157, 2019. ISSN 18641245.

TSAI, Chih Fong; SUNG, Ya Ting. Ensemble feature selection in high dimension, low sample size datasets: Parallel and serial combination approaches. **Knowledge-Based Systems**, Elsevier B.V., v. 203, p. 106097, 2020. ISSN 09507051. Disponível em: <https://doi.org/10.1016/j.knosys.2020.106097>.

TSAO, Yu Chung et al. Imperfect economic production quantity models under predictive maintenance and reworking. **International Journal of Systems Science: Operations and Logistics**, v. 7, n. 4, p. 347–360, 3 2020. ISSN 23302682. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/23302674.2019.1590663>.

UHLMANN, Iracyanne Retto; FRAZZON, Enzo Morosini. Production rescheduling review: Opportunities for industrial integration and practical applications. **Journal of Manufacturing Systems**, v. 49, p. 186–193, 2018. ISSN 02786125. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0278612518304308>.

USTUNDAG, Alp; CEVIKCAN, Emre. Industry 4.0: Managing The Digital Transformation. **Springer Series in Advanced Manufacturing**, n. January, p. 1–283, 2018. ISSN 1860-5168. Disponível em: <https://link.springer.com/content/pdf/10.1007/978-3-319-57870-5.pdf%0Ahttp://link.springer.com/10.1007/978-3-319-57870-5%0Ahttp://www.springer.com/series/7113%0Ahttps://www.researchgate.net/publication/322172971%0Ahttp://link.springer.com/10.1007/978-3-31>.

VERMA, Pushpak; CHANDRA, Tej Bahadur; DWIVEDI, A. K. Network security in big data: Tools and techniques. **Advances in Intelligent Systems and Computing**, v. 433, n. 8, p. 255–262, 2016. ISSN 21945357.

VOGL, Gregory W.; WEISS, Brian A.; HELU, Moneer. A review of diagnostic and prognostic capabilities and best practices for manufacturing. **Journal of Intelligent Manufacturing**, Springer New York LLC, v. 30, n. 1, p. 79–95, 2019. ISSN 15728145. Disponível em: <http://link.springer.com/10.1007/s10845-016-1228-8>.

WAN, Shan et al. A collaborative machine tool maintenance planning system based on content management technologies. **International Journal of Advanced Manufacturing Technology**, v. 94, n. 5-8, p. 1639–1653, 2 2018. ISSN 14333015. Disponível em: <http://link.springer.com/10.1007/s00170-016-9829-0>.

WANG, Jinjiang et al. Deep learning for smart manufacturing: Methods and applications. **Journal of Manufacturing Systems**, The Society of Manufacturing Engineers, v. 48, p. 144–156, 2018. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2018.01.003>.

____. A new paradigm of cloud-based predictive maintenance for intelligent manufacturing. **Journal of Intelligent Manufacturing**, v. 28, n. 5, p. 1125–1137, 2017. ISSN 15728145. Disponível em: <http://link.springer.com/10.1007/s10845-015-1066-0>.

WANG, Junliang; ZHENG, Peng; ZHANG, Jie. Big data analytics for cycle time related feature selection in the semiconductor wafer fabrication system. **Computers and Industrial Engineering**, Elsevier, v. 143, n. December 2019, p. 106362, 2020. ISSN 03608352. Disponível em: <https://doi.org/10.1016/j.cie.2020.106362>.

WANG, Shiyong et al. Implementing Smart Factory of Industrie 4.0: An Outlook. **International Journal of Distributed Sensor Networks**, v. 2016, 2016. ISSN 15501477. Disponível em: <https://www.engineeringvillage.com/share/document.url?mid=inspec_606eee4915482d89354M635110178163171&database=ins>.

WEN, Pengfei et al. A generalized remaining useful life prediction method for complex systems based on composite health indicator. **Reliability Engineering and System Safety**, Elsevier Ltd, v. 205, n. December 2019, p. 107241, 2021. ISSN 09518320. Disponível em: <https://doi.org/10.1016/j.ress.2020.107241>.

WU, Ji Yan et al. A joint classification-regression method for multi-stage remaining useful life prediction. **Journal of Manufacturing Systems**, Elsevier Ltd, v. 58, n. PA, p. 109–119, 2021. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2020.11.016>.

XIA, Min et al. Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks. **IEEE/ASME Transactions on Mechatronics**, IEEE, v. 23, n. 1, p. 101–110, 2018. ISSN 10834435.

XIA, Tangbin et al. An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation. **Computers in Industry**, Elsevier B.V., v. 115, p. 103182, 2020. ISSN 01663615. Disponível em: <https://doi.org/10.1016/j.compind.2019.103182>.

