IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Big Data Life Cycle in Shop-floor – Trends and Challenges

TERRIN PULIKOTTIL[1,2], LUIS A. ESTRADA-JIMENEZ[1,2], JOSÉ JOAQUÍN PERALTA ABADÍA[3], ANGELA CARRERA-RIVERA[3], AGAJAN TORAYEV[4], HAMOOD UR REHMAN[4], FAN MO[4], SANAZ NIKGHADAM-HOJJATI [1,2](MEMBER, IEEE) AND JOSE BARATA[1,2] (Member, IEEE)

[1]UNINOVA—Centre of Technology and Systems (CTS), FCT Campus, Caparica, Portugal
[2]Faculdade de Ciencias e Tecnologia, Departamento de Engenharia Electrotecnica, Universidade Nova de Lisboa, Caparica, Portugal
[3]Faculty of Engineering, Mondragon University, Arrasate, Spain
[4]University of Nottingham, Nottingham, UK

Corresponding authors: Luis A. Estrada-Jimenez (e-mail: lestrada@uninova.pt) - Fan Mo (e-mail: fan.mo@nottingham.ac.uk)

This research is supported by the Digital Manufacturing and Design Training Network (DiManD) project funded by the European Union through the Marie Skłodowska-Curie Innovative Training Networks (H2020-MSCA-ITN-2018) under grant agreement no. 814078.

ABSTRACT Big data is defined as a large set of data that could be structured or unstructured. In manufacturing shop-floor, big data incorporates data collected at every stage of the production process. This includes data from machines, connecting devices, and even manufacturing operators. The large size of the data available on the manufacturing shop-floor presents a need for the establishment of tools and techniques along with associated best practices to leverage the advantage of data-driven performance improvement and optimization. There also exists a need for a better understanding of the approaches and techniques at various stages of the data life cycle.

In the work carried out, the data life-cycle in shop-floor is studied with a focus on each of the components - Data sources, collection, transmission, storage, processing, and visualization. A narrative literature review driven by two research questions is provided to study trends and challenges in the field. The selection of papers is supported by an analysis of n-grams. Those are used to comprehensively characterize the main technological and methodological aspects and as starting point to discuss potential future research directions. A detailed review of the current trends in different data life cycle stages is provided. In the end, the discussion of the existing challenges is also presented.

INDEX TERMS Big Data, Data Life Cycle, Intelligent Manufacturing, Machine Learning, Literature Review

I. INTRODUCTION

The evolution of data storing and analyzing has been a key factor in the development of manufacturing processes. During the pre-industrial revolution, low quantities of data were stored and were mostly transmitted verbally, which led to low production volumes and low quality products. Thereafter, during the first industrial revolution, two kinds of data were being recorded, i.e. machine and worker data. Worker data (attendance and performance) and machine data helped to improve productivity and maintenance, respectively. The mass production model introduced in the second industrial revolution also shifted the job of data processing to educated managers. Scientific methods and statistical models helped in all stages of manufacturing from production planning to inventory management [1]. With the introduction of IT in manufacturing, computer systems, such as CAM and FEA, and information systems, such as MES and ERP, helped in product creation, process optimization, and management. The merge between data and manufacturing in the information age has helped in the shift from dedicated production to flexible production. The extension of IT with unified communication, i.e. ICT further enhanced the role of data in manufacturing.

The concept of SM has emerged as a new paradigm focused on responding in real time to constant changing demand and conditions in factories, supply networks, and customer needs [2]. Three key SM technologies include: (i) CPS (physical assets integrated with computational capabilities), (ii) IoT (highly connected devices with embedded sensors), and big data [3]. The big data age has arisen with the massive

Acronyms

ARIMA	Auto-Regressive Integrated Moving Average
CAD	Computer-Aided Design
CAM	Computer-Aided Manufacturing
CATIA	Computer-Aided Three-Dimensional Interactive Application
CNC	Computer Numeric Control
CNN	Convolutional Neural Network
CPS	Cyber Physical System
CSV	Comma Separated Values
CSS	Cascading Style Sheets
DCS	Distributed Control System
DDBS	Distributed DataBase System
DevOps	Development Operations
ERP	Enterprise Resource Planning
FEA	Finite Element Analysis
HDFS	Hadoop Distributed File System
HMI	Human-Machine Interface
HTML	HyperText Markup Language
IT	Information Technology
ICT	Information & Communication Technologies
IoT	Internet of Things
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
MES	Manufacturing Execution System
ML	Machine Learning
MQTT	Message Queuing Telemetry Transport
NFC	Near Field Communication
OEE	Overall Equipment Effectiveness
OPC	Open Platform Communications
OPC-UA	OPC Unified Architecture
OS	Operating System
OWL	Web Ontology Language
PLC	Programmable Logic Controller
RDBMS	Relational DataBase Management System
RDF	Resource Description Framework
RFID	Radio-Frequency IDentification
RMS	Root Mean Square
RQ	Research Question
SCADA	Supervisory Control And Data Acquisition
SM	Smart Manufacturing
SQL	Structured Query Language
SWRL	Semantic Web Rule Language
TCP/IP	Transmission Control Protocol/Internet Protocol
TSDB	Time Series DataBase
VPN	Virtual Private Network
WiFi	wireless fidelity
WoS	Web of Science
XML	Extensible Markup Language

use of mobile and smart devices, the great availability of IoT devices, and cloud computing, when traditional methods were not sufficient for adequate information processing [4]. In general, big data refers to the storage and analysis of data sets that are characterized by large *volume* and *variety* of sources, high *velocity* of generation and processing, and *value* generation from its analysis [5].

In the age of big data technologies, various data sources generate manufacturing data, which are collected from connected software solutions, sensors, and IoT devices. On a high level, manufacturing data may be categorized into management, equipment, user, product, operational, and process data [1], [6]. On a low level, manufacturing data may be categorized into structured, semi-structured, and unstructured data [7]. Structured data have clear relationships between their attributes and is the simplest data type to store and organize, usually represented as tables. Unstructured data comprise most manufacturing data, has no associated data model, and cannot be organized using tables or spreadsheets. Examples of unstructured data include images, audio, text, video. Semi-structured data do not reside in relational databases but have an organizational structure that makes them easier to analyze. Examples of semi-structured data include XML, JSON, and HTML.

The collection and processing of the data in the shop-floor is critical, as most manufacturing operations are carried out there. The advent of IoT and new industrial protocols have supported the acquisition of the information from manufacturing cells, products, transport systems, and people [8]. Thus, many data-driven SM applications have emerged recently, e.g., smart design, smart planning and process optimization, material distribution and tracking, process monitoring, quality control, and smart equipment maintenance [1]. This SM applications rely on transforming primary data to information, making manufacturing processes more intelligent. Examples of shop-floor data include energy consumption, quality test, equipment status, equipment parameters, resource loading, delivery time, and material data [1]. However, despite the benefits foreseen by the usage and processing of data in the shop-floor, challenges in SM need to be considered.

The 5Vs characteristics of big data are widely acknowledged as challenges of big data in manufacturing, including: (i) volume (level of data size), (ii) velocity (ingesting or processing big data in streams or batches, in real time or non-real time), (iii) variety (dealing with complex big data formats, schemas, semantic models and information), (iv) value (analysing data to deliver added-value to some events), (v) and veracity (validate data consistency and trustworthiness) [9]. In addition, cybersecurity is an important aspect in manufacturing. Since big data platforms connect physical spaces with cyber spaces, the danger of not considering cybersecurity might swiftly spread to physical parts of manufacturing systems [7].

Influx of big data generated from multitude of production systems (data sources) on the shop-floor complicates decision making. Combined with multiple data sources, varied transmission protocols and storage requirements for production systems on shop-floor further complicate decision making. As such, the increasing size of data on the shop-floor requires accurately classifying data for reliable decision making. This study aims to develop a homogeneous approach to gathering and utilizing data on shop-floor in manufacturing environments, based on influences and insights of a literature review. Therefore, the complete data life cycle is reviewed.

A need for reviewing the data life cycle in the shop-floor is identified, as research in this field has focused on other aspects of big data, i.e. applications, manufacturing systems and processes, decision making, economics, supply chain, business management, and product life cycle (see Table 1). This work focuses mainly on big data life cycle in the shop-floor, where increasing complexities of data life cycle management requires a detailed review. The effective

use of data sources for generating big data for objective completion is studied. Needs, requirements, and methods for data collection and data transmission are also reviewed. Special focus is given to homogenising data acquired, as multiple production systems operate on several protocols and technologies, generating heterogeneous data. Thereafter, data storage, data processing, and data visualisation applied to shop-floor in manufacturing is reviewed. Finally, the review builds on the aforementioned aspects of the data life cycle to elaborate on data application.

This contribution leverages the data life cycle for capturing big data in shop-floor. Specifically, the suitability and adaptation of big data life cycle to shop-floor in manufacturing is the main goal in this contribution. This study, addressing the need for big data on shop-floor, establishes the approach for data acquisition, processing, and utilisation for decision making. Challenges towards real-time data-driven manufacturing are also elaborated.

The rest of this paper is organized as follows. *Section II* presents the data life cycle to have an uniform terminology for big data in shop-floor. *Section III* presents the methodology used to understand current trends and future challenges of big data in shop-floor. *Section IV* presents the results of the literature review, based on data life cycle presented in *Section II*. *Section V* presents a discussion of the results and existing challenges in implementing big data in shop-floor. Finally, *Section VI* presents the conclusions, as well as on outlook on future work.

II. DATA LIFE CYCLE

Big data, and data in general, requires to be structured into specific content formats and context to be useful for users [16]. Big data is useful for automating processes in manufacturing, as it enables machines to communicate among themselves and enables users to extract information and knowledge. As such, research has focused on the data life cycle and how to extract knowledge from varied, heterogeneous data sources, enabling informed decision making. In this context, the data life cycle in shop floor has been presented as consisting of seven stages. This list was developed considering similar works (Table 2) to have a simplified uniform terminology. Furthermore, Figure 1 presents a visual representation of the seven stages of the data life cycle.

1) **Data sources:** Data sources generate big quantities of data across all the manufacturing value chain and product life cycle, bringing the concept of big data to the shop-floor. According to Demchenko et al. [9], big data is characterized by the 5Vs model, i.e. high volume (big quantities of data), variety (data have different formats and sources), velocity (data is rapidly generated), variety (heterogeneous data in varied formats), and value (data has value, which needs to be extracted and analyzed). In this regard, the 5Vs model applies to big data sources in the shop-floor. Data sources includes manufacturing information systems, industrial IoT technologies, internet sources (e.g. e-commerce platforms and social networks), smart products, and governmental public data [1].
2) **Data collection:** After data sources generate data, data collection is performed. The collection is performed mainly by IoT technologies, by means of smart sensor nodes equipped with sensing devices, such as accelerometers and temperature sensors, and the data is then transmitted using standardized communication protocols [23]. Data collection may be performed at different frequencies, referred to as sampling frequency or sampling rate, based on the processing power of sensor nodes and the requirements of the variables being measured. In addition to shop floor data sources, other data collection sources, such as third-party application program interfaces or web crawling of internet sources, may be used to collect data, further enriching and expanding the context of data collected during the process.
3) **Data transmission:** Data transmission maintains the communications between the elements involved in the data life cycle, e.g. manufacturing systems and manufacturing resources. Defining standardized means of transmission, communication and application protocols define how the elements communicate data among each other, for example data transmission rate and communication range, ensuring real-time, secure, and scalable data transmission [24]. As with data collection, data may be transmitted at different frequencies, based on the requirements of the monitoring strategy, such as real-time data transmission or batch data transmission.
4) **Data storage:** Data obtained during data collection must be stored securely and integrally. Nevertheless, data sources have different formats and may be structured, semi-structured, and unstructured [25]. As stated in [26], the second design principle of knowledge discovery in big data is that one size does not fit all, and several different storage types must be considered. Besides structured data storage, object-based storage provides a flexible solution for storing semi-structured and unstructured data, thus covering the integrity requirement of data storage. In addition, by means of cloud computing, data storage may achieve cost effectiveness and high-processing power, as well as security, scalability and heterogeneity.
5) **Data processing:** Data processing builds upon data storage and refers to the operations required to extract information, i.e. knowledge from heterogeneous data sources. By processing raw data, hidden information and patterns may be revealed, providing stakeholders with valuable information for decision making. Different processing techniques and tools may be used depending on analysis to be done on the data. Big data may be processed efficiently by means of data cleaning, data reduction, data analysis, and data mining techniques, owing to advances in artificial intelligence,

TABLE 1: Existing review papers about big-data on manufacturing

Publication	Main Focus
Cui *et al.*, 2020 [7]	Identified key drivers for big data applications, essential components for big data ecosystem, research domains and future directions based on manufacturing requirements and available big data capabilities.
Wang *et al.*, 2021 [10]	Presents big data frameworks, key technologies and application considering concepts of model and data driven methods for manufacturing systems. The work also sheds light on the current challenges and future research opportunities.
Chunquan Li *et al.*, 2021 [11]	Provides analysis for big data-driven decision making considering the practicability of intelligent technologies in manufacturing. The work also provides a conceptual framework, challenges and future research directions.
Mageto, 2021 [12]	Presents an argumentation model considering the elements of big data analytics including security and economics for understanding the impact of big data on the sustainable manufacturing supply chain.
Belhadi *et al.*, 2019 [13]	Proposes a novel model with focus on big data analytics on manufacturing processes considering ongoing research and phosphate manufacturers as case study.
Sahoo, 2021 [14]	Presents a bibliometric, visual and factorial analysis to understand the research clusters in business management of big data analytics in manufacturing.
Ren *et al.*, 2019 [15]	Proposes a conceptual framework considering the product life cycle and sustainable manufacturing based on big data, as well as the potential applications and research directions.

