Information Fusion and Data Science

Series Editor: Henry Leung

Tania Cerquitelli · Nikolaos Nikolakis ·
Niamh O'Mahony ·
Enrico Macii · Massimo Ippolito ·
Sotirios Makris *Editors*

Predictive Maintenance in Smart Factories

Architectures, Methodologies, and Use-cases

Springer

Information Fusion and Data Science

Series Editor

Henry Leung, University of Calgary, Calgary, AB, Canada

This book series provides a forum to systematically summarize recent developments, discoveries and progress on multi-sensor, multi-source/multi-level data and information fusion along with its connection to data-enabled science. Emphasis is also placed on fundamental theories, algorithms and real-world applications of massive data as well as information processing, analysis, fusion and knowledge generation.

The aim of this book series is to provide the most up-to-date research results and tutorial materials on current topics in this growing field as well as to stimulate further research interest by transmitting the knowledge to the next generation of scientists and engineers in the corresponding fields. The target audiences are graduate students, academic scientists as well as researchers in industry and government, related to computational sciences and engineering, complex systems and artificial intelligence. Formats suitable for the series are contributed volumes, monographs and lecture notes.

More information about this series at http://www.springer.com/series/15462

Tania Cerquitelli · Nikolaos Nikolakis ·
Niamh O'Mahony · Enrico Macii ·
Massimo Ippolito · Sotirios Makris
Editors

Predictive Maintenance in Smart Factories

Architectures, Methodologies, and Use-cases

Springer

Editors
Tania Cerquitelli
Department of Control and Computer Engineering
Politecnico di Torino
Turin, Italy

Nikolaos Nikolakis
Laboratory for Manufacturing Systems and Automation
University of Patras
Patras, Greece

Niamh O'Mahony
Dell EMC Research Europe
Cork, Ireland

Enrico Macii
Interuniversity Department of Regional and Urban Studies and Planning
Politecnico di Torino
Turin, Italy

Massimo Ippolito
COMAU S.p.A.
Grugliasco, TO, Italy

Sotirios Makris
Laboratory for Manufacturing Systems and Automation
University of Patras
Patras, Greece

ISSN 2510-1528 ISSN 2510-1536 (electronic)
Information Fusion and Data Science
ISBN 978-981-16-2939-6 ISBN 978-981-16-2940-2 (eBook)
https://doi.org/10.1007/978-981-16-2940-2

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

In the last years, Information Technology services and components are becoming pervasive inside production lines, devices, and control systems, transforming traditional manufacturing and shop floor environments into the fully digital, interconnected factory of the future.

The real-time collection of a growing amount of data from factories is paving the way to the creation of *manufacturing intelligence* platforms. *Predictive maintenance*, that is, the ability to predict equipment critical conditions before their occurrence disrupts the production line, is among the most important expressions of manufacturing intelligence. By leveraging innovative computing paradigms from the Internet of Things to cloud computing and data-driven methodologies, effective predictive diagnostic tools can be built and delivered as-a-service. Ultimately, this enables industrial stakeholders to take control of their assets and design smart and efficient maintenance plans that increase productivity while at the same time reducing costs. All industries, regardless of sector and size, can benefit from such solutions; however, integrating data-driven methodologies in shop floor environments is still a complex task, which significantly hinders their adoption, especially from small and medium manufacturing enterprises.

This book presents the fruitful outcomes of the European project "SERENA," involving fourteen partners, including renowned academic centers, IT companies, and industries at the international level. The project addresses the design and development of a plug-n-play end-to-end cloud architecture, enabling predictive maintenance of industrial equipment to be easily exploited by small and medium manufacturing companies with minimal data-analytics experience.

This book offers design principles for data-driven algorithms, architectures, and tools for enhancing intelligence in industrial environments. The book is divided into two parts to ease readability, tailored to readers of different backgrounds. First Part presents more technical content, while Second Part real industrial use cases.

The integrated solution developed with the SERENA project overcomes some of the most challenging issues associated with predictive maintenance (e.g., architecture configuration and deployment, collection of large-scale datasets with a known failure status, and real-time assessment of predictive model performance) by exploiting self-configuring tools and techniques that make their adoption feasible by a wide range of industries. The solution addresses the complete data lifecycle from data collection to value exploitation. It integrates a wide range of innovative data-driven methodologies that have been designed specifically to streamline the prognostics of IoT industrial components, the characterization of equipment health status and operating conditions to generate early warnings, and the forecasting of future equipment degradation. The data-driven models are not only capable of self-configuring, thus requiring limited data analytics expertise, but once deployed, are capable of continuously self-assessing their performance, raising a warning when their predictions are no longer accurate, and the model may need to be retrained to cope, e.g., with evolving operating conditions. All these components are integrated using a scalable methodology-as-a-service approach on a mixed cloud-edge analytics platform.