XIA, Tangbin; XI, Lifeng. Manufacturing paradigm-oriented PHM methodologies for cyber-physical systems. **Journal of Intelligent Manufacturing**, v. 30, n. 4, p. 1659–1672, 2019. ISSN 15728145. Disponível em: <http://link.springer.com/10.1007/s10845-017-1342-2>.

XU, Gaowei et al. Data-driven fault diagnostics and prognostics for predictive maintenance: A brief overview. In: **IEEE International Conference on Automation Science and Engineering**. IEEE, 2019. v. 2019-Augus, p. 103–108. ISBN 9781728103556. ISSN 21618089. Disponível em: <https://ieeexplore.ieee.org/document/8843068/>.

YAN, Jihong et al. Big-data-driven based intelligent prognostics scheme in industry 4.0 environment. **2017 Prognostics and System Health Management Conference, PHM-Harbin 2017 - Proceedings**, 2017. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8079310/>.

YANG, Chuangyan et al. A novel based-performance degradation indicator RUL prediction model and its application in rolling bearing. **ISA Transactions**, Elsevier Ltd, v. 121, p. 349–364, 2022. ISSN 00190578. Disponível em: <https://doi.org/10.1016/j.isatra.2021.03.045>.

YANG, Jilei. Fast TreeSHAP: Accelerating SHAP Value Computation for Trees. **arXiv preprint arXiv:2109.09847**, 2021. Disponível em: <http://arxiv.org/abs/2109.09847>.

YANG, Jian Bo et al. Deep convolutional neural networks on multichannel time series for human activity recognition. **IJCAI International Joint Conference on Artificial Intelligence**, v. 2015-Janua, n. Ijcai, p. 3995–4001, 2015. ISSN 10450823.

YANG, Tianji; ZHENG, Zeyu; QI, Liang. A method for degradation prediction based on Hidden Semi-Markov models with mixture of Kernels. **Computers in Industry**, Elsevier, v. 122, 2020. ISSN 01663615. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0166361520305297>.

YOKOYAMA, Atsushi. Innovative Changes for Maintenance of Railway by Using ICT-To Achieve "smart Maintenance". **Procedia CIRP**, Elsevier B.V., v. 38, p. 24–29, 2015. ISSN 22128271. Disponível em: <http://dx.doi.org/10.1016/j.procir.2015.07.074>.

YU, Wennian; KIM, Il Yong; MECHEFSKE, Chris. Analysis of different RNN autoencoder variants for time series classification and machine prognostics. **Mechanical Systems and Signal Processing**, Elsevier Ltd, v. 149, p. 107322, 2021. ISSN 10961216. Disponível em: <https://doi.org/10.1016/j.ymssp.2020.107322>.

ZEILER, Matthew D.; FERGUS, Rob. Visualizing and understanding convolutional networks. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 8689 LNCS, n. PART 1, p. 818–833, 2014. ISSN 16113349.

ZHANG, Kang et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. **Cell**, Elsevier Inc., v. 181, n. 6, p. 1423–1433, 2020. ISSN 10974172. Disponível em: <https://doi.org/10.1016/j.cell.2020.04.045>.

ZHANG, Qinghua et al. Vibration sensor based intelligent fault diagnosis system for large machine unit in petrochemical industries. **International Journal of Distributed Sensor Networks**, v. 2015, n. i, 2015. ISSN 15501477.

ZHANG, Xuewen et al. Time-Series Regeneration with Convolutional Recurrent Generative Adversarial Network for Remaining Useful Life Estimation. **IEEE Transactions on Industrial Informatics**, v. 17, n. 10, p. 6820–6831, 2021. ISSN 19410050.

ZHANG, Yuyan et al. Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning. **Journal of Manufacturing Systems**, Elsevier, v. 48, n. April, p. 34–50, 2018. ISSN 02786125. Disponível em: <https://doi.org/10.1016/j.jmsy.2018.04.005>.

ZHANG, Yue; SUZUKI, Genki; SHIOYA, Hiroyuki. Prediction and Detection of Sewage Treatment Process Using N-BEATS Autoencoder Network. **IEEE Access**, IEEE, v. 10, n. October, p. 112594–112608, 2022. ISSN 21693536.

ZHAO, Dezun; WANG, Tianyang; CHU, Fulei. Deep convolutional neural network based planet bearing fault classification. **Computers in Industry**, Elsevier B.V., v. 107, p. 59–66, 2019. ISSN 01663615. Disponível em: <https://doi.org/10.1016/j.compind.2019.02.001>.

ZHENG, Pai et al. Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives. **Frontiers of Mechanical Engineering**, v. 13, n. 2, p. 137–150, 6 2018. ISSN 20950241. Disponível em: <http://link.springer.com/10.1007/s11465-018-0499-5>.