TABLE 2: Relevant data life cycles proposed in literature

Ref.	Stages	Summary
Levitin *et al.*, 1993 [17]	Define view (planning), implement view (sources), obtain values (collect and store), assess and analyze (data quality check), update records (processing), present results, assess and analyze (updated data quality check), use data, assess and analyze (results check), and delete data.	The authors propose two data lifecycles : acquisition and usage. When merged, the data lifecycle is comprised of three main activities: (i) data generation/collection, (ii) data storage, and (iii) data processing/utilization. The complete data lifecycle starts with planning phases, i.e. define views of data to be acquired and implement them in the sources. Data is then obtained and assessment and analysis of the data is performed. Following, data collected is update by adding, modifying or deleting data (roughly, data processing) and presented to users. The processed data is once again assessed and analyzed before being used for data analysis. Finally, the usage is assessed and analyzed, deleting unnecessary resulting data.
Yoon *et al.*, 2000 [18]	Create metadata, create metadata structure, use data, manipulate data, refine data, and refine metadata.	The authors build upon the work of [17], adding additional metadata stages. Metada creation begins by defining data architecture and data model structures (views). Metadata structuring implements the views and associates them to data sources. Data creation collects and stores the data, as well as performs quality checks. Then, data utilization focuses on using and presenting the data and is followed by data manipulation, where data is processed by altering data forms or values. Data assessment determines the manipulated data suitability for current and future use. If data is not suitable, it may be corrected either by data refinement or metada refinement stages.
Borgman *et al.*, 2007 [19]	Design experiment, calibrate, capture, derive, integrate, analyze, publish, and preserve.	The authors propose a big data lifecycle comprised of eight stages. Experiment design begins by reusing historic data to design new experiments. Before deploying the system, sensors are calibrate to known values. Then, the system is deployed and data is captured from data sources. Afterwards, deriviations processes the data and reveals hidden data or salient features and data from different sources is integrated. With the processed data, data analysis is performed to generate knowledge and information. Results are then published and the data is stored for preservation.
Fisher *et al.*, 2012 [20]	Acquire data, choose architecture, shape data into architecture, code/debug, and reflect.	The authors clustered data analyst tasks for big data into a life cycle comprised of five stages. Big data sources are identified and data is acquired. Based on cost and performance, an architecture (computing platform) for processing and analyzing data is chosen, such as cloud computing. Thereafter, data is uploaded to the chosen platform and is then cleaned, processed and reshaped (when needed). Finally, the data analytics approach is coded and debugged and the results are reflected (visualized) by the user to extract knowledge and information.
Khan *et al.*, 2014 [21]	Raw data, collect data, filter and classify, analyze data, store data, share and publish, and retreave and discover data.	The authors propose a big data lifecycle comprised of seven stages (eight including raw data). Collection starts by obtaining raw data from data sources. Data is filtered, classified and mined to identify salient features. Thereafter, data analysis is used to understand the patters and correlations in the data and to develop methods to accurately predict future observations. During storing, big data datasets are stored and managed with reliability, availability, and accessibility and are shared and published. Finally, retrieving and discoverying of the data is made possible for further analysis and historic queries.
Chi *et al.*, 2016 [22]	Select data application, identify big data, deploy big data, select innovative data method, visualize big data, and interpret big data.	The authors discuss a big data lifecycle comprised of six stages. First, design and planning of the business need is done during data application selection. Then, big data sources are identified, big data collectors are deployed and collected data is stored. With the stored data, innovative data methods are to be developed to process and extract knowledge from big data datasets. Finally, big data and obtained results are visualized and interpreted by users.

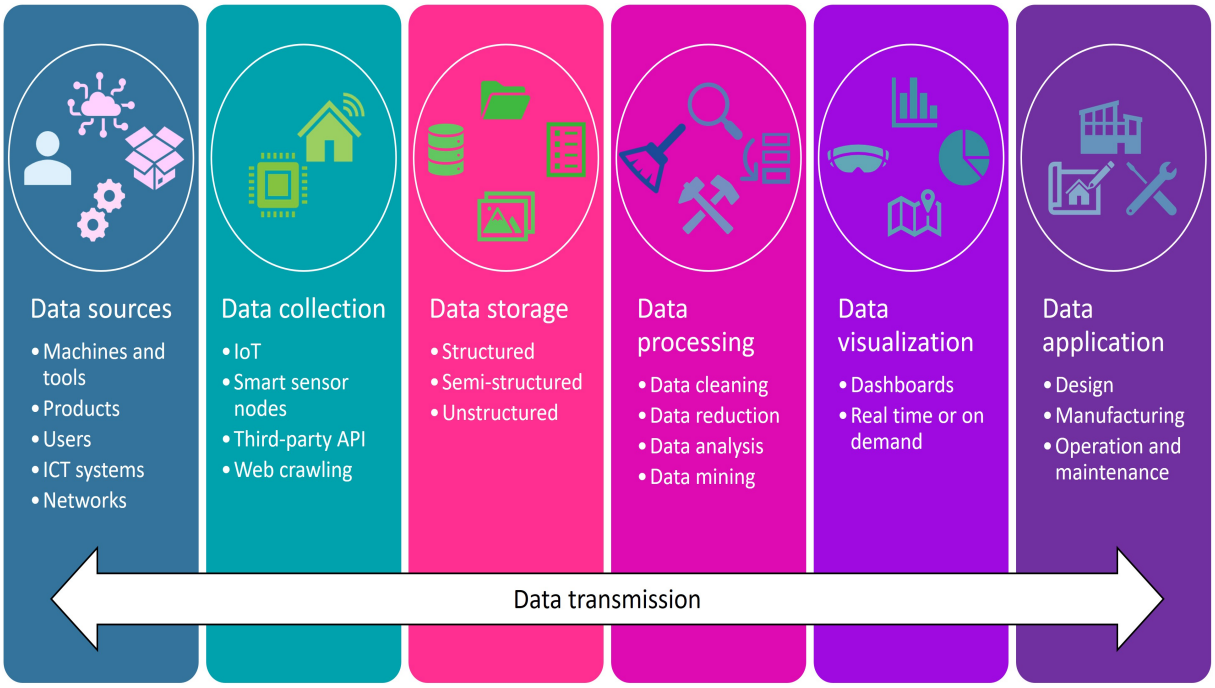

FIGURE 1: Proposed data life cycle

cloud computing and IoT [1]. As such, the first design principle of knowledge discovery in big data is that data processing should be supported by a variety of data processing methods and analysis environments [26].

6) **Data visualization:** Data visualization provides the means to visually understand the information extracted during data processing. Data may be visualized in dashboards, including statements, charts, graphs and augmented reality [27], and data may be queried in real time or on demand, based on the users needs, enabling decision making based on historical or real-time data. In addition, data visualization should be accessible and easy to understand, as stated in the third design principle of knowledge discovery in big data [26]. As such, popular open standards and lightweight architectures should be used for presenting results, as well as exposing the results using application program interfaces for third-party software integration.
7) **Data application:** Data application refers to data analytics performed during the entire product life cycle, providing stakeholders with tools for decision making. Data analytics may be applied during the design phase, translating customer needs into product features and quality requirements [28]. Thereafter, during production, data analytics monitor the production process and lead to informed decision making regarding the manufacturing process, improving product quality and reducing production costs [10]. Finally, during product operation and maintenance, data analytics may be used to predict possible faults and to provide preventive maintenance, elongating the life cycle of the product and improving relationships with costumers [10].

III. OBJECTIVE AND METHOD

The work focuses on understanding the trends and challenges in implementing big data on shop-floor applications, emphasizing their data life cycle. For this purpose, a narrative literature review was carried out supported by the extraction of n-grams that allow the preliminary exploration of related trending research. The following research questions guide hereafter the development of this review.

- **RQ1:** What are the recent trends in big data life cycle in shop-floor?
- **RQ2:** What are the main challenges and future research directions in big data life cycle in shop-floor?

A. METHODOLOGY: NARRATIVE REVIEW

Narrative reviews contemplate the identification of several key studies that describe a problem of interest to have a general overview of a field [29]. Despite having a less rigorous approach compared to a more systematic one, in this paper we support the selection of references of interest by extracting monograms, bigrams, trigrams, and quatrograms, related to the main RQs and objective of the work.

B. OBJECTIVES

This review's objectives are twofold and is aligned with the research questions presented above.

- In terms of *RQ1*: What are the recent trends in big data life cycle in shop-floor? The main interest is to briefly characterize technologies and approaches, considering

the seven stages of the data life cycle presented in section II of this review.
- In terms of *RQ2*: What are the main challenges and future research directions in big data life cycle in shop-floor? The main interest is to discuss the general challenges mentioned in literature, establishing a baseline of potential future research directions.

C. STUDY IDENTIFICATION, SCREENING AND INCLUSION

A set of keywords has been chosen considering relevant terminology in the area. Core concepts reflected here are "manufactur*", "factory", "factories" and "shop floor". Those are accompanied by the keyword "big data". Group 1 and Group 2 are linked with the operator *AND*, whereas internally they are linked by the operator *OR*. This resulted in the following string:

- ("manufactur*" OR "factory" or "factories" or "shop floor") AND ("big data")

The research string was applied in the electronic database *Web of Science (WOS)* as it is a well-known and large academic scientific repository. Fundamental consideration to select studies were:

- Works published after 2012.
- Review papers were excluded from the search.

A set of 4912 papers was obtained from this search. From this point, the strategy was first the extraction of monograms, bigrams, trigrams and quatrograms to have a brief notion of characterization of relevant terms in the field and based on such characterization the selection of main works of interest. The notion of data life cycle and the consideration of relevant application in shop-floor operations were additional considerations for the manual selecttion of papers of interest.

In the end, a total of 61 articles were chosen, and further analyzed to answer each of the RQs. Fig. 2 presents the methodology used in this review.

D. RESULTS

Fig. 3 presents the result of monograms, bigrams, trigrams, quatrograms from the set of papers collected. In general, we should highlight the presence of technological enablers like internet of things, cyber-physical systems, artificial intelligence (neural networks), digital twin models, cloud computing and other which are supporting the implementation of big data in the shop floor. Other representative key words are related to specific applications e.g. predictive maintenance, energy optimization, product quality, process monitoring, anomaly detection and decision making process. From another perspective, cloud computing and edge computing are also highlighted as computation infrastructure to treat the data. Various of these properties are used as a baseline to characterize the data shop-floor data life cycle in the next section of this review.

IV. RECENT TRENDS IN SHOP-FLOOR BIG DATA LIFE CYCLE

This section explains the results of the narrative review of publications related to big data life cycle. The section is divided into different stages of data life cycle. The results presented are a collective overview of studies presented in the last decade on each of these stages related to big data in manufacturing shop-floor.