All proposed solutions are thoroughly compared and discussed with respect to the state-of-the-art and state-of-the-practice methodologies and architectures. Furthermore, methodological approaches and architectures proposed in the context of SERENA have been thoroughly experimentally evaluated on real data sets collected in a variety of industrial environments, including robotics, white goods, manufacturing of elevator cabins, metrological equipment, and steel production. An in-depth discussion of the experimental results is also included for each proposed use case.

Perspectives and new opportunities to address open issues on predictive maintenance conclude each chapter with some interesting future research activities.

We hope readers will find the book of interest and that the content will inspire researchers and practitioners to develop creative, successful, and technological-advanced research!

Turin, Italy — Tania Cerquitelli
Patras, Greece — Nikolaos Nikolakis
Cork, Ireland — Niamh O'Mahony
Turin, Italy — Enrico Macii
Grugliasco, Italy — Massimo Ippolito
Patras, Greece — Sotirios Makris
April 2021

Acknowledgements The editors would like to thank all people who contributed to the SERENA project and inspired the content included in the book.

This research has been partially funded by the European project "SERENA—VerSatilE plug-and-play platform enabling REmote predictive mainteNAnce" (Grant Agreement: 767561).

Contents

Methodologies and Enabling Technologies

Industrial Digitisation and Maintenance: Present and Future 3
Massimo Ippolito, Nikolaos Nikolakis, Tania Cerquitelli,
Niamh O'Mahony, Sotirios Makris, and Enrico Macii

A Hybrid Cloud-to-Edge Predictive Maintenance Platform 19
Angelo Marguglio, Giuseppe Veneziano, Pietro Greco,
Sven Jung, Robert Siegburg, Robert H. Schmitt, Simone Monaco,
Daniele Apiletti, Nikolaos Nikolakis, Tania Cerquitelli, and Enrico Macii

Data-Driven Predictive Maintenance: A Methodology Primer 39
Tania Cerquitelli, Nikolaos Nikolakis, Lia Morra, Andrea Bellagarda,
Matteo Orlando, Riku Salokangas, Olli Saarela, Jani Hietala,
Petri Kaarmila, and Enrico Macii

Services to Facilitate Predictive Maintenance in Industry4.0 75
Nikolaos Nikolakis, Tania Cerquitelli, Georgios Siaterlis,
Andrea Bellagarda, Enrico Macii, Guido Coppo, Salvatore Andolina,
Florian Bartholomauss, Martin Plutz, and Kosmas Alexopoulos

Industrial Use-cases

Predictive Analytics in Robotic Industry 99
Simone Panicucci, Lucrezia Morabito, Chiara Napione,
Francesco Ventura, Massimo Ippolito, and Nikolaos Nikolakis

Remote Condition Monitoring and Control for Maintenance Activities in Coordinate Measurement Machine: A Data-Driven Approach for the Metrology Engineering Industry 121
Andrea Gómez Acebo, Antonio Ventura-Traveset,
and Silvia de la Maza Uriarte

Estimating Remaining Useful Life: A Data-Driven Methodology for the White Goods Industry 149
Pierluigi Petrali, Tania Cerquitelli, Paolo Bethaz, Lia Morra, Nikolaos Nikolakis, Claudia De Vizia, and Enrico Macii

Predictive Analytics in the Production of Elevators 165
Valeria Boldosova, Jani Hietala, Jari Pakkala, Riku Salokangas, Petri Kaarmila, and Eetu Puranen

Predictive Maintenance in the Production of Steel Bars: A Data-Driven Approach 187
Paolo Bethaz, Xanthi Bampoula, Tania Cerquitelli, Nikolaos Nikolakis, Kosmas Alexopoulos, Enrico Macii, and Peter van Wilgen

In-Situ Monitoring of Additive Manufacturing 207
Davide Cannizzaro, Antonio Giuseppe Varrella, Stefano Paradiso, Roberta Sampieri, Enrico Macii, Massimo Poncino, Edoardo Patti, and Santa Di Cataldo

