The Role of Big Data in Industry 4.0: A Systematic Literature Review

Dejana Gladić[1[0000-0001-8898-9738]], Jelena Petrovački[1[0000-0001-6145-4223]], Sonja Ristić[1[0000-0003-0059-6062]], Darko Stefanović[1[0000-0001-9200-5092]] and Milan Čeliković[1[0000-0002-9164-1262]]

[1] University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia
{dejanagladic, jelenapetrovacki, sdristic, darko.stefanovic, milancel}@uns.ac.rs

Abstract. Industry 4.0 integrates technologies, organizational concepts and management principles in order to establish the flexible production, enabling mass customization of products. Smart factories, characteristic for Industry 4.0, produce large volume of various data. The data has to be handled and analyzed effectively and efficiently in order to adequately support decision making processes and production system's adaptability to frequent changes. Big data technologies are aimed to enable storage of large volumes of structured, semi-structured and unstructured data alongside cost-effective, innovative forms of information processing for enhanced insight and decision making. The goal of this paper is to conduct a systematic literature review on the role of big data in Industry 4.0 with the focus on different data sources, technologies used for handling data and benefits of data analysis.

Keywords: Industry 4.0, Big Data, Smart Manufacturing.

1 Introduction

We are witnessing a constant growth in the amount of data that is generated every day in various ways. In addition to structured data, there is an increasing need for storing and analyzing semi-structured and unstructured data. Consequently, the IT community is faced with challenges such as how to design appropriate systems to handle the data effectively, and how to analyze data in order to extract relevant meaning and information. The umbrella term big data represents any technology that adequately handles large amounts of generated data, including data collection, network, storage, computing and analytical methods. The main characteristics of big data are the 3V's – volume, velocity and variety that respectively refer to: the large amount of data; the speed of its generation and analysis; and the diversity of the nature and structure of data [1]. In our previous work [2], research was carried out on different types of NoSQL databases according to the domains of their application in different economic sectors, due to their benefit for storage of both structured and unstructured data. One of the conclusions was that NoSQL databases are broadly used in the domain of industry and Industrial Internet of Things (IIoT). IIoT is aimed at connecting all industrial assets with information systems and business processes. That was the motivation to extend our research toward the application of big data technologies within the framework of Industry 4.0, where IIoT has a very important role.

Industry 4.0 relies on the adoption of digital technologies which can be divided into four "smart" dimensions – Smart Working, Smart Manufacturing, Smart Supply Chain and Smart Products. The smart environment defined through these dimensions makes it possible to increase the quality and flexibility, leading to the faster creation of customized products with a rational consumption of resources [3].

Contemporary industrial systems include cyber-physical systems (CPS), IIoT, big data, and artificial intelligence (AI) [4]. By accessing the data generated within such smart industrial systems, real-time insight into the production process is enabled, thus achieving one of the main goals of Industry 4.0 – data analysis of the production system, which results in taking timely actions to adapt the analyzed system to the desired state. Therefore, an adequate way to store and analyze data generated in the previously described way will enable modeling of predictive functions, making conclusions about indicators of key performance of the production system and analysis in real-time [5].

For the aforementioned reasons, the goal of this paper is to conduct a systematic literature review (SLR) on the topic of using big data in Smart Manufacturing processes as part of Industry 4.0. In that way it will be possible to define the framework and summarize the knowledge from previously conducted researches, in order to undertake new research activities in the domain of big data usage within Industry 4.0.

The rest of the paper is organized as follows: theoretical foundations are presented in Section 2; the methodology used to conduct the systematic literature review as well as the results of the review are described in Section 3; considerations of the previously presented results are presented in Section 4; and conclusions and future work are given in Section 5.

2 Theoretical foundations

Industry 4.0 is changing the way companies create, produce and improve their products and services by relying on digital technologies. The application of various convergent technologies enables the collection and analysis of data in real time in order to provide useful information for the further execution and control of business processes [3]. The main goal is the integration of information and communication systems with other business resources such as machines, devices and people [6]. The integration of all business resources creates a smart network that enables organizations to manage and automate processes more easily, predict unplanned outcomes in real time, and more efficiently adapt to changes and new requirements for production or service provision [7].