A. DATA SOURCES

Different applications in the context of smart manufacturing require different data sources (Figure 4). They are mostly based on the utilization of IoT devices i.e. sensors that collect data from machines, shop-floor, products, people and environmental variables. Other important data sources are the ones provided by heterogeneous product requirements, specially in product driven manufacturing applications.

For decision making activities, examples of data sources include customer requirement documents, datasets, and CAD models. These sources are multi-modal with different forms and, hence, require separate processing methods. Another example is information embedded in CAD models. In this context, Collada can be used as the data format to describe CAD models. If CAD models are designed in CATIA V5, then converters from CATIA V5 to Collada can be used to obtain Collada models [30].

Devices used to monitor energy in shop-floor include smart meters, current and voltage clamps, and machine-integrated devices that provide out-of-the-box instantaneous power consumption [31]. Industrial robots, for example, can provide power consumption for each joint of the robot directly from a robot controller [32]. Experimental data regarding actuation torques and servo drive voltages, used directly to derive input power of plants, can be captured with energy sensors, such as clamps [33]. Alternatively, single-phase and 3-phase smart plugs have become popular for monitoring the energy consumption of manufacturing equipment on the shop-floor [34].

Human data can also provide additional context information to current shop-floor situations. This data provide a better user experience for operators, improving productivity and decision quality. Human data can be divided into human attribute data and state data. Human attribute data are comprised by demographic and characteristic information that does not change or changes sporadically (e.g. age, profession, education status, and skills). This data may be used for "user modelling" to deliver information or services according, for instance, to the proficiency, skills, and interest of the user. Human state data refers to a collection of all kinds of data that may be used to model abstract human characteristics, such as behaviour and comfort [35]. Traditional IoT devices may acquire data about the state of operators (e.g. current position and vital functions). For instance, wearable trackers measure human performance under stressful or difficult conditions, analyzing the data and sending warnings when needed [36]. Furthermore, operators can use portable smart devices (e.g.

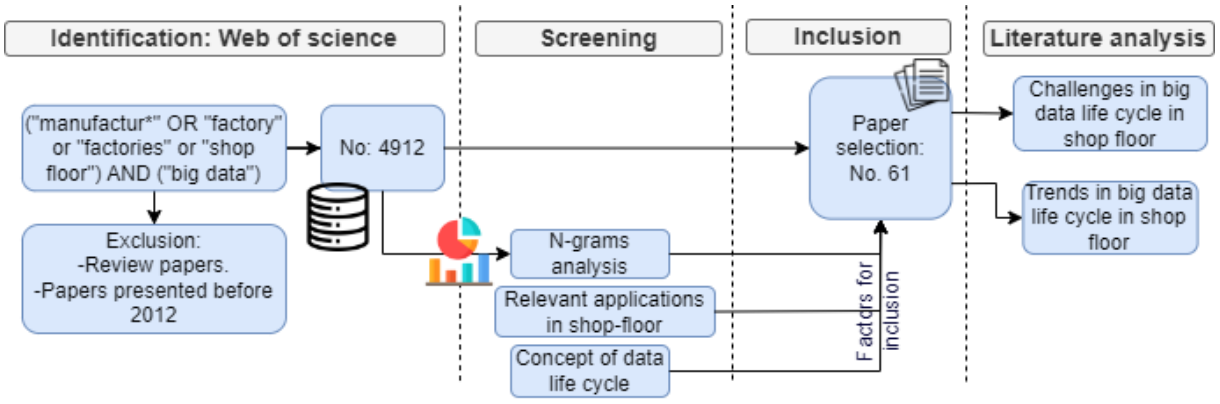

FIGURE 2: Methodology used for the narrative literature review

smartphone, smartwatch, and tablet) with NFC readers to check into a location and receive information about relevant parts of the production system equipped with NFC or RFID tag. [36]. The behaviour can also be inferred through interactions that users have with machines or applications, capturing the interactions with plugins or applications, such as Google Analytics and Matomo. Acquired data can be uploaded to cloud services using IoT technology, where it is processed and analyzed to deliver personalized information to operators and supervisors, informing about potential issues.

Most applications for data-driven automation rely on optimal decision making, considering status of machines and conveyors (availability) [37]. Smart sensors have been used to track equipment and people e.g. RFID tags [38]–[41]. Smart sensors have also been used to monitor best conditions of machines, e.g., in terms of temperature [42]. In addition, information of images (quality control) has been used as a decision factor for autonomous reconfiguration and adaptation processes [43].

Data-based maintenance sensors that have been used in literature include vibration [44], [45], acoustic emission [44], [45], temperature [40], [44], current [44], [45], velocity [40], pressure [40], and forces [46], implemented in various parts of the machine. The sensor may exist in the machine [47] or may be installed as add-on sensors dependent on the application. PLC controllers provide process-related data, such as cutting speed, feed, and depth of cut [44]. Application-specific data sources also contribute in monitoring and maintenance activities. For example, 3D laser scanners have been used to evaluate tool flank wear [45]. Other sources have used device status (such as alarms and logs) [47] and historical failure data [48] logged after quality inspection, aiding in identifying product failure patterns. RFID tags also have been used to identify defective products, comparing with the failure data [40].

Accuracy and quality of data play a vital role in successful implementation of intelligent systems, depending on the effectiveness of data sources. However, data gaps and incompatibility in system applications may be found. To overcome them, proper calibration of data sources is needed. Data sources consist of automation system resources (such as sensors, actuators, PLC, SCADA, DCS, and CNC systems), identification systems (such as RFID, AutoID, barcodes, and vision systems), communication standards between production resources (such as fieldbus and wired and wireless communication), with accompanying data exchange standards (such as OPCUA, MTConnect, and MQTT).

Automation technologies allow a significant reduction of human participation on the shop-floor during production operation. On the one hand, there are processes that may not be automated, mainly due to infeasibility of economic outcome. Specific production processes may involve manual work to be carried out in different manners. The employee carrying out the work may enter the information to a management support system. Nevertheless, the information accumulated from employees through this approach is highly unreliable and cannot be used for machine adaptation. On the otherhand, production systems may perform automated data acquisition without human intervention. Data accumulated in this manner can be used for decision making. However, interfaces and processing of the data may be necessary. Most common data sources in automated production systems for machine adaptation have been identified to be control and measurement devices, measurement instruments (such as sensors and transducers), PLCs (and other control mechanisms), and robots.

B. DATA COLLECTION

The data collection techniques for decision-making are dependent on the data sources. In case of customer requirement, natural language processing techniques, such as named entity recognition [49], relation extraction [50], and attribute extraction [51], have been used. If data come from datasets, deep learning techniques and sampling techniques have been used to collect data [52].

There are mainly two types of data collection techniques, manual data acquisition and automatic data acquisition. Manual data acquisition techniques are employee dependent and are gathered through a manufacturing support system. How-

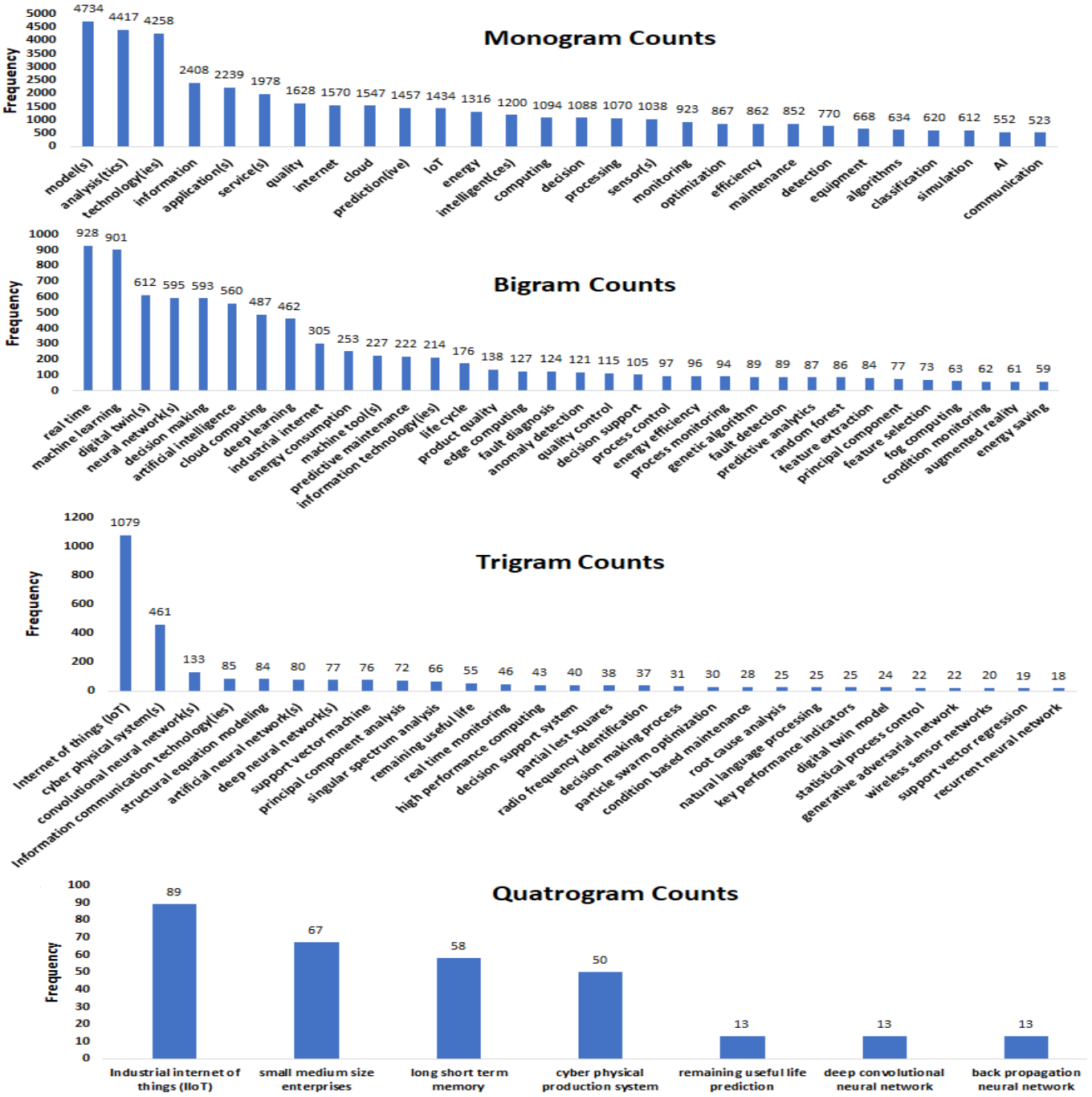

FIGURE 3: Relevant N-grams on the related publication

ever, they are highly inconsistent and unreliable [53]. Automated data collection is performed by automated systems like sensors, measuring, and control devices that correspond to changes in physical processes [54].

Data collection in shop-floor depends on the nature of the data, i.e. structured and unstructured [55]. Multiple frameworks are in-place that incorporate data collection strategies for structured and unstructured data [55]. Data collection for machine adaptation is a six-step process involving initialisation, configuration, capturing, analysing, and focusing [56].