Glossary 229

Index 233

Contributors

Andrea Gómez Acebo TRIMEK, Madrid, Spain

Kosmas Alexopoulos Laboratory for Manufacturing Systems and Automation, University of Patras, Patras, Greece

Salvatore Andolina SynArea Consultants s.r.l., Turin, Italy

Daniele Apiletti Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Xanthi Bampoula Laboratory for Manufacturing Systems and Automation, University of Patras, Patras, Greece

Florian Bartholomauss oculavis GmbH, Aachen, Germany

Andrea Bellagarda Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Paolo Bethaz Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Valeria Boldosova Prima Power, Seinäjoki, Finland;
University of Vaasa, Vaasa, Finland

Davide Cannizzaro Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Tania Cerquitelli Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Guido Coppo SynArea Consultants s.r.l., Turin, Italy

Claudia De Vizia Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Santa Di Cataldo Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Pietro Greco Research & Innovation, Engineering Ingegneria Informatica S.p.A., Palermo, Italy

Jani Hietala VTT Technical Research Centre of Finland, Espoo, Finland

Massimo Ippolito COMAU S.p.A., Grugliasco, TO, Italy

Sven Jung Fraunhofer Institute for Production Technology IPT and Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, Aachen, Germany

Petri Kaarmila VTT Technical Research Centre of Finland, Espoo, Finland

Enrico Macii Interuniversity Department of Regional and Urban Studies and Planning, Politecnico di Torino, Turin, Italy

Sotirios Makris Laboratory for Manufacturing Systems and Automation, University of Patras, Patras, Greece

Angelo Marguglio Research & Innovation, Engineering Ingegneria Informatica S.p.A., Palermo, Italy

Simone Monaco Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Lucrezia Morabito COMAU S.p.A., Grugliasco, TO, Italy

Lia Morra Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Chiara Napione COMAU S.p.A., Grugliasco, TO, Italy

Nikolaos Nikolakis Laboratory for Manufacturing Systems and Automation, University of Patras, Patras, Greece

Matteo Orlando Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Niamh O'Mahony Dell EMC Research Europe, Cork, Ireland

Jari Pakkala Prima Power, Seinäjoki, Finland

Simone Panicucci COMAU S.p.A., Grugliasco, TO, Italy

Stefano Paradiso Stellantis, Turin, Italy

Edoardo Patti Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Pierluigi Petrali Whirlpool EMEA, Biandronno, VA, Italy

Martin Plutz oculavis GmbH, Aachen, Germany

Massimo Poncino Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Eetu Puranen KONE, Hyvinkää, Finland

Olli Saarela VTT Technical Research Centre of Finland, Espoo, Finland

Riku Salokangas VTT Technical Research Centre of Finland, Espoo, Finland

Roberta Sampieri Stellantis, Turin, Italy

Robert H. Schmitt Fraunhofer Institute for Production Technology IPT and Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, Aachen, Germany

Georgios Siaterlis Laboratory for Manufacturing Systems & Automation, University of Patras, Patras, Greece

Robert Siegburg Fraunhofer Institute for Production Technology IPT and Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, Aachen, Germany

Silvia de la Maza Uriarte TRIMEK, Madrid, Spain

Peter van Wilgen VDL Weweler bv, Apeldoorn, The Netherlands

Antonio Giuseppe Varrella Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Giuseppe Veneziano Research & Innovation, Engineering Ingegneria Informatica S.p.A., Palermo, Italy

Francesco Ventura Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy

Antonio Ventura-Traveset TRIMEK, Barcelona, Spain

Acronyms

ADC	Analog to Digital Converter
AI	Artificial Intelligence
AIC	Akaike's Information Criterion
AR	Augmented Reality
BA	Blowing Agents
CART	Classification and Regression Tree
CBM	Condition-Based Maintenance
CM	Condition Monitoring
CMM	Coordinate Measurement Machine
CMMS	Computerised Maintenance Management System
CPS	Cyber-Physical System
CRIS	Common Relational Information Model
DH	Decision Horizon
DL	Deep Learning
DS	Descriptor Silhouette
ECS	Elastic Cloud Storage
ERP	Enterprise Resource Planning
FMEA/FMECA	Failure Mode and Effects (or and Criticality) Analysis
GMM	Gaussian Mixture Mode
GPIO	General-Purpose Input Output
GPU	Graphical Processing Unit
HDFS	Hadoop Distributed File System
HMI	Human-Machine Interface
ICT	Information and Communication Technologies
IIoT	Industrial Internet of Things
IoT	Internet of Things
IRI	Internationalised Resource Identifier
ISTEP	Integrated Self-Tuning Engine for Predictive maintenance
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
KNN	K-Nearest Neighbors
KPI	Key Performance Indicator