The theoretical framework of Industry 4.0 technologies given in [3] assumes two layers of technologies. The first layer is represented by front-end technologies, which includes four smart dimensions such as Smart Working, Smart Manufacturing, Smart Supply Chain and Smart Products. The focus of the presented SLR is on smart manufacturing dimension that refers to flexible manufacturing, dynamic reconfiguration, and production optimization, adapting to rapid changes in production due to customers' needs. The second layer comprises base technologies for creating smart production systems, like: IoT [5], cloud services [3], CPS [1], big data and data analytics. They enable the connection of front-end technologies into a complete integrated production system [3]. Concerning aforementioned base technologies, this SLR is focused on the role of big data in Industry 4.0 and data analytics [1].

3 Methodology

When performing the SLR, the approach of Barbara Kitchenham [8] is followed. According to this approach, activities are divided into three phases: (i) planning the review, (ii) conducting the review, and (iii) reporting the review.

3.1 Planning the review

At the beginning of the research on this topic, a search for already existing literature reviews is made. There are various existing SLR studies relating to the usage of big data in Industry 4.0 [9]–[12]. Some of them focus on a specific area in manufacturing, while current SLR is based on the larger scope of Industry 4.0. Lepasepp et al. [9] studied ways to manage big data collections that would ensure and optimize the manufacturing of medical devices, while Tesch da Silva et al. [11] did the similar research but in the field of energy consumption in Industry 4.0. Zheng et al. [13] deals with the impact of technologies enabled by Industry 4.0 on various processes of the production system. The emphasis is on the fact that the needs for technologies are different depending on the nature of the process in which they are used.

In [10] and [12], a systematic review was carried out, which aims to analyze the performance of the system under the influence of the application of different technologies. Bueno et al. [10] focused on testing the performance of the production system during the application of smart concepts, including big data analysis, over the process of planning and production control.

The aim of this systematic literature review is to identify the entire flow of data within the framework of Industry 4.0, starting from the source of raw data, through the means of their transmission, transformation, integration, storage and presentation, as well as the benefits that enable the analysis of such data. In order to fulfill the aim of the planned research, adequate research questions are defined, as follows: **(*RQ1*)** Which are the most common sources of data in Industry 4.0?; **(*RQ2*)** What are the benefits of data analysis within Industry 4.0?; **(*RQ3*)** What are the most commonly used technologies for handling big data?; **(*RQ3.1*)** Which big data management problems are still not solved using existing technologies?; and **(*RQ3.2*)** Is the application of Data Lake evident and what are the ways of its implementation?

The Scopus and Web of Science (WoS) index databases are searched using the keywords used in a generated query:

“Industry 4.0” AND “Smart Manufacturing” AND “Big Data” AND “data”.

Exclusion and inclusion to eliminate or accept a paper are given in Table 1.

Table 1. Exclusion/Inclusion criteria

	Exclusion criteria	Inclusion criteria
1.	Non-English papers should be excluded.	Paper relates to Industry 4.0 and manufacturing.
2.	Non-conference papers and non-journal articles should be excluded.	Paper lists and describes the sources of data generated in the manufacturing processes of Industry 4.0.
3.	Duplicate papers in search results should be excluded.	Paper describes the benefits of analyzing collected and adequately stored data.
4.	Papers that are not related to Smart Manufacturing processes should be excluded.	Paper mentions the technologies used for data management.
5.	Non-electronic papers should be excluded.	

3.2 Conducting the review

The initial search resulted in a total of 465 papers that meet the defined criteria. After eliminating papers that appear in both index databases, that number is reduced to 343. Further selection of papers is made based on: the title of the paper, the abstract, and their content. The numbers of selected papers are shown in Table 2. The total number of selected papers is reduced to 42.

Table 2. Initial search of WOS and SCOPUS

	WoS	SCOPUS
Initial search	228	237
Number of papers after eliminating duplicates	228	115
Number of selected papers based on title	171	74
Number of selected papers based on abstract	95	42
Number of selected papers based on their content	32	10
In total:	**42**	

In order to define descriptive statistics and answer previously defined research questions, data about year of publication, publication type, keywords, described data sources, types of described data analysis and technologies used for data handling were extracted from all the selected papers.