Cui et. al. [7] stated that almost half of big data collection applications were distributed in monitoring (25%) and predictive applications (24%), characterized for real-time process and non-real-time process, respectively. Real-time process data analysis in manufacturing refers to methods where data from production lines are acquired, processed, and delivered to operators. Thus, it is possible to timely detect anomalies or to quickly know the status of the shop floor, production, machines, and personnel [57]. This is one of the basic needs for operators on the shop-floor, who require a

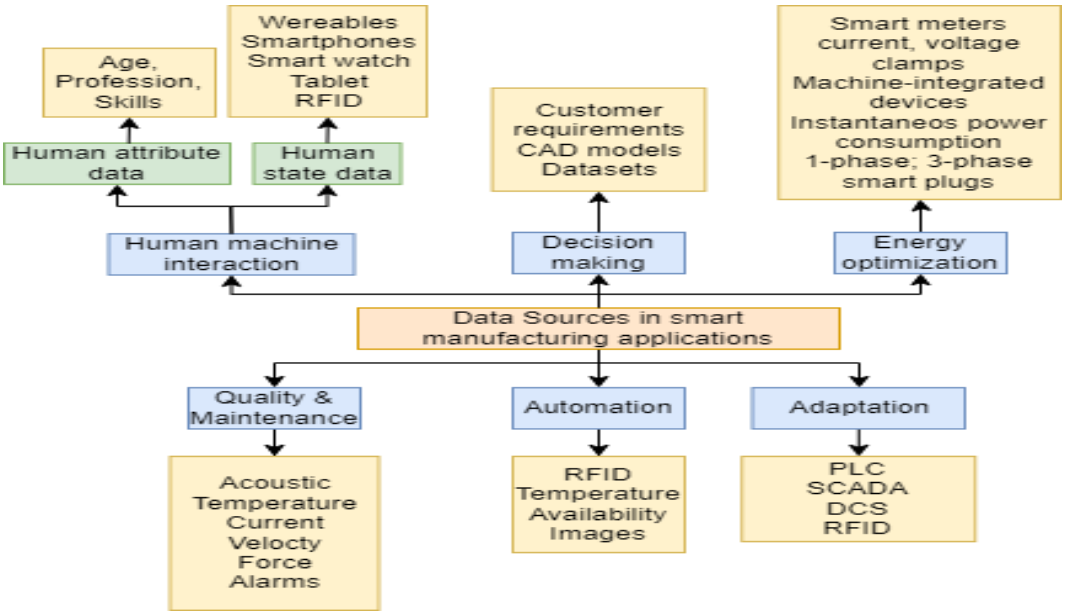

FIGURE 4: Data sources in smart manufacturing applications

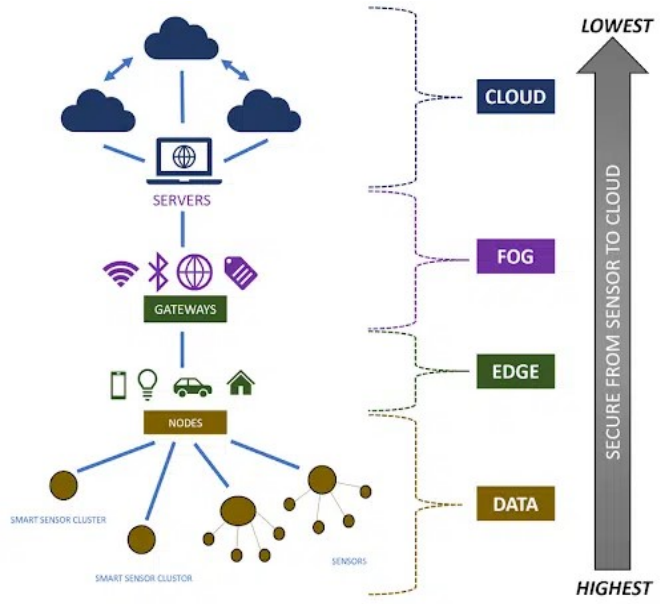

FIGURE 5: Data Collection in Manufacturing applications

synthesized and centralized view of multiple data sources, which could be highly dispersed. Nevertheless, predictive applications do not necessarily require a real-time data collection and focuses on extracting patterns and trends based on historical process data for optimization and management innovation [57].

Although real-time data collection is preferred, in practice, it is seldom the case for maintenance-related data. Add-on sensors, such as temperature, vibration, pressure, force, and process data from PLC controllers (cutting speed, feed, and depth of cut), may provide near real-time data. Device status and logs have been periodically collected and stored [47]. Wear information has been collected after a predefined amount of time to accurately analysis the wear (e.g. tool wear is measured every 20min in [44]). Process parameters and performance metrics (historic data) have been collected after each production run/shift [40], such as maintenance history and failure records [48]. Almost all data relevant for monitoring or maintenance are time series, being assigned time stamps during collection. Data collection techniques (Figure 5) include support for RESTful/configurable application layer protocols, OPC unified architectures, and distributed data acquisition (e.g. Flume [47]).

Automation activities rely on event-driven data collection techniques e.g. time driven, quantity driven, operation driven [58]. Event driven approaches allow the storage of manufacturing information after a specific time interval. These techniques are also useful to query manufacturing services for process automation purposes. Optimal decision making usually require storage of historical data and the comparison with a real-time monitoring data collection [40].

For time-driven data collection, energy data from manufacturing equipment has been studied. Energy is usually monitored in given time intervals, such as every 15 minutes, monitoring total energy consumption. However, some applications, such as profiling the robotic motions and understanding the parameters affecting the energy consumption, requires real-time energy data sampled in few milliseconds [59].

C. DATA TRANSMISSION

Data transmission protocols includes sockets, OPC-UA, MQTT, TCP/IP (such as PLC simulator), or other communication protocols (Figure 6). Data transmission protocols depend on the application domain and may be dynamically

chosen. Data transmission is used as the communication channel between different devices, including IoT devices, workstations, and digital twins. When workstations in manufacturing environments use different operating systems, OPC-UA is a suggested solution. Cloud-based systems have also been recommended, as modularity among components of the pipeline is promoted [60].

The transmission of data for further processing depends on the logging frequency of the data. High-frequency data may be stored first in storage devices of monitoring solutions. Thereafter, collected data are transmitted manually in batch to processing computers via Ethernet connections. Some monitoring solutions also offer transmitting data via WiFi. Transmitting energy data via WiFi has the benefit of transport flexibility and high transmission distance. However, WiFi comes with shortcomings, such as high latency and transmission unreliability. Hence, industrial standards such as Modbus and Profinet have been used for mission-critical applications [59], [61].

Process automation may require connecting manufacturing resources to the Internet. Generally, the connection has been done by Ethernet [37] and wireless communications [38]. Data transmission has also been implemented using industrial standards with higher reliability, such as OPC-UA, Modbus, and Profibus [62]. IoT communication has been used to perform data transmission using publish/subscribe messaging, e.g. MQTT protocol [43], for event-driven process automation purposes.

Real-time data may be transmitted using WiFi, Zigbee, and 4G through Internet and using VPNs. Non-real time data may be transmitted through technologies or application like Apache Sqoop and Data/X [40]. Production and sensor data with high frequency have been transmitted through Ethernet to a local server and then, after feature extraction, have been sent to cloud servers in the Internet using WiFi protocol [44].

The introduction of IoT in the shop-floor has increased the transmission of low-frequency sensor information directly from the source through WiFi from various sources. This has also had an impact on the latency of the system response. Data transmission rates play a vital role that depend on the manufacturing application. To incorporate multiple data formats, standards, and needs for machine adaptation, a combination of technologies is proposed in this study to assists in data transmission. To this end, a data transmission framework is necessary to improve data transmission across the production domain.

D. DATA STORAGE

Common data formats to store machine information are XML and JSON files [38]. Different data types include structured (formatted as tables), semi-structured (such as XML, JSON, and HTML) and unstructured data (such as documents, images, audio, video, text, and emails) [58]. Table 3 oresents data storage types and technologies used in manufacturing shop-floor. Unstructured data are first processed to extract relevant information internally before being stored

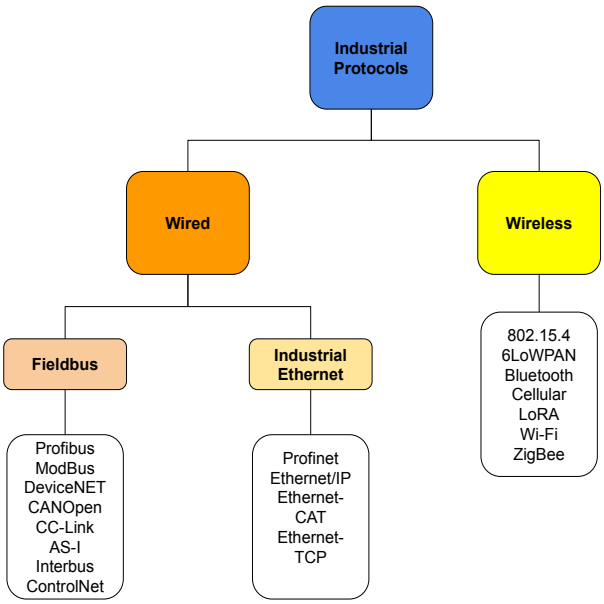

FIGURE 6: Industrial protocols for Data Transmission

in databases. For example, tool wear information has been extracted from wear images using image processing software and converted into flank/crater wear values along with their time stamps [45].

Depending on the data type, data may be stored using several techniques. Traditionally, RDBMS and DDBS have been used for structured data. RDBMS are characterized by well-defined schemas and relationships. For example, basic user information may be stored in traditional database systems such as MySQL, PostgreSQL, and SQLite. RDBMS have been used for user interaction data storage. For instance, Matomo, an user analytics platform, captures user interaction streams (e.g. clicks and page views) in MySQL and MariaDB databases. However, RDBMS offer limited scalability.

NO SQL databases (e.g. MongoDB and Cassandra) have proven to be better approaches for semi-structured (JSON, XML) and unstructured (audio, video) data. In addition, XML has been used to transform structured data to semi-structured data [40]). HDFS may also be used for dealing with unstructured data. Some examples of these kind of databases include:

- Cassandra to store event data of automation controller.
- MongoDB (document NoSQL database) to store machine data.
- TSDBS, such as OpenTSDB and InfluxDB, to store and access sensor time-series data.

Data models are also used to represent manufacturing data. Data models are comprised of two parts: (i) run time conditions (process knowledge and time-sensitive dimension) and (ii) process model (production requirements of products). Once data models are defined, knowledge graphs may be used to store data. There are two main types of storage for knowledge graphs: RDF-based storage and graph-based

storage. An important design principle of RDF-based storage is the ease of data distribution and sharing, while graph-based storage focuses on efficient graph queries and search. The Neo4j system is a widely used graph database [63]. It has an active community, and the system itself is efficient in querying. However, it lacks of support for quasi-distribution.

Smart manufacturing applications have used distributed file systems (for data-at-rest) and databases (for data-at-motion) for storage [37]. Historical data are ingested from databases to predict production planning performance, safety critical aspects, and network designs. In addition, Hadoop and MapReduce techniques may be used to reduce the storage space required for big data.

Production and sensor data from the machines have been initially stored in industrial computers connected to machines, which are then processed internally using feature extraction to understand the states of the machines. Thereafter, the data have been sent to cloud servers for managed and storage in a database, acting as remote server for data storage [44].

Automation applications relying in storage of manufacturing information, as well as services, have increased the responsiveness and interoperability of the shop-floor and thus, the automation capacity. The choice of storage solutions greatly affects the application. High-frequency big-data files require special solutions such as Hadoop and Spark that can deal with the high volume property of big data. Data have been recorded in regular time intervals, resulting in time-series data [64]. To this end, special database solutions for storing time-series data, such as InfluxDB, may be used. Also, relational database methods have been used for their reliability. Futhermore, some monitoring solutions have stored the collected energy in storage devices using CSV files.

TABLE 3: Data storage types and technologies used in manufacturing shop-floor

Data Storage Type	Data Storage Technologies
Relational database	MySQL, SQLlite, Oracle DB, SQL server, ProgresSQL
NoSQL database	Column-based: HBase,Cassandra Document-based: MongoDB Key value-based: Redis Graph-based: Neo4j
NewSQL database	VoltDB
Other data storage types	Time series data base : OpenTSDB Search engine : Solr, Elasticsearch, SparkSQL Data warehouse : Hive, Kylin ETL (Extract, Transform and Load) : Pig Others : HDFS, Clustrix, NuoDB

E. DATA PROCESSING

When data are collected and transformed into usable form, data processing takes place. Data processing must be done appropriately to avoid having detrimental impact on the final product, or data output. It is typically performed by data scientists or teams of data scientists. Different techniques can be used for data processing. Figure 7 presents the traditional data processing process performed in the shop-floor.

