Heliyon 10 (2024) e26503

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Review article

Digital twin: Data exploration, architecture, implementation and future

Md. Shezad Dihan *, Anwar Islam Akash, Zinat Tasneem, Prangon Das, Sajal Kumar Das, Md. Robiul Islam, Md. Manirul Islam, Faisal R. Badal, Md. Firoj Ali, Md. Hafiz Ahamed, Sarafat Hussain Abhi, Subrata Kumar Sarker, Md. Mehedi Hasan

Department of Mechatronics Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh

ARTICLE INFO

Keywords:
Digital twin
Data analysis
Manufacturing
Urbanization
Medical
Agriculture
Robotics
Military and aviation

ABSTRACT

A Digital Twin (DT) is a digital copy or virtual representation of an object, process, service, or system in the real world. It was first introduced to the world by the National Aeronautics and Space Administration (NASA) through its Apollo Mission in the '60s. It can successfully design a virtual object from its physical counterpart. However, the main function of a digital twin system is to provide a bidirectional data flow between the physical and the virtual entity so that it can continuously upgrade the physical counterpart. It is a state-of-the-art iterative method for creating an autonomous system. Data is the brain or building block of any digital twin system. The articles that are found online cover an individual field or two at a time regarding data analysis technology. There are no overall studies found regarding this manner online. The purpose of this study is to provide an overview of the data level in the digital twin system, and it involves the data at various phases. This paper will provide a comparative study among all the fields in which digital twins have been applied in recent years. Digital twin works with a vast amount of data, which needs to be organized, stored, linked, and put together, which is also a motive of our study. Data is essential for building virtual models, making cyber-physical connections, and running intelligent operations. The current development status and the challenges present in the different phases of digital twin data analysis have been discussed. This paper also outlines how DT is used in different fields, like manufacturing, urban planning, agriculture, medicine, robotics, and the military/aviation industry, and shows a data structure based on every sector using recent review papers. Finally, we attempted to give a horizontal comparison based on the features of the data across various fields, to extract the commonalities and uniqueness of the data in different sectors, and to shed light on the challenges at the current level as well as the limitations and future of DT from a data standpoint.

1. Introduction

In the 1960s, NASA pioneered the idea of examining a physical object using a digital twin. For exploratory missions, NASA recreated its spacecraft on Earth to mirror the systems in space. Despite the fact that DT is not a new idea, it is now considered one

* Corresponding author.
E-mail address: hanzad.oo9@gmail.com (M.S. Dihan).

https://doi.org/10.1016/j.heliyon.2024.e26503
Received 26 March 2023; Received in revised form 13 February 2024; Accepted 14 February 2024
Available online 21 February 2024

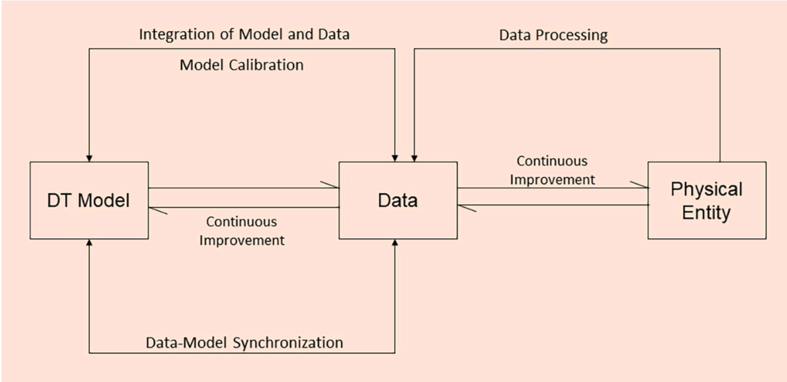

Fig. 1. The interrelationship among multiple aspects of a DT system.

of the topmost research topics among academics all over the globe. “A digital twin is a virtual representation of a physical system (and its associated environment and processes) that is updated through the exchange of information between the physical and virtual systems” [1]. It is a virtual representation of a physical system or object that replicates, simulates, predicts, and gives prognostic data analysis throughout the entire life cycle of the physical entity. A DT is the digital representation of a physical thing or process, and as such, it is a living, intelligent, and constantly developing model. It does this by replicating the lifetime of its physical counterpart in order to monitor, manage, and improve the functioning of its systems. It continually anticipates future states (such as flaws, damages, and failures) and enables creating and testing fresh configurations to apply preventative maintenance procedures [2].

A basic DT system consists of a physical object and a virtual model connected through bidirectional data flow. DT technology depends on collecting a lot of data and figuring out how it fits together. The information sent from the physical component to the virtual model is in its raw form and must be processed before it can be converted into useful data [3]. Digital Twins require the integration of numerous enabling technologies, whose technical level of intelligence and development must also be considered, such as the CPS (Cyber Physical System) concept from the system engineering and control perspective, IOT (Internet of Things) from the networking and IT (Information Technology) perspective, and Digital Twin from the computational modeling (Machine Learning (ML)/Artificial Intelligence (AI)) perspective [1]. Because of this, DT is closely linked with extensive modeling driven by advanced ML/Deep Learning (DL) and big data analytic techniques. Fig. 1 comprehensively depicts the interrelationship among several components of a DT system.

The data obtained from tangible items, sensors, Internet of Things (IoT) devices, and many origins form the fundamental basis of a digital twin. The provided data encompasses details pertaining to the present condition, actions, and efficacy of the tangible entity. Data can exhibit many types, encompassing numerical, category, time series, and other variations. The abundance and variety of the data enhance the accuracy and authenticity of the digital twin. The models within a digital twin system serve as virtual representations of their corresponding physical elements. The aforementioned models are mathematical abstractions or simulations that aim to represent or simulate real-world objects or systems. They imitate the behavior and traits of the corresponding physical entity. A range of modeling methodologies can be employed, encompassing physics-based models [4,5], data-driven models [6,7], and machine learning models [8]. The selection of a modeling technique is contingent upon the particular application and the data that is accessible. Data obtained from tangible items is utilized for training models that rely on data analysis. These models utilize previous data to acquire knowledge, provide forecasts, simulate the behavior, and provide valuable insights into the future performance of the physical system. Data is frequently employed for the purpose of calibrating and validating models. Comparing the predictions generated by the model with empirical data obtained from the actual world can enhance the model's accuracy [9]. This ensures that the digital twin closely approximates the physical entity it represents. One essential attribute of digital twin systems is the reciprocal data exchange between the physical entity and the simulated model [10]. The virtual model is consistently updated using real-time data from its physical counterpart. Modifications or adaptations performed on the virtual model can influence the tangible system, enabling the opportunity to conduct testing and enhance performance within a secure setting. Data obtained from physical entities sometimes exists in a raw state and may necessitate pre-processing to eliminate noise, outliers, or redundant information. Data processing techniques render the data appropriate for input into a model. Subsequently, the processed data is employed to revise or improve the models, maintaining their fidelity in accurately depicting the existing condition of the physical system. In order to uphold the precision of a digital twin, it is imperative to ensure the synchronization of both data and models [11]. This implies that the data utilized for model inputs should accurately represent the current state of the physical object, while the model outputs should correspond with observable behaviors. The interplay between data and models within digital twin systems facilitates a perpetual process of enhancement and refinement. With the accumulation and analysis of further data, there is an opportunity to enhance and revise models, leading to the development of a digital replica that exhibits increased precision and enhanced capabilities as time progresses.

This makes it possible to make predictions and forecasts in real time. A DT system can be considered a copy of a physical target system. It uses a model to simulate the different ways of how biological systems work. Using extensive data analysis in intelligent manufacturing systems can help system administrators and engineers find weak spots. Also, system administrators and engineers can update systems to improve supply chain and product performance based on what they learn from analyzing big data. Through the

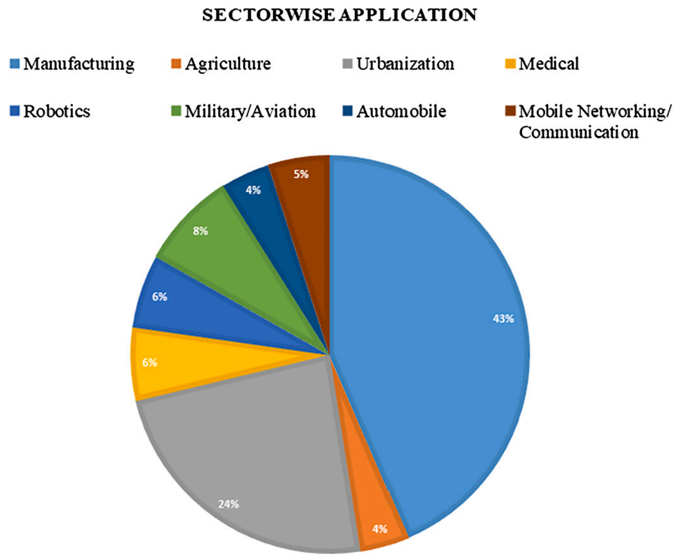

Fig. 2. Sector-wise percentage of digital twin-based data analysis paradigm observed in recent papers.

use of sensors, real-world information is fed into digital models for use in simulation, validation, and real-time fine-tuning, and the information gleaned from the simulation is used to inform and enhance the real-world implementation and value-creation processes in response to the alterations [12]. Still, there are severe risks to changing or updating the whole system. So, a good strategy is to make a digital copy of an existing physical system to simulate real situations in that physical system. This is called a "Digital Twin" (DT). The development and uses of DT, however, provide new trends and needs due to the ongoing growth and upgrading of application requirements. For instance, if we analyze Fig. 2, in recent years, the applications of DT have steadily moved from their original focus on the military and aerospace industries to the realms of daily life [13]. Much research and development work has been done on intelligent cities and DT applications as part of the urbanization system. Like smart manufacturing, smart cities are made up of many IoT domains that work together to solve the complex problems that cities face. These domains include cloud/edge computing, extensive data collection and analysis, and edge computing. These are all essential techniques for driving efficiency and optimization. Digital twins and autonomous cognitive systems could help the agriculture industry deal with the growing problems of managing resources and meeting food demand. Agricultural production systems must change to produce more while using fewer resources. Autonomous intelligent technologies and digital twin could help solve the problems of managing resources and meeting the growing demand for food in agriculture. However, many things make it hard to get the data needed to help doctors make decisions, such as ethical and financial issues.

The first step should be to use advanced modeling strategies or tools like SysML, Modelica, SolidWorks, 3DMAX, and AutoCAD to make high-precision twin models that match the real thing. Then, a digital twin should make from these plans. Next, health IoT and mobile network protocols should be used to link data so that real-time interaction between real and virtual things can continue. On the other hand, robotics infrastructure depends a lot on simulation and technology based on simulation. The term "Digital Twin" (DT) is often used to describe these simulations at a deep level similar to the system they were made to the model. In the aerospace industry, simulations imitate the continuous history of flights. This gives much information about what the plane has been through and can be used with various feature-based simulation techniques to predict future serviceability and violations. We attempted to enumerate and represent in keywords every essential technology needed to construct sector-specific DT systems in Fig. 3.

Setting pertinent standards to enable inter-operation in developing applications that enable digital twins is vital. Interoperation and interconnection among various businesses or fields will be unavoidable in the process [43]. In Table 1, we have tried to summarize the current status of the data analytical study regarding digital twins. DT applications rely on the collaboration of full or all parts from a variety of various domains in order to achieve their ultimate goal of establishing a closed-loop of an intelligent decision-making optimization system [43]. It is clear from the table that the study regarding this manner is largely distinguished among individual sectors. This paper will try to unify them and present a comparative study. With their data-driven foundation, this paper hypothesizes that Digital Twin offers a transformative approach to system optimization, resource management, and decision-making across various sectors. By exploring the nuances of data analysis in Digital Twins and addressing the challenges and security concerns, we aim to provide a comprehensive understanding of the current state of the technology and outline pathways for future development and innovation. Through a rigorous examination of data analysis techniques, sector-specific implementations, and potential solutions to emerging challenges, this paper seeks to contribute to the growing body of knowledge in the domain of Digital Twins, inspiring researchers, practitioners, and decision-makers to harness the full potential of this revolutionary concept.

The remaining parts of this work are structured in the following manner. The data analysis process is reviewed in Section 2. Sector-wise implementation of data analysis for a digital twin is elaborated in Section 3. Then, present data processing issues and multi-faceted challenges in data analysis methodologies were addressed in Section 4. Finally, the paper was concluded in Section 5

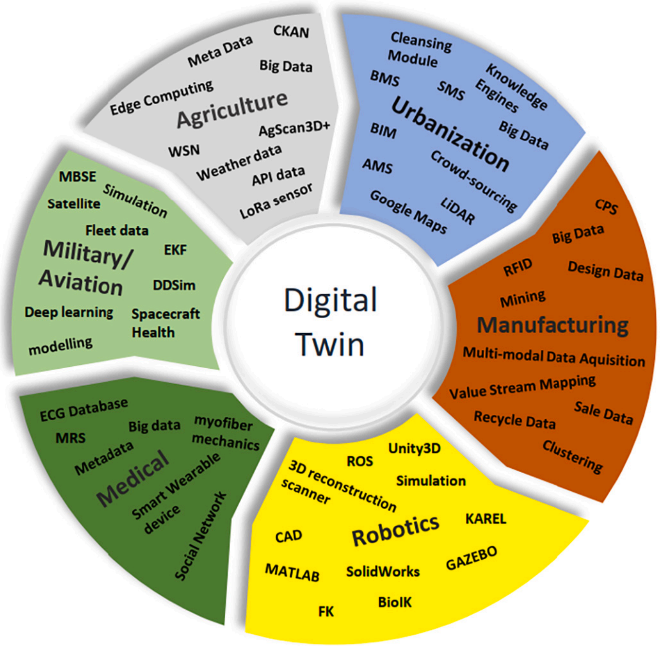

Fig. 3. Keywords used in every sector regarding the digital twin.

Table 1
A comparison to recently published literature reviews in this field.

SI No.	Author	References	Date of Publishment	Country	Manufacturing	Agriculture	Urbanization	Medical	Robotics	Military/ Aviation
1	Fei Tao et al	[12] [14] [15] [16] [17] [18]	2017-2022	China	✓	×	✓	×	✓	×
2	Lihui Wang et al	[19] [20] [21]	2019-2021	Sweden	✓	×	×	×	✓	×
3	Meng Zhang et al	[22] [16] [15] [23] [24]	2018-2021	China	✓	×	×	×	×	×
4	A.Y.C Nee et al	[22] [25] [16]	2018-2021	Singapore	✓	×	×	×	×	×
5	Tianliang Hu et al	[26] [27] [28] [29]	2018-2022	China	✓	×	×	×	✓	×
6	SKA Rahim et al	[30]	2021	Malaysia	×	✓	×	×	×	×
7	Petr Skobelev et al	[31]	2021	Russia	×	✓	×	×	×	×
8	Timon Höbert *et al*	[32]	2019	Austria	✓	×	×	×	✓	×
9	AF Mendi et al	[33] [34] [35]	2020-2021	Turkey	×	×	✓	✓	×	✓
10	Rogelio Gámez Díaz et al	[36] [37]	2019-2020	Canada	×	×	×	✓	×	×
11	BR Barricelli et al	[38]	2019	Italy	×	×	×	✓	×	×
12	M Blackburn et al	[39] [40]	2017-2018	USA	×	×	×	×	×	✓
13	Calin Boje et al	[41] [42]	2020-2022	Luxembourg	×	×	✓	✓	×	×
14	**Current Study**	-	-	**Bangladesh**	✓	✓	✓	✓	✓	✓

2. Data analysis process

2.1. Data collection

DT data structures mainly consist of physical entity data, virtual model data, service-related data, and domain-based knowledge [44]. We use sensors, embedded systems, offline measurement, and sampling inspection methods [45] for physical entities to gather data. In the case of virtual models, it is modeling and real-time simulation. For service, related data is collected through service construction and maintenance. DT depends on experts, crowdsourcing, and historical data for domain-based knowledge. Comprehensive data is necessary to enhance the effectiveness and improve the accuracy, efficacy, and adaptability of DT-based services. Complete data from the actual and virtual worlds should be the foundation of any efficient DT system. Current data generation technology available for comprehensive data generation is multidimensional modeling technology, which produces geometric, physical, and behavioral data. Another one is transfer learning (metamodel) technology, which generates sampled data, and lastly, highly efficient simulation technology that does not generate data typically but enhances the efficiency of generated data. The physical devices implemented in the real world to collect the data are sensors, IoT devices, mobile devices, and wearable devices (Augmented Reality (AR), Virtual Reality (VR) devices). These peripherals work based on an integrated computing infrastructure process [46], including cloud and fog computing edge computing. Big data also plays a significant role in the effective data generation process. Big data can be defined as the ability to quickly extract hidden values and information from an enormous amount of data. Lack of precision in collecting physical assets is one example of a barrier associated with data collection, as is the gathering of data in real-time automation.