ZHONG, Ray Y. et al. Intelligent Manufacturing in the Context of Industry 4.0: A Review. **Engineering**, v. 3, n. 5, p. 616–630, 2017. ISSN 20958099. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2095809917307130>.

ZHOU, Keliang; LIU, Taigang; ZHOU, Lifeng. Industry 4.0: Towards future industrial opportunities and challenges. **2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015**, IEEE, p. 2147–2152, 2016.

ZIEGEL, Eric R. **System Reliability Theory: Models, Statistical Methods, and Applications**. [S.l.]: John Wiley {\&} Sons, Inc., 2004. 495–496 p. ISSN 0040-1706.

ZIO, Enrico. Some challenges and opportunities in reliability engineering. **IEEE Transactions on Reliability**, v. 65, n. 4, p. 1769–1782, 12 2016. ISSN 00189529.

ZOU, Yisheng et al. A method for predicting the remaining useful life of rolling bearings under different working conditions based on multi-domain adversarial networks. **Measurement: Journal of the International Measurement Confederation**, Elsevier Ltd, v. 188, n. October 2021, p. 110393, 2022. ISSN 02632241. Disponível em: <https://doi.org/10.1016/j.measurement.2021.110393>.

APPENDIX

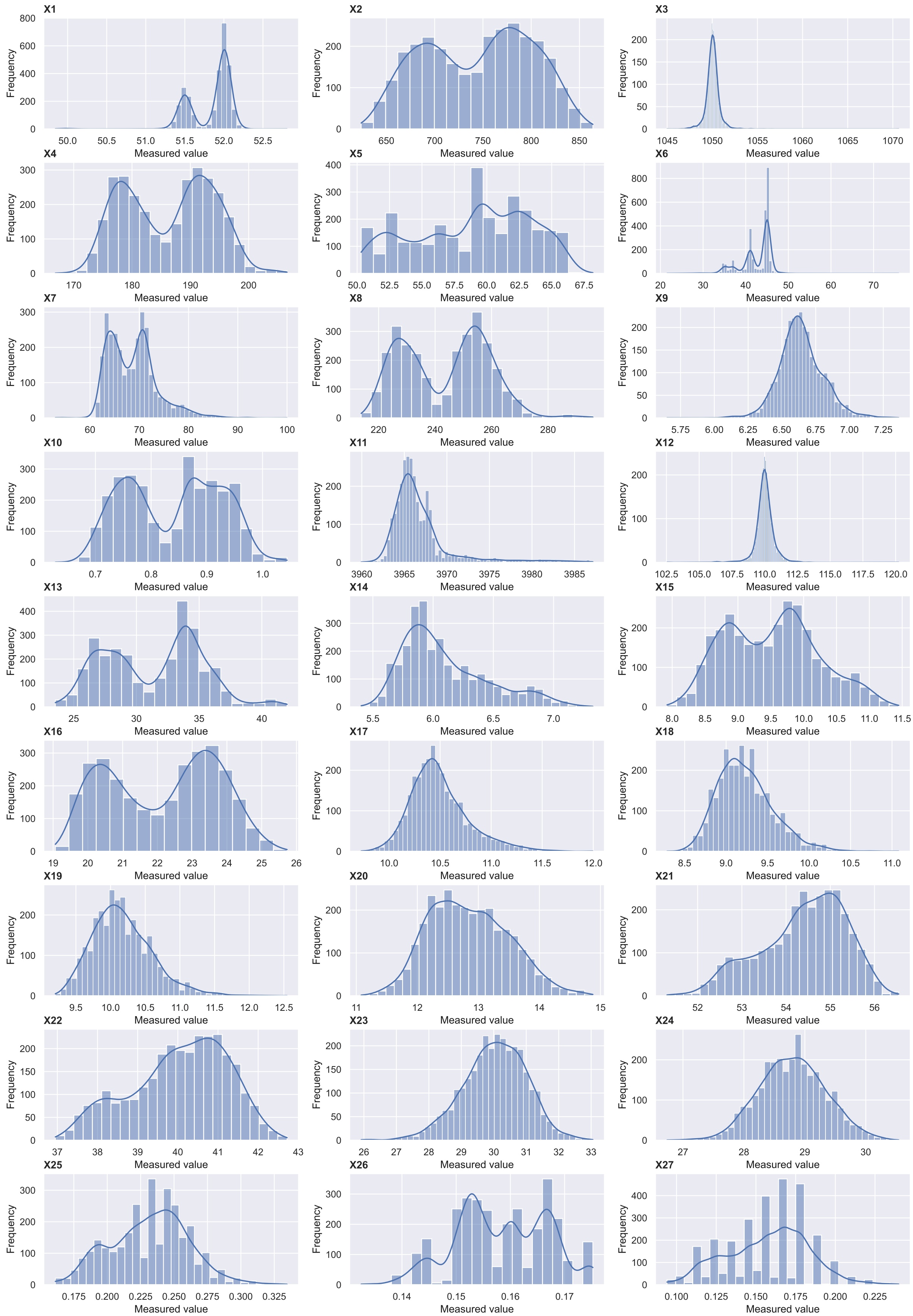
A The trend of features over time analyzed