LSTM	Long Short-Term Memory
MAAPE	Mean Arctangent Absolute Percentage Error
MES	Manufacturing Execution System
ML	Machine Learning
MNA	Maximum Number of Alternatives
OEM	Original Equipment Manufacturer
OSA-EAI	Open Systems Architecture for Enterprise Application Integration
PdM	Predictive Maintenance
PLC	Programmable Logic Controller
PSBB	Punching-Shearing-Buffering-Bending
PvM	Preventive Maintenance
RCFA	Root Cause Failure Analysis
RMSE	Root Mean Square Error
RNNs	Recurrent Neural Networks
RPCA	Reverse Proxy Certification Authority
RUL	Remaining Useful Life
SERENA	VerSatilE plug-and-play platform enabling REmote predictive mainteNAnce
SME	Small and medium-sized enterprise
SOA	Service-Oriented Architecture
SPI	Serial Peripheral Interface
SR	Sampling Rate
SSL	Secure Sockets Layer
TLS	Transport Layer Security
TPU	Tensor Processing Unit
UI	User Interface
VPN	Virtual Private Network
VR	Virtual Reality
XML	eXtensible Markup Language

Methodologies and Enabling Technologies

Industrial Digitisation and Maintenance: Present and Future

Massimo Ippolito, Nikolaos Nikolakis, Tania Cerquitelli, Niamh O'Mahony, Sotirios Makris, and Enrico Macii

Abstract In recent years various maintenance strategies have been adopted to maintain industrial equipment in an operational condition. Adopted techniques include approaches based on statistics generated by equipment manufacturers, human knowledge, and intuition based on experience among others. However, techniques like those mentioned above often address only a limited set of the potential root causes, leading to unexpected breakdown or failure. As a consequence, maintenance costs were considered a financial burden that each company had to sustain. Nevertheless, as technology advances, user experience and intuition are enhanced by artificial intelligence approaches, transforming maintenance costs into a company's strategic asset. In particular, for manufacturing industries, a large volume of data is generated on a shop floor as digitisation advances. Combining information and communication technologies (ICT) with artificial intelligence techniques may create insight over production processes, complement or support human knowledge, revealing undetected anomalies and patterns that can help predict maintenance actions. Consequently, the

M. Ippolito
COMAU S.p.A., Grugliasco, TO, Italy
e-mail: massimo.ippolito@comau.com

N. Nikolakis (✉) · S. Makris
Laboratory for Manufacturing Systems and Automation, University of Patras, Patras, Greece
e-mail: nikolakis@lms.mech.upatras.gr

S. Makris
e-mail: makris@lms.mech.upatras.gr

T. Cerquitelli
Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy
e-mail: tania.cerquitelli@polito.it

N. O'Mahony
Dell EMC Research Europe, Cork, Ireland
e-mail: Niamh.Omahony@dell.com

E. Macii
Interuniversity Department of Regional and Urban Studies and Planning,
Politecnico di Torino, Turin, Italy
e-mail: enrico.macii@polito.it

T. Cerquitelli et al. (eds.), *Predictive Maintenance in Smart Factories*,
Information Fusion and Data Science,
https://doi.org/10.1007/978-981-16-2940-2_1

company yields a reduction of unexpected breakdowns, production stoppages, and production costs. The outcomes are significant but selecting an appropriate data-driven method that can generate helpful and trustworthy results is challenging. It is mainly affected by the quality of the available data and the capability to understand the process under analysis correctly. This chapter reviews architectures for data management and data-driven methodologies for enabling predictive maintenance policies. Then follows the presentation of integrated solutions for predictive analytics to conclude with the main challenges identified and future outlook.

1 Introduction

The advent of the Industry 4.0 brought a revolution in the manufacturing sector, introducing new challenges and opportunities related to Big Data analysis through machine learning techniques. Thanks to sensors installed directly on machinery, companies are now able to continually monitor production activity, collecting a massive amount of data that describe the machinery's behavior through time. The effective processing and analysis of the collected data can provide accurate and valuable results that reveal underlying patterns, identify anomalies not easily detected by a human eye, support the deterioration of the machinery, and support the company's decision-making process about its maintenance activities.