3.3 Results if the review

In this section, descriptive statistics based on relevant information extracted from 42 selected papers is presented. More than a half of papers are published at conferences (54.8%), while a slightly smaller number of papers (45.2%) were published in scientific journals. Based on the graph shown in Fig. 1 (a), we can see a trend in the growth of the number of published papers by year, starting from 2016. Considering that none of the selection criteria was related to the year of publication, we conclude that before

2016 there are no papers that correspond to the topic, and that more serious research on the topic was started in the past 5 years. We can see a peak in the number of published papers in 2019, but in the following years there is also a constant growth in research on this topic.

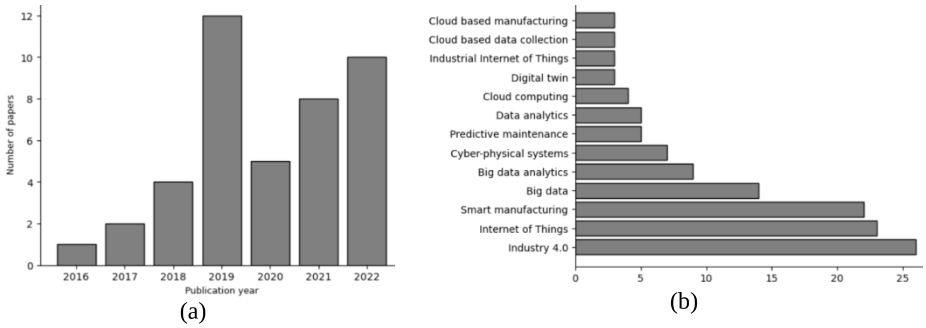


Fig. 1. Year-wise publication/Keyword analysis

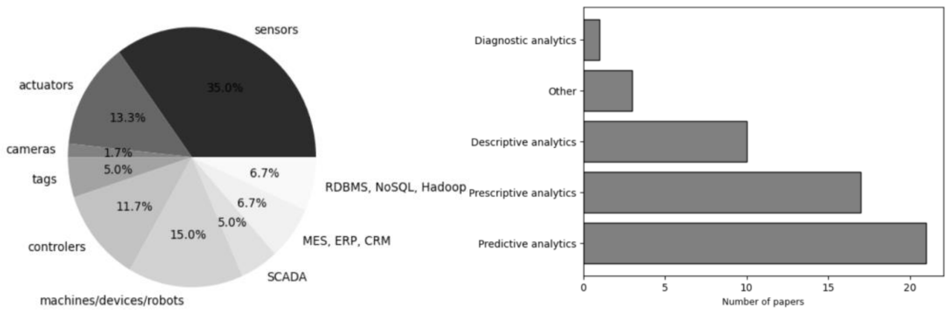
When it comes to the analysis of the keywords of the selected papers, it is important to emphasize that only keywords that appear at least 3 times are taken into account. It is observed that the result justifies the previously defined query for the search of index databases and that based on the papers with these keywords, adequate answers to the research questions can be obtained. The repetition numbers of the most frequent keywords are presented in Fig. 1 (b).

During the analysis of the papers, the parts of the paper that provide an answer to the corresponding research question are recorded. In Table 3, it can be seen in which papers the answer to which previously defined research question can be found.

Table 3. Division of papers in relation to the research questions they answer

RQ	Paper	%
RQ1	[3], [14]–[21], [4], [22]–[27], [28]–[35]	54,8
RQ2	[3], [14]–[16], [19], [22], [4], [23]–[25], [27], [30], [31], [33], [34], [35]–[42], [43]–[45]	61,9
RQ3	[3], [15], [17], [18], [20]–[22], [26]–[32], [35], [36], [38], [40]–[42], [44], [46]–[53]	69

The descriptive statistics, shown in Fig. 2 (a), aims to provide an answer to the question ***RQ1***. When analyzing the most common data sources in smart manufacturing systems, two types of data sources are observed: (i) machines and devices with built-in sensors that provide completely raw, unprocessed data directly from manufacturing processes, like sensors, actuators, cameras, tags and controllers; and (ii) information and communication systems that provide data already collected from production, integrated and stored, which have been partially processed and can be used for further analysis, like manufacturing execution systems (MES), enterprise resource planning (ERP), customer relationship management (CRM) systems, database systems, Apache Hadoop and Supervisory Control and Data Acquisition (SCADA). The first type of data sources collectively takes 81.6%, where the sensors take 35%.