B Summary statistics of the features

Feature	count	mean	std	min	25%	50%	75%	max
X1	17521	50.44	4.80	-0.86	50.89	51.05	51.93	53.29
X2	17521	713.79	101.63	131.07	694.34	726.25	760.09	911.69
X3	17521	993.81	175.74	-0.48	1019.39	1020.16	1039.81	1428.00
X4	17521	176.95	32.86	-6.06	178.10	181.99	186.45	206.58
X5	17521	58.15	8.82	-0.60	54.86	59.63	63.44	94.91
X6	17521	35.28	7.69	-2.00	34.79	35.12	37.99	95.28
X7	17521	63.49	13.71	0.00	63.55	67.05	69.29	100.00
X8	17521	258.55	100.17	-30.00	232.24	242.29	251.81	1530.00
X9	17521	6.48	0.60	0.07	6.24	6.49	6.61	10.20
X10	17521	0.77	0.14	-0.02	0.74	0.79	0.84	1.33
X11	17521	3,842	695.45	-100.00	3965.27	3967.17	3968.40	4694.31
X12	17521	107.61	13.94	15.80	109.65	109.97	110.30	132.52
X13	17521	31.23	3.77	5.97	28.89	31.69	33.78	97.61
X14	17521	5.98	1.18	0.00	5.86	6.14	6.41	68.01
X15	17521	9.39	2.06	0.00	8.82	9.34	10.21	81.36
X16	17521	17.42	7.98	0.00	19.27	20.52	21.61	75.54
X17	17521	9.85	1.83	0.00	9.84	10.21	10.44	66.66
X18	17521	8.67	1.77	0.00	8.55	9.20	9.43	50.76
X19	17521	9.79	1.89	0.00	9.66	10.24	10.58	33.19
X20	17521	12.45	2.42	0.00	12.18	12.89	13.51	45.28
X21	17521	55.01	10.16	0.00	54.40	56.27	57.69	62.87
X22	17521	39.23	7.21	0.00	38.64	40.17	42.31	44.96
X23	17521	29.43	5.26	0.00	29.58	30.33	31.02	34.84
X24	17521	27.42	4.94	0.00	27.68	28.34	28.89	31.68
X25	17521	0.19	0.17	-1.00	0.17	0.23	0.25	0.84
X26	17521	0.12	0.16	-1.00	0.13	0.15	0.16	0.19
X27	17521	0.12	0.16	-1.00	0.11	0.16	0.18	0.24

Source: Prepared by the author.

C Frequency distribution of features



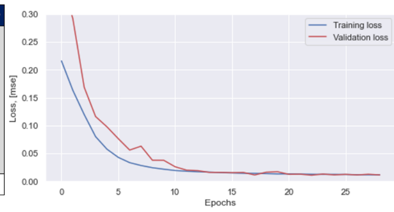
D Summary of tests performed on the primary hyperparameters of the autoencoder model

Model 1

Layers	Dropout Rate	Kernel size	Filter	Max Pooling 1D	Dense	Activation	Loss	Optimizer
First convolutional layer	0.3	9	128	2		Elu		
Second convolutional layer	0.3	7	64	2		Elu		
Third convolutional layer	0.3	5	32	2		Elu		
Fourth convolutional layer	0.3	5	32	2		Elu		
Fifth convolutional layer	0.3	7	64	2		Elu		
Sixth convolutional layer	0.3	9	128			Elu		
First layer fully connected	0.3				100	ReLU		
Optimizer							mse	adam

Minimum validation loss: 0.01015

Epoch 34: early stopping

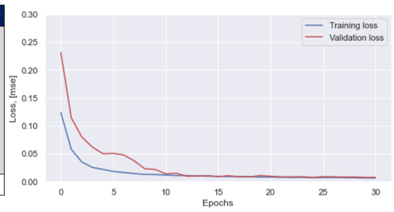


Model 2

Layers	Dropout Rate	Kernel size	Filter	Max Pooling 1D	Dense	Activation	Loss	Optimizer
First convolutional layer	0.1	7	128	2		Elu		
Second convolutional layer	0.1	5	64	2		Elu		
Third convolutional layer	0.1	3	32	2		Elu		
Fourth convolutional layer	0.1	3	32	2		Elu		
Fifth convolutional layer	0.1	5	64	2		Elu		
Sixth convolutional layer	0.1	7	128			Elu		
First layer fully connected	0.1				100	ReLU		
Optimizer							mse	adam