In this context, innovative ICT architectures and data-driven methodologies are needed to effectively and efficiently support predictive diagnostics, allowing industrial stakeholders to plan maintenance operations by using data-driven methodologies-as-a-service efficiently. Although industries need such innovative solutions, the widespread adoption of data-driven methodologies remains challenging in industrial environments. The challenges are further reinforced for small and medium-sized enterprises (SMEs) considering the capital cost required to digitize their processes and benefit from them. Nevertheless, several efforts are underway towards making the use of advanced digital solutions more straightforward and more affordable for SMEs, requiring less capital investment from their end. Nevertheless, the benefits mentioned above are already recognized, and companies adopt similar advanced solutions in their production processes. In fact, one of the twelve "Manufacturing Lighthouses" factories identified by the World Economic Forum is an SME [1], showing how it is rather a matter of paradigm shift than of investments. Stakeholders of the private and public sector recognise the potential benefits of the Industry 4.0 revolution in manufacturing; in many countries national platforms and private-public partnerships have been created to increase awareness, support development of new use-cases and enable collaboration between research institutes and private organisations. Results are encouraging, as 70% of industrial organisations are either piloting Industry 4.0 solutions in manufacturing or using these technologies on a large scale. SMEs, however, can significantly benefit from novel Industry 4.0 technologies by increased flexibility of their production processes under such a smarter production paradigm [2].

The collection of a growing amount of data in factories paved the way to create intelligence over those data and benefit. Predictive maintenance, the ability to identify equipment critical conditions before their actual occurrences, is one type of intelligent practices that has received increasing attention in recent years. Despite the existence of various other maintenance strategies, already used widely in practice, maintenance needs can cause costly disruptions in the manufacturing process. In this context the condition based approach is considered as an effective [3], but also complex to implement and integrate to a production system [4] alternative.

With predictive analytics, however, repair and maintenance activities can be managed more smartly. For example, if needed, they can be prioritised and allocated to pre-planned outages based on real-time probabilities of expected or potential future failures, thus bringing the maintenance costs down. Hence, the predictive maintenance strategy can safely be supported, resulting in savings in both time and costs, due to reduced production downtime [5]. In case this is combined with logistics and maintenance stock parts, then the overall production costs can be reduced even further. Last but not least, predictive maintenance techniques such as vibration and thermal monitoring along with Reliability techniques such as Failure Modes and Effects Analysis (FMEA) and Root Cause Failure Analysis (RCFA) [6, 7] can result in bottom-line savings through early detection and maintenance, preventing failures from disturbing the production process.

However, a purely technology-driven approach to these solutions may easily result in failure; without a clear business objective, the deployment of numerous Internet of Things devices within an industrial plant could lead to an investment without significant return, and equally important, without discovering significant insights from collected data to create business value. The kind of desired data-driven insights tend to determine the appropriate ICT solutions to be adopted by a company. Hence, digitisation and adoption of advanced technologies needs to go hand in hand with a wider transformation of an enterprise's business model and practices.

Considering the emergent need for data-driven analytics and the increased adoption of the as-a-service models, the European project SERENA, funded under the Horizon 2020 framework, introduces a platform for facilitating manufacturers' maintenance needs with a clear focus on supporting predictive maintenance strategies. As part of it, versatile industrial use cases could be analysed considering two discrete needs of the manufacturing sector; equipment providers and equipment consumers or else factories.

After the identification of the industrial needs, we identified data-driven techniques in the context of:

- enabling condition monitoring for providing advanced services to customers, or
- enhancing existing preventive maintenance strategies applied on production equipment or lines with predictive ones.

With respect to enabling the aforementioned techniques with the least disruption of the companies' practices, the as-a-service model was adopted, leading to the design and implementation of a web platform that could be scalable, resilient and deployed in different ways: *at a cloud*, *on-premise, or*, *hybrid*, while facilitating the provision of latest research and development services in the context of condition monitoring and predictive maintenance to inherently heterogeneous industrial sectors. In particular the industrial sectors considered in the context of the SERENA project [8] include the following manufacturing companies: (1) Robotic equipment manufacturer, (2) Metrology equipment manufacturer, (3) Elevators producer, (4) Steel parts producer, (5) White goods manufacturing company.