2.2. Data storage

Physical entity data, virtual model data, and service data from different application objects, situations, and scenarios should be unified and stored so they can be shared, reused, and exchanged. Data must first be represented in a standard way to reach this goal. This includes describing the data format, structure, encapsulation, sampling frequency, historical data accumulation, interface, communication protocol, etc. Then, to eliminate unsuitable application scenarios where DT can't be used because there isn't enough data, the necessary limits are set regarding how much historical data is collected, what kind of data is collected, and how often samples are taken. The relevant application criteria are met when certain restrictions can't be fulfilled. For the qualified ones, data from different situations (like design, production, and maintenance) with other formats, structures, and encapsulations are converted based on a single template. Next, a familiar interface and communication protocol are used to send data from different objects, like a robot and a machine tool. After that, the data can be modeled using the correct modeling language. Other modeling languages and methods have been used in the literature to manage data and information modeling for products and systems. These include Unified Modeling Language (UML) [47], Systems Modeling Language (SysML) [48], Ontology Language [49], and others [44]. However, each modeling language has its meaning, which makes it hard to exchange data and formats that are compatible with other languages. New mathematical approaches based on Category theory [50] could provide a complete foundation for modeling, interoperability, and integration. Based on this, the data can then be saved. Since digital twins rely on a huge data set, creating a repository for all digital twin data will be a challenge in the future.

2.3. Data association

Association ties between Digital Twin Data (DTD) are mined to aid knowledge discovery. First, data from the physical entity, virtual model, and service preprocessing to remove unnecessary and worthless data via data filtering, reduction, and feature extraction [44]. The second step is to perform temporal and spatial alignments. The least squares method, for example, can be used to synchronize data in time and translate data into the same spatial coordinate system. Then, using Pearson correlation analysis [51], K-means [52], Apriori algorithm [53], and other techniques, the relationships (e.g., causation, similarity, and complementation) among data are mined. Finally, a sophisticated network can develop to express these relationships completely. Numerous data variables are handled as nodes in the network, while data association relations are treated as edges. Further knowledge can be deduced using statistical, clustering, and classification methods, and the supposed ability can be expressed as a knowledge graph. Two types of data linkages, close and complementary, are highly critical to enable future data fusion. The former describes the relationship between data with similar attributes, values, or shifting trends. In contrast, the latter describes the relationship between multi-modal data from diverse sources that can explain the same quality or behavior from various perspectives. Related technologies for data association include spatial-temporal data alignment, data mining, knowledge reasoning, knowledge representation, and so on. Due to spatial-temporal data alignment technology, DTD is synchronous in time and the same coordinate system in space. Data mining algorithms (for example, Pearson correlation analysis, K-means, and the Apriori algorithm can reveal clustering groupings and association linkages among DTD. Knowledge reasoning technology can extract new knowledge from current knowledge or relations and groups. Knowledge representation technology (a knowledge graph) can visually depict understanding, knowledge carriers, and knowledge interactions. Future challenges in data association will come from the difficulties of ensuring the accuracy of data filtering for the massive dataset.

2.4. Data fusion

Living work on data fusion is mostly about combining data from the real world (e.g., sensor data with data entered by hand), and there have been few endeavors to integrate data from the real world and the virtual world. On the other hand, it combines all data related to physical entities, virtual models, services, and domain knowledge. The following are parts of DTD fusion. Suppose environmental changes, sensor failures, or human interference mess up biological entity-related data. In that case, methods like the weighted average method [54], the Dempster–Shafer theory [55], and the Kalman filter [56] can combine the physical entity-related data with similar virtual model-related data, and service-related data reduce the information entropy. By doing this, the randomness and fuzziness [57] of the data can be cut down. In the same way, if the virtual model and service-related data don't match reality,

Table 2
Popular mathematical data fusion ontology.

Sl No.	Fusion Method	Reference	Governing Equation	Parameter Description
1	Bayesian Inference	[59]	$y^T_{MAP} = \left[\frac{1}{\tau^2} z^T H^T + \tilde{\mu}^T \Sigma^{-1}\right] \left[\frac{1}{\tau^2} H H^T + \Sigma^{-1}\right]^{-1}$	H = Output operator T = number of realizations in the proper orthogonal decomposition with constraints method z = Quantities of interest measurements y = Field data τ^2 = Measurement noise Σ = Variance-covariance matrix $\tilde{\mu}$ = Mean of multivariate normal distribution
2	Proper Orthogonal Decomposition (POD)	[59]	$a^*, \lambda^* = \arg\min_{a,\lambda}\left(J_{a,\lambda}\right) = \frac{1}{2}\left(\Phi_k a - \tilde{\mu}\right)\left(\Phi_k^T a - \tilde{\mu}\right) + \lambda^T\left(H^T \Phi_k a - z\right)$	J = Cost function λ = Lagrange parameter Φ = left singular vector a = proper orthogonal decomposition basis expansion coefficients
3	Dempster-Shafer theory or DS evidence theory	[60]	$S'' = S'_1 \oplus S'_2 \oplus \cdots \oplus S'_R = \left[m''\left(A_1\right) \quad m''\left(A_2\right) \cdots m''\left(A_P\right)\right]^T$	S'' = Basic probability distribution of the final quadratic fusion applying DS synthesis $(S'_1, S'_2, \ldots S'_r)$ = Basic probability distributions of the $r_{th} file$ $(A_1, A_2, .., A_P)$ = Identification frame $m''(A_P)$ = Basic probability distribution for A_P
4	Convolutional Neural Network (CNN)	[61]	$\text{Fea} = f^p\left(f^c\left(f^p\left(f^c\left(f^p\left(f^c\left(X,\theta_1\right)\right),\theta_2\right)\right),\theta_3\right)\right)$	Fea = Feature Extracted by CNN $\theta_1, \theta_2, \theta_3$ = First, second, last convolution layer f^p = Max-pooling layer f^c = Convolutional layer
5	Particle Filtering Framework	[62]	$p\left(\mathbf{x}_k \mid z_{1:k}\right) \approx \sum_{i=1}^{N} w_k^{(i)} \delta\left(\mathbf{x}_k - \mathbf{x}_k^{(i)}\right)$, $\sum_{i=1}^{N} w_k^{(i)} = 1$	$p\left(\mathbf{x}_k \mid z_{1:k}\right)$ = Posterior probability density function $\mathbf{x}_k$ = State vector $\mathbf{z}_{1:k} = z_1, ., z_k$ = Set of measurements δ = Dirac condition $w_k^{(i)}$ = Weight i = 1,2,3,...,N

we can use methods like the Bayesian method and neural network [58] to merge the data with similar physical entity-related data to make it more accurate and reliable. For complimentary multi-modal data from different parts of DTD, we can use methods like neural network [58] and weighted average method [54] to increase the variety of information. Table 2 summarizes the widely used mathematical data fusion ontology along with its guiding formula.

2.5. Data sorting

It is necessary to unify the data linked to physical entities, virtual models, and services gathered from various application objects, situations, and scenarios before storing it for sharing, reusing, and exchanging. To do this, data must first be represented consistently, describing the data's format, structure, encapsulation, sampling frequency, historical data accumulation, interface, communication protocol, etc. To filter those unqualified application scenarios where DT is not applicable owing to insufficient data, relevant limits are established regarding historical data accumulation, data type, sampling frequency, etc. The accompanying application conditions are unqualified if a certain constraint cannot be met. Data from multiple situations (such as design, production, and maintenance) with varying formats, structures, and encapsulations are transformed for the qualifying ones based on a standard template. A single interface and communication protocol convey data from several devices (such as a robot and machine tool). The data can then be modeled using an appropriate modeling language. To handle data and information modeling for goods and systems, a broad range of modeling languages and techniques are utilized in the literature, including unified modeling language (UML) [63], systems modeling language (SysML) [64], ontology language [65], etc. However, the semantics of each modeling language differ, which restricts the sharing and compatibility of data and formats. New, reliable mathematical techniques based on Category theory—a methodology for data from digital twins may provide a new method for building a solid foundation.

2.6. Data coordination

DTD should be set up to allow real-time interactions so that its different parts can emit each other. First, selecting the correct data from various aspects of the DTD that carry valuable data to support the message transmission between any two elements is essential.

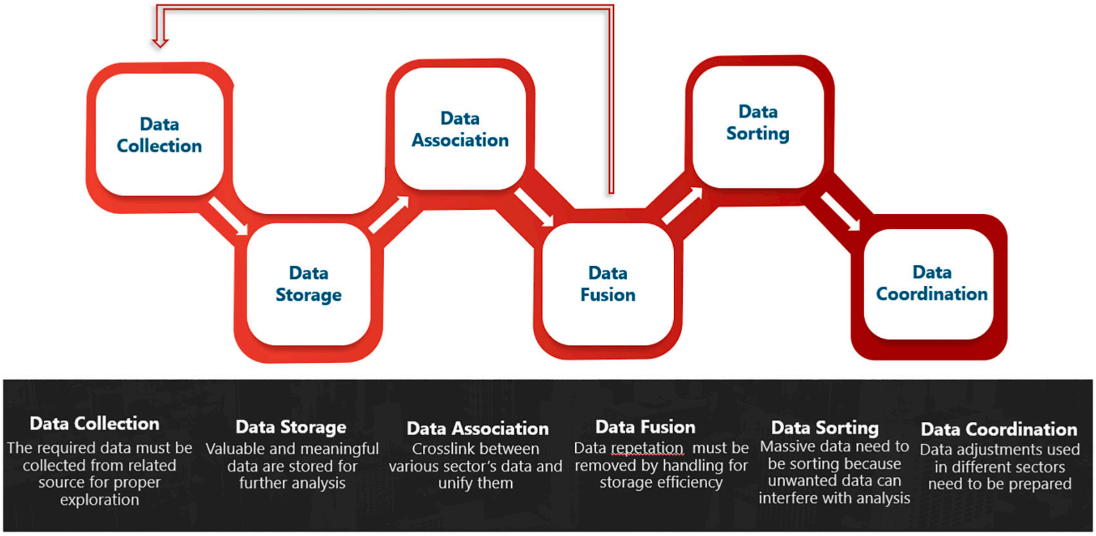

Fig. 4. Data analysis process for building a successful digital twin.

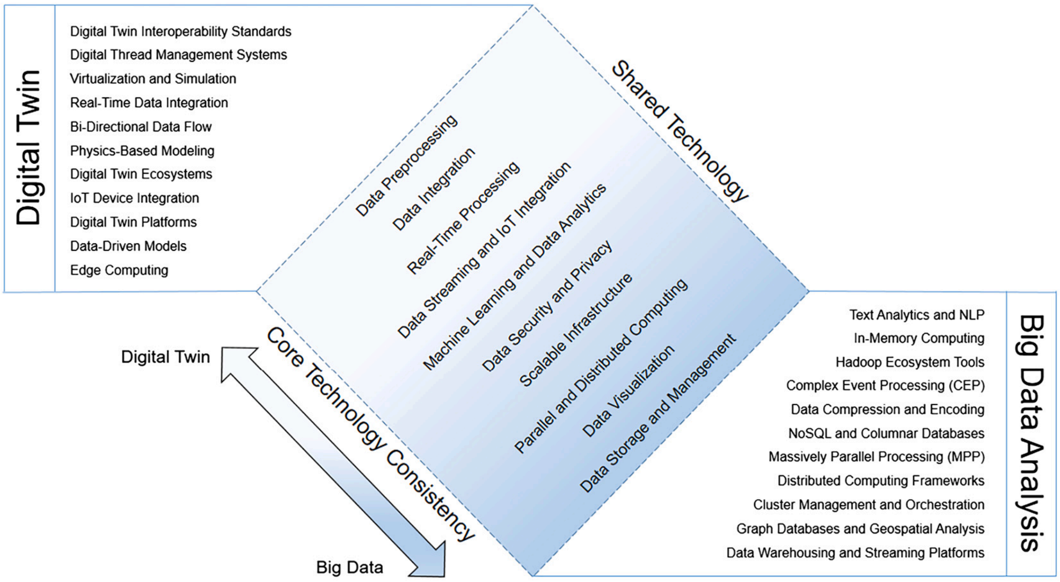

Fig. 5. Analytical consistency between digital twins and big data.

Taking a piece of equipment as an example, the actual states collected by sensors, which would show how the equipment works, and the simulated states produced by virtual models, which show what the expected forms are, can be chosen in advance to send messages between data about the physical entity and data about the virtual model. Second, to help with data transmission, the data is further processed by cleaning, dimensionality reduction, and compaction algorithms [66]. These algorithms help eliminate noise, duplicates, and redundant data. Then, in DT, the data is sent through the sensing devices, software, and database communication interfaces. Third, the Euclidean distance [67] between two parts is calculated in real-time to see how well their connection data matches up. When the distance between two corresponding parts exceeds a predefined threshold, it means there is inconsistency or a contradiction between them. To fix this, the correct parameters of the virtual model should change, should update service configurations, or should change the behavior of the physical entity. Time savings for DTD analysis dataset linking may prove challenging to achieve. In short, DTD interaction is vital to keeping the physical entity, virtual model, and service of DT in sync with each other. We attempted to illustrate in Fig. 4 how different individual analysis methods function as a unit for the overall data analysis process.

2.7. Digital twin and traditional big data analysis

Data analysis in digital twin systems exhibits many similarities to conventional big data analysis. However, it also presents notable distinctions stemming from the unique characteristics and goals associated with digital twins. Fig. 5 gives us a general sense of it. The fundamental technologies employed in these domains generally exhibit consistency, but their specific implementation and emphasis may differ. Digital twin systems heavily depend on the utilization of sophisticated virtualization and simulation technology. These technologies facilitate the creation of intricate, dynamic, and real-time virtual representations of physical systems, allowing for precise duplication and ongoing surveillance of the corresponding actual entities. Real-time data processing is a prevalent practice across multiple sectors [68]. However, digital twin systems emphasize the seamless and ongoing integration of data from physical entities, sensors, and Internet of Things (IoT) devices [69,70]. Real-time synchronization is crucial for the maintenance of an accurate virtual representation. Digital twins frequently utilize physics-based modeling, a technique that models physical phenomena by applying scientific principles. This methodology can provide precise and reliable depictions of physical systems and their interconnections. The notion of a digital thread encompasses preserving thorough digital documentation of an item or system over its entire existence. The comprehensive digital depiction facilitates the monitoring, examination, and enhancement of all stages, from initial design to eventual retirement. Digital thread management solutions facilitate the efficient administration of digital thread data, ensuring consistency and traceability over the entire lifespan of the physical entity [71]. Digital twin systems often depend on a diverse range of Internet of Things (IoT) sensors and devices for the purpose of acquiring real-world data, thus establishing their significance within the technology stack [72]. Utilizing specialized platforms and software tailored for digital twin development and administration represents a distinctive facet of technological advancement. These technologies enable digital replicas' development, incorporation, and live tracking. Data-driven models are widely utilized in the field of data analytics [73]. However, their significance becomes particularly evident inside digital twin systems, as they play a critical role in updating virtual models by incorporating real-time data. Machine learning and artificial intelligence (AI) algorithms dynamically adjust models in response to prevailing circumstances. The unique characteristic of digital twin systems lies in the bidirectional data flow between the physical system and its corresponding virtual model, wherein modifications made in the virtual representation can influence the physical counterpart. Edge computing technologies are important in digital twin systems due to their ability to facilitate real-time data processing and decision-making at the network edge [74–76]. This capability effectively reduces latency and improves responsiveness. The development of interoperability standards for digital twins is distinct within the area, as they do not constitute a technology in themselves. These standards assure the interoperability of digital twins across diverse systems, sectors, and domains. The incorporation of many technologies, such as the Internet of Things (IoT), cloud computing, big data analytics, and machine learning, inside a unified digital twin ecosystem is a notable characteristic of this technological advancement.