Minimum validation loss: 0.008168

Epoch 31: early stopping

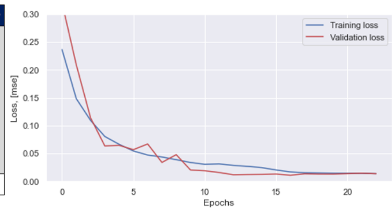


Model 3

Layers	Dropout Rate	Kernel size	Filter	Max Pooling 1D	Dense	Activation	Loss	Optimizer
First convolutional layer	0.3	9	256	2		Elu		
Second convolutional layer	0.3	7	128	2		Elu		
Third convolutional layer	0.3	5	64	2		Elu		
Fourth convolutional layer	0.3	5	64	2		Elu		
Fifth convolutional layer	0.3	7	128	2		Elu		
Sixth convolutional layer	0.3	9	256			Elu		
First layer fully connected	0.3				100	ReLU		
Optimizer							mse	adam

Minimum validation loss: 0.01079

Epoch 23: early stopping

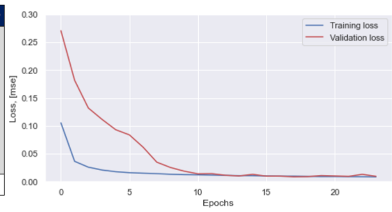


Model 4

Layers	Dropout Rate	Kernel size	Filter	Max Pooling 1D	Dense	Activation	Loss	Optimizer
First convolutional layer	0.1	7	128	2		ReLU		
Second convolutional layer	0.1	5	64	2		ReLU		
Third convolutional layer	0.1	3	32	2		ReLU		
Fourth convolutional layer	0.1	3	32	2		ReLU		
Fifth convolutional layer	0.1	5	64	2		ReLU		
Sixth convolutional layer	0.1	7	128			ReLU		
First layer fully connected	0.1				100	ReLU		
Optimizer							mse	adam

Minimum validation loss: 0.00846

Epoch 24: early stopping



E Summarize the selected articles sorted by the journal

Article	Publisher	Name
(LEE et al., 2014)	Elsevier	Procedia CIRP
(LEE et al., 2015)	Elsevier	Procedia CIRP
(LEE et al., 2015)	Elsevier	Industrial Agents: Emerging Applications of Software Agents in Industry
(ZIO, 2016)	IEEE	Transactions on Reliability
(FUMAGALLI et al., 2016)	Elsevier	IFAC-Papers OnLine
(SANDENGEN et al., 2016)	Atlantis Press	Advanced Manufacturing and Automation
(BOUSDEKIS; MENTZAS, 2017)	Springer	IFIP Advances in Information and Communication Technology
(LEE; ZHANG; NG, 2017)	Springer	Advances in Manufacturing
(WANG et al., 2017)	Springer	Journal of Intelligent Manufacturing
(CANITO et al., 2017)	Springer	Advances in Intelligent Systems and Computing
(CHO et al., 2018)	Springer	IFIP Advances in Information and Communication Technology
(HE et al., 2018)	SAGE	Advances in Mechanical Engineering
(KIANGALA; WANG, 2018)	Springer	International Journal of Advanced Manufacturing Technology
(KLEIN; BERGMANN, 2018)	CEUR	Workshop Proceedings
(KU, 2018)	Springer	Wireless Personal Communications
(MOURTZIS; VLACHOU, 2018)	Elsevier	Journal of Manufacturing Systems
(NEMETH et al., 2018)	Elsevier	Procedia CIRP
(RIMPAULT; BALAZINSKI; CHATELAIN, 2018)	MPDI	Journal of Manufacturing and Materials Processing
(SYAFRUDIN et al., 2018)	MPDI	Sensors (Switzerland)
(ZHENG et al., 2018)	Springer	Frontiers of Mechanical Engineering
(PÉREZ-LARA et al., 2018)	Springer	Best Practices in Manufacturing Processes
(SCHREIBER et al., 2019)	Elsevier	Procedia CIRP