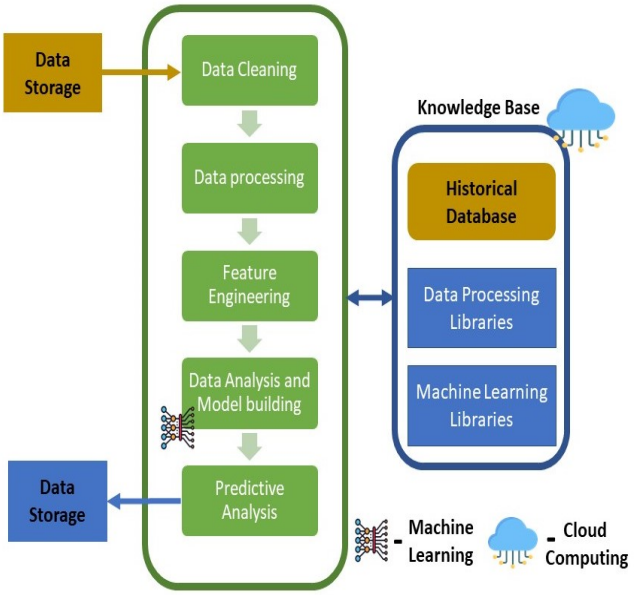

FIGURE 7: Data processing for big data shop-floor

Data processing is a computationally intensive task. First, data should be resampled to match the recorded timestamps. Resampling methods such as averaging, forward filling, or backward filling have been used in literature [65]. Averaging method takes an average value within a pre-defined time interval and replaces the missing values with the average value in the data. In forward- and backward-filling methods, missing timestamps are filled with values before and after the missing timestamp, respectively. Once data has been processed, it has been fed into application-dependent algorithms such as ARIMA, Seasonal ARIMA, Bayesian Optimization, clustering, neural networks [66], genetic algorithms [67] and parameter identification methods [68].

Several approaches exist for data processing in decision making. Several studies have used a method based on multi-neural collaboration to extract knowledge and the extracted knowledge has been classified according to labels. An ontology model and schema layer of the knowledge graph has been defined and the knowledge has been represented with fuzzy comprehensive evaluation [69]. Knowledge has been directly described as production rules [70] and as knowledge graph [71]. Owing to the wide range of knowledge sources, the knowledge base that has been constructed according to the two steps above has high redundancy. To this end, latent semantic analysis, similarity calculations and attribute weighting may be used to eliminate redundancy in the knowledge. First, the entity triples in the preliminary knowledge base have been mapped with the Protege ontology library, and then the semantic web rule language (SWRL) has been used to represent the empirical rule knowledge. Finally, the data layer has been instantiated to construct the final knowledge base [72].

As for the data processing in HMI, in addition to using several data mining and machine learning techniques, the

TABLE 4: Relevant references for big data shop-floor

Reference	Data Sources	Data Collection	Data Transmission	Data Storage	Data Processing	Data Visualization
Wang et. al., 2016 [37]	Machine, conveyors, products	NA	Ethernet, Wireless communication	HDFS	Hadoop, Hive(SQL), Yarn	Web pages
Rehman et. al., 2021 [43]	Images	Historical Data	Gateway, Internet, MQTT	Cloud Storage	Machine Learning services	User Interface
Tsuda et. al., 2014 [42]	Extrusion process data, Temperature and pressure sensor, machine controllers	NA	TCP-IP	Oracle, S7	Classification models	GUI: MATLAB, QuickCog
Yan et. al., 2017 [45]	Sensors - Vibration signals, acoustical signals, power and 3D laser scanner - Tool flank wear	Vibration (every min, 1MHz), Wear (every 20min), Power (every 1min)	NA	BMP format as unstructred data	Envelop analysis, statistical feature extraction, ANN model	NA
Lin et. al., 2017 [38]	Sensors, RFID, metrology, processing and machining data	NA	Wi-Fi, ethernet, 6LoWPAN, ZigBee/WSN, REST/SOAP	XML, JSON, Hadoop	Sparq, Impala, Hive	NA
Villalonga et. al., 2020 [73]	Machining data, Vibration Signals	NA	Profibus, Ethernet	Global Warehouse	Machine learning	NA
Lu et. al., 2019 [58]	CNC controllers, sensors, mechanical actuators, machine tool data	Time, quantiy & Operator driven event	Cloud via HTTP request	XMill	Data packaging techniques	SSPNET Web APP
Zhong et. al., 2015 [39]	Raw material, internal logistic opeartor, tuples	NA	NA	NA	Saptio-temporal pattern recognition, logistic knowledge interpretation, machine learning regression, structural insight abalysis	NA
Zhang et. al., 2017 [40]	RFID reader, sensors - pressure, velocity & temperature	Real time	non-real time data	Internet, 4G, WLAN, Sqoop Apache, Data/X	DDBS, XML, NoSQL/HDFS & Storm, Hadoop, Machine Learning	NA
Zhang et. al., 2017 [41]	RFID tags, Smart meters	NA	NA	Storm, Hadoop, Mapreduce, DDBS	Data cleaning, reduction, Clustering, Association, Classification, Prediction	NA
Wan et. al., 2017 [47]	Alarms, device logs, device status, machine center with embedded equipment	Nodes with Restful protocol , OPC unified arch with real-time, periodic & aperiodic data	Zigbee, Wifi, Industrial switch & routing	HDFS based on Hadoop, Sqoop, aMysql monitoring database	Correlation analysis, STORM cluster, Hadoop cluster using MapReduce batch calcuation	Large screen and mobile services, visual analysis reports
Lee et. al., 2015 [44]	Saw machining data, PLC controller, add on sensor - vibration, acoustic emission, temperature & current	Real time	Ethernet, WiFi	Cloud server	Time & frequency domain feature extraction, adaptive clustering	Web & iOS-based User interface
Bonnard et. al., 2021 [74]	PLCs	NA	OPC-UA, API REST, Ethernet, Wireless solutions, Modbus, Profibus	Cassandra	Spark, SparJAVA, Machine Learning	Web, mobile, dashboard
Nakata et. al., 2017 [48]	Failure data	Historical data	NA	NA	Distributed clustering method, scalable K-Means++, Apriori, FPGrowth, Deep Learning Classification (CNN)	Single View integrated failure map pattern monitoring & cause identification Screen
Zhang et. al., 2020 [75]	Smart meter, sensors, RFID tags	Distributed perception	Ethernet, RS232, Modbus	HDFS, HBASE	Feature extraction, clustering algorithm, data association analysis, anomaly detection	NA
Ji et. al., 2017 [76]	Machine state and parameters	Real-time	NA	Local network & cloud	Cluster analysis, Factor analysis, Analysis of correlation and dependence, Regression analysis, A/B testing, Data mining	NA
Wang et. al., 2020 [77]	RFID, Excel & video	NA	NA	DDBS, NoSQL	Data reduction, data transformation, data cleaning, data integration	NA
Cui et. al., 2020 [7]	Files & Web	OPC-UA and MTConnect	NA	RDBMS, NoSQL, NewSQL	Online analytic processing,online transaction processing	Zeppelin, Matplotlib, Tableau, D3, GraphX
Tao et. al., 2018 [1]	Manufacturing information System (MES, ERP, CRM, SCM, and PDM), IoT sensors, RFID	Real-time using IoT sensors	NA	Object-based storage, cloud storage	Data cleaning, data reduction	Statements, chart, diagrams, graphs, and virtual reality
Zambal et. al., 2018 [78]	Sensors, Digital Twin	NA	NA	HDF5 storage	Digital twin data defect analysis	NA
Kahvenci et. al., 2022 [79]	AGV, PLC data	Real time	OPC-UA, Modbus, MTConnect	InfluxDB	RESTful APIs, Operational KPI Calculation	Amazon Quicksight and PowerBI on Azure, Grafana, Kibana, Splunk and Custom dashboards
Yu et. al., 2019 [80]	Sensor data from multiple manufacturing plants	Real time data	OPC/UA, SFTP	Apache Hive Central Data Warehouse and MapR Database	Backend APIs, Apache Spark cluster computing, PCA predicting model for anomaly detection	Databoard (React Javascript framework)
Saez et. al., 2018 [81]	Sensors, Machines	NA	OPC/UA, MTConnect, TCP, HTTP	RDBMS, Mysql, SQL, Postgre, NO-SQL, MariaDB, Node4j, InfluxDB	Streaming: Spark, Micro Batching, Batching: Spark-ML and R	Grafana dashboards
Reuter et. al., 2016 [53]	Production feedback data	Plans, Ad-hoc control interventions, resources' efficiency & job status	NA	NA	Naïve Bayes Classifier (NBC), the Association Rule Induction (ARI) algorithm, and the k-Nearest Neighbor (kNN) algorithm	NA
Robertson et. al., 2017 [54]	Simulated data source	Manual, spreadsheet, database, automated	NA	NA	NA	NA
Azad et. al., 2020 [55]	Manufacturing equipment	Automated	NA	NA	SQL database, NoSQL database, and graph database	NA
Lu et. al., 2019 [58]	Event generators, event channels	Data packaging	Event procedding	cloud	cloud data analytics	cloud apps
Gadaleta et. al., 2021 [82]	Robot, differential probe, current	Data collection in datasets	Acquisition module Data Translation DT9826	Local on robot	Variation analysis under parameter effects	Plotting and heat maps
Wan et. al., 2017 [47]	Alarm, log and equipment status	Adjust sampling frequency and configure application layer protocols	Wired and wireless transmission	Cloud	Real-time active maintenance and offline analysis and prediction in the cloud is provided	Visual presentation through digital twin etc.
Nakata et. al., 2017 [48]	Yield analysis identifying the cause of failure from wafer failure map patterns	Utilization of Pattern Mining. Manufacturing history	NA	NA	Distributed clustering method scalable K-Means	Plots failure map pattern monitoring, failure cause identification and failure recurrence monitoring.

development of analytic solutions requires selecting the right strategy according to diverse scenarios. Streaming, large-batch, and small-batch analytics are the three main processing strategies for big data [81]. Streaming is a processing technique for real-time analysis of data streams, particularly necessary when data arrives at high velocity. Large-batch processing is the most traditional form of processing where big data volumes are collected, representing large periods of time (e.g. hours, day, week) and being analysed with complex machine learning models. For batch processing, real-time data processing is not a priority. Small-batch processing (also known as micro-batch) is the process of small cumulus of data on a small time window (e.g. seconds, minutes).

Data processing can be also used for automation. Intelligent decision making for process automation and self-organization requires the analysis of machine status and energy consumption. This makes necessary the use of machine learning techniques. Some examples for process automation include neural networks, support vector machines, and k-nearest neighbours [42]. Negotiation based approaches with machine learning have been used for choosing proper routing and transportation of products, e.g. for storing or scrapping [43]. Genetic algorithms have also been used under the scope of ML. For process automation, genetic algorithms find optimal production resources e.g. the ones with minimum energy consumption or the ones that require less production time. In general, classical machine learning techniques are enough for this type applications.

In maintenance sector, feature extraction of the time series data from sensors like vibration/forces include both time-domain and frequency domain feature extraction. Time domain features include RMS, peak, mean, standard deviation, skewness, kurtosis, and crest factor [44]. Frequency domain features include main frequency, harmonics, frequency band energy percentage. Before feature extraction of high-frequency data, noise reduction should be performed to the signal. Data and pattern mining models for maintenance (e.g. Apriori [40] or FPGrowth [48]) could be used for knowledge and rules generation. Generated knowledge along with production data could aid in fault diagnosis and prediction. Correlation analysis has provided internal relationships between device and faults [47].

Traditional and Deep machine Learning techniques have been used for data analytics. Clustering algorithms have been identified to be the most common machine learning algorithm for preliminary grouping of sensor data and for creating labels according to their process state [44], [48]. Clustering algorithms have been followed by classification algorithms based on traditional machine learning (e.g. k-means in [48]) or deep learning (e.g. CNN in [46]). Technologies that have been used for data analysis in maintenance include STORM [40] (distributed computing), STORM cluster [47] (resource scheduling), Hadoop [40] (offline prediction - considering both current status and historical information).

The collected data needs to be processed to generate insights. Primary steps in data processing involve cleaning the data to remove noisy and incorrect format issues. Streamlink (Flink, Storm), micro-batching (Spark) and batching data processing (MapReduce) provide technologies to clean and process big data volumes. Manufacturing applications like complex event processing by Storm, and detecting deviations by Flink, prediction and quality control by MapReduce are some examples where these technologies are used to process manufacturing data. Knowledge can be generated by harvesting big data technologies on the generated big data. Apache Hive-Mind based platforms have aided knowledge generation for predictive maintenance. Hadoop and OWL technologies can manage knowledge of intelligent applications for smart manufacturing applications.