The discipline of big data analysis utilizes a variety of technologies that are frequently unique to this domain since they are designed to address the special issues and goals associated with managing and extracting valuable insights from large-scale information. Technologies such as Hadoop and Apache Spark play a pivotal role in analyzing large-scale data [77–80]. The utilization of clusters of computers permits the distribution of data processing, hence facilitating the study of extensive datasets that would be unfeasible to manage on a solitary system. NoSQL databases, including MongoDB, Cassandra, and HBase, are employed in the realm of big data analysis for the purpose of storing and effectively managing unstructured or semi-structured data [81–84]. These systems provide the ability to scale and adapt to manage a wide range of data kinds effectively. Data warehousing technologies such as Amazon Redshift and Google BigQuery are specifically engineered to accommodate the storage and analysis of vast quantities of structured data [85,86]. They enhance the efficiency of query execution for intricate analytical activities. Columnar databases, exemplified by Apache Cassandra, have been specifically designed to enhance the performance of analytical workloads. Data is stored in a column-wise format, which proves to be particularly advantageous for queries that require aggregations and reporting [87]. Neo4j, a type of graph database, is commonly employed for the purpose of analyzing interconnected data [88]. This characteristic renders it well-suited for various applications, including but not limited to social network analysis, fraud detection, and recommendation systems. Massively Parallel Processing (MPP) databases, such as Teradata [89] and Snowflake, have been purposefully engineered to cater to the demands of high-performance analytics [90,91]. Data and processing duties are distributed among numerous nodes in order to enhance the speed of query execution. The study of big data frequently depends on a range of technologies offered by the Hadoop ecosystem, including Pig for data processing, Hive for querying, and HBase for NoSQL data storage [92,93]. Technologies such as Apache Kafka play a crucial role in the management of real-time data streams, a task that has become increasingly significant in big data applications for processing and analyzing data as it is received [94]. Complex Event Processing (CEP) technologies, such as Apache Flink and Esper, are employed in the realm of real-time data analysis to facilitate the identification and analysis of patterns and trends within streaming data [95,96]. Text analytics and natural language processing (NLP) technologies play a vital role in the extraction of valuable insights from unstructured text data, including but not limited to social media content and consumer reviews. The utilization of compression and encoding methods is necessary in order to achieve efficient storage and processing of large-scale data, particularly in the context of columnar data formats like Apache Parquet. In-memory databases such as SAP HANA and Apache Ignite facilitate rapid data retrieval by storing it in random-access memory (RAM) [97,98], hence enhancing query execution speed for real-time analytical processes. Geospatial technologies are utilized for the examination of location-centric data, which holds significant value in various domains such as GPS navigation, logistics optimization, and urban planning. The study of big data frequently necessitates the utilization of specialized tools for data visualization that possess the capability to effectively manage the intricate nature and extensive magnitude of sizable datasets. These tools facilitate the interpretation and presentation of findings. Technologies such as Kubernetes and Docker Swarm are employed for the purpose of managing and orchestrating the deployment and scaling of large-scale data clusters [99,100].

The core technology consistency between data analysis in digital twin systems and traditional big data analysis resides in the fundamental techniques, tools, and infrastructure employed for processing and deriving insights from data. Both digital twin systems and traditional big data analysis necessitate efficient data storage and administration solutions. The aforementioned components encompass databases, data lakes, distributed file systems, and storage systems hosted on cloud platforms. Technologies such as

Hadoop HDFS, Apache Cassandra, and NoSQL databases have the potential to be employed in several scenarios [101,77,88]. The data preparation technique holds significant importance in the realms of digital twin and big data analysis. The process encompasses the tasks of data cleansing, transformation, and standardization in order to prepare the data for further analysis adequately. Both data cleaning, data imputation, and feature engineering are widely utilized techniques in the field [102–104]. Data integration plays a crucial role in both digital twin systems and conventional big data analysis. The integration of data from diverse sources, such as Internet of Things (IoT) devices, sensors, big data analysis, and digital twin systems, is a common necessity. Machine learning and data analytics are essential components in both domains. Both digital twin and traditional big data analysis employ algorithms for classification, regression, clustering, and anomaly detection. Popular libraries and frameworks can be utilized, including TensorFlow, scikit-learn, and Apache Spark MLlib. Real-time processing is a characteristic that sets digital twin systems apart. However, it can also be integrated into conventional big data analysis, particularly in scenarios where there is a requirement to process and analyze real-time data streams, such as social media feeds or financial transactions [12,16]. Data visualization tools and techniques are commonly employed in both digital twin and traditional big data analysis to depict data findings visually. This facilitates comprehension of intricate patterns, trends, and anomalies present in the data. The importance of scalability is significant in both sectors. Digital twin systems frequently necessitate scalable infrastructure to effectively handle the ongoing influx of real-time data, whilst conventional big data analysis necessitates scalability to manage substantial datasets successfully. Both areas exhibit shared issues pertaining to the aspects of data security and privacy. In both digital twin systems and big data analysis, the implementation of techniques such as encryption, access control, and secure data transmission has significant importance. Both digital twin systems and traditional big data analysis might potentially derive advantages from using parallel and distributed computing frameworks to enhance the efficiency of data processing. Both cases can utilize technologies such as Apache Hadoop and Apache Spark. Data streaming and IoT integration play a crucial role in the functioning of digital twin systems, and they also hold significance in specific big data scenarios. Real-time data processing from Internet of Things (IoT) devices is a prevalent practice in both domains, often necessitating the utilization of comparable technologies such as Apache Kafka or MQTT for efficient data streaming. The essential similarity in technology between data analysis in digital twin systems and traditional big data analysis lies in the shared utilization of core data analytics tools and methodologies, data management infrastructure, machine learning algorithms, and data pretreatment approaches. Nevertheless, there exists a distinction in the particular utilization and emphasis of these technologies, as digital twin systems prioritize real-time integration and simulation to generate virtual duplicates of physical systems. At the same time, standard big data analysis is centered around larger aims in data analytics.

2.8. Modern day cloud services and digital twin data analysis

Cloud services are playing an increasingly important role in the way that businesses and organizations collect, store, and manage data. One of the key benefits of cloud services is that they allow organizations to collect and store large amounts of data from a wide range of sources, including sensors, devices, and applications, in a centralized location. This data can then be easily accessed and analyzed by different departments or teams, enabling faster and more informed decision-making. In addition to data collection and storage, cloud services also offer powerful tools for data association, fusion, sorting, and coordination. With the ability to link data from different sources, organizations can better understand their operations and identify patterns and trends that may be absent from individual data sources. Data fusion is particularly important in industrial settings, where organizations may have data from a range of sensors and devices that need to be combined to provide a complete picture of a process or operation. By fusing data from different sources, organizations can gain insights into the performance of individual machines, as well as broader trends across their entire production line. Cloud services also offer powerful tools for data sorting and coordination, enabling organizations to manage and analyze their data in a way that is organized and efficient. This is particularly important for businesses that have large amounts of data, as it allows them to quickly identify and extract the information they need without having to sift through large volumes of irrelevant data. Overall, cloud services are playing a crucial role in helping organizations collect, store, and manage their data in a way that is efficient, effective, and scalable. With powerful tools for data association, fusion, sorting, and coordination, organizations can better make sense of their data and use it to drive innovation and growth. The cloud services offered by contemporary tech giants for data-related purposes such as analysis, storage, and other uses are highlighted in Table 3.

These services are a key enabler of digital twin technology, which is a virtual model of a physical system or process that allows for real-time monitoring, analysis, and optimization. Digital twins are created by collecting and integrating data from a wide range of sensors and devices and using this data to create a digital representation of the physical system. Cloud services offer several key benefits for digital twin technology. First of all, they provide a centralized location for storing and managing the data required to create and maintain digital twins. This allows for easy access to data from different sources and ensures that the data is secure and backed up. Second, cloud services offer powerful tools for data processing and analysis, which are critical for digital twin technology. With the ability to process large volumes of data in real time, cloud services enable organizations to quickly identify patterns and anomalies in the data and use this information to optimize their operations. Third, cloud services enable collaboration and data sharing across different departments and teams. This is particularly important for digital twin technology, as it allows different stakeholders to access and analyze the same data and work together to improve the performance of the physical system. Finally, cloud services offer the scalability and flexibility required to support digital twin technology as it grows and evolves. With the ability to easily add or remove resources as needed, cloud services allow organizations to scale their digital twin capabilities as their needs change over time. Ultimately, cloud services are a critical component of digital twin technology, providing the infrastructure, tools, and scalability required to create and maintain virtual models of complex physical systems.

Table 3
Cloud services provided by the modern tech giants for storage, analysis, and other data-related purposes [105–107].

Tech Giants	Cloud Platform	Launch year	Geographical Regions	Availability Zones	Data Center	Key offerings	Unique Feature
Microsoft	Microsoft Azure	2006	78	164	200	Compute, storage, database, analytics, networking, machine learning, and AI, mobile, developer tools, IoT, security, enterprise applications, blockchain.	Microsoft's Windows operating system Windows and database SQL Server, Microsoft's mixed reality technology (products for HoloLens), Microsoft's TFS and VSTS, Office suite, Sharepoint, and Power BI, Office 365, Microsoft Cognitive Services
Amazon	Amazon Web Service (AWS)	2010	26	84	300	Compute, storage, mobile, data management, messaging, media services, Content Delivery Network (CDN), machine learning and AI, developer tools, security, blockchain, functions, IoT.	Virtual Private Cloud, EC2, AWS Data Transfer, Simple Storage Service, DynamoDB, Elastic Compute Cloud, AWS Key Management Service, AmazonCloudWatch, Simple Notification Service, Relational Database Service, Route 53, Simple Queue Service, CloudTrail, and Simple Email Service.
Google	Google Cloud	2008	34	103	147	Compute, storage, databases, networking, big data, cloud AI, management tools, Identity and security, IoT, Application Programming Interface (API) platform	Healthcare and Life Sciences, Hybrid and Multi-cloud, Management Tools, Media and Gaming, Migration, Networking, Security and Identity, Serverless Computing, G Suite, Google Maps Platform, Google Hardware, Google Identity, Chrome Enterprise, Android Enterprise, Apigee, Firebase, and Orbitera
IBM	IBM Cloud	2011	6	19	60	Cloud Computing, storage, Networking, Cloud Analytics, Artificial Intelligence (AI), Machine Learning, IoT, and Mobile	Integration, Migration, Private Cloud, and Vmware, IBM Watson, Apache Spark & Hadoop services, IBM Weather API
Oracle	Oracle Cloud	2016	39	21	44	Compute, storage, open source databases, data lakehouse, digital media service, application integration, Machine lear-ning and AI, analytics and BI, Containers & function	Governance, Load Balancing, DNS Monitoring, Ravello, and FastConnect, CX, HCM, ERP, SCM, EPM
Alibaba	Alibaba Cloud	2009	24	74	25	Storage, security, Enterprise Applications & Cloud Commu-nication, Analytics, Artificial Intelligence, Media Services, Hybrid Cloud, Container & Middleware, Developer Services, IoT	Elastic Computing, Storage and CDN, Networking, Database Services, Security, Monitoring and Management, Domains and Websites, Analytics and Data Technology, Application Services, Media Services, Middleware, Cloud Communication, Apsara Stack

3. Sector-wise implementation of data analysis for a digital twin

3.1. Manufacturing

First things first, let's get a fundamental understanding of certain categorizations for basic manufacturing data gathered from recent review publications. Volatile and non-volatile data [108]. Information stored on a live network that is discarded when a system is turned off is known as volatile data. Nonvolatile data refers to a sort of digital information that is consistently kept inside a file system on some form of electronic media, and it maintains the condition it was in even after the power is turned off. From

Table 4
Data management and analysis characteristics in manufacturing sector [12,16,102].

Data Collection	Multi-modal data acquisition technology, API (Application Programming Interface), SDK (software development kit), web crawler, RFID, Sensors, gauges, and readers, cameras, scanners	Data Types	Manufacturing Resource, Management data, computer-aided systems data (e.g., CAD, CAE, and CAM), Internet data including social networks and e-commerce platform, product lifecycle data
Data Storage	Cloud storage, data clustering storage technology, MongoDB database, MySQL Database, HTML/JSON	Networking Technology	4G, 5G, NB-IoT, LoRaWAN, Sigfox, Bluetooth, 802.11 ah, 802.11n, ZigBee, Z-Wave, and WirelessHART
Communication Protocol	MQTT, OPC, OPC-UA, TCP/IP Connection, Diablo, Siemens vendor protocol, IIoT, MTConnect, Modbus TCP, RESTful API, UDP & data from ROS, AutomationML	Dataset/Algorithm	Big Data, fuzzy sets, rule-based reasoning, intelligent algorithm, class diagram, XML, UML, KEPWARE, Zigbee
Data Analysis Technology	Value stream mapping, Augmented Reality(AR), Technomatix Plant Simulation, machine learning, forecasting models, virtual-real bidirectional mapping	Modeling Technologies	3D modeling, multi-granularity/scale data planning, interpretable-operable-traceable heterogeneous data fusion, ISO-compliant data model

a structural point of view, data can also be divided into three categories: structured data, semi-structured data, and unstructured data [12]. Structured data consists of columns and rows in a database, whereas information that does not consist of structured data but yet retains some level of structure is referred to as semi-structured data and unstructured data refers to information that has not been arranged in a consistent fashion or doesn't adhere to a standard data model. Manufacturing data can be classified as static property data, real-time data, and measurement data [109]. The basic characteristics of a physical part are referred to as the part's static properties. Examples of static properties include information about machines, cutting tools, workpieces, and the physical surroundings. This portion of the data could stand in for the "physical" component of the Digital Twin. Real-time data refers to data that is collected at various stages of the manufacturing process and made available to use immediately and effectively or for further processing. Measuring data are the measurements acquired from various measurement equipment throughout the production process [109]. For a successful DT application from a data perspective, it must be handled in the following manner. First, raw data needs to be collected and stored. Then, we need to apply a process algorithm to enable data interaction of various data types, formats, and classes. Further, association and fusion technology is applied for efficient data analysis. Finally, iterative process like evolution, servitization, etc., is adapted to ensure further improvement in the analysis process, time, cost, and end-user service [102].

We need to start by acquiring data from a physical world whose volume will be huge by nature. Radiofrequency identification (RFID), sensors, cameras, and scanner hardwires are used for data acquisition at different stages of manufacturing production lines. From a software point of view, Value stream mapping, multimodal data acquisition technology, Augmented Reality (AR), Virtual Reality (VR), and the Internet of Things (IoT) are the current cutting-edge technologies for data acquisition. After that, data needs to be stored in a hierarchical fashion to give a well-structured data description for the purpose of data collection, exploitation, and subsequent use [26]. This technique reduces data loss and damage attributes and offers opportunities to add more data to storage when needed. Cloud storage is primarily the most up-to-date technology for data holding from an economic perspective, as it is one of the main issues for any manufacturing process. Cloud storage is not always suitable from a data security, faster access, and retrieval perspective. Although DT consists of both physical and virtual parts, the physical part is not its strongest section. Rather, it emphasizes the virtual part. That's where big data comes in place. Big data-driven DT-based technology is the smartest way of implementing intelligent and sustainable manufacturing [110]. Big data technology is ideal for retrieving more valuable and comprehensive knowledge from ever-increasing volumes of data, and it is more suitable to apply when the dataset is complex and consists of a number of types, classes, and formats. The information that makes up big data comes from the Internet, information systems, and physical entities; these are all things produced by actions that occur in the real world. The information that makes up a digital twin comes not only from the real world but also from computer simulations. The digital twin requires model data connections in order to collect data from the virtual world, but big data in manufacturing does not require this. Also, big data is two-dimensional and carries a vast amount of knowledge compared to DT, whereas DT data is three-dimensional and dynamic [12]. So, a DT-based manufacturing system is expected to be designed in collaboration with big data to reduce data redundancy, clustering, and blending when data interaction happens. We already know DT is a changing and always-evolving automated system that is supported by the continuous flow of data in both directions (Physical and Virtual entities). To ensure this constant bidirectional flow of information, we need to continuously feed the system with resource and ingredient data, logistics data, concept data, design data, performance data, trial and error data, market data, supply and demand data, customer review, retailer data, distribution, and discharge data. Most of this data relates to product lifecycle management [16], which is essential for successful servitization. The technologies that are available and being used for manufacturing DT development in the current years for overall data handling are 3D scanning, AutomationML, MTConnect, OLE for Process Control (OPC), OPC UA (Unified Architecture), Extensible Markup Language (XML), Kepware, ZigBee, Message Queuing Telemetry Transport (MQTT), MySQL, Application Programming Interface (API), UML/SD, MTComm, Standard Internet Protocol, Industrial Internet of Things (IIoT) Protocol, Wireless Communication, TCP/IP Connection [111,112]. Table 4 provides a synopsis of the entire discussion above.