Continued on next page

Appendix E – continued from previous page

Article	Publisher	Name
(SÉNÉCHAL; TRENTESAUX, 2019)	Elsevier	Environmental Impact Assessment Review
(TSAO et al., 2020)	Taylor Francis	International Journal of Systems Science: Operations and Logistics
(ADU-AMANKWA et al., 2019)	Springer	International Journal of Advanced Manufacturing Technology
(BIANCHINI; PELLEGRINI; ROSSI, 2019)	Springer	International Journal of System Assurance Engineering and Management
(CAO et al., 2019)	Elsevier	Procedia CIRP
(SALAZAR et al., 2019)	Springer	International Journal of Advanced Manufacturing Technology
(GODREAU et al., 2019)	Springer	Journal of Intelligent Manufacturing
(HÄCKEL et al., 2019)	Springer	Business Research
(LI et al., 2019)	Springer	Frontiers of Information Technology & Electronic Engineering
(ROSSIT; TOHMÉ; FRUTOS, 2019)	Elsevier	Journal of Industrial Information Integration
(PALAU; DHADA; PARLIKAD, 2019)	Springer	Journal of Intelligent Manufacturing
(THOPPIL; VASU; RAO, 2019)	Springer	Journal of Failure Analysis and Prevention
(VOGL; WEISS; HELU, 2019)	Springer	Journal of Intelligent Manufacturing
(XIA; XI, 2019)	Springer	Journal of Intelligent Manufacturing
(KŁOS; PATALAS-MALISZEWSKA, 2019)	Springer	Advances in Intelligent Systems and Computing

Source: Prepared by the author.

F Summarize the selected articles sorted by the conference

Article	Publisher	Name
(LEE et al., 2014)	INDIN	Proceedings of International Conference on Industrial Informatics
(FERREIRO et al., 2016)	PHM Society	European Conference of the Prognostics and Health Management Society
(MYERS; TICKEM; EVANS, 2016)	IET	7th IET Conference on Railway Condition Monitoring
(SIAFARA et al., 2017)	IEEE	43rd Annual Conference of the IEEE Industrial Electronics Society
(TERRISSA et al., 2016)	IEEE	Colloquium in Information Science and Technology
(CANIZO et al., 2017)	IEEE	International Conference on Prognostics and Health Management
(TAN et al., 2017)	IEEE	10 Annual International Conference, Proceedings
(MADHIKERMI et al., 2018)	Aalto	2nd International Conference on System Reliability and Safety
(RUIZ-SARMIENTO et al., 2018)	Cranfield Collec	Frontiers in Artificial Intelligence and Applications
(BALOGH et al., 2018)	IEEE	22nd International Conference on Intelligent Engineering Systems
(SEZER et al., 2018)	Cranfield Collec	International Conference on Engineering, Technology and Innovation
(PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018)	IEEE	International Enterprise Distributed Object Computing Workshop
(FORDAL; RØDSETH; SCHJØLBERG, 2019)	Springer	Lecture Notes in Electrical Engineering
(LAZAROVA-MOLNAR; MOHAMED, 2019)	Elsevier	Procedia CIRP
(SHIHUNDLA; MPOFU; ADENUGA, 2019)	Elsevier	Procedia CIRP
(XU et al., 2019)	IEEE	IEEE 15th International Conference on Automation Science and Engineering

Continued on next page

Appendix F – continued from previous page

Article	**Publisher**	**Name**
(MOHAMED; AL-JAROODI, 2019)	IEEE	9th Annual Computing and Communication Workshop and Conference

Source: Prepared by the author.

G Methodological application of the study

Purpose of the study	Articles
Case Study	(BIANCHINI; PELLEGRINI; ROSSI, 2019; CANIZO et al., 2017; CHO et al., 2018; FERREIRO et al., 2016; FUMAGALLI et al., 2016; GODREAU et al., 2019; HÄCKEL et al., 2019; HE et al., 2018; KIANGALA; WANG, 2018; KLEIN; BERGMANN, 2018; KŁOS; PATALAS-MALISZEWSKA, 2019; KU, 2018; LEE et al., 2015; LI et al., 2019; MADHIKERMI et al., 2018; MOURTZIS; VLACHOU, 2018; PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018; RIMPAULT; BALAZINSKI; CHATELAIN, 2018; RUIZ-SARMIENTO et al., 2018; SCHREIBER et al., 2019; SÉNÉCHAL; TRENTESAUX, 2019; SEZER et al., 2018; SHIHUNDLA; MPOFU; ADENUGA, 2019; SYAFRUDIN et al., 2018; TAN et al., 2017; THOPPIL; VASU; RAO, 2019; TSAO et al., 2020; WANG et al., 2017; XIA; XI, 2019)
Architecture	(ADU-AMANKWA et al., 2019; BALOGH et al., 2018; BOUSDEKIS; MENTZAS, 2017; CANITO et al., 2017; CAO et al., 2019; SALAZAR et al., 2019; LEE; ZHANG; NG, 2017; LEE et al., 2015; MOHAMED; AL-JAROODI, 2019; MYERS; TICKEM; EVANS, 2016; PÉREZ-LARA et al., 2018; ROSSIT; TOHMÉ; FRUTOS, 2019; PALAU; DHADA; PARLIKAD, 2019; SANDENGEN et al., 2016; SIAFARA et al., 2017; TERRISSA et al., 2016; ZHENG et al., 2018)
Trends	(FORDAL; RØDSETH; SCHJØLBERG, 2019; LAZAROVA-MOLNAR; MOHAMED, 2019; LEE et al., 2014, 2014; NEMETH et al., 2018; VOGL; WEISS; HELU, 2019; XU et al., 2019; ZIO, 2016)

Source: Prepared by the author.