F. DATA VISUALIZATION

Data visualization is an integral part of data analysis which uses tables and graphs for presenting quantitative and qualitative information, and is used by users to interact with the data [83]. However, few state-of-the-art works describe methods for data visualization in the context of smart manufacturing automation and big data. Data visualization is usually implemented in the form of dashboards, a type of graphical user interface that consolidates big data (e.g. sensor, operational, and maintenance data). Dashboards are used to monitor and access production status and, in some cases, are used as a direct interface between the customer and the shop floor. They are often interactive and users can filter and query data, zoom in/out, and scroll. Many of the visualizations contained in dashboards show changes over time and are updated as new data is released, thus displaying real-time data updated every few seconds or minutes. In general, data visualization can include [84]:

- Different types of charts and graphs, tables, time trends, etc.
- Interactive widgets (e.g. knobs, dimers, and keypads) used to interact with CPS, IoT devices and applications, based on current data analysis.
- Visualization of geo-referenced data (machines in different locations, operators location tracking, external sensors)

From the technological perspective, research has prefered the use of Python programming language to develop machine learning models. Therefore, for data visualization, Python libraries such as Seaborn or Matplotlib have been chosen to develop charts and graphs. Matplotlib has been used to visualize a heat map and to find the correlation between the variables involved in milling tool wear (Figure 8.**a**) [85].

Depending on the tools and technology used (e.g. SQL databases, graph databases), visualisation methods integrated into the development environment have been used [63]. However, these options are not intuitive or designed for end-users. At the moment, multiple platforms and frameworks can produce analytics applications and visualizations easily with very aesthetically pleasing results. Grafana is one of the most popular open-source platforms for interactive data

visualization. [79] has used Grafana to create a dashboard for visualising energy data at the workstation level to show operational KPI and power consumption trends (Figure 8.**b**). Similarly, [81] has developed dashboards using Grafana and Amazon QuickSight for its compatibility with Spark to display the results of small-batch processing for the detection of anomalies on CNC Machines (Figure 8.**c**). Other similar products include Qlikview, Tableau, Kibana, and Splunk.

Although these platforms are claimed for their ease of use; the target users are data scientists and engineers, business analysts, or DevOps engineers. For end-users (i.e customers, operators, supervisors) customized applications accessible through mobile devices or web interfaces using browsers [62] is the best option. In [44], a Web and iOS-based user interface has been used in real-time for decision-making on the assessment of health. In [47], the manufacturing data processed has been sent to backstage supporters and the diagnosis or prognosis reports have been visualized on large screens through a web application (Single View integrated failure map pattern and cause [48]) or sent to mobile devices of the maintenance personnel (Figure 8.**d**). These kinds of applications require software development. Javascript is the most used web programming language for reactive applications, with multiple frameworks such as React, AngularJs, and NodeJs. There are specific Javascript libraries that allow the development of interactive visualizations such as CanvasJS or ChartJS. [86] has developed a web application for historical analysis and real-time tracking of the assembly line performance. The web is created with a combination of HTML5, CSS, JavaScript, the JavaScript Data-Driven Documents (D3) library, the Three.js, and several JavaScript framework and utility libraries including Underscore.js, Backbone.js, and JQuery (Figure 8.**e**).

It is important to consider that manufacturing processes involve several types of users where multiple variables intervene (e.g. expertise, role, and age). Therefore, users have different perceptions of visual data presentation and interactive data analysis [57]. User-centered design as a methodology can help to understand the requirements and needs of determined roles in the industry.

V. DISCUSSION ON THE RECENT TRENDS AND CHALLENGES

Table 4 provides a brief overview of the relevant references for big data shop-floor reviewed in this work. Different manufacturing applications require different data sources. Data sources comprise mostly smart sensors and IoT devices that convert physical variables into digitized measurable units. Smart decision making in product driven manufacturing applications rely on specifications of production requirements. Manufacturing automation concepts are based on logic-based or negotiation based approaches. In particular, it has been identified that data-driven automation has been considered less, making this as an opportunity for future research.

Some applications rely fundamentally in data acquisition and number of sensors placed in shop-floor machines and resources. Two examples are maintenance and energy optimization. One the one hand, maintenance has relied on acoustic emissions, temperature, velocity, pressure, and other variables to understand health status of machines. On the other hand, energy optimization application have relied mostly on measurement of electrical variables, e.g. smart meters, current and voltage clamps, and single-phase and 3-phase smart plugs. With the advent of human-centre manufacturing applications, the acquisition of data from operators has become a trend in current research, specially data used to model human characteristics, such as behaviour and comfort. Wearable trackers can measure human performance under stressful or difficult conditions. Consideration should be given to data sources that contain collection of data that should not be used due to regulations i.e General Data Protection Regulation.

Data collection may be performed with either manual or automatic data acquisition. Main trade-offs happen in form, consistency and reliability of the data. Data collection is dependent on the type of data source and comes from sources, such as IoT devices, evaluations, simulations, and predictions, in structured or unstructured formats. Data collection has been usually accompanied by an underlying framework that leverages step-wise processes to gather desired data for decision-making. Predictive maintenance, monitoring, energy consumption, and event-driven automation applications require data to be collected as per specific requirements. These requirements include real-time, time-driven, and periodic data collection, as well as application-specific criterion.

Data transmission may be performed with sockets, OPC-UA, MQTT, TCP/IP (such as PLC simulator), or other communication protocols depending on the application domain and can be dynamically chosen. Data transmission is the middleware between digital twins and the shop-floor. Moreover, it is the communication channel between devices in digital twins and their physical counterpart. The introduction of IoT on the shop floor has increased the transmission of low-frequency sensor information directly from sources through wireless communication. This has had impact on the latency of the response of the system. Industrial wireless communication devices include industrial switches, industrial routing, and wireless access points.

As manufacturers becomes increasingly reliant on sensors and various data sources, data storage has become an increasingly important concern. In particular, the ability to store big data has been given special attention. A trend has been identified in manufacturers, moving from traditional RDBMS databases to NoSQ and NewSQL databases when considering scalabilty. Moreover, a need has been identified to develop techniques to not only store data in a structured manner but also filter redundant data and delete data which is no longer relevant. This could greatly reduces storage costs and complexity. However, it has been recognized that there are few studies considering this aspect.

Data processing techniques have been widely used in manufacturing. With the development of IoT, 5G and 6G, and cloud computing technologies, the data quantity from manu-

FIGURE 8: Data visualization. a) [85] Seaborn visualization b) [79] Grafana visualization c) [81] Grafana and QuickSight d) [47] web and mobile apps e) [86] HTML5, CSS, JavaScript web application

facturing systems has increased rapidly. With industrial big data, achievements beyond expectations have been made in product design, manufacturing, and maintenance processes. Data processing has been a core technology to empower intelligent manufacturing systems.

Finally, visualization has been identified to usually be a neglected aspect in research. As presented in the results, multiple scholars prefer Python libraries for simple static visualization. However, to provide adequate commercial implementations of big data applications, visualization is as essential as the other stages. The capability of applications to further exploit data from user behaviour, improving the visualization aspect in manufacturing, needs further research. Furthermore, there is a lack of standardization that requires researchers and engineers to identify generic abstractions for industrial data and understand different users groups. Thus, new frameworks for visualization applications may be developed.

CHALLENGES

Challenges found in literature have been compiled in this study from the results and discussion of the review process. Although some of the challenges below are application-specific, they were found quite often in the reviewed literature.

- Data measurement solutions usually come with inherent measurement errors. Although these errors are relatively small, transferability has been affected. For instance, the same sensor for the same equipment performing the same application can yield different energy consumption values. Noisy and non-deterministic measurement values challenge data-processing and decision-making algorithms.
- Frequency of collected data is identified as another challenge in literature. Sampling at a high rate produces big data that is difficult to transmit and process in real-time. However, some applications require high-frequency data, such as energy parameter profiling applications. Therefore, trade-offs should be considered in data collection on the shop-floor.
- Data acquisition systems, incorporating all information gathered during the production process, are needed to collect data, discover knowledge, and share it among all stakeholders.
- Real-time processing, analysis, production reporting, and monitoring of data-driven sources must be implemented for real-time analysis of sensor data.
- Reliable data and valuable knowledge is needed to support optimized decision-making of product life-cycle management.
- Data heterogeneity must be processed in shop-floor systems comprised of multi-source heterogeneous data and complex processes, such as fault prediction using traditional signal processing techniques considering the 5V challenges posed by industrial big data.
- Data visualization disgned should be improved for hu-

man interaction. Visual and task complexity must be consider for data visualization, such as complex dashboards and unorganized big data. In addition, a high number of steps to realize a task may cause mistakes and reduce the performance of operators.
- The lack of implementations of cybersecurity and data privacy remains a challenge in shop-floor systems, in particular for big data analytics.
- Governance of big data handles data integrity, quality, provenance, retention, processing, and analysis in the full data life cycle. Governance of industrial big data should consider the issues of cybersecurity and data privacy as well.

VI. CONCLUSION

In this work, a basis for the development of an homogeneous approach to gather and use big data on the shop-floor in manufacturing environments has been presented. A literature review of research regarding big data in manufacturing has been performed, targeting the complete data life cycle. In this regard, the needs, requirements and methods for the seven stages of the big-data life cycle in manufacturing have been presented and discussed. Therefore, approaches for data acquisition, processing and utilisation for decision making in shop-floor in manufacturing have been established and challenges in each stage have been elaborated.

As results of this study, approaches have been identified in each stage of the big-data life cycle in manufacturing, focusing on maintenance, automation, quality, decision making, energy optimization, user interaction, and adaptability. Data sources, such as sensors, documents and models, have been identified and elaborated, detailing their usage and benefits, as well as possible drawbacks. Thereupon, data collection techniques have been presented, i.e. manual data acquisition and automatic data acquisition, describing the benefits and drawbacks of each. Furthermore, a separation between monitoring and predictive applications has been described, highlighting the effect that the intended application has in data collection. Having presented data collection techniques, data transmission protocols and techniques have been studied. Techniques and protocols for data transmission have been presented, as well as the cases in which each may be used. Following, data storage possibilities have been presented. Since data may be structured, semi-structured and unstructured, storage options have been discussed for each type of data structure, as well as the methods to integrate data in different formats and from different sources. In the context of data processing, several approaches towards data processing have been presented, as well as leading technologies for big data processing. In general, artificial intelligence and statistical approaches have been identified as the main contributors in this stage. Finally, data visualization methods, an integral part of data analysis, have been described in the context of smart manufacturing automation and big data. Several platforms and frameworks for data visualization have been reviewed and programming languages suitable for creating dashboards and visualization applications have been described.

A discussion of the trends and challenges obtained from the review process has been presented. It has been identified that the primary data sources include smart sensors and IoT devices. Nevertheless, human-centered manufacturing applications have included data acquisition from operators, allowing modelling of behaviour and comfort. An important consideration that has been highlighted, regardless of the source of the data, is data privacy and restrictions that may apply due to regulations.

Regarding data transmission, several protocols have been identified and their usage will depend on the technologies being used and the application. Data format, data size, transmission distance and transmission rates have a determining effect on which protocols to use and how to integrate the data being sent. In data storage, moving from traditional structured data storage, such as RDBMS, to unstructured and semi-structured data storage, such as NoSQL and NewSQL, has been identified as the leading trend. In addition, it has been identified that there is a lack of focus on irrelevant data filtering and deletion, which might help to reduce cost and processing power in applications where there are economical or storage constraints.

In general, this research has identified several challenges in literature. Challenges involve possible errors in the collected data, which may lead to inaccurate measurements, as well as the challenges regarding the handling of varied sampling frequencies and the impact on the transmission technologies used. Furthermore, challenges regarding heterogeneity of data have been identified, where the integration of varied data sources could represent a challenge during data storage, processing, and visualization, deriving in incorrect analysis of data or complexity in understanding the data obtained during the data life cycle. Finally, cybersecurity and data privacy have been identified as important challenges, as several studies have lacked attention in this regard.

Future work will focus on developing a consolidated framework and methodology for big-data life cycle. Based on the findings of this review, it is expected that this work will serve as basis for future frameworks for big-data life cycle on the shop floor.

ACKNOWLEDGEMENT

Some icons used in certain figures were provided by www.flaticon.com and made by Becris and Freepik.

REFERENCES

[1] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Data-driven smart manufacturing," Journal of Manufacturing Systems, vol. 48, pp. 157–169, 2018.
[2] B. Wang, F. Tao, X. Fang, C. Liu, Y. Liu, and T. Freiheit, "Smart manufacturing and intelligent manufacturing: A comparative review," Engineering, vol. 7, no. 6, pp. 738–757, 2021.
[3] H. S. Kang, J. Y. Lee, S. Choi, H. Kim, J. H. Park, J. Y. Son, B. H. Kim, and S. Do Noh, "Smart manufacturing: Past research, present findings, and future directions," International journal of precision engineering and manufacturing-green technology, vol. 3, no. 1, pp. 111–128, 2016.