Table 5
Features of data management and analysis in the urbanization sector [117,114].

Data Collection	Radio-frequency identification (RFID), and image-based techniques, distributed sensor systems, wireless communica-tion, and mobile access, LiDAR scanner, Crowd-sourcing	Data Types	Geographical position data, Data from outdoor surveillance cameras, Data from open sources,
Data Storage	Knowledge Engines (KEs), DynamoDB	Networking Technology	Gigabit Ethernet, EPON, and GPON are used for wired data transmission, 2G (GSM), 3G, LTE, 5G technologies.
Communication Protocol	Modbus and HART, wireless protocols of the field level are Bluetooth; WirelessHART; ZigBee (IEEE 802.15.4); Z-Wave, near-field communication(NFC), MQTT, CoAP, HTTP.	Dataset/ Algorithm	Building information modeling (BIM), building management system (BMS), asset management system (AMS), space management system (SMS), NoSQL database, Natural Language Processing (NLP)
Data Analysis Technology	Data Cleansing Module (DCM), machine learning techniques, artificial general intelligence (AGI), radiological imaging diagnosis, Strategy game, data mining techniques, Crowdsourcing	Vehicle Assistance	Unmanned Aerial Vehicle (UAV), Unmanned maritime vehicle (UMV)

3.2. *Urbanization*

The term "urbanization" describes the enclaving of large human groups. As a result of this concentration, land is being transformed for use in residential, commercial, economic, corporate, industrial, and transportation projects. It can encompass both the peri-urban or suburban outskirts that surround heavily populated regions. So, in order to ensure the performance of the city DTs, effective and hierarchical model/data storing, integration, and query design are the most crucial tasks. Complex and vast volumes of data are gathered, necessitating large-scale data storage and management systems. Here, data/model visualization, cloud computing, and storage may be utilized to handle data in a dynamic and efficient manner at the city and building levels. Different knowledge engines fundamentally power these operations. Domain knowledge is crucial for the development of knowledge engines. Integration of heterogeneous data sources supports effective data querying and analysis, supports decision-making processes in O&M (operation and maintenance) management as well as advances in building information modeling (BIM) is likely to aid the reduction of the time taken for updating databases in operations and maintenance (O&M) phases by 98 percent [113]. Examples of data-gathering methods include contactless data collection, wireless communication, distributed sensor systems, radio-frequency identification (RFID), and image-based methods (e.g., WiFi environment). Building DTs also entails layers for data gathering, transmission, digital modeling, data/model integration, and service layer. Data acquisition layer examples include employing IoT devices, wireless sensor networks, or rapid response (QR) codes (e.g., space utilization and workplace design). This layer could make use of a variety of communication technologies, including access network technologies with short-range coverage (like WiFi, Zigbee, Near Field Communication (NFC), mobile-to-mobile (M2M), and Zwave) and wider coverage (like 3G, 4G, long-term evolution (LTE), 5G, and low-power wide-area networks (LP-WAN)). Important data like Information about the traffic flow of the city, Information about physical parameters (air temperature and humidity, the number of suspended particles and chemical composition of air, noise pollution, radiation level, the chemical composition of water, etc. linked to the geographical position), Data from outdoor surveillance cameras, Data from open sources are need to be collected [114]. Another important aspect of a smart city is the safe deployment of autonomous vehicles. For that, longitudinal and lateral control of autonomous vehicles, such as car following, lane changing, lane keeping, trajectory tracking, and collision avoidance, must be ensured [115]. So, data regarding this manner must be sorted as well.

A crucial element of the digital twin of a city should be the Data Cleansing Module (DCM). The use of artificial intelligence may also assist academics in recognizing trends among the massive amounts of digital data created by cities and infrastructure systems. However, the development of artificial general intelligence (AGI), despite making significant progress on some fronts (such as assisting radiological imaging diagnosis and playing computer strategy games), is still in its infancy and is a long way from being able to solve real-world policy problems as complex as traffic congestion. As data are gathered, processed, evaluated, and utilized to assist decision-making, the amount of the data would drop, but the data value would grow. The progress of data science, especially the methods of machine learning, will complement the ideas that are already in place regarding cities and infrastructure, and together, they will add to the fundamental information that is required for the development of digital twins [116]. Another popular way for gathering and analyzing data during catastrophes is crowdsourcing, which is cost-effective and quick [117]. Data collecting and/or decision-making methods that rely on combining the opinions of many people in order to produce better decisions than would be possible with only the facts at hand are referred to as crowdsourcing. There are a number of ways in which crowdsourcing may be used, not only to gather high-quality catastrophe scenario information but also to develop machine learning algorithms in the Digital Twin framework by supplying annotated photos and social media postings. Table 5 provides a summary of the discussion on this approach.

Table 6
Characteristics of the agriculture sector's data management and analysis techniques [118,120,119,121–125].

Data Collection	IoT, Microsoft Azure database, Raspberry PI, Comprehensive Knowledge Archive Network (CKAN)	Data Types	Real-time/streaming data, Microclimate historical data, environmental historical data, previous climate control strategies and crop treatments data, API data (energy, weather), live sensor data, manual records, training data
Data Storage	phpMyAdmin, MySQL server, Stark data server, CUED server, Google drive, personal computer, MongoDB, MyPHP	Networking Technology	LoRa(Long Range) based Wireless Sensor, Network (WSN), Low-Power Wide-Area Networks (LPWAN)
Communication Protocol	MQTT, ModBus network protocol	Dataset/ Algorithm	MobileNet, UNet models, machine learning algorithm, XGBoost
Data Analysis Technology	Visual Studio, Python, Edge Computing, external simulation, dynamic augmentation, AgScan3D+	Automotive Assistance	Unmanned Ground Vehicle (UGV), drone

3.3. Agriculture

Digital twins and cognitive autonomous systems provide a possible solution to the agriculture sector's increasing resource management and food consumption challenges. It is imperative that agricultural production systems evolve in order to increase output while simultaneously minimizing the amount of resources used [118]. Autonomous intelligent technologies and digital provide a possible solution to agriculture's ever-increasing food demand and resource management challenges. We have already established the fact that data is the bridge between the physical part and the virtual part of a DT system. DT evaluates the various crop treatments and climate control plans it receives from the data layer using both current and historical information. The agricultural and food industries have been influenced by digitalization, which also enables the deployment of technology and sophisticated data processing methods in the agricultural industry [119]. For a successful DT from a data perspective in agricultural applications, we need to gather comprehensive data. Then, we must store them for further processing, like cleaning, evaluation, fusion, and clustering. Various network and communication protocols are implemented for cross-linking of data. The following dataset algorithm and data analysis are implemented to discover the hidden value. It is also noticed that automotive assistance is strongly advised in this sector. The data gathered by connecting devices (the Internet of Things system) were used to inspire the construction of a virtual environment that included decision-making tools and models, which was then used to provide feedback to the physical system [119]. Table 6 includes traits of data management and analysis methods used in the agriculture sector.

3.4. Medical

Several constraints, including ethical and financial barriers, hinder the collection of the necessary data to aid clinical decision-making. It has been shown that the combination of mechanical and statistical models may be useful in assisting with diagnosis, therapy, and assessment of prognosis [126]. Initially, using sophisticated modeling strategies or tools like SysML, Modelica, SolidWorks, 3DMAX, and AutoCAD to create high-precision twin models that coincide with the physical entity should be the first step. These designs should be used to create a digital twin. Next, data linkage should be done using health IoT and mobile network protocols in order to sustain the interaction between physical and virtual items in real-time. This may be accomplished by maintaining a constant connection. The analysis process differs in data service and simulation from other sectors. User activity on social media is analyzed to determine the user's sentiment. To put it another way, social media may be considered a different kind of sensor. The AI-Inference Engine can perform the necessary analysis on the data, which is now accessible [36]. The characteristics of data management and analysis techniques utilized in the medical field are listed in Table 7.

3.5. Robotics

Unlike other sectors, the robotics infrastructure relies greatly on simulation and simulation-based technology. Table 8 provides us with an overview of it. The design and optimization of an assembly station, robot workshop, and sensors may be done considerably more effectively and adaptably with the help of these computer simulations. Enhancing conventional simulation approaches using the testbed-based methodology is highly beneficial. In reality, the amount of processing power has significantly expanded, and it is now easily accessible in all regions of the planet. Because of this, the simulation method is now capable of solving issues that are more complicated, integrating the concepts of many designs, and precisely predicting the mechanisms. The term "Digital Twin" (DT) is often used to refer to these deep-level simulations, which correlate to the system that they modeled [128]. Game engines are now an emerging and the most efficient technology to integrate with current state-of-the-art robotic technology to make a successful DT system [129]. Computer programming is also important for data analysis in this field of research, unlike other sectors, as a simulation-based approach is vital for designing a robotic digital twin system.

Table 7
Data characteristics used in medical sector [126,127,36].

Data Collection	Physical people, medical equipments, smart wearable devices, GPS, gyroscope, accelerometer, camera, microphone, light, proximity, heart rate, Social networks	Data Types	Mobile health monitor data, omics, clinical reports, clinical & experimental records, medical images, medical examination results, medical records after diagnosis in medical institutions
Data Storage	CloudDTH, PTB Diagnostic ECG Database	Modeling Technology	SysML, Modelica, SolidWorks, 3DMAX and AutoCAD, Arduino and Raspberry Pi
Communication Protocol	IoT and mobile internet technologies, Bluetooth, USB	Data Service	data mining service, real-time monitoring service and medication reminder service
Data Analysis Technology	Unsupervised machine learning, Deep learning neural network, Kalman filter, AI-Inference Engine	Simulations	Patient-specific electromechanical computer simulations, myofiber mechanics simulations, Automatic cardiac MR segmentation

Table 8
Data characteristics commonly found in robotics sector [129,128,130,131].

Data Collection	HTC Vive VR, Versatile Simulation Database–VSD, PLCs, controllers, 3D reconstruction scanner, Sick microScan 3 Core scanner, three-dimensional sensor data (point clouds), FUNK_Sampling_Pointcloud	Data Types	Sensor and image data, Simulation and result data, component and layout data, training data, historical data, geometry data
Designing Technology	CAD software, such as SolidWorks or CATIA, graphics software like Blender/Maya 3D, Raspberry Pi 3 B+, Siemens S7-1200 PLC,	Programming/ Simulation	KAREL programming language, GAZEBO and V-REP, ROBCAD from SIMSOL, MATLAB /Simulink, Virtual Environment and Robotic Simulation–VEROSIM, robot operating system (ROS), Human Industrial Robot Interaction Tool (HIRIT)
Modeling Technology	Unity 3D, ROBOTRAN	Dataset/ Algorithm	Forward Kinematics (FK), BioIK, testbed-based algorithm, peg-in-hole insertion algorithm, three-dimensional object recognition algorithm, agglomerative hierarchical clustering
Data Analysis Technology	Virtual Reality, Augmented Reality, KUKA LWR4, RoboDK, Artificial Neural Networks (ANN), Density Based Spatial Clustering Analysis with Noise (DBSCAN)	Communication Protocol	Modbus TCP/IP, Dot Net-based API, JOpenShowVar, FeedForward Network (FFN)

3.6. Military/aviation

Also, in the aerospace industry, simulations imitate the continuous time history of flights, yielding an enormous amount of information on simulations to recognize what the aircraft has been through and project upcoming serviceability and infringements using a range of features-based simulation techniques. This is done in order to acknowledge what simulations are used [128]. But still, modeling technologies are of prime utility in designing defense infrastructure worldwide with the integration of digital twin systems. The components that make up contemporary airplanes are getting more complicated to put together. An airplane is made up of a large number of different components, each of which has its own set of characteristics. When these components are coupled, additional characteristics may be derived from their interaction. The dynamic properties are powerful, and the condition of the components changes notably and quickly over time. The flying environment of the aircraft is unclear, and the chance of complex systems incurring inadvertent damage in an inherently unpredictable environment grows, leaving the aircraft more susceptible to harm. Maintaining the dependability of a system is not a simple task. Routine maintenance- There is a lack of accurate estimation of the current state of a complex system, which makes it prone to too frequent inspection and maintenance or premature failure of the system due to untimely maintenance, leading to high repair costs and inadequate durability of the aircraft. When dealing with a complex system, there needs to be a more accurate evaluation of the current state of the system [132]. Table 9 provides a comprehensive overview of military/aviation data features.

3.7. Categorization of data analysis phase

The utilization of data analysis from digital twin technology in diverse systems may be classified into two main stages: system design and system operation. Digital twins are of paramount importance in both phases as they significantly contribute to the improvement of efficiency, process optimization, and the practicality of physical solutions. Table 10 presents a concise summary of the research findings. Now, let us dig into an in-depth review of the technologies, processes, and/or methods employed inside various sectors.

Table 9
Military/Aviation data characteristics [34,128,133,132].

Data Generation	Satellite, Condition Based Maintenance Plus Structural Integrity (CBM + SI), sensor, signal and equipment operation	Data Types	Fleet data, Prognostics and Health Management data, aero-engine data, civil aviation domain knowledge, Machine data, prototype data, mechanical structure data, historical operation data, detailed spacecraft design data, spacecraft in-orbit control data, spacecraft manufacturing data
Data Simulation	Computational Fluid Dynamics (CFD), Computer-Aided Engineering (CAE), Finite Element Methods (FEM) and Monte Carlo simulation, Structural Health Monitoring (SHM), Damage and Durability Simulator (DDSim)	Data Architecture	CPS architecture, cloud-based CPS (C2PS), Dynamic Data Driven Application System (DDDAS), Iso-Geometric Analysis (IGA),
Software Support	Siemens Product Lifecycle Management (PLM) software, Teamcenter® portfolio, NX ™ software, Simcenter ™ solution, and Tecnomatix® portfolio, ABAQUS	Dataset/ Algorithm	Fault isolation algorithm, industrial Digital Mock-Up (iDMU) dataset, Savitzky-Golay filter algorithm, NASA's turbofan aero-engine simulation dataset
Data Analysis Technology	Machine learning and predictive models, IoT, blockchain, artificial intelligence, and 5G, fatigue mechanics, Extended Kalman Particle Filters (EKF), deep learning methods, comprehensive probabilistic damage tolerance analysis	Modeling Technology	Airframe Digital Twin (ADT), numerical simulation-based DT models, Multi-dimensional modeling, virtual environment mapping, structural conceptual model, long short-term memory (LSTM) neural network model, Model-Based Systems Engineering (MBSE)

Table 10
Phases devoted to data analysis.