(a) (b)

Fig. 2. Most common sources of data in Industry 4.0 / Types of data analysis

In order to provide an answer to the research question ***RQ2*** the papers are classified based on the performed type of data analytics as: descriptive, predictive, prescriptive or diagnostic. For a certain number of papers it was not possible to conclude which type of data analytics is meant and they are classified as "Other".

Descriptive and diagnostic analytics are based on reports and queries that explain the previous behavior of the observed system. The focus of these analytics is on historical data review, answering the questions *What happened*? and *Why did it happen*?, respectively [14]. Predictive and prescriptive analytics are performed to answer the questions *What will happen*? and *What to do*?, respectively [14]. Prescriptive analytics uses quantitative optimization and stochastic simulation to support data-driven decision making [54]. Based on the results presented in Fig. 2 (b), the most researches use data analytics to predict the future behavior of the observed industrial system. Descriptive analytics is presented in a slightly smaller number of papers, while diagnostic analytics is described in only one paper.

There is a wide range of available tools and technologies used to manage large data sets. To answer the research question ***RQ3*** the technologies are presented and classified according to the following layers: data acquisition, data transfer, data integration, data storage, data analytic and data visualization layer (Table 4). For each layer, the technologies are sorted starting with the most used technology. For some layers there is a wide range of different technologies that are mentioned, while for some other layers there are just few tools that are dominantly used. The most frequently mentioned data storage system is Hadoop Distributed File System (HDFS). Apache Spark, especially MLib library, and Apache Hadoop's MapReduce are the most widely used for data analytics. Dominant usage of various Apache tools throughout all layers of the data management system architecture is evident.

4 Discussion

The conducted literature review provides: an overview of the most commonly used data sources; technologies used to collect raw data, their purification and elimination of incomplete data, integration and storage; and, at the very end, data analysis and visual representation of the obtained information. In addition, the paper presents the types of data analysis that are carried out in industrial systems, as well as their importance, that is, the reason why adequate data handling is important in all the previously mentioned stages.

Table 4. Big data technologies

Data acquisition	SCADA, Human Machine Interface (HMI), Manufacturing execution system (MES) , Distributed control system (DCS), Apache Flume, Siemens PLC, Apache Hadoop, Savvy Smart Box, Python, R, Scala, ARTEMIS, NodeRED
Data transfer	Apache Kafka, 4G network, 5G network, Bluetooth, WiFi, Z Wave, WiMax, Cellular, NFC, ZigBee, RFID, REST and API, Logstash, AVRO, protocols: MQTT, AMQP, MODBUS
Data integration	MapReduce, Apache Storm stream, Python (NumPy and Pandas libraries), Data Wrangler, Scala (slick), R (dplyr)
Data storage	HDFS, MongoDB, Cassandra, InfluxDB, Elasticsearch, Cloud infrastructure, CouchDB, MySQL, MariaDB, Apache HBase, Enterprise DB, Time-Series DB, Apache Hive Central Data Warehouse, SQL Server, KEP OPC Server, MachinesList, ERP, MapR Binary DB
Data analytic	Apache Spark, Hadoop MapReduce, Apache Flink, Mixpanel's data analytics tool, Google Analytics tool, R, TimeFlow, NodeXL, Microsoft HDInsight
Data visualisation	Python, Grafana, Smart.viz, Kibana, Scala, R (ggplot2 package), Tabelau, Apache Drill, Apache Zeppelin

Data collected from different devices are found in different formats, such as batch files, signal stream, sensor data, which makes further steps in the data handling process difficult. It is important to emphasize that there is another group of data that is important for decision-making within the industrial system, namely user-generated data, obtained through questionnaires, surveys, interviews, observations, and documents [19]. This type of data also affects decision-making, but is beyond the scope of this research. The raw data from the devices needs to be temporarily saved as it is, and based on the previously presented results. It can be concluded that the SCADA data collection system is most often used for this purpose. Data transmission is provided through various network and communication technologies and protocols. Manual data collection is still prevalent, but it has been shown that this method of data collection does not lead to effective decision-making [55].