[4] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Big data technologies: A survey," Journal of King Saud University-Computer and Information Sciences, vol. 30, no. 4, pp. 431–448, 2018.

[5] B. Furht and F. Villanustre, "Introduction to big data," in Big data technologies and applications. Springer, 2016, pp. 3–11.

[6] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," Expert Systems with Applications, vol. 166, p. 114060, 2021.

[7] Y. Cui, S. Kara, and K. C. Chan, "Manufacturing big data ecosystem: A systematic literature review," Robotics and computer-integrated Manufacturing, vol. 62, p. 101861, 2020.

[8] H. Yang, S. Kumara, S. T. Bukkapatnam, and F. Tsung, "The internet of things for smart manufacturing: A review," IISE Transactions, vol. 51, no. 11, pp. 1190–1216, 2019.

[9] Y. Demchenko, P. Grosso, C. De Laat, and P. Membrey, "Addressing big data issues in scientific data infrastructure," in 2013 International conference on collaboration technologies and systems (CTS). IEEE, 2013, pp. 48–55.

[10] J. Wang, C. Xu, J. Zhang, and R. Zhong, "Big data analytics for intelligent manufacturing systems: A review," Journal of Manufacturing Systems, 2021.

[11] C. Li, Y. Chen, and Y. Shang, "A review of industrial big data for decision making in intelligent manufacturing," Engineering Science and Technology, an International Journal, 2021.

[12] J. Mageto, "Big data analytics in sustainable supply chain management: A focus on manufacturing supply chains," Sustainability, vol. 13, no. 13, p. 7101, 2021.

[13] A. Belhadi, K. Zkik, A. Cherrafi, M. Y. Sha'ri et al., "Understanding big data analytics for manufacturing processes: insights from literature review and multiple case studies," Computers & Industrial Engineering, vol. 137, p. 106099, 2019.

[14] S. Sahoo, "Big data analytics in manufacturing: a bibliometric analysis of research in the field of business management," International Journal of Production Research, pp. 1–29, 2021.

[15] S. Ren, Y. Zhang, Y. Liu, T. Sakao, D. Huisingh, and C. M. Almeida, "A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions," Journal of cleaner production, vol. 210, pp. 1343–1365, 2019.

[16] J. Lee, E. Lapira, B. Bagheri, and H.-a. Kao, "Recent advances and trends in predictive manufacturing systems in big data environment," Manufacturing letters, vol. 1, no. 1, pp. 38–41, 2013.

[17] A. V. Levitin and T. C. Redman, "A model of the data (life) cycles with application to quality," Information and Software Technology, vol. 35, no. 4, pp. 217–223, 1993.

[18] V. Y. Yoon, P. Aiken, and T. Guimaraes, "Managing organizational data resources: quality dimensions," Information Resources Management Journal), vol. 13, no. 3, pp. 5–13, 2000.

[19] C. L. Borgman, J. C. Wallis, M. S. Mayernik, and A. Pepe, "Drowning in data: digital library architecture to support scientific use of embedded sensor networks," in Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, 2007, pp. 269–277.

[20] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," interactions, vol. 19, no. 3, pp. 50–59, 2012.

[21] N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. Mahmoud Ali, M. Alam, M. Shiraz, and A. Gani, "Big data: survey, technologies, opportunities, and challenges," The scientific world journal, vol. 2014, p. 712826, 2014.

[22] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," Proceedings of the IEEE, vol. 104, no. 11, pp. 2207–2219, 2016.

[23] Y. Zhang, G. Zhang, J. Wang, S. Sun, S. Si, and T. Yang, "Real-time information capturing and integration framework of the internet of manufacturing things," International Journal of Computer Integrated Manufacturing, vol. 28, no. 8, pp. 811–822, 2015.

[24] J. J. Peralta Abadía, C. Walther, A. Osman, and K. Smarsly, "A systematic survey of internet of things frameworks for smart city applications," Sustainable Cities and Society, p. 103949, 2022.

[25] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," International journal of information management, vol. 35, no. 2, pp. 137–144, 2015.

[26] E. Begoli and J. Horey, "Design principles for effective knowledge discovery from big data," in 2012 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture. IEEE, 2012, pp. 215–218.

[27] S. Mittal, M. A. Khan, D. Romero, and T. Wuest, "Smart manufacturing: Characteristics, technologies and enabling factors," Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, vol. 233, no. 5, pp. 1342–1361, 2019.

[28] Y. Liu, J. Jin, P. Ji, J. A. Harding, and R. Y. Fung, "Identifying helpful online reviews: a product designer's perspective," Computer-Aided Design, vol. 45, no. 2, pp. 180–194, 2013.

[29] G. Demiris, D. P. Oliver, and K. T. Washington, Behavioral intervention research in hospice and palliative care: Building an evidence base. Academic press, 2018.

[30] M. Milivojević, I. Antolović, and D. Rančić, "Using collada and x3d for webgl based 3d data visualization."

[31] M. Yao, Z. Shao, and Y. Zhao, "Review on Energy Consumption Optimization Methods of Typical Discrete Manufacturing Equipment," in Intelligent Robotics and Applications, X.-J. Liu, Z. Nie, J. Yu, F. Xie, and R. Song, Eds. Cham: Springer International Publishing, 2021, pp. 48–58.

[32] Paryanto, M. Brossog, M. Bornschlegl, and J. Franke, "Reducing the energy consumption of industrial robots in manufacturing systems," The International Journal of Advanced Manufacturing Technology, vol. 78, no. 5, pp. 1315–1328, May 2015. [Online]. Available: https://doi.org/10.1007/s00170-014-6737-z

[33] D. Meike, M. Pellicciari, and G. Berselli, "Energy Efficient Use of Multirobot Production Lines in the Automotive Industry: Detailed System Modeling and Optimization," IEEE Transactions on Automation Science and Engineering, vol. 11, no. 3, pp. 798–809, Jul. 2014, conference Name: IEEE Transactions on Automation Science and Engineering.

[34] K. Ding, Y. Zhang, F. T. Chan, C. Zhang, J. Lv, Q. Liu, J. Leng, and H. Fu, "A cyber-physical production monitoring service system for energy-aware collaborative production monitoring in a smart shop floor," Journal of Cleaner Production, vol. 297, p. 126599, 2021.

[35] R. J. Machchhar, C. N. K. Toller, A. Bertoni, and M. Bertoni, "Data-driven value creation in smart product-service system design: State-of-the-art and research directions," Computers in Industry, vol. 137, p. 103606, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016636152200001X

[36] I. Zolotová, P. Papcun, E. Kajáti, M. Miškuf, and J. Mocnej, "Smart and cognitive solutions for operator 4.0: Laboratory h-cpps case studies," Computers & Industrial Engineering, vol. 139, p. 105471, 2020.

[37] S. Wang, J. Wan, M. Imran, D. Li, and C. Zhang, "Cloud-based smart manufacturing for personalized candy packing application," The Journal of Supercomputing, vol. 74, no. 9, pp. 4339–4357, 2018.

[38] Y.-C. Lin, M.-H. Hung, H.-C. Huang, C.-C. Chen, H.-C. Yang, Y.-S. Hsieh, and F.-T. Cheng, "Development of advanced manufacturing cloud of things (amcot)—a smart manufacturing platform," IEEE Robotics and Automation Letters, vol. 2, no. 3, pp. 1809–1816, 2017.

[39] R. Y. Zhong, G. Q. Huang, S. Lan, Q. Dai, X. Chen, and T. Zhang, "A big data approach for logistics trajectory discovery from rfid-enabled production data," International Journal of Production Economics, vol. 165, pp. 260–272, 2015.

[40] Y. Zhang, S. Ren, Y. Liu, and S. Si, "A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products," Journal of cleaner production, vol. 142, pp. 626–641, 2017.

[41] Y. Zhang, S. Ren, Y. Liu, T. Sakao, and D. Huisingh, "A framework for big data driven product lifecycle management," Journal of Cleaner Production, vol. 159, pp. 229–240, 2017.

[42] M. Kohlert and A. König, "Large, high-dimensional, heterogeneous multi-sensor data analysis approach for process yield optimization in polymer film industry," Neural Computing and Applications, vol. 26, no. 3, pp. 581–588, 2015.

[43] H. U. Rehman, T. Pulikottil, L. A. Estrada-Jimenez, F. Mo, J. C. Chaplin, J. Barata, and S. Ratchev, "Cloud based decision making for multi-agent production systems," in EPIA Conference on Artificial Intelligence. Springer, 2021, pp. 673–686.

[44] J. Lee, H. D. Ardakani, S. Yang, and B. Bagheri, "Industrial big data analytics and cyber-physical systems for future maintenance & service innovation," Procedia cirp, vol. 38, pp. 3–7, 2015.

[45] J. Yan, Y. Meng, L. Lu, and L. Li, "Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance," IEEE Access, vol. 5, pp. 23 484–23 491, 2017.

[46] G. Martínez-Arellano, G. Terrazas, and S. Ratchev, "Tool wear classification using time series imaging and deep learning," The International Journal of Advanced Manufacturing Technology, vol. 104, no. 9, pp. 3647–3662, 2019.

[47] J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas, and A. V. Vasilakos, "A manufacturing big data solution for active preventive maintenance," IEEE Transactions on Industrial Informatics, vol. 13, no. 4, pp. 2039–2047, 2017.

[48] K. Nakata, R. Orihara, Y. Mizuoka, and K. Takagi, "A comprehensive big-data-based monitoring system for yield enhancement in semiconductor manufacturing," IEEE Transactions on Semiconductor Manufacturing, vol. 30, no. 4, pp. 339–344, 2017.

[49] V. Yadav and S. Bethard, "A survey on recent advances in named entity recognition from deep learning models," arXiv preprint arXiv:1910.11470, 2019.

[50] N. Peng, H. Poon, C. Quirk, K. Toutanova, and W.-t. Yih, "Cross-sentence n-ary relation extraction with graph lstms," Transactions of the Association for Computational Linguistics, vol. 5, pp. 101–115, 2017.

[51] L. Sun, "Research on product attribute extraction and classification method for online review," in 2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII). IEEE, 2017, pp. 117–121.

[52] S. Durga, R. Nag, and E. Daniel, "Survey on machine learning and deep learning algorithms used in internet of things (iot) healthcare," in 2019 3rd international conference on computing methodologies and communication (ICCMC). IEEE, 2019, pp. 1018–1022.

[53] C. Reuter and F. Brambring, "Improving data consistency in production control," Procedia CIRP, vol. 41, pp. 51–56, 2016.

[54] N. Robertson and T. Perera, "Automated data collection for simulation?" Simulation Practice and Theory, vol. 9, no. 6-8, pp. 349–364, 2002.

[55] P. Azad, N. J. Navimipour, A. M. Rahmani, and A. Sharifi, "The role of structured and unstructured data managing mechanisms in the internet of things," Cluster computing, vol. 23, no. 2, pp. 1185–1198, 2020.

[56] K. Duh, G. Neubig, K. Sudoh, and H. Tsukada, "Adaptation data selection using neural language models: Experiments in machine translation," in Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2013, pp. 678–683.

[57] F. Zhou, X. Lin, C. Liu, Y. Zhao, P. Xu, L. Ren, T. Xue, and L. Ren, "A survey of visualization for smart manufacturing," Journal of Visualization, vol. 22, no. 2, pp. 419–435, 2019.

[58] Y. Lu and X. Xu, "Cloud-based manufacturing equipment and big data analytics to enable on-demand manufacturing services," Robotics and Computer-Integrated Manufacturing, vol. 57, pp. 92–102, 2019.

[59] M. Gadaleta, G. Berselli, M. Pellicciari, and F. Grassia, "Extensive experimental investigation for the optimization of the energy consumption of a high payload industrial robot with open research dataset," Robotics and Computer-Integrated Manufacturing, vol. 68, p. 102046, Apr. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S073658452030257X

[60] S.-J. Shin, "An opc ua-compliant interface of data analytics models for interoperable manufacturing intelligence," IEEE Transactions on Industrial Informatics, vol. 17, no. 5, pp. 3588–3598, 2020.