	System Design Phase	System Operation Phase
Manufaturing	•Product and Process Simulation •Quality Control	•Real-time Process Monitoring •Predictive Maintenance
Urbanization/ Smart Cities	•Population and Infrastructure Simulation •Environmental Impact Assessment	•Real-time Traffic and Infrastructure Monitoring •Emergency Response and Disaster Management
Agriculture	•Crop and Soil Simulation •Water Resource Management	•Real-time Monitoring of Crop Health •Precision Farming
Healthcare/ Medical	•Patient Data Analysis •Drug Discovery	•Real-time Monitoring •Telemedicine
Robotics	•Simulation and Modeling •Kinematics and Dynamics Analysis	•Sensor Data Processing •Path Planning and Collision Avoidance
Military/ Aviation	•Aircraft and Weapon Simulation •Cybersecurity	•Real-time Monitoring •Mission Planning and Execution

Manufacturing: Technologies such as finite element analysis (FEA) and computational fluid dynamics (CFD) are employed in the simulation of product designs and production processes [134,135]. Data-driven models utilize previous data in order to optimize designs. Statistical process control (SPC) methodologies are utilized in the context of quality control [136]. Methods such as Six Sigma and Lean manufacturing employ past quality data to facilitate design enhancements [137]. The utilization of Industrial Internet of Things (IIoT) technology facilitates the collecting of sensor data in real-time [138]. Data analytics techniques, including anomaly detection, regression analysis, and machine learning, are employed in the realm of monitoring and process optimization [139,140]. The implementation of predictive maintenance is contingent upon the utilization of data analytics and machine learning techniques. Condition-based monitoring, reliability analysis, and predictive algorithms leverage sensor data to anticipate occurrences of equipment failures.

Urbanization/Smart Cities: Geographic Information Systems (GIS) play a crucial role in the simulation of population and infrastructure [141]. Spatial analytics and scenario modeling are two prominent data analysis methodologies [142,143]. Environmental modeling and simulation technologies integrate historical climate and pollutant data. Environmental impact assessment models are among the various methodologies utilized for data analysis [144]. Urban traffic management systems utilize real-time data obtained from a variety of sources, including cameras, sensors, and GPS technology. Data analytics encompasses several techniques and methodologies for the analysis of traffic flow, prediction of congestion, and implementation of adaptive traffic control strategies. Geographic data analytics and real-time data integration technologies are employed in the context of emergency response. Some examples of these tools and platforms encompass spatial analysis tools and crisis mapping platforms.

Agriculture: Crop modeling software utilizes historical data on crops and soil to make predictions and simulations [145]. Data analysis techniques commonly employed in the field of crop production optimization encompass statistical analysis, regression analysis, and predictive modeling. Water management software utilizes previous data to facilitate the process of irrigation planning [146]. Methods such as hydrological modeling and irrigation scheduling heavily depend on the examination of data [147]. Remote sensing technology and agricultural sensors offer the capability to acquire and analyze real-time data. Data analytics encompasses several applications, such as image processing, machine learning techniques for disease identification, and crop health assessment. The implementation of precision agriculture is contingent upon the utilization of sensor data and Global Positioning System (GPS) technology. Data analysis approaches commonly employed in precision agricultural procedures encompass data fusion, geostatistics, and variable-rate technologies.

Healthcare/Medical: Clinical decision support systems are designed using historical patient data [148]. The methodologies employed for data analysis encompass patient risk stratification, predictive modeling, and the study of electronic health record (EHR) data. Data analysis approaches such as bioinformatics and cheminformatics play a crucial role in the field of drug discovery. The utilization of molecular modeling and data mining techniques facilitates the identification of prospective therapeutic candidates. Patient monitoring devices and wearables have the capability to collect and transmit health data in real time. The procedures of data analytics encompass the analysis of vital signs, the discovery of trends, and the implementation of early warning systems. Telehealth platforms leverage data analytics to facilitate remote patient consultations [149]. Diagnostic algorithms and image analysis techniques are employed in the field of remote healthcare to facilitate the process of diagnosing medical conditions and providing treatment recommendations.

Robotics: Using software such as ROS (Robot Operating System) is prevalent in the simulation of robot behaviors and settings [150]. The process of data analysis entails the comparison of simulated data with actual data in order to validate and enhance the accuracy and reliability of the results. Analytical methodologies and algorithms are employed to forecast the behavior of robots by leveraging their design and physical attributes. Robotic systems have the capability to produce substantial quantities of sensor data. Real-time data interpretation employs many data analytics approaches, such as signal processing, computer vision, and machine learning. Algorithms and data analysis techniques are utilized to improve the trajectories of robots, considering real-time sensor data in order to avoid collisions effectively.

Military/Aviation: Aircraft and weapon system simulations employ both historical and design data. Data analysis approaches encompass many methods to analyze and simulate aircraft performance, stress analysis, and the effectiveness of the weapon system. These techniques often entail the utilization of computational fluid dynamics (CFD) simulations [151], finite element analysis (FEA) [152], and Monte Carlo simulations [153,154]. The utilization of data analytics and machine learning techniques is prevalent in the domain of cybersecurity threat analysis. The utilization of anomaly detection and behavior analysis, in conjunction with machine learning algorithms and pattern recognition techniques, aids in the identification of possible risks. Sensors and real-time data are key components utilized in military and aviation systems. Data analysis encompasses various applications in the aerospace and defense sectors, such as the identification of potential threats, aiding in decision-making processes, and facilitating predictive maintenance for aircraft and defense systems. The analysis of mission data in real time is conducted to inform decision-making processes. The procedures of data analytics encompass the optimization of mission planning, the planning of routes, and the provision of tactical decision support.

4. Present challenges, issues, security concerns, potential solutions, recommendations pertaining to data processing, and future research direction

4.1. Data processing difficulties and future development recommendations

Data processing in various businesses might provide distinct obstacles and necessitate specific requirements due to their unique needs and characteristics. We made an effort to compile them in Table 11 and make recommendations based on them. Digital twins in the manufacturing industry facilitate the generation of substantial volumes of real-time data derived from sensors, machinery, and production lines [155]. The management and analysis of this data with a high rate of velocity can provide significant difficulties. The integration of data from many sources, including Internet of Things (IoT) devices and legacy systems, might present inherent complexities [156–158]. The presence of data integration challenges can impede the achievement of a cohesive perspective on the manufacturing process. The assurance of data quality and accuracy from various sensors and devices is of utmost importance. The presence of inaccurate data has the potential to result in faulty simulations and subsequent decision-making. The implementation of predictive maintenance models that leverage both historical and real-time data is crucial. This encompasses difficulties pertaining to the identification of anomalies and the recognition of patterns. The objective is to create and use sophisticated analytics and machine learning models to facilitate predictive maintenance. It is best to allocate resources towards the implementation of anomaly detection algorithms in order to proactively identify and address potential issues before they manifest so that system downtime is taken into account. Developing real-time data platforms is essential for effectively managing the high-velocity data produced by sensors and devices. One possible approach to address the task at hand is to incorporate data streaming and processing tools such as Apache Kafka or Azure Stream Analytics. It is advisable to allocate resources toward the acquisition of data integration solutions that possess the capability to establish smooth connections among legacy systems, Internet of Things (IoT) devices, and contemporary data sources. The integration of middleware and Internet of Things (IoT) platforms can be employed to establish connections between disparate data silos. The implementation of data quality frameworks is essential for the purpose of monitoring and enhancing the accuracy of data [159]. The implementation of data validation standards and data cleansing processes is crucial in ensuring the accuracy and integrity of data.

Urbanization digital twins encompass a wide range of data kinds, such as traffic data, environmental sensor data, and social data from multiple sources. The process of integrating and analyzing this diverse range of data can present inherent complexities. Smart city initiatives can encompass expansive geographical regions, leading to the accumulation of substantial volumes of data. There is a

Table 11
Data processing challenges and suggestions for future development.

	Problem faced while processing data	Suggestions for the next step of development
Manufaturing	•Data Volume and Velocity •Data Integration •Data Quality •Predictive Maintenance	•Implement Advanced Analytics •Real-Time Data Platforms •Data Integration Solutions •Data Quality Assurance
Urbanization/ Smart Cities	•Data Variety •Scalability •Privacy and Security •Real-Time Analysis	•Scalable Infrastructure •Robust Privacy and Security •Leverage Real-Time Analytics Platforms •Data Visualization
Agriculture	•Data Fusion •Data Interoperability •Data Accuracy •Data Visualization	•IoT Integration •Data Standardization •Data Accuracy •Farmers' Education
Healthcare/ Medical	•Data Security and Privacy •Data Integration •Data Ethics •Clinical Validation	•Strengthen Data Security •Interoperability •Data Ethics and Governance •Clinical Validation
Robotics	•Simulation Realism •Data Noise •Interoperability	•Simulation Fidelity •Data Noise Reduction •Hardware and Software Integration
Military/ Aviation	•Security •Data Accuracy •Legacy Systems	•Cybersecurity •Calibration and Validation •Legacy Systems Integration

need for data processing and storage systems that are capable of scaling. The management of sensitive data derived by surveillance cameras and Internet of Things (IoT) devices while simultaneously assuring the preservation of privacy and security is a significant area of apprehension. Numerous intelligent urban applications necessitate the utilization of real-time data analysis to effectively carry out duties such as traffic control, emergency response, and resource optimization. The objective is to design and implement a scalable infrastructure that can effectively manage the substantial volume of data produced by sensors in smart cities, encompassing storage and processing capabilities [160]. The examination of cloud-based options for achieving elasticity is warranted. To ensure the safeguarding of sensitive data, it is imperative to incorporate rigorous privacy and security protocols, such as encryption and access controls. It is imperative to adhere to data protection standards in order to maintain compliance. Utilize real-time analytics technologies to expedite decision-making processes in domains such as traffic management and public safety. The implementation of edge computing is proposed as a means to achieve low-latency data processing. Develop user-friendly data visualization dashboards and tools to facilitate data interpretation and use by city authorities and residents.

Integrating data from various sources, such as Internet of Things (IoT) sensors, satellite imaging, and weather forecasts, is a significant challenge for the agricultural sector. The integration of data in an efficient manner is of utmost importance in facilitating well-informed decision-making. Data can be sourced from various devices and technologies. The task of ensuring interoperability and maintaining consistent data formats may pose challenges. The presence of erroneous data has the potential to result in suboptimal decision-making within the field of precision agriculture. The maintenance of data accuracy, particularly in remote and field settings, is of utmost importance. User-friendly data visualization tools are typically necessary for farmers to interpret complicated data in agriculture digital twins. This study aims to explore the potential for further integration and use of Internet of Things (IoT) sensors and devices within the field of precision agriculture. The objective is to achieve a smooth and efficient amalgamation of data derived from many sources, including soil sensors, drones, and weather stations. This study aims to create data standardization protocols that can effectively establish uniform data formats and promote interoperability across diverse agricultural technology [161]. To boost the accuracy of data, it is advisable to make investments in sensor technology that is both accurate and reliable, as well as in data validation methods. The primary objective is to impart knowledge to farmers regarding digital twin technologies and the utilization of data-driven decision-making processes, with the ultimate aim of facilitating the extensive adoption of these practices.

Ensuring the security and confidentiality of patient data is of utmost importance. The task of maintaining security and privacy compliance during the data processing process poses a substantial difficulty. Healthcare systems frequently employ a diverse range of both legacy and contemporary technology, which may provide challenges in terms of interoperability and data sharing. The integration of electronic health records is a prevalent challenge. The processing of healthcare data necessitates the careful evaluation of ethical factors, including but not limited to data ownership, consent, and responsible utilization. It is imperative to thoroughly evaluate the accuracy and clinical validity of the data utilized for the purposes of diagnosis and treatment. Enhance the robustness of data security through the implementation of sophisticated encryption techniques, comprehensive identification and access control protocols, and strict adherence to healthcare legislation such as the Health Insurance Portability and Accountability Act (HIPAA) [162]. Enhance healthcare data interoperability by leveraging HL7 FHIR standards and open application programming interfaces (APIs) to provide a seamless exchange of data among various systems [163]. Develop robust data ethics and governance policies encompassing a wide range of considerations, including permission, data ownership, and appropriate utilization of patient data. It is imperative to collaborate with healthcare professionals to guarantee clinical validation and adherence to stringent criteria for diagnosis and treatment of digital twin data.

When it comes to the field of robotics, the task of attaining a high level of accuracy in simulations that accurately depict real-life situations is a complex endeavor that necessitates the utilization of sophisticated modeling techniques and substantial computational capabilities. It is imperative to ensure that simulated data appropriately represents the real-world environment. The mitigation of noise in data derived from simulations poses a significant challenge. Robotics systems are comprised of many hardware and software components. A difficulty often encountered involves the need to provide seamless collaboration and data sharing among different entities. Further advancements in simulation realism can be achieved through the allocation of resources toward the development and implementation of high-fidelity modeling and simulation technologies. This includes the utilization of gaming engines to create immersive and authentic settings. The objective is to incorporate noise reduction methodologies into the process of generating simulated data to achieve a high level of fidelity in virtual sensors that accurately emulate real-world behavior [164,15]. The objective is to optimize interoperability among diverse robotics hardware and software components to streamline the process of data sharing and integration.

The preservation of the confidentiality and integrity of critical military and aviation data is of utmost importance. The perpetual difficulty lies in safeguarding against cyber threats and mitigating the risk of data breaches [165]. The acquisition of high-precision data is of utmost importance in ensuring the accuracy of simulations and predictions. The calibration and validation processes play a crucial role in ensuring the correctness of data. The process of incorporating digital twin technology into existing systems within the military and aviation domains can present intricate challenges, necessitating the establishment of compatibility and the seamless transfer of data. To safeguard critical military and aviation data, it is imperative to give utmost importance to cybersecurity by implementing advanced threat detection mechanisms, intrusion prevention systems, and secure communication protocols. To ensure the correctness of data in simulations and real-time systems, it is imperative to establish and adhere to meticulous calibration and validation procedures. This study aims to devise effective techniques and advanced technologies to facilitate the smooth integration of new systems with existing legacy systems. The primary focus is ensuring compatibility and seamless data migration between the two systems.

4.2. Blockchain technology in conjunction with federated learning techniques to address security issues

Blockchain is a decentralized and transparent ledger technology that serves as the foundation for cryptocurrencies such as Bitcoin [166]. However, it has a diverse array of uses that extend beyond the realm of digital currency. The system may be described as a decentralized and tamper-resistant mechanism that is designed to record and authenticate transactions. In contrast, Federated learning is a method in the field of machine learning that aims to enhance privacy by enabling model training to be conducted on distributed devices or servers, hence ensuring the localization and privacy of data [167]. This technology facilitates the collaborative training of models while avoiding the concentration of sensitive data in a single area. The integration of blockchain and federated learning methodologies has the potential to bolster data security across digital twin systems across diverse industries. The digital twin systems used in many industries utilize a mix of blockchain and federated learning technologies to safeguard data, uphold data privacy, and facilitate collaborative analysis and model enhancement [168]. Each of the aforementioned technologies has distinct strengths and applications, and the selection of a certain technology is contingent upon the needs and limitations of the digital twin system within a given sector. Fig. 6 illustrates the potential use of a comprehensive security system that combines a generalized blockchain and federated learning approach across several industries.

Manufacturing: The first step involves the acquisition of data from many sources inside a manufacturing plant, including sensors, Internet of Things (IoT) devices, and digital twins. Leverage blockchain technology for the purpose of establishing a decentralized and tamper-resistant ledger. Every data point is documented and stored as a transaction or block on the blockchain. Cryptographic hashing is used as a means to ensure the integrity of data [169]. Hyperledger Fabric has tremendous potential as a suitable choice for implementing digital twin systems in the manufacturing industry [170]. Permissioned networks may be established to facilitate supply chain management, monitor the origin of products, and preserve the integrity of data [171]. The data collected by sensors is assigned a timestamp and then appended to the blockchain. The objective is to deploy intelligent contracts on the blockchain. Smart contracts have the capability to autonomously carry out predetermined actions in response to certain data triggers. The ability to manage access rights and regulate data sharing is under their hands. The implementation of Hyperledger Fabric entails the establishment of a consortium network, the development of smart contracts for the purpose of product tracking, and the establishment of consensus procedures. This technology may be used by manufacturers to monitor and trace the whole life cycle of their goods. The use of blockchain technology for access control enables the restriction of data access to only authorized entities. In order to get access to data, it is important for users to provide digital signatures. All instances of unauthorized access attempts are recorded and stored on the blockchain. The implementation of security monitoring systems is essential for the detection of anomalous actions. All instances of irregularities or breaches in security are documented and stored on the blockchain. It is advisable to use encryption techniques to safeguard confidential information prior to its inclusion in the blockchain. The management of decryption keys is rigorously regulated by smart contracts.