H Used techniques by the purpose of the study

Purpose of the study	Techniques used	Articles
Case Study	Communication	(LEE et al., 2015; MOURTZIS; VLACHOU, 2018; PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018; SÉNÉCHAL; TRENTESAUX, 2019; SHIHUNDLA; MPOFU; ADENUGA, 2019)
	Ingestion	(BIANCHINI; PELLEGRINI; ROSSI, 2019; FERREIRO et al., 2016; FUMAGALLI et al., 2016; GODREAU et al., 2019; KIANGALA; WANG, 2018; LI et al., 2019; MADHIKERMI et al., 2018; RUIZ-SARMIENTO et al., 2018; SCHREIBER et al., 2019; SYAFRUDIN et al., 2018; THOPPIL; VASU; RAO, 2019; WANG et al., 2017)
	Analysis	(CANIZO et al., 2017; CHO et al., 2018; HE et al., 2018; KLEIN; BERGMANN, 2018; KŁOS; PATALAS-MALISZEWSKA, 2019; RIMPAULT; BALAZINSKI; CHATELAIN, 2018; SEZER et al., 2018; TSAO et al., 2020; XIA; XI, 2019)
	Storage	(HÄCKEL et al., 2019; KUEHN, 2018; TAN et al., 2017)
Architecture	Communication	(BOUSDEKIS; MENTZAS, 2017; ADU-AMANKWA et al., 2019; CANITO et al., 2017; SALAZAR et al., 2019; LEE et al., 2015; LEE; ZHANG; NG, 2017; MOHAMED; AL-JAROODI, 2019; PÉREZ-LARA et al., 2018; PALAU; DHADA; PARLIKAD, 2019; SANDENGEN et al., 2016; ZHENG et al., 2018)
	Ingestion	(ROSSIT; TOHMÉ; FRUTOS, 2019)
	Analysis	(CAO et al., 2019; MYERS; TICKEM; EVANS, 2016; SIAFARA et al., 2017)
	Storage	(BALOGH et al., 2018; TERRISSA et al., 2016)
Trends	Communication	(FORDAL; RØDSETH; SCHJØLBERG, 2019; LEE et al., 2014; VOGL; WEISS; HELU, 2019)
	Ingestion	(LAZAROVA-MOLNAR; MOHAMED, 2019; XU et al., 2019; ZIO, 2016)
	Analysis	(NEMETH et al., 2018)
	Storage	(LEE et al., 2014)

Source: Prepared by the author.

I Main applied techniques and focus of the articles

Techniques Used	Main Focus	Articles
Communication	Predictive Maintenance	(ADU-AMANKWA et al., 2019; LEE et al., 2015; SANDENGEN et al., 2016; PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018)
	Diagnosis and Prognostic	(LEE et al., 2014; PALAU; DHADA; PARLIKAD, 2019; VOGL; WEISS; HELU, 2019)
	Condition-based Operation	(LEE et al., 2015; MOURTZIS; VLACHOU, 2018; SHIHUNDLA; MPOFU; ADENUGA, 2019)
	Communication with Equipment	(SALAZAR et al., 2019; LEE; ZHANG; NG, 2017; ZHENG et al., 2018)
	Proactive Maintenance	(BOUSDEKIS; MENTZAS, 2017; CANITO et al., 2017)
	Value Chain	(PÉREZ-LARA et al., 2018)
	Reliability of Systems	(MOHAMED; AL-JAROODI, 2019)
	Improve Operational Availability	(FORDAL; RØDSETH; SCHJØLBERG, 2019)
	Maintenance for Sustainability	(SÉNÉCHAL; TRENTESAUX, 2019)
Ingestion	Predictive Maintenance	(FERREIRO et al., 2016; KIANGALA; WANG, 2018; ROSSIT; TOHMÉ; FRUTOS, 2019; RUIZ-SARMIENTO et al., 2018; SYAFRUDIN et al., 2018; THOPPIL; VASU; RAO, 2019; WANG et al., 2017; SCHREIBER et al., 2019)
	Condition-based Operation	(FUMAGALLI et al., 2016; GODREAU et al., 2019; LI et al., 2019; MADHIKERMI et al., 2018)
	Reliability of Systems	(LAZAROVA-MOLNAR; MOHAMED, 2019; ZIO, 2016)
	Preventive Maintenance	(BIANCHINI; PELLEGRINI; ROSSI, 2019)
	Diagnosis and Prognostic	(XU et al., 2019)
Analysis	Predictive Maintenance	(CANIZO et al., 2017; CAO et al., 2019; CHO et al., 2018; HE et al., 2018; KLEIN; BERGMANN, 2018; KŁOS; PATALAS-MALISZEWSKA, 2019; MYERS; TICKEM; EVANS, 2016; SEZER et al., 2018; RIMPAULT; BALAZINSKI; CHATELAIN, 2018)
	Diagnosis and Prognostic	(SIAFARA et al., 2017; XIA; XI, 2019)
	Prescriptive maintenance	(NEMETH et al., 2018)
	Economic Production Quantity	(TSAO et al., 2020)
Storage	Predictive Maintenance	(BALOGH et al., 2018; KU, 2018; TERRISSA et al., 2016)
	Diagnosis and Prognostic	(LEE et al., 2014)
	Value Chain	(TAN et al., 2017)
	Productivity of the Manufacturing	(HÄCKEL et al., 2019)