[61] M. Gadaleta, M. Pellicciari, and G. Berselli, "Optimization of the energy consumption of industrial robots for automatic code generation," Robotics and Computer-Integrated Manufacturing, vol. 57, pp. 452–464, Jun. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0736584518301856

[62] R. Bonnard, K. M. M. Vieira, S. ARANTES, R. Lorbieski, C. NUNES, and A. P. Mattei, "A big data/analytics platform for industry 4.0 implementation in smes," CIGI QUALITA, 2019.

[63] D. Fernandes and J. Bernardino, "Graph databases comparison: Allegrograph, arangodb, infinitegraph, neo4j, and orientdb." in Data, 2018, pp. 373–380.

[64] C.-A. Chen, M. Won, R. Stoleru, and G. G. Xie, "Energy-efficient fault-tolerant data storage and processing in mobile cloud," IEEE Transactions on cloud computing, vol. 3, no. 1, pp. 28–41, 2014.

[65] M. Zhang, Y. Zuo, and F. Tao, "Equipment energy consumption management in digital twin shop-floor: A framework and potential applications," in 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC). IEEE, 2018, pp. 1–5.

[66] M. Zhang and J. Yan, "A data-driven method for optimizing the energy consumption of industrial robots," Journal of Cleaner Production, vol. 285, p. 124862, Feb. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652620349064

[67] J. Yan and M. Zhang, "A transfer-learning based energy consumption modeling method for industrial robots," Journal of Cleaner Production, vol. 325, p. 129299, Nov. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652621034843

[68] A. Liu, H. Liu, B. Yao, W. Xu, and M. Yang, "Energy consumption modeling of industrial robot based on simulated power data and parameter identification," Advances in Mechanical Engineering, vol. 10, no. 5, p. 1687814018773852, May 2018, publisher: SAGE Publications. [Online]. Available: https://doi.org/10.1177/1687814018773852

[69] L. Guo, F. Yan, T. Li, T. Yang, and Y. Lu, "An automatic method for constructing machining process knowledge base from knowledge graph," Robotics and Computer-Integrated Manufacturing, vol. 73, p. 102222, 2022.

[70] Y. Zhang, X. Luo, H. Zhang, and J. W. Sutherland, "A knowledge representation for unit manufacturing processes," The International Journal of Advanced Manufacturing Technology, vol. 73, no. 5, pp. 1011–1031, 2014.

[71] B. Zhou, J. Bao, J. Li, Y. Lu, T. Liu, and Q. Zhang, "A novel knowledge graph-based optimization approach for resource allocation in discrete manufacturing workshops," Robotics and Computer-Integrated Manufacturing, vol. 71, p. 102160, 2021.

[72] H. L. Nguyen, D. T. Vu, and J. J. Jung, "Knowledge graph fusion for smart systems: A survey," Information Fusion, vol. 61, pp. 56–70, 2020.

[73] A. Villalonga, G. Beruvides, F. Castano, and R. E. Haber, "Cloud-based industrial cyber–physical system for data-driven reasoning: A review and use case on an industry 4.0 pilot line," IEEE Transactions on Industrial Informatics, vol. 16, no. 9, pp. 5975–5984, 2020.

[74] R. Bonnard, M. D. S. Arantes, R. Lorbieski, K. M. M. Vieira, and M. C. Nunes, "Big data/analytics platform for industry 4.0 implementation in advanced manufacturing context," The International Journal of Advanced Manufacturing Technology, vol. 117, no. 5, pp. 1959–1973, 2021.

[75] C. Zhang, Z. Wang, K. Ding, F. T. Chan, and W. Ji, "An energy-aware cyber physical system for energy big data analysis and recessive production anomalies detection in discrete manufacturing workshops," International Journal of Production Research, vol. 58, no. 23, pp. 7059–7077, 2020.

[76] W. Ji and L. Wang, "Big data analytics based fault prediction for shop floor scheduling," Journal of Manufacturing Systems, vol. 43, pp. 187–194, 2017.

[77] Y. Wang, S. Wang, B. Yang, L. Zhu, and F. Liu, "Big data driven hierarchical digital twin predictive remanufacturing paradigm: Architecture, control mechanism, application scenario and benefits," Journal of Cleaner Production, vol. 248, p. 119299, 2020.

[78] S. Zambal, C. Eitzinger, M. Clarke, J. Klintworth, and P.-Y. Mechin, "A digital twin for composite parts manufacturing: Effects of defects analysis based on manufacturing data," in 2018 IEEE 16th international conference on industrial informatics (INDIN). IEEE, 2018, pp. 803–808.

[79] S. Kahveci, B. Alkan, A. Mus'ab H, B. Ahmad, and R. Harrison, "An end-to-end big data analytics platform for iot-enabled smart factories: A case study of battery module assembly system for electric vehicles," Journal of Manufacturing Systems, vol. 63, pp. 214–223, 2022.

[80] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, and Y. Liu, "A global manufacturing big data ecosystem for fault detection in predictive maintenance," IEEE Transactions on Industrial Informatics, vol. 16, no. 1, pp. 183–192, 2019.

[81] M. Saez, S. Lengieza, F. Maturana, K. Barton, and D. Tilbury, "A data transformation adapter for smart manufacturing systems with edge and cloud computing capabilities," in 2018 IEEE International Conference on Electro/Information Technology (EIT). IEEE, 2018, pp. 0519–0524.

[82] M. Gadaleta, G. Berselli, M. Pellicciari, and F. Grassia, "Extensive experimental investigation for the optimization of the energy consumption of a high payload industrial robot with open research dataset," Robotics and Computer-Integrated Manufacturing, vol. 68, p. 102046, 2021.

[83] M. Mani and S. Fei, "Effective big data visualization," in Proceedings of the 21st International Database Engineering & Applications Symposium, 2017, pp. 298–303.

[84] G. W. Young and R. Kitchin, "Creating design guidelines for building city dashboards from a user's perspectives," International Journal of Human-Computer Studies, vol. 140, p. 102429, 2020.

[85] S. Vijay, B. Kuraichen et al., "Data driven prognostics of milling tool wear: A machine learning approach," in 2021 International Conference on Computational Performance Evaluation (ComPE). IEEE, 2021, pp. 002–007.

[86] P. Xu, H. Mei, L. Ren, and W. Chen, "Vidx: Visual diagnostics of assembly line performance in smart factories," IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 1, pp. 291–300, 2017.

TERRIN PULIKOTTIL completed his Bachelor in Manufacturing Engineering from College of Engineering Guindy, Anna University Chennai, India in 2012. After his under graduation, He worked in Carborundum Universal Ltd, India as technical deputy manager, for two years (2012-14) responsible for supporting production and QA. He then completed his Master from Politecnico di Milano, Italy in Mechanical Engineering with specialization in Industrial Production in 2016, after which he worked as a Research Fellow in National Research Council of Italy in Institute for Intelligent Industrial Technologies and Systems for Advanced Manufacturing (STIIMA) for three years (2017-20). He is currently working as a Researcher in UNINOVA from June 2020 on H2020 DIMAND project.

LUIS A. ESTRADA-JIMENEZ received the B.Sc. degree in Electronic and Control engineering from Escuela Politecnica Nacional in Ecuador in 2016 and the M.Sc. degree in Mechatronics Engineering at the University of Oviedo in Spain in 2019. His master's thesis was developed at the company FESTO in Germany at the department of modular automation. He is currently a Ph.D. student in Electrical and Computer Engineering at the Nova University of Lisbon in Portugal. He also works at the UNINOVA research institute. Currently, his main role is as an Early Stage Researcher at the Digital Manufacturing and Design Training Network (DIMAND), funded by the European Union. His research interests include self-organization and automation in smart manufacturing systems and the application of artificial intelligence in industrial environments.

JOSÉ JOAQUÍN PERALTA ABADÍA is a researcher with artificial intelligence, computer science, and business administration background. He received a B.Sc. degree in computer sciences and an MBA at Universidad de Costa Rica in Costa Rica. Thereafter, he received an M.Sc. degree in artificial intelligence at Universidad Politécnica de Madrid in Spain. Currently, José Joaquín is a PhD candidate at Mondragon Unibertsitatea in Spain, funded by the H2020 DIMAND project. His research field is in artificial intelligence applied in manufacturing for process optimization.

ANGELA CARRERA-RIVERA received a B.Sc. degree in Information Systems from Escuela Superior Politecnica del Litoral in Ecuador and a M.Sc. degree in Information Technology Engineering at the University of Melbourne in Australia. She is currently a Ph.D. candidate at Mondragon University in the Faculty of Computer Science.

AGAJAN TORAYEV is a researcher with a strong background in applied mathematics, informatics, and computer science. He holds a Diploma in Applied Mathematics and Informatics from Magtymguly Turkmen State University and completed his Master of Science in Computer, specializing in Intelligent Systems, Machine Learning, and Deep Learning, at the University of Bonn. Currently, Agajan is pursuing a PhD in Manufacturing Engineering at the University of Nottingham, where he works as a researcher on the H2020 DIMAND project. His current research topic is the Optimal Manufacturing Configurations Selection for rapidly changing requirements.

HAMOOD UR REHMAN graduated from NED University of Engineering & Technology in Pakistan with B.E. in Industrial and Manufacturing Engineering. He went on to work in various organisations focused on manufacturing before going to complete his MSc. in Mechanical Engineering Modelling at the Budapest University of Technology and Economics. He worked as a lecturer at NED University after his master's. He is currently working at TQC Ltd. (Automation and Test Solutions) as Robotics and Control Systems Engineer while doing his PhD at the University of Nottingham, United Kingdom. His research interests include robotics, automation, digital manufacturing and self-configuration in smart manufacturing systems.

FAN MO earned his Bachelor's degree in Engineering from Tongji University in Shanghai, China, and obtained his Master's degree in Vehicle and Engine Technology from the University of Stuttgart in Germany. He can speak English and German fluently. During and after his studies, he gained practical experience through internships and full-time positions at BMW, Daimler, and Volkswagen in Germany and China. He is pursuing his Ph.D. at the institute of advanced manufacturing at the University of Nottingham in the United Kingdom. Meanwhile, he is working as a Marie Curie Early Stage Researcher supported by DiManD Innovative Training Network (ITN) project funded by the European Union through the Marie Skłodowska-Curie Innovative Training Networks. His research interests include robotics, knowledge graph, artificial intelligence, and multi-agent programming.

SANAZ NIKGHADAM-HOJJATI is a Senior Researcher at UNINOVA institute, Nova University of Lisbon. She received her Ph.D. in Information Technology Management (Business Intelligence) from I.A.U, in 2017. She has also worked as a post-Doc researcher during 2018-2019 at Nova School of Science and Technology, Nova University of Lisbon. Her research interests include Computational Creativity, affective computing, Business Intelligence, Human behaviour, and Emerging technologies, ICT, and Innovation Management. She has published several books and academic papers in a number of peer-reviewed journals and presented various academic papers at conferences. She has led and participated in several European Union projects, Portuguese and Iranian National projects. In addition, she has worked as a university invited professor, and also she is the director of the WoSTEM (Women In Science, Technology, Engineering, and Mathematics) program in UNINOVA.

JOSÉ BARATA (Member, IEEE) received the Ph.D. degree in robotics and integrated manufacturing from the NOVA University of Lisbon, in 2004. He is a Professor with the Department of Electrical Engineering, NOVA University of Lisbon, and a Senior Researcher with the UNINOVA—Instituto de Desenvolvimento de Novas Tecnologias. He has participated in more than 15 international research projects involving different programs, including NMP, IST, ITEA, and ESPRIT. Since 2004, he has been leading the UNINOVA participation in EU projects, namely, EUPASS, self-learning, IDEAS, PRIME, RIVERWATCH, ROBO-PARTNER and PROSECO. In the last years, he has participated actively researching SOA-based approaches for the implementation of intelligent manufacturing devices, such as within the Inlife Project. He has authored or coauthored over 100 original papers in international journals and international conferences. His main research interest includes intelligent manufacturing, with an emphasis on complex adaptive systems, involving intelligent manufacturing devices. He is a member of the IEEE Technical Committee on Industrial Agents (IES), Self-Organization and Cybernetics for Informatics (SMC), and Education in Engineering and Industrial Technologies (IES).