Urbanization: Collect data from diverse urban sensors, surveillance systems, and other relevant sources. Incorporate blockchain technology in order to establish a ledger that is resistant to alteration. One potential use is the utilization of blockchain technology to provide safe and verifiable timestamps for data. Utilize blockchain technology for the purposes of access control and identity management. Ethereum demonstrates suitability for the implementation of smart city applications, particularly within areas like

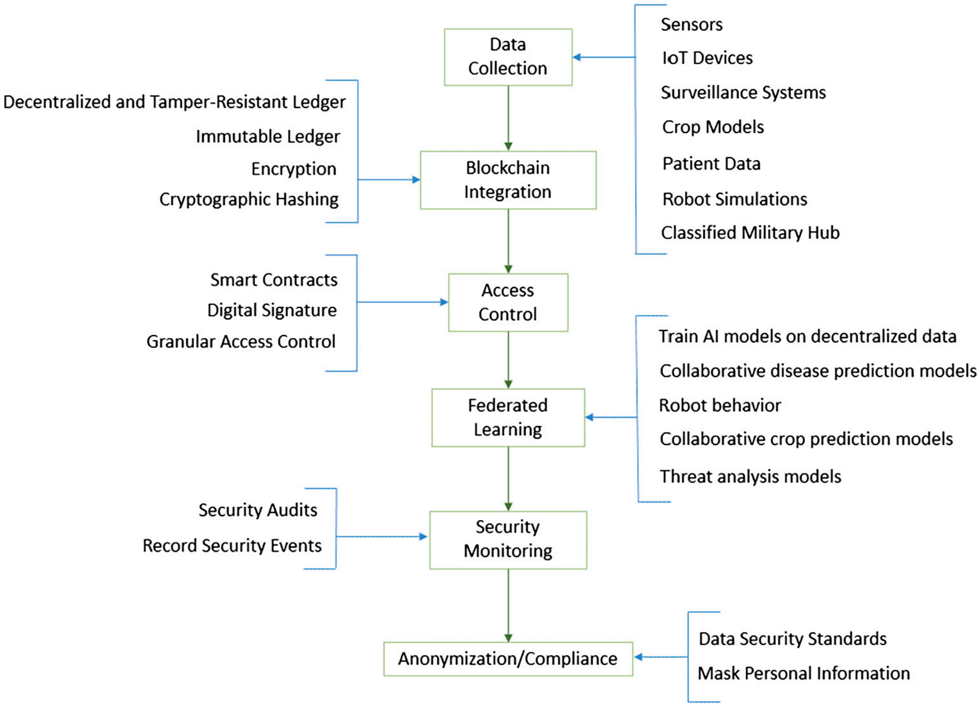

Fig. 6. Blockchain-federated learning security concept for digital twins.

transportation and energy. The system has the capability to process transactions pertaining to smart contracts as well as public services [172]. Smart contracts establish and delineate the permissions pertaining to data access. Utilise federated learning methodologies for the purpose of training artificial intelligence models using data that is distributed in a decentralized manner. TensorFlow is an open-source machine learning framework developed by Google. It can be used for Federated learning in the context of smart cities, enabling the use of the Federated approach for various applications such as traffic optimization and environmental monitoring [173]. The data stays stored on the local devices, while only the changes to the model are exchanged. Incorporate methodologies for the anonymization of personal data. It is essential to adhere to privacy rules in order to maintain compliance. The process of documenting security audits on the blockchain is proposed. It is essential to establish and preserve a comprehensive record of data access and sharing activities. The objective is to design and implement intelligent contracts for applications such as automated traffic management while also using federated learning techniques for data analysis on distributed edge devices.

Agriculture: The safeguarding of data generated by agricultural sensors and crop models is of paramount importance. The use of blockchain technology may be employed to facilitate the recording of data obtained from diverse agrarian sensors. The use of encryption is crucial in ensuring the security of data. Smart contracts provide the parameters governing the authorization of individuals to access agricultural data [174,175]. Data retrieval is restricted to those who have been granted authorization. Corda can potentially be used within the agricultural sector, establishing robust and safeguarded supply chain networks [176]. The use of traceability measures in the agricultural sector can effectively guarantee the capacity to track and verify the origin and movement of agricultural goods. The platform in question is a blockchain-based solution designed specifically for enterprises. This technology enables the safe exchange of data and execution of transactions across various entities. Every node inside the network retains its own copy of data, guaranteeing privacy. Instead of providing raw data, it is recommended to provide updates on the model. To maintain compliance with privacy regulations, it is necessary to anonymize sensitive data. Utilise strategies to obfuscate personal data. Federated learning methodologies, such as the use of PySyft, have the potential to safeguard confidential agricultural data while concurrently enhancing the accuracy and efficacy of crop prediction models. PySyft is a publicly accessible framework designed to preserve privacy in the field of machine learning [177]. Federated learning is used to facilitate the training of models using decentralized data, including the crucial aspect of preserving data on local devices. Secure multi-party computation is applied.

Medical: One effective strategy for safeguarding patient data inside digital twins is the use of encryption techniques. Access is restricted to healthcare practitioners who have been granted authorization. So, a blockchain-based ledger system can be developed to facilitate the storage and management of healthcare data [178]. It is essential to meticulously document all instances of data exchanges, including comprehensive records of those who have had access to the data. Utilize smart contracts to provide fine-grained access restrictions. Individuals have the ability to authorize or withdraw permission for others to access their personal data. Ethereum has the potential to be used as a secure platform for the storing of healthcare records and clinical data, therefore guaranteeing the confidentiality of patients' information. The platform is characterized as a public blockchain with the capability of executing smart contracts, hence ensuring data security via its decentralized network of nodes [179]. Miners are responsible for validating transactions and then incorporating them into the blockchain. Differential privacy approaches and homomorphic encryption play a crucial role in ensuring privacy preservation in the context of federated learning within the healthcare domain. Homomorphic encryption is a cryptographic technique that enables the execution of calculations on encrypted data while preserving the confidentiality of the underlying data. Federated learning employs this technique to ensure data confidentiality throughout the process of sharing calculations [180]. Local hospitals engage in the practice of training models using patient data while maintaining strict confidentiality and refraining from disclosing such data. This proposal suggests the implementation of Ethereum-based systems to store electronic health records (EHRs) and integrate federated learning algorithms with privacy-preserving measures to provide collaborative illness prediction and diagnosis. To maintain compliance, it is essential to adhere to healthcare data rules, such as the Health Insurance Portability and Accountability Act (HIPAA).

Robotics: The acquisition and storage of data obtained from robot simulations and sensors in a secure manner. The use of blockchain technology may be employed for the purpose of recording simulation data and robot telemetry [181,182]. The use of encryption and cryptographic hashing techniques is recommended. The purpose of this inquiry is to provide a comprehensive definition of access rights pertaining to robot data via the use of smart contracts. Hashgraph has the potential to provide robust security measures for safeguarding data and facilitating transactions inside collaborative robotics systems [183]. This tool is well-suited to preserve a comprehensive record of tasks and interactions. Hashgraph employs a directed acyclic graph (DAG) framework to ensure data security. The system uses a consensus mechanism in order to authenticate and arrange transactions. The direct communication between all nodes contributes to the enhancement of security. It is important to restrict the sharing of data only to authorized institutions. This study aims to use federated learning techniques to enhance the behavioral performance of robots. The use of secure multi-party computation (MPC) is of utmost importance in the context of aggregating model updates while simultaneously ensuring the preservation of data privacy for robots [184]. The concept of secure multi-party computing (MPC) enables many entities to engage in a collaborative calculation process while ensuring the confidentiality of their own data. The technique is used in federated learning to securely combine model updates while maintaining data privacy by avoiding the exchange of raw data. This study proposes the integration of Hashgraph technology into multi-robot systems to document interactions. Additionally, it suggests the use of federated learning with secure multi-party computation (MPC) to enhance robot performance while ensuring the privacy of sensitive data. The surveillance of robotic operations to detect potential security risks is crucial. Documenting security incidents on the blockchain is a recommended approach.

Military/Aviation: The safeguarding of confidential military and aviation data inside digital twins is of paramount importance. The use of blockchain technology is recommended for the purpose of documenting alterations and managing accessibility to data. It is essential to use encryption techniques to safeguard classified military information prior to its storage. Smart contracts establish stringent access limitations [185]. Capture and document all instances of data access activities. Hyperledger Fabric has the potential to be used across the military and aviation sectors to provide secure data storage and ensure data integrity. This technology is well-suited for the management and security of mission data. Federated learning may be employed in the development of threat analysis models. Disseminate model upgrades across diverse military groups. Differential privacy and homomorphic encryption play a crucial role in safeguarding sensitive mission data inside federated learning systems. Hyperledger Fabric can be used to establish robust ledgers for mission data, ensuring enhanced security in military and aviation domains. Furthermore, the integration of federated learning algorithms, along with privacy-preserving mechanisms, enables secure data analysis. It is essential to guarantee adherence to defense data security requirements.

4.3. Comparison of data management and analysis strategies across different sectors

Rather than focusing on finding applications for the digital twin, the key difficulty is figuring out how to get the data required to transform it into a truly global platform. Typically, private corporations have ownership of the data. Moreover, there are situations when the information may be stored at the federal level. In many cases, there is either no real-time data accessible or data is only accessible from a small number of sensors. The next problem is figuring out whether the data we obtained is in a useful format and can be readily incorporated into the centralized database of the digital twin. In addition, collaboration and joint innovation are essential components in the process of making a digital twin operational on a national and regional level and transforming it into a practical instrument for use by public authorities.

Data processing methods for individual research fields are the most challenging part of the data analysis process on DT. So, our first and most crucial challenge will be picking out the right approach for data analysis for an individual sector in terms of collection, storing, association, fusion, coordination, etc. The sources of the digital twin system are quite small. This system has yet to spread among the people. So, the work done using digital twins is not so vast. In the mentioned terms, there will always be a challenge regarding process technology individually. For example, there is no clear indication of how we will increase our storage system for ever-expanding data sets. Both virtual memory and physical memory space are going to be very big issues for the constantly increasing data sets. Cloud storage could be a solution, but cyber security, access problems, transmission speed, etc., come to the point. But in a way, we can eradicate the existing problems of one process method through another by integrating them. For that, we need to arrange to cross-link. As a result, the whole process might become extremely complex, and the repetitiveness of data sets might happen, which will further increase the storage space problem. In another example, sorting large data sets into indefinitely extending data sets is a challenge for data analysis. However, the main challenge will be to integrate different data sorting algorithms and establish a clear and transparent connection between different architectures successfully and in the most efficient way. In the infusion process, the major meaning of data might get modified, giving the system a wrong message. As a result, the purpose of DT might be hampered.

Table 12
A comparative evaluation of the data management and data analysis paradigms across several sectors.

	Manufacturing	Urbanization	Agriculture	Medical	Robotics	Military/ Aviation
Integrating Technology	•Cyber Physical System (CPS) •Big Data	•Building Infrastructure Management (BIM) •Big Data	•Metadata •Big Data	•Metadata •Big Data	•Unity3D •Robot Operating System (ROS)	•Model-Based Systems Engineering (MBSE)
Degree of Programming/ Simulation	Large	Less	Less	Medium	A lot	A lot
Automotive Assistance	Less needed	Needed a lot	Needed a lot	Not Needed	Not Needed	Less needed
Degree of Modeling Technology Requirement	Less	Medium	Medium	Large	A lot	A lot
Degree of networking and Inter-communication among physical & virtual entity	A lot	A lot	Medium	Less	Medium	Medium
Dataset Size	Very large	Very large	Large	Less	Medium	Large
Crowdsourcing & Social Networking	Medium	Needed a lot	Medium	Needed a lot	Less	Less
State of the Art Data Collection Technology	Multi-modal data acquisition technology, RFID	LiDAR scanner, Crowd-sourcing, RFID, Satellite, Social Network	Comprehensive Knowledge Archive Network (CKAN)	Smart Wearable Devices, Personal log, Social Network	3D reconstruction scanner, Sick microScan3 Core scanner	Satellite, Data provided by simulation software

In the servitization process, we need to facilitate human-machine relations at the consumer level. For that, we need to ensure digital twin-based knowledge from primary to at least intermediate difficulty among the mass population who will be the end-users or consumers of the DT service. And it will be a huge challenge for DT analysis. Finally, all the challenges discussed will differ from the solution point of view for individual research. For example, manufacturing data storage problems will contain different techniques for storage problems in medical data. We have to find this dissimilarity, a challenge, and propose an individual approach to solving it.

If we analyze Table 12, we can say that big data is the state-of-the-art technology to develop a digital twin system. Incorporating data analytics into the IoT's rich digital twin ecosystem offers an important digital twin platform and infrastructure for a wide range of use cases, including invasive surgical diagnosis in healthcare, fault detection, diagnosis, analysis, as well as forecasting in manufacturing, and intelligent transport systems in smart city technologies, to name just a few. Although sector-wise, there are some individual technologies that support the development within a specific field; big data is the common ground for every sector. From a programming and modeling point of view, robotics, military, and aviation sectors rely heavily on software and simulation technology, whereas the other sections don't as much. Smart devices, Internet of Things (IoT) gadgets, Cyber networks, Big data, and Blockchain are some of the fundamental technologies employed in the twin technology. Crowd-sourcing and social media networks are emerging technologies in manufacturing, urbanization, and medical sectors that are very easy to apply. However, sectors like robotics, military, and aviation don't encourage the application of such technology as there are huge security issues concerning data breaches. Often, projects related to robotics, the military, and the aviation sector contain classified information while developing them. Automated systems that can be attacked from a distance, IoT gadgets, and cloud platforms are all prime targets for hackers.

4.4. Benefit to the community and where to go with future studies

The many contributions of this research have the potential to generate significant beneficial transformations in society. Enhancing the comprehension of Digital Twins and data analytics provides enterprises with the expertise necessary to enhance efficiency and foster innovation in many sectors, such as manufacturing, urban planning, agriculture, healthcare, and the military. These improvements are based on responsible data practices that emphasize ethical considerations and privacy concerns, therefore cultivating a culture that places high importance on maintaining the accuracy and security of data. Furthermore, the analysis places significant emphasis on user-centric technology, aiming to enhance the accessibility of digital systems and empower decision-makers via the provision of well-informed insights. It facilitates the enhancement of workforce development and encourages cooperation across different disciplines, hence propelling advancements and cultivating a society that is more competent and resilient. The primary

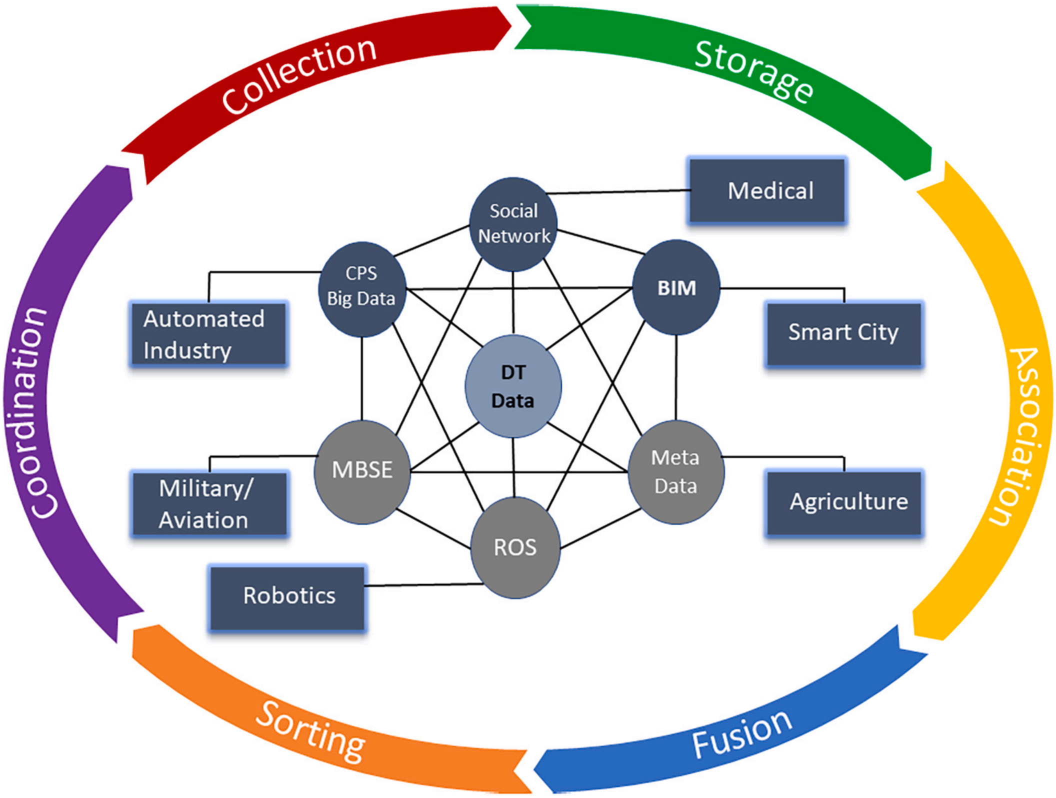

Fig. 7. A schematic for comprehending data analysis operation across several industries.

objective is to prioritize security measures to safeguard important military and aviation systems. The collective impact of these efforts results in the establishment of a society characterized by a digital landscape that is more sustainable, well-informed, and safe, positioning it for a more promising future that benefits all individuals.