Raw data needs to be cleaned, transformed into a suitable format and integrated, to be ready for the storage process. The MapReduce tool founded by the Apache company is used in the largest number of reviewed papers.

It can be concluded that the widest assortment of technologies has been identified for the purposes of data storage. The most common is the Apache HDFS. In addition, various relational and NoSQL databases are used. NoSQL databases are broadly used, as expected, primarily due to the possibility of storing heterogeneous data, for which traditional databases are not suitable.

In relation to data processing of a large amount of data Apache Spark and Apache Flink are the core technologies used. MapReduce-based technologies are extensively used to enable parallel computation on several computing nodes [18]. Data analytics is crucial in order to detect possible irregularities in critical process parameters in real time, and to make timely reactions aimed at correcting their. In this way, companies manage to meet the requirements and norms of customer expectations.

The first three research questions defined in Section 3 are successfully and extensively answered. There was not enough information to extract and analyze the answers to the research questions ***RQ3.1*** and ***RQ3.2***, but certain conclusions can be successfully drawn. Valero et al. in the paper [53] mention the challenges of implementing a smart environment that are not completely overcome by the application of existing technologies. The first problem is the security of data transmission and storage on the internet, which should be addressed by introducing more secure protection mechanisms such as end-to-end data encryption and private cloud. Another problem relates to the existence of data transfer and communication standards such as the MTConnect standard and Modbus to which all devices should be adapted to enable their interoperability. The remaining challenges are the adaptation of the smart environment to existing systems and the collaboration between machines and people [53].

A data lake is a centralized repository that stores large amounts of structured, unstructured, and semi-structured data. In essence, data lake represents the evolution of the data warehouse system as it adapts to the requirements of big data and the well-known 3V model of big data [56]. In the selected papers, the concept of data lake is rarely mentioned, and there is no comprehensive answer to ***RQ3.2***. It can be concluded that this is a gap that requires further research. As described in [56] when implementing a data lake it is necessary to answer challenges such as successfully designing and engineering a storage architecture that allows to achieve maximum effectiveness and efficiency of the data lake. Data lakes store raw data in just-as-it-is form, and before the actual analysis, it is necessary to research such data and create metadata describing it in order to reveal its semantics and facilitate further analysis [56]. Among other things, it is another problem where there is room for finding and improving technologies that facilitate work with such systems.

5 Conclusion

Big data analytics become a key base for competitiveness, productivity growth and innovation of industrial enterprises. In a big data environment, data sets are significantly larger and may be too complex for traditional data analysis software. For this reason, the choice of appropriate technologies is the basis for data analysis that contributes to the successful functioning of a system.

The purpose of this literature review is to show the most commonly used data sources, ways of collecting such data, their transformation, storage and integration, as well as visualization and generation of information necessary for making decisions that will improve the operation of the industrial system. The results of this research can be used as a guide when introducing a new industrial system or adapting an already existing industrial system to Industry 4.0 concepts.

In further work, it is necessary to carry out more research, which includes the development and testing of new methods and techniques for the analysis of big data.

Also, it is necessary to research and apply the techniques found in order to determine which the best choice is when it comes to quality control of the process as well as improving the experience and satisfaction of the users of the end product.

References

[1] L. Duan and L. Da Xu, “Data Analytics in Industry 4.0: A Survey,” *Inf. Syst. Front.*, 2021

[2] J. Petrovački, D. Gladić, S. Havzi, D. Stefanović, and S. Ristić, “NoSQL Databases Deployment in Various Application Domains : A Systematic Literature Review.”

[3] A. G. Frank, L. S. Dalenogare, and N. F. Ayala, “Industry 4.0 technologies: Implementation patterns in manufacturing companies,” *Int. J. Prod. Econ.*, vol. 210, pp. 15–26, Apr. 2019.