To this end and to address the requirements for remote condition-monitoring of industrial assets, the following functionalities have been identified as key enablers of facilitating a transition to predictive maintenance approaches in the industry:

- Smart acquisition mechanisms at the edge for collecting data from heterogeneous assets (e.g., robots, machines, welding guns, PLCs, external sensors)
- Increased connectivity for remote data management of the huge data volumes generated by shopfloor data sources
- A software platform supporting predictive maintenance as-a-service as a result of data-driven methodologies,
- Technology-neutral middleware with security capabilities
- Versatile deployments, addressing different needs
- Scalability of functionalities and data processing capabilities

The approaches mentioned above enabled the creation of the SERENA system where each item mentioned above has been implemented into software services and seamlessly integrated into its leading platform.

This chapter is organised as follows: Sect. 2 presents a literature review in the context of intelligent manufacturing with a specific focus on data management architectures, data-driven methodologies for analytics and integrated solutions. Next the key challenges identified are presented in Sect. 3, to conclude with Sect. 4 presenting the trends in the context of predictive maintenance along with the vision of the SERENA project.

2 Literature Review

In the last decade a rich landscape of research and development activities has been carried out in the context of Industry 4.0. Research works can be classified into four main complementary categories:

- Maintenance approaches, discussed in Sect. 2.1.
- Data management architectures, discussed in Sect. 2.2.
- Data analytics methodologies and algorithms are presented in Sect. 2.3, tackling (i) anomaly detection, (ii) predictive analytics, and (iii) Remaining Useful Life (RUL) estimation.
- Integrated solutions are described in Sect. 2.4.

2.1 Maintenance Approaches

Maintenance activities have been traditionally reactive, with corrective operations performed after a failure was observed. This, however, caused significant losses for the production system [9]. As a response, Preventive Maintenance (PvM) approaches were established [10]. PvM approaches objective is to proactively perform fixed in time routine maintenance operations or inspections in order to prevent breakdowns, considering the importance of an uninterrupted production process [11, 12]. In this context, frequent causes of failures are identified in advance, either from experiments or out of experience. Nevertheless, the modelling of a machine's degradation is a complex process [13].

Thus, Condition Monitoring (CM) approaches were introduced providing huge amounts of data of the actual operational status of a machine, known as Big Data [14–16]. However, the identification of the correct set of parameters which are related to the operational status of the machine and its degradation process is not an easy case, especially when considering the number and heterogeneity of equipment on a shopfloor [17–19].

For this reason, Predictive Maintenance (PdM) approaches were introduced allowing the examination of these large-scale data sets [20]. PdM approaches, find out correlations and patterns in order to predict and prevent conditions that can be harmful to the operational lifetime of the production equipment. The ultimate ambition though, is the prolongation of the maintenance procedure and simultaneously the estimation of Remaining Useful Life (RUL) of the industrial equipment [21]. RUL estimation can be achieved through statistical methods and probabilistic models that apply to the available data without depending on physics of the underlying degradation process [22]. The main difference can be set upon the targeted outcome. For statistical learning, the aim is understanding the data correlation and inferences, while in Machine Learning (ML) the outcome of the evaluation is important without requiring a clear understanding of the underlying data and their interaction [23]. Purpose of statistical learning can be the mathematical analysis of the data values in order to discover relations among them and draw inferences to approximate the reality, while, ML approaches are based upon statistical methods in order to teach a machine approximate to real conditions [24, 25].

However, the performance of PdM applications depends on the appropriate choice of the ML method, as various techniques have been investigated until now [26, 27]. Bayesian networks can be used for diagnosing and predicting faults in large manufacturing dataset with little information on the variables, presenting computational learning and speed issues [28]. On the other hand, Convolutional Neural Networks present an excellent performance and a low computational cost [29]. However, Recurrent Neural Networks (RNNs) contain feedback loops and are more suitable for sequential data such as time-series data [30–32]. Nevertheless, since RNNs present some issues on long-term RUL predictions, Long Short-Term Memory (LSTM) networks are preferred [33], presenting, however lack of efficiency because of sensitivity to dataset changes [34]. As a consequence, hybrid models combine more than one network in order to overcome these kind of defects [35]. LSTMs-Autoencoder networks, for example, are presented in [36] as a deep learning approach for RUL estimation of a hot rolling milling machine. Finally, Transformer-based approaches have recently received increased attention for forecasting time series data as they seem to outperform the other ML models [37, 38].