The study presents a thorough roadmap for future research initiatives that aim to use the capabilities of Digital Twins fully. The strategies outlined above provide the foundation for a future characterized by disruptive changes, with a particular focus on cross-sector synergy, enhanced data analytics, security advances, scalability, and human-machine interaction. The integration of Digital Twins into environmental and legal frameworks is of paramount relevance, as it aligns with the principles of sustainability and ethical concerns. Moreover, the implementation of education and training programs serves to guarantee the presence of a proficient and capable workforce. By adopting these research pathways, the academic community may make valuable contributions to the continuous development of Digital Twins, which have the potential to fundamentally transform our engagement with physical and virtual environments and tackle significant global issues.

5. Conclusion

Data analysis is driving technology for a successful digital twin system. As data is the core divider between a successful and unsuccessful system, care should be given to the proper channeling of data structure. However, the digital twin is an ever-increasing popular technology among researchers. The emergence of digital twins entails new standards for data in terms of collection, extraction, fusion, interface, cyclic optimization, credibility, and application. As a result, data-related technology is getting more complex day by day, and more content is being added to the currently existing database every day. As such, more value, volume, veracity, and velocity are needed for the applied technology. So, before going into more profound development of the digital twin system, we need first to find out the differences between various data structure applications among the different fields of research. Our work here attempts to achieve that intention. In that instance, we attempted to depict everything in Fig. 7, which can be interpreted as oversimplifying the entire procedure. We tried to differentiate a digital twin system from a data perspective among various fields like manufacturing, urbanization, agriculture, medical, robotics, military, and aviation. Critical technologies for digital twins and data analysis, including CPS, big data analytics, BIM, MBSE, ROS, metadata management, and social networks. The preceding technologies facilitate the acquisition, storage, fusion, analysis, and representation of data in manners that were before unattainable. The use of digital twins and data analysis has the potential to enhance operational efficiency, facilitate informed decision-making, and foster innovation in several industry sectors. Here are some particular instances of the current integration of different driving technologies. General Electric (GE) employs digital twins, big data analytics, and MBSE techniques to facilitate the design and optimization of their jet engines. The Mayo Clinic is using digital twins, big data, and social network data in order to formulate individualized treatment strategies for patients diagnosed with cancer. The city of Pittsburgh is undertaking the use of digital twins, big data analytics, and ROS to foster the development of an autonomous vehicle system. The National Renewable Energy Laboratory is using digital twins, big data analytics, and BIM techniques to enhance the design and operational efficiency of solar energy systems. As the progression and refinement of these technologies persist, it is foreseeable that there will be a proliferation of inventive and pioneering implementations of digital twins and data analysis in forthcoming times.

This study presents a novel set of advancements that have the potential to transform digital twins significantly in several industries. The versatility of digital twins is highlighted by their capacity to be used across several sectors, thanks to their reliance on data-driven decision-making. This universal notion may be effectively utilized in diverse fields such as manufacturing, urban planning, healthcare, robotics, military/aviation, and other domains. The significance of data analysis and artificial intelligence in facilitating real-time decision-making is underscored, especially in industries such as manufacturing, where it enables predictive maintenance and enhances operational efficiency. The incorporation of blockchain technology, data coordination, and contemporary hierarchical data storage solutions contributes to the improvement of security, data organization, and accessibility, therefore effectively tackling significant difficulties. Moreover, the significance of machine learning and artificial intelligence in the field of data analysis highlights their contribution to the automation process and the development of Digital Twins into intelligent and self-governing systems. These preceding advancements jointly establish the trajectory of Digital Twins, enabling their extensive implementation, ensuring their security, and accommodating their customization to certain sectors. The investigation began with the tabular depiction of a categorical circumstance. The analysis made it very evident that the realm of digital twins is primarily concerned with industrialization and urbanization. However, the twin technique is starting to gain traction in a number of other scientific sub-fields as well. As a result, some theoretical discussion on certain common data evaluation procedures was presented despite the fact that each particular industry is different. The many ways data analysis is used across industries were detailed throughout the review. It is hoped that this publication will lay the groundwork for future work that will assist researchers in selecting the appropriate research methodology or methodologies in accordance with the domains in which they work from the perspective of the data.

CRediT authorship contribution statement

Md. Shezad Dihan: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anwar Islam Akash:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zinat Tasneem:** Validation, Supervision, Resources, Investigation, Conceptualization. **Prangon Das:** Supervision. **Sajal Kumar Das:** Supervision. **Md. Robiul Islam:** Supervision. **Md. Manirul Islam:** Supervision. **Faisal R. Badal:** Supervision, Software. **Md. Firoj Ali:** Supervision. **Md. Hafiz Ahamed:** Supervision. **Sarafat Hussain Abhi:** Supervision. **Subrata Kumar Sarker:** Supervision. **Md. Mehedi Hasan:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E. VanDerHorn, S. Mahadevan, Digital twin: generalization, characterization and implementation, Decis. Support Syst. 145 (2021) 113524.
[2] B.R. Barricelli, E. Casiraghi, D. Fogli, A survey on digital twin: definitions, characteristics, applications, and design implications, IEEE Access 7 (2019) 167653–167671.
[3] A. Agrawal, M. Fischer, V. Singh, Digital twin: from concept to practice, preprint, arXiv:2201.06912, 2022.
[4] S. Liao, Toward a digital twin of metal additive manufacturing: Process optimization and control enabled by physics-based and data-driven models, Ph.D. thesis, Northwestern University, 2023.
[5] A. Werner, J. Lentes, N. Zimmermann, Digitaler zwilling zur vorausschauenden instandhaltung in der produktion–physikbasierte modellierung und simulation zur optimierung datengetriebener modelle (Digital twin for predictive maintenance in production–physics-based modelling and simulation to optimize data-driven models), in: Stuttgart Symposium for Product Development SSP 2021, Scientific Conference, Stuttgart, May 20, 2021, Fraunhofer Society, 2021.
[6] F. Grumbach, P. Reusch, Data-driven generation of digital twin models for predictive-reactive job shop scheduling, Research proposal, Center for Applied Data Science (CfADS), FH Bielefeld, 33330 Gütersloh, Germany, September 2021, https://doi.org/10.13140/RG.2.2.24021.70884.
[7] N. Pillai, Enabling data-driven predictive maintenance for s&c through digital twin models and condition monitoring systems, PWI J. 139 (4) (October 2021).
[8] A. Sharma, E. Kosasih, J. Zhang, A. Brintrup, A. Calinescu, Digital twins: state of the art theory and practice, challenges, and open research questions, J. Ind. Inf. Integr. (2022) 100383.
[9] J. Argota Sánchez-Vaquerizo, Getting real: the challenge of building and validating a large-scale digital twin of Barcelona's traffic with empirical data, ISPRS Int. J. Geo-Inf. 11 (1) (2022) 24.
[10] A. Fuller, Z. Fan, C. Day, C. Barlow, Digital twin: enabling technologies, challenges and open research, IEEE Access 8 (2020) 108952–108971.
[11] Y. Pan, T. Qu, N. Wu, M. Khalgui, G. Huang, Digital twin based real-time production logistics synchronization system in a multi-level computing architecture, J. Manuf. Syst. 58 (2021) 246–260.
[12] Q. Qi, F. Tao, Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison, IEEE Access 6 (2018) 3585–3593.
[13] Q. Qi, F. Tao, T. Hu, N. Anwer, A. Liu, Y. Wei, L. Wang, A. Nee, Enabling technologies and tools for digital twin, J. Manuf. Syst. 58 (2021) 3–21.
[14] X. He, Q. Ai, J. Wang, F. Tao, B. Pan, R. Qiu, B. Yang, Situation awareness of energy Internet of thing in smart city based on digital twin: from digitization to informatization, IEEE Int. Things J. (2022).
[15] T. Wang, J. Li, Y. Deng, C. Wang, H. Snoussi, F. Tao, Digital twin for human-machine interaction with convolutional neural network, Int. J. Comput. Integr. Manuf. 34 (7–8) (2021) 888–897.

[16] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, F. Sui, Digital twin-driven product design, manufacturing and service with big data, Int. J. Adv. Manuf. Technol. 94 (9) (2018) 3563–3576.
[17] F. Tao, Y. Zhang, Y. Cheng, J. Ren, D. Wang, Q. Qi, P. Li, Digital twin and blockchain enhanced smart manufacturing service collaboration and management, J. Manuf. Syst. (2020).
[18] J. Cheng, H. Zhang, F. Tao, C.-F. Juang, Dt-ii: digital twin enhanced industrial Internet reference framework towards smart manufacturing, Robot. Comput.-Integr. Manuf. 62 (2020) 101881.
[19] X.V. Wang, L. Wang, Digital twin-based weee recycling, recovery and remanufacturing in the background of industry 4.0, Int. J. Prod. Res. 57 (12) (2019) 3892–3902.
[20] S. Liu, X.V. Wang, L. Wang, Digital twin-enabled advance execution for human-robot collaborative assembly, CIRP Ann. (2022).
[21] Y. Jeong, E. Flores-García, D.H. Kwak, J.H. Woo, M. Wiktorsson, S. Liu, X.V. Wang, L. Wang, Digital twin-based services and data visualization of material handling equipment in smart production logistics environment, in: IFIP International Conference on Advances in Production Management Systems, Springer, 2022, pp. 556–564.
[22] F. Tao, F. Sui, A. Liu, Q. Qi, M. Zhang, B. Song, Z. Guo, S.C.-Y. Lu, A.Y. Nee, Digital twin-driven product design framework, Int. J. Prod. Res. 57 (12) (2019) 3935–3953.
[23] M. Zhang, F. Sui, A. Liu, F. Tao, A. Nee, Digital twin driven smart product design framework, in: Digital Twin Driven Smart Design, Elsevier, 2020, pp. 3–32.
[24] F. Tao, N. Anwer, A. Liu, L. Wang, A.Y. Nee, L. Li, M. Zhang, Digital twin towards smart manufacturing and industry 4.0, J. Manuf. Syst. 58 (2021) 1–2.
[25] M. Zhang, F. Tao, A. Nee, Digital twin enhanced dynamic job-shop scheduling, J. Manuf. Syst. 58 (2021) 146–156.
[26] T. Kong, T. Hu, T. Zhou, Y. Ye, Data construction method for the applications of workshop digital twin system, J. Manuf. Syst. 58 (2021) 323–328.
[27] Y. Wei, T. Hu, Y. Wang, S. Wei, W. Luo, Implementation strategy of physical entity for manufacturing system digital twin, Robot. Comput.-Integr. Manuf. 73 (2022) 102259.
[28] W. Luo, T. Hu, C. Zhang, Y. Wei, Digital twin for cnc machine tool: modeling and using strategy, J. Ambient Intell. Humaniz. Comput. 10 (3) (2019) 1129–1140.
[29] W. Luo, T. Hu, Y. Ye, C. Zhang, Y. Wei, A hybrid predictive maintenance approach for cnc machine tool driven by digital twin, Robot. Comput.-Integr. Manuf. 65 (2020) 101974.
[30] O. Elijah, S.K.A. Rahim, A.A. Emmanuel, Y.O. Salihu, Z.G. Usman, A.M. Jimoh, Enabling smart agriculture in Nigeria: application of digital-twin technology, in: 2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), IEEE, 2021, pp. 1–6.
[31] P. Skobelev, A. Tabachinskiy, E. Simonova, T.-R. Lee, A. Zhilyaev, V. Laryukhin, Digital twin of rice as a decision-making service for precise farming, based on environmental datasets from the fields, in: 2021 International Conference on Information Technology and Nanotechnology (ITNT), IEEE, 2021, pp. 1–8.
[32] T. Hoebert, W. Lepuschitz, E. List, M. Merdan, Cloud-based digital twin for industrial robotics, in: International Conference on Industrial Applications of Holonic and Multi-Agent Systems, Springer, 2019, pp. 105–116.
[33] T. Erol, A.F. Mendi, D. Doğan, The digital twin revolution in healthcare, in: 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), IEEE, 2020, pp. 1–7.
[34] A.F. Mendi, T. Erol, D. Dogan, Digital twin in the military field, IEEE Internet Comput. (2021).
[35] A.F. Mendi, A digital twin case study on automotive production line, Sensors 22 (18) (2022) 6963.
[36] R. Martinez-Velazquez, R. Gamez, A. El Saddik, Cardio twin: a digital twin of the human heart running on the edge, in: 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE, 2019, pp. 1–6.
[37] R. Gámez Díaz, Q. Yu, Y. Ding, F. Laamarti, A. El Saddik, Digital twin coaching for physical activities: a survey, Sensors 20 (20) (2020) 5936.
[38] B.R. Barricelli, E. Casiraghi, J. Gliozzo, A. Petrini, S. Valtolina, Human digital twin for fitness management, IEEE Access 8 (2020) 26637–26664.
[39] T.D. West, M. Blackburn, Is digital thread/digital twin affordable? A systemic assessment of the cost of dod's latest Manhattan project, Proc. Comput. Sci. 114 (2017) 47–56.
[40] T.D. West, M. Blackburn, Demonstrated benefits of a nascent digital twin, Insight 21 (1) (2018) 43–47.
[41] C. Boje, A. Guerriero, S. Kubicki, Y. Rezgui, Towards a semantic construction digital twin: directions for future research, Autom. Constr. 114 (2020) 103179.
[42] C. Boje, A. Marvuglia, Á.J.H. Menacho, S. Kubicki, A. Guerriero, T. Navarrete, E. Benetto, A pilot using a building digital twin for lca-based human health monitoring, in: Proc. of the Conference CIB W78, vol. 2021, 2021, pp. 11–15.
[43] K. Wang, Y. Wang, Y. Li, X. Fan, S. Xiao, L. Hu, A review of the technology standards for enabling digital twin, Digital Twin 2 (4) (2022) 4.
[44] M. Zhang, F. Tao, B. Huang, A. Liu, L. Wang, N. Anwer, A. Nee, Digital twin data: methods and key technologies, Digital Twin 1 (2) (2022) 2.
[45] H.F. Dodge, H.G. Romig, A method of sampling inspection, Bell Syst. Tech. J. 8 (4) (1929) 613–631.
[46] H.R. Abdulshaheed, I. Al_Barazanchi, M.S.B. Sidek, Survey: benefits of integrating both wireless sensors networks and cloud computing infrastructure, Sustain. Eng. Innov. (ISSN 2712-0562) 1 (2) (2019) 67–83.
[47] S. Uke, R. Thool, Uml based modeling for data aggregation in secured wireless sensor network, Proc. Comput. Sci. 78 (2016) 706–713.
[48] R. Brahmi, M. Hammadi, N. Aifaoui, J.-Y. Choley, Interoperability of cad models and sysml specifications for the automated checking of design requirements, Proc. CIRP 100 (2021) 259–264.
[49] M. Manaa, J. Akaichi, Ontology-based modeling and querying of trajectory data, Data Knowl. Eng. 111 (2017) 58–72.
[50] D. Legatiuk, Mathematical modelling by help of category theory: models and relations between them, Mathematics 9 (16) (2021) 1946.
[51] I. Jebli, F.-Z. Belouadha, M.I. Kabbaj, A. Tilioua, Prediction of solar energy guided by Pearson correlation using machine learning, Energy 224 (2021) 120109.
[52] X. Wang, Z. Wang, M. Sheng, Q. Li, W. Sheng, An adaptive and opposite k-means operation based memetic algorithm for data clustering, Neurocomputing 437 (2021) 131–142.
[53] M. Tian, L. Zhang, P. Guo, H. Zhang, Q. Chen, Y. Li, A. Xue, Data dependence analysis for defects data of relay protection devices based on apriori algorithm, IEEE Access 8 (2020) 120647–120653.
[54] S. Wang, X. Guo, Y. Tie, I. Lee, L. Qi, L. Guan, Weighted hybrid fusion with rank consistency, Pattern Recognit. Lett. 138 (2020) 329–335.
[55] D. Mourtzis, E. Vlachou, M. Doukas, N. Kanakis, N. Xanthopoulos, A. Koutoupes, Cloud-Based Adaptive Shop-Floor Scheduling Considering Machine Tool Availability, ASME International Mechanical Engineering Congress and Exposition, vol. 57588, American Society of Mechanical Engineers, 2015, p. V015T19A017.
[56] Z. Zheng, H. Qiu, Z. Wang, S. Luo, Y. Lei, Data fusion based multi-rate Kalman filtering with unknown input for on-line estimation of dynamic displacements, Measurement 131 (2019) 211–218.
[57] B. Rebiasz, Fuzziness and randomness in investment project risk appraisal, Comput. Oper. Res. 34 (1) (2007) 199–210.
[58] D. Guan, Y. Cao, J. Yang, Y. Cao, M.Y. Yang, Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection, Inf. Fusion 50 (2019) 148–157.
[59] S.A. Renganathan, K. Harada, D.N. Mavris, Aerodynamic data fusion toward the digital twin paradigm, AIAA J. 58 (9) (2020) 3902–3918.
[60] B. He, X. Cao, Y. Hua, Data fusion-based sustainable digital twin system of intelligent detection robotics, J. Clean. Prod. 280 (2021) 124181.
[61] K. Huang, S. Wu, Y. Li, C. Yang, W. Gui, A multi-rate sampling data fusion method for fault diagnosis and its industrial applications, J. Process Control 104 (2021) 54–61.
[62] T. Germa, F. Lerasle, N. Ouadah, V. Cadenat, Vision and rfid data fusion for tracking people in crowds by a mobile robot, Comput. Vis. Image Underst. 114 (6) (2010) 641–651.
[63] G. Booch, The Unified Modeling Language User Guide, Pearson Education India, 2005.
[64] L. Balmelli, et al., An overview of the systems modeling language for products and systems development, J. Object Technol. 6 (6) (2007) 149–177.