[4] H. S. Park and R. A. Febriani, “Modelling a platform for smart manufacturing system,” in *29th International Conference on Flexible Automation and Intelligent Manufacturing(FAIM2019), June 24-28, Limerick, Ireland*, 2019, vol. 38, pp. 1660–1667

[5] E. A. Hinojosa-Palafox, O. M. Rodriguez-Elias, J. A. Hoyo-Montano, and J. H. Pacheco-Ramirez, “Towards an Architectural Design Framework for Data Management in Industry 4.0,” in *Proceedings - 2019 7th International Conference in Software Engineering Research and Innovation, CONISOFT 2019*, Oct. 2019, pp. 191–200

[6] M. Khan, X. Wu, X. Xu, and W. Dou, “Big Data Challenges and Opportunities in the Hype of Industry 4.0,” 2017.

[7] M. Y. Santos *et al.*, “A Big Data system supporting Bosch Braga Industry 4.0 strategy,” *Int. J. Inf. Manage.*, vol. 37, no. 6, pp. 750–760, Dec. 2017

[8] B. Kitchenham, “Procedures for Performing Systematic Reviews,” 2004.

[9] T. K. Lepasepp and W. Hurst, “A systematic literature review of industry 4.0 technologies within medical device manufacturing,” *Futur. Internet*, vol. 13, no. 10, Oct. 2021

[10] A. Bueno, M. Godinho Filho, and A. G. Frank, “Smart production planning and control in the Industry 4.0 context: A systematic literature review,” *Comput. Ind. Eng.*, vol. 149, Nov. 2020.

[11] F. S. Tesch da Silva, C. A. da Costa, C. D. Paredes Crovato, and R. da Rosa Righi, “Looking at energy through the lens of Industry 4.0: A systematic literature review of concerns and challenges,” *Comput. Ind. Eng.*, vol. 143, May 2020.

[12] M. Andronie, G. Lăzăroiu, M. Iatagan, I. Hurloiu, and I. Dijmărescu, “Sustainable cyber-physical production systems in big data-driven smart urban economy: A systematic literature review,” *Sustain.*, vol. 13, no. 2, pp. 1–15, Jan. 2021.

[13] T. Zheng, M. Ardolino, A. Bacchetti, and M. Perona, “The applications of Industry 4.0 technologies in manufacturing context: a systematic literature

review," *Int. J. Prod. Res.*, vol. 59, no. 6, pp. 1922–1954, 2021.

[14] J. Vater, L. Harscheidt, and A. Knoll, "Smart Manufacturing with Prescriptive Analytics A review of the current status and future work," 2019.

[15] K. Nagorny, S. Scholze, J. Barata, and A. W. Colombo, "An approach for implementing ISA 95-compliant big data observation, analysis and diagnosis features in industry 4.0 vision following manufacturing systems," in *IFIP Advances in Information and Communication Technology*, 2016, vol. 470, pp. 116–123.

[16] F. Tao and Q. Qi, "New IT driven service-oriented smart manufacturing: Framework and characteristics," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 49, no. 1, pp. 81–91, Jan. 2019.

[17] S. Kahveci, B. Alkan, M. H. Ahmad, 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," *J. Manuf. Syst.*, vol. 63, pp. 214–223, Apr. 2022.

[18] G. Guerreiro, R. Costa, P. Figueiras, D. Graça, and R. Jardim-Gonçalves, "A self-adapted swarm architecture to handle big data for 'factories of the future,'" in *Conference on Manufacturing Modelling, Management and Control (IFAC MIM)*, Sep. 2019, vol. 52, no. 13, pp. 916–921

[19] H. Han and S. Trimi, "Towards a data science platform for improving SME collaboration through Industry 4.0 technologies," *Technol. Forecast. Soc. Change*, vol. 174, Jan. 2022.

[20] S. Ren, J. S. Kim, W. S. Cho, S. Soeng, S. Kong, and K. H. Lee, "Big Data Platform for Intelligence Industrial IoT Sensor Monitoring System Based on Edge Computing and AI," in *3rd International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2021*, Apr. 2021.