Source: Prepared by the author.

J The subject of study of selected articles with equipment focus

Study subject	Case Study	Architecture	Trends
Productivity and optimization	(GODREAU et al., 2019; LI et al., 2019; PREUVENEERS; JOOSEN; ILIE-ZUDOR, 2018; RUIZ-SARMIENTO et al., 2018; SYAFRUDIN et al., 2018; THOPPIL; VASU; RAO, 2019)		(LAZAROVA-MOLNAR; MOHAMED, 2019; XU et al., 2019)
Optimization of the manufacturing processes	(KLEIN; BERGMANN, 2018; RIMPAULT; BALAZINSKI; CHATELAIN, 2018)		
Productivity	(SEZER et al., 2018)	(BOUSDEKIS; MENTZAS, 2017)	
Reliability modeling of components and safe systems	(BIANCHINI; PELLEGRINI; ROSSI, 2019)	(CAO et al., 2019)	
Human resources optimization		(MYERS; TICKEM; EVANS, 2016)	
Reliability modeling of components			(ZIO, 2016)
Total	**10**	**3**	**3**

Source: Prepared by the author.

K The subject of study of selected articles with equipment and productivity focus

Study subject	Case Study	Architecture	Trends
Productivity and optimization	(KIANGALA; WANG, 2018; KŁOS; PATALAS-MALISZEWSKA, 2019; KU, 2018; LEE et al., 2015; MOURTZIS; VLACHOU, 2018; WANG et al., 2017)	(BALOGH et al., 2018; PALAU; DHADA; PARLIKAD, 2019)	(FORDAL; RØDSETH; SCHJØLBERG, 2019; NEMETH et al., 2018)
Optimization of the manufacturing processes	(CHO et al., 2018)	(ADU-AMANKWA et al., 2019)	(VOGL; WEISS; HELU, 2019)
Human resources optimization	(SCHREIBER et al., 2019)	(LEE; ZHANG; NG, 2017)	
Reliability modeling of components and safe systems	(CANIZO et al., 2017; FUMAGALLI et al., 2016)		
Equipment Life Cycle Optimization	(HE et al., 2018; SÉNÉCHAL; TRENTESAUX, 2019)		
Optimization of human resources and eco-efficiency		(SIAFARA et al., 2017)	
Security and connectivity		(TERRISSA et al., 2016)	
Process and product quality	(TAN et al., 2017)		
Process productivity			(LEE et al., 2014)
Equipment prognosis			(LEE et al., 2014)
Total	**13**	**6**	**5**

Source: Prepared by the author.

L The subject of study of selected articles with business focus

Study subject	Case Study	Architecture	Trends
Productivity and optimization	(SHIHUNDLA; MPOFU; ADENUGA, 2019; TSAO et al., 2020; XIA; XI, 2019)	(LEE et al., 2015; ROSSIT; TOHMÉ; FRUTOS, 2019; ZHENG et al., 2018)	
Equipment Life Cycle Optimization	(HÄCKEL et al., 2019)		
Security and connectivity		(MOHAMED; AL-JAROODI, 2019)	
Reliability modeling of components and safe systems		(SALAZAR et al., 2019)	
After-sales	(MADHIKERMI et al., 2018)		
Supply chain		(PÉREZ-LARA et al., 2018)	
Optimization of the manufacturing processes		(SANDENGEN et al., 2016)	
Human resources optimization		(CANITO et al., 2017)	
Total	**5**	**8**	**0**

Source: Prepared by the author.