2.2 Data Management Architectures

An emerging challenge of modern industries is to effectively collect, process and analyse large amounts of data, even in real time. To this aim, various data management architectures have been proposed in years [39–46] based on Big Data frameworks. These approaches may differ in some aspects and may complement each other or even demonstrate similarities in their methods. In turn this may address the need for efficient data storage, effective communication, and knowledge extraction mechanisms tailored to the individual characteristics of each industrial case. In particular, the work in [39] presents how recent trends in Industry 4.0 solutions influence the development of manufacturing execution systems, while in [40] authors present a framework facilitating scalable and flexible data analytics towards realising real-time supervision systems for manufacturing environments.

Early software solutions for production systems were mainly based on monolithic architectures providing low flexibility, adaptability, and scalability, thus increasing the complexity to upgrade as well as the maintenance costs. A step towards more flexible architectures was based on Service-Oriented Architectures (SOA), overcoming some previous drawbacks, such as maintainability and flexibility. Unfortunately, these systems could not guarantee high levels of flexibility and modularity required by modern business models. With the advent of cyber-physical production systems, new business models focusing on the as-a-service strategy emerged. As a consequence, versatile strategies based on micro-services architectures were proposed [47]. As discussed in [48], monolithic applications can be decomposed into micro-services to implement lightweight and scalable strategies for achieving better management of services distribution and operation. Nevertheless, an emerging concern is related to the data stewardship and ownership, especially as data are slowly getting linked to value. Data management architectures, at the beginning, were mostly based on cloud resources to enable on-demand services and scalability to future needs. However, this structure gave the cloud provider complete authority over the data. One step towards more privacy-preserving data was represented by fog computing engines [49], enabling on-demand computation to selectively exploit edge devices or cloud resources based on the kinds of jobs to be executed. A review of the recent fog computing trends and applications is provided in [50], where among other challenges the application service, resources management, and communication among layers while enabling security and privacy features, are highlighted.

A parallel research effort was devoted to the main challenges associated with effectively integrating existing production software, legacy systems and advanced

Internet of Things (IoT) technologies [41]. To this aim, the enabling technologies are based on the virtualization strategy and a cloud-based micro-service architecture. The latter technology is mainly exploited in different research activities to deal with big data applications [42, 43].

A step towards more heterogeneous architectures is discussed in [51] where virtualisation technologies combined with IEC61499 function blocks enabled the holistic orchestration of a production station. As a result, the configuration or re-configuration is highly dependent on software resulting in increased automation levels and flexibility.

Furthermore, the authors in [44] presented a Big Data analytics framework, capable of providing a health monitoring application for an aerospace and aviation industry. The studies in [45, 46] use open source technologies such as Apache Spark and Kafka to implement a scalable architecture capable of processing data both online and offline.

2.3 Data-Driven Analytics: Methodologies and Algorithms

The advent of the Industry 4.0 brought a revolution in the manufacturing sector, introducing new challenges and opportunities related to data analysis as enabled by machine learning, data mining and statistical processing techniques. Advanced sensors installed directly on the machinery, allow the constant monitoring of the production activity, collecting a huge amount of data that describes the behaviour and performance of monitored equipment over time. In addition, the effective processing and analysis of collected data, can facilitate a company's decision making process. Among the most beneficial approaches that the collection and analysis of industrial data has brought in a manufacturing context, there are the *predictive maintenance* strategies. Such approaches allow the analysis of a huge amount of data with different algorithms and addressing different objectives and requirements. Predictive maintenance strategies can be implemented through different data analytics methodologies, including *anomaly detection*, *predictive analytics*, and *RUL*, discussed in Sects. 2.3.2, 2.3.2, and 2.3.3, respectively.

To guarantee the maintenance process chain, the assessment of the current equipment condition is an important issue to be addressed. To this end, predictive maintenance strategies focus on identifying possible malfunctions over time and estimating the *RUL*, for assessing the operational life time of a machine.

2.3.1 Anomaly Detection

Different anomaly detection strategies have been proposed based on a joint exploitation of raw data and smart trend filtering [52]. The main drawback of the existing strategies is the need of raw data of proper quality that include both normal and abnormal working conditions. Unfortunately, in many real-life settings such data are unavailable. Methods, to detect noise and outliers, for data quality improvements are discussed by [53].