[21] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, and Y. Liu, "Implementation of Industrial Cyber Physical System: Challenges and Solutions," 2019.

[22] T. Cerquitelli *et al.*, "Manufacturing as a Data-Driven Practice: Methodologies, Technologies, and Tools," *Proc. IEEE*, vol. 109, no. 4, pp. 399–422, Apr. 2021.

[23] Q. Qi and F. Tao, "Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison," *IEEE Access*, vol. 6, pp. 3585–3593, Jan. 2018.

[24] D. Sanghavi, S. Parikh, and S. A. Raj, "Industry 4.0: Tools and implementation," *Manag. Prod. Eng. Rev.*, vol. 10, no. 3, pp. 3–13, 2019.

[25] S. K. Jagatheesaperumal, M. Rahouti, K. Ahmad, A. Al-Fuqaha, and M. Guizani, "The Duo of Artificial Intelligence and Big Data for Industry 4.0: Review of Applications, Techniques, Challenges, and Future Research Directions," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 12861–12885, Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.02425.

[26] S. Ma, W. Ding, Y. Liu, S. Ren, and H. Yang, "Digital twin and big data-driven sustainable smart manufacturing based on information management systems for energy-intensive industries," *Appl. Energy*, vol. 326, Nov. 2022.

[27] A. Bousdekis and G. Mentzas, "Enterprise Integration and Interoperability for

Big Data-Driven Processes in the Frame of Industry 4.0," *Front. Big Data*, vol. 4, Jun. 2021.

[28] 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," *Int. J. Adv. Manuf. Technol.*, vol. 117, no. 5–6, pp. 1959–1973, Nov. 2021.

[29] F. Xing, G. Peng, T. Liang, S. Zuo, and S. Li, "Managing Changes Initiated by Industrial Big Data Technologies: A Technochange Management Model," in *Lecture Notes in Computer Science*, 2019, vol. 11587 LNCS, pp. 75–87

[30] P. Mallioris, G. Kokkas, A. Styliadis-Heinz, I. Margaritis, F. Stergiopoulos, and D. Bechtsis, "Conceptual Framework of Predictive Maintenance in a Canning Industry," in *HAICTA 2022, September22–25, 2022, Athens, Greece*, 2022, pp. 253–259.

[31] A. Nigussie Turi and X. Samuel Li, "Data-Driven Decision-Making in Digital Entrepreneurship," *Int. J. Ind. Syst. Eng.*, vol. 16, pp. 122–127, 2022, [Online]. Available: https://www.researchgate.net/publication/361778286.

[32] E. Hossain, I. Khan, F. Un-Noor, S. S. Sikander, and M. S. H. Sunny, "Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review," *IEEE Access*, vol. 7, pp. 13960–13988, 2019.

[33] K. Haricha, A. Khiat, Y. Issaoui, A. Bahnasse, and H. Ouajji, "Towards smart manufucturing: Implementation and benefits," in *Procedia Computer Science*, 2020, vol. 177, pp. 639–644

[34] M. Fay and N. Kazantsev, "When Smart Gets Smarter: How Big Data Analytics Creates Business Value in Smart Manufacturing," 2018, [Online]. Available: https://www.researchgate.net/publication/328873095.

[35] R. Badarinath and V. V. Prabhu, "Advances in internet of things (Io T) in manufacturing," in *IFIP Advances in Information and Communication Technology*, 2017, vol. 513, pp. 111–118

[36] A. Bashar Fakhri *et al.*, "Industry 4.0: Architecture and Equipment Revolution," *Comput. Mater. Contin. Novemb. 2020*, 2020.

[37] A. Corallo, A. M. Crespino, M. Lazoi, and M. Lezzi, "Model-based Big Data Analytics-as-a-Service framework in smart manufacturing: A case study," *Robot. Comput. Integr. Manuf.*, vol. 76, Aug. 2022.

[38] N. Iftikhar, T. Baattrup-Andersen, F. E. Nordbjerg, E. Bobolea, and P. B. Radu, "Data analytics for smart manufacturing: A case study," in *8th International Conference on Data Science, Technology and Applications*, 2019, pp. 392–399

[39] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, and Y. Liu, "A global manufacturing big data ecosystem for fault detection in predictive maintenance," *IEEE Trans. Ind. Informatics*, vol. 16/1, pp. 183–192, 2020.

[40] K. H. Chang, C. C. Tsai, C. H. Wang, C. J. Chen, and C. M. Lin, "Optimizing the energy efficiency of chiller systems in the semiconductor industry through big data analytics and an empirical study," *J. Manuf. Syst.*, vol. 60, pp. 652–661, Jul. 2021.

[41] R. Sahal, S. H. Alsamhi, K. N. Brown, D. O'shea, C. McCarthy, and M. Guizani, "Blockchain-empowered digital twins collaboration: Smart transportation use case," *Machines*, vol. 9, no. 9, Sep. 2021.

[42] M. O. Gökalp, E. Gökalp, K. Kayabay, A. Koçyiğit, and P. E. Eren, "Data-driven manufacturing: An assessment model for data science maturity," *J. Manuf. Syst.*, vol. 60, pp. 527–546, Jul. 2021.

[43] A. Sebbar, K. Zkik, A. Belhadi, A. Benghalia, M. Boulmalf, and M. D. Ech-Cherif El Kettani, "BDA-enabler Architecture Based on Cloud Manufacturing: the Case of Chemical Industry," *Int. J. Supply Oper. Manag.*, vol. 9, no. 3, pp. 251–263, Aug. 2022.

[44] A. Sharma and H. Pandey, *Big Data and Analytics in Industry 4.0*. 2020.

[45] A. Bagozi, D. Bianchini, V. De Antonellis, A. Marini, and D. Ragazzi, "Enabling big data exploration in dynamic contexts," 2018.

[46] P. Zheng *et al.*, "Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives," *Front. Mech. Eng.*, vol. 13, no. 2, pp. 137–150, Jun. 2018.

[47] P. Bellavista *et al.*, "Design Guidelines for Big Data Gathering in Industry 4.0 Environments," 2019.

[48] B. Suryajaya, C.-C. Chen, M.-H. Hung, Y.-Y. Liu, J.-X. Liu, and Y.-C. Lin, "A Fast Large-size Production Data Transformation Scheme for Supporting Smart Manufacturing in Semiconductor Industry," 2017.

[49] A. A. C. Vieira, L. Dias, M. Y. Santos, G. A. B. Pereira, and J. Oliveira, "Bypassing data issues of a supply chain simulation model in a big data context," in *International Conference on Industry 4.0 and Smart Manufacturing (ISM 2019)*, 2020, vol. 42, pp. 132–139

[50] G. Terrazas, N. Ferry, and S. Ratchev, "A cloud-based framework for shop floor big data management and elastic computing analytics," *Comput. Ind.*, vol. 109, pp. 204–214, Aug. 2019.

[51] A. Kirmse, F. Kuschicke, and M. Hoffmann, "Industrial big data: From data to information to actions," in *4th International Conference on Internet of Things, Big Data and Security*, 2019, pp. 137–146

[52] R. M. Holom, K. Rafetseder, S. Kritzinger, and H. Sehrschön, "Metadata management in a big data infrastructure," in *International Conference on Industry 4.0 and Smart Manufacturing*, vol. 42, 2020, pp. 375–382

[53] M. R. Valero, S. T. Newman, and A. Nassehi, "Link4Smart: A New Framework for Smart Manufacturing Linking Industry 4.0 Relevant Technologies," in *55th CIRP Conference on Manufacturing Systems*, 2022, vol. 107, pp. 1594–1599

[54] C. H. Wang, H. Y. Cheng, and Y. T. Deng, *Using Bayesian belief network and time-series model to conduct prescriptive and predictive analytics for computer industries*, vol. 115. 2018.

[55] A. Tsanousa *et al.*, "A Review of Multisensor Data Fusion Solutions in Smart Manufacturing: Systems and Trends," *Sensors*, vol. 22, no. 5, Mar. 2022.

[56] A. Cuzzocrea, "Big data lakes: Models, frameworks, and techniques," *IEEE International Conference on Big Data and Smart Computing*, 2021, pp. 1–4.