2.3.2 Predictive Analytics

The topic of predictive maintenance in a big data environment is also addressed in [54], where, with the purpose of monitoring the operation of wind turbines, a data-driven solution deployed in the cloud for predictive model generation is presented.

The spread of predictive maintenance in recent years is wide. In [55] a systematic review of the literature is performed to analyze academic articles about predictive maintenance and machine learning in an Industry 4.0 context, from 2015 to 2020. In these scenarios, historical data set has a fundamental role in order to obtain real time and satisfactory results even on new data. For this reason, steps such as features engineering and data labelling (if the label is not already present) are particularly important and can greatly influence the quality and accuracy of results obtained, as well as the selection of the prediction model.

2.3.3 Estimation of the Remaining Useful Life

The estimation of a component's RUL can be defined as the period from a given starting time up to the point it deteriorates to a level that is no longer operating within the limits evaluated by its producer and for a specific process. This evaluation as well as a preventive maintenance plan, are usually result of extensive tests conducted by the Original Equipment Manufacturer (OEM). These tests, even though covering a wide area of parameters and values to provide the best possible estimations, cannot possibly address all possible use cases and degradation factors that may impact the behaviour of a machine. Thus, it is not that following the preventive maintenance plan provided by an OEM leads to over-maintenance, which is associated with increased cost. On the other hand, if the maintenance deviates a lot from that preventive maintenance plan, it may be the cause of unexpected breakdowns and thus increased production costs. Hence, the estimation of the degradation of machine under its operating environment becomes a factor that could reduce costs, given a certain level of accuracy in its estimation. Usually, the RUL value is adopted to represent and quantify the estimation of a machine's degradation.

In this context, several methodologies to estimate the status of machine's degradation, in terms of a machine's RUL, have been devised [56]. State-of-the-art methods [57–59] are mainly based on complex statistical analysis to map a value with physical meaning for the maintenance personnel, responsible for further actions.

An adaptive skew-Wiener process model is discussed in [60] and validated in a use case concerning the ball bearings in rotating electrical machines. In the same study, the need to incorporate stochastic phenomena is highlighted. In this context, a review on Wiener-process-based methods for degradation data analysis, modelling and RUL estimation is provided in [61].

Moreover, a state space model for prognostics using a mathematically derived Kalman filtering approach is discussed in [62]. The proposed model is shown to result in lower RUL estimation errors in comparison to other approaches defined in the same study and based on a use case concerning on a battery RUL prediction.

Nevertheless, mathematical modelling is a complex process. However, with the rapid development of information and communication technology, data-driven approaches are gaining momentum, promising increased accuracy in RUL estimation. An approach for estimating the RUL of rolling element bearings is discussed in [63]. Relevance vector machines were used with different kernel parameters while experimental results support the effectiveness of the proposed hybrid approach.

Considering the enormous amount of data generated in manufacturing shop floors, Artificial Intelligence (AI) and in particular deep learning approaches have been widely investigated with promising results [64, 65]. Deep recurrent neural networks are discussed in [66, 67]. In both studies the need for a large dataset and sufficient training data is highlighted. LSTM recurrent neural network is another widely used AI method with reports validating its performance [68, 69].

In addition, novel approaches based on data-driven techniques are investigated, such as transformer-based neural networks [70] and digital twin models [71].

2.4 Integrated Solutions

A parallel research effort has been devoted to proposing integrated solutions providing both an architectural engine and data-driven methodologies [43, 46]. The work in [46] proposed a distributed engine, enabling predictive maintenance based on a dynamic combination of native interpretable algorithms to provide human-readable knowledge to the end user.

However, integrated solutions have been customized either to a specific use case or to a given data analytics task, such as predictive analytics driven by interpretable models [46], forecasting activities in the smart-city context [43]. More general solutions should be devised with the final aim to easily integrate them in existing industrial solutions.

3 Key Challenges Identified

The decreasing cost of electronics in tandem with their ever-increasing capabilities made sophisticated embedded electronics cost-effective for various applications. As a result, complete historical trends can be stored in digital means and processed by advanced techniques, revealing underlying patterns and detecting anomalies. Thus, data-driven techniques pave the way for novel strategies bringing additional benefits for adopters. In the case of manufacturers, this can bring down production costs via multiple approaches, predictive analytics, and the resulting adoption of predictive