[65] X. Su, L. Ilebrekke, A comparative study of ontology languages and tools, in: International Conference on Advanced Information Systems Engineering, Springer, 2002, pp. 761–765.
[66] E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra, Locally adaptive dimensionality reduction for indexing large time series databases, in: Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data, 2001, pp. 151–162.
[67] P.-E. Danielsson, Euclidean distance mapping, Comput. Graph. Image Process. 14 (3) (1980) 227–248.
[68] H. Haße, B. Li, N. Weißenberg, J. Cirullies, B. Otto, Digital twin for real-time data processing in logistics, in: Artificial Intelligence and Digital Transformation in Supply Chain Management: Innovative Approaches for Supply Chains, in: Proceedings of the Hamburg International Conference of Logistics (HICL), vol. 27, Epubli GmbH, Berlin, 2019, pp. 4–28.
[69] R. Minerva, G.M. Lee, N. Crespi, Digital twin in the iot context: a survey on technical features, scenarios, and architectural models, Proc. IEEE 108 (10) (2020) 1785–1824.
[70] M.J. Kaur, V.P. Mishra, P. Maheshwari, The convergence of digital twin, iot, and machine learning: transforming data into action, in: Digital Twin Technologies and Smart Cities, 2020, pp. 3–17.
[71] S. Liu, Y. Lu, X. Shen, J. Bao, A digital thread-driven distributed collaboration mechanism between digital twin manufacturing units, J. Manuf. Syst. 68 (2023) 145–159.
[72] C.L. Stergiou, K.E. Psannis, Digital twin intelligent system for industrial iot-based big data management and analysis in cloud, Virtual Real. Intell. Hardw. 4 (4) (2022) 279–291.
[73] C. Adrian, R. Abdullah, R. Atan, Y.Y. Jusoh, Expert review on big data analytics implementation model in data-driven decision-making, in: 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), IEEE, 2018, pp. 1–5.
[74] Y. Dai, Y. Zhang, Adaptive digital twin for vehicular edge computing and networks, J. Commun. Inf. Netw. 7 (1) (2022) 48–59.
[75] T. Do-Duy, D. Van Huynh, O.A. Dobre, B. Canberk, T.Q. Duong, Digital twin-aided intelligent offloading with edge selection in mobile edge computing, IEEE Wirel. Commun. Lett. 11 (4) (2022) 806–810.
[76] J. Wang, Y. Liu, S. Ren, C. Wang, S. Ma, Edge computing-based real-time scheduling for digital twin flexible job shop with variable time window, Robot. Comput.-Integr. Manuf. 79 (2023) 102435.
[77] K. Sharmila, S. Kamalakkannan, R. Devi, C. Shanthi, Big data analysis using apache hadoop and spark, Int. J. Recent Technol. Eng. 8 (2) (2019) 167–170.
[78] P. Sewal, H. Singh, A critical analysis of apache hadoop and spark for big data processing, in: 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), IEEE, 2021, pp. 308–313.
[79] A.M. Altriki, O. Alarafee, Techniques management big data apache hadoop and apache spark and which is better in structuring and processing data, 2020.
[80] T. Tekdogan, A. Cakmak, Benchmarking apache spark and hadoop mapreduce on big data classification, in: Proceedings of the 2021 5th International Conference on Cloud and Big Data Computing, 2021, pp. 15–20.
[81] R. Kumar, S. Charu, S. Bansal, Effective way to handling big data problems using nosql database (mongodb), J. Adv. Database Manag. Syst. 2 (2) (2015) 42–48.
[82] R. Sreekanth, G.M. Rao, S. Nanduri, Big data electronic health records data management and analysis on cloud with mongodb: a nosql database, Int. J. Adv. Eng. Global Technol. 3 (7) (2015) 943–949.
[83] A. Chebotko, A. Kashlev, S. Lu, A big data modeling methodology for apache cassandra, in: 2015 IEEE International Congress on Big Data, IEEE, 2015, pp. 238–245.
[84] D. Chrimes, H. Zamani, et al., Using distributed data over hbase in big data analytics platform for clinical services, Comput. Math. Methods Med. 2017 (2017).
[85] M.H. Ali, M.S. Hosain, M.A. Hossain, Big data analysis using bigquery on cloud computing platform, Australian J. Eng. Innov. Tech. 3 (1) (2021) 1–9.
[86] M. Serik, G. Nurbekova, M. Mukhambetova, Optimal organisation of a big data training course: big data processing with bigquery and setting up a dataproc hadoop framework, World Trans. Eng. Technol. Educ. 19 (4) (2021) 417–422.
[87] K. Ghane, Big data pipeline with ml-based and crowd sourced dynamically created and maintained columnar data warehouse for structured and unstructured big data, in: 2020 3rd International Conference on Information and Computer Technologies (ICICT), IEEE, 2020, pp. 60–67.
[88] S. Uzunbayir, Relational database and nosql inspections using mongodb and neo4j on a big data application, in: 2022 7th International Conference on Computer Science and Engineering (UBMK), IEEE, 2022, pp. 148–153.
[89] K. Awada, M.Y. Eltabakh, C. Tang, M. Al-Kateb, S. Nair, G. Au, Cost estimation across heterogeneous sql-based big data infrastructures in teradata intellisphere, in: EDBT, 2020, pp. 534–545.
[90] N. Golov, L. Rönnbäck, Big data normalization for massively parallel processing databases, Comput. Stand. Interfaces 54 (2017) 86–93.
[91] D. Chen, Y. Hu, C. Cai, K. Zeng, X. Li, Brain big data processing with massively parallel computing technology: challenges and opportunities, Softw. Pract. Exp. 47 (3) (2017) 405–420.
[92] J. Ramsingh, V. Bhuvaneswari, An insight on big data analytics using pig script, Int. J. Emerg. Trends Technol. Comput. Sci. 4 (6) (2015) 2278–6856.
[93] E.L. Lydia, M.B. Swarup, Big data analysis using hadoop components like flume, mapreduce, pig and hive, Int. J. Sci., Eng. Comput. Technol. 5 (11) (2015) 390.
[94] B.R. Hiraman, et al., A study of apache kafka in big data stream processing, in: 2018 International Conference on Information, Communication, Engineering and Technology (ICICET), IEEE, 2018, pp. 1–3.
[95] G.P. Davidson, D.D. Ravindran, Technical review of apache flink for big data, Int. J. Aquat. Sci. 12 (2) (2021) 3340–3346.
[96] M. Ku, E. Choi, D. Min, An analysis of performance factors on esper-based stream big data processing in a virtualized environment, Int. J. Commun. Syst. 27 (6) (2014) 898–917.
[97] J. Oláh, E. Erdei, J. Popp, Applying big data algorithms for sales data stored in sap hana, An. Univ. Oradea 453 (July 2017).
[98] S. Abuayeid, L. Alarabi, Comparative analysis of spark and ignite for big spatial data processing, Int. J. Adv. Comput. Sci. Appl. 12 (9) (2021).
[99] V.K. Vennu, S.R. Yepuru, A performance study for autoscaling big data analytics containerized applications: Scalability of apache spark on kubernetes, 2022.
[100] N. Singh, Y. Hamid, S. Juneja, G. Srivastava, G. Dhiman, T.R. Gadekallu, M.A. Shah, Load balancing and service discovery using docker swarm for microservice based big data applications, J. Cloud Comput. 12 (1) (2023) 1–9.
[101] S. Mohanty, T.W. Elmer, S. Bakhtiari, R.B. Vilim, A review of sql vs nosql database for nuclear reactor digital twin applications: With example mongodb based nosql database for digital twin model of a pressurized-water-reactor steam-generator, in: ASME International Mechanical Engineering Congress and Exposition, vol. 85697, American Society of Mechanical Engineers, 2021, p. V013T14A003.
[102] M. Zhang, F. Tao, B. Huang, A. Liu, L. Wang, N. Anwer, A. Nee, Digital twin data: methods and key technologies, Digital Twin 1 (2) (2022) 2.
[103] F. Tao, B. Xiao, Q. Qi, J. Cheng, P. Ji, Digital twin modeling, J. Manuf. Syst. 64 (2022) 372–389.
[104] Y. Cheng, K. Chen, H. Sun, Y. Zhang, F. Tao, Data and knowledge mining with big data towards smart production, J. Ind. Inf. Integr. 9 (2018) 1–13.
[105] M. Chand, Top 10 cloud service providers in 2021, https://www.c-sharpcorner.com/article/top-10-cloud-service-providers/.
[106] Discover the next generation cloud platform, https://www.oracle.com/cloud/.
[107] Product introduction – alibaba cloud documentation center, https://www.alibabacloud.com/help/en/apsaradb-for-redis/latest/features.
[108] T.H.-J. Uhlemann, C. Schock, C. Lehmann, S. Freiberger, R. Steinhilper, The digital twin: demonstrating the potential of real time data acquisition in production systems, Procedia Manuf. 9 (2017) 113–120.
[109] Z. Zhu, C. Liu, X. Xu, Visualisation of the digital twin data in manufacturing by using augmented reality, Proc. CIRP 81 (2019) 898–903.
[110] B. He, K.-J. Bai, Digital twin-based sustainable intelligent manufacturing: a review, Adv. Manuf. 9 (1) (2021) 1–21.
[111] W. Kritzinger, M. Karner, G. Traar, J. Henjes, W. Sihn, Digital twin in manufacturing: a categorical literature review and classification, IFAC-PapersOnLine 51 (11) (2018) 1016–1022.

[112] C. Cimino, E. Negri, L. Fumagalli, Review of digital twin applications in manufacturing, Comput. Ind. 113 (2019) 103130.
[113] Q. Lu, A.K. Parlikad, P. Woodall, G.D. Ranasinghe, X. Xie, Z. Liang, E. Konstantinou, J. Heaton, J. Schooling, Developing a digital twin at building and city levels: a case study of West Cambridge campus, J. Manag. Eng. 36 (3) (2020).
[114] S. Ivanov, K. Nikolskaya, G. Radchenko, L. Sokolinsky, M. Zymbler, Digital twin of city: concept overview, in: 2020 Global Smart Industry Conference (GloSIC), IEEE, 2020, pp. 178–186.
[115] A. Biswas, M.O. Reon, P. Das, Z. Tasneem, S. Muyeen, S.K. Das, F.R. Badal, S.K. Sarker, M.M. Hassan, S.H. Abhi, et al., State-of-the-art review on recent advancements on lateral control of autonomous vehicles, IEEE Access (2022).
[116] L. Wan, T. Nochta, J. Schooling, Developing a city-level digital twin–propositions and a case study, in: International Conference on Smart Infrastructure and Construction 2019 (ICSIC) Driving Data-Informed Decision-Making, ICE Publishing, 2019, pp. 187–194.
[117] C. Fan, C. Zhang, A. Yahja, A. Mostafavi, Disaster city digital twin: a vision for integrating artificial and human intelligence for disaster management, Int. J. Inf. Manag. 56 (2021) 102049.
[118] J.D. Chaux, D. Sanchez-Londono, G. Barbieri, A digital twin architecture to optimize productivity within controlled environment agriculture, Appl. Sci. 11 (19) (2021) 8875.
[119] A. Nasirahmadi, O. Hensel, Toward the next generation of digitalization in agriculture based on digital twin paradigm, Sensors 22 (2) (2022) 498.
[120] P. Angin, M.H. Anisi, F. Göksel, C. Gürsoy, A. Büyükgülcü, Agrilora: a digital twin framework for smart agriculture, J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl. 11 (4) (2020) 77–96.
[121] M. Jans-Singh, K. Leeming, R. Choudhary, M. Girolami, Digital twin of an urban-integrated hydroponic farm, Data-Centric Eng. 1 (2020).
[122] A. Ghandar, A. Ahmed, S. Zulfiqar, Z. Hua, M. Hanai, G. Theodoropoulos, A decision support system for urban agriculture using digital twin: a case study with aquaponics, IEEE Access 9 (2021) 35691–35708.
[123] P. Moghadam, T. Lowe, E.J. Edwards, Digital twin for the future of orchard production systems, Proceedings 36 (1) (2020) 92, Multidisciplinary Digital Publishing Institute.
[124] M. Moshrefzadeh, T. Machl, D. Gackstetter, A. Donaubauer, T.H. Kolbe, Towards a distributed digital twin of the agricultural landscape, J. Digit. Landsc. Archit. 5 (2020) 173–186.
[125] A. Kampker, V. Stich, P. Jussen, B. Moser, J. Kuntz, Business models for industrial smart services–the example of a digital twin for a product-service-system for potato harvesting, Proc. CIRP 83 (2019) 534–540.
[126] J. Corral-Acero, F. Margara, M. Marciniak, C. Rodero, F. Loncaric, Y. Feng, A. Gilbert, J.F. Fernandes, H.A. Bukhari, A. Wajdan, et al., The 'digital twin'to enable the vision of precision cardiology, Eur. Heart J. 41 (48) (2020) 4556–4564.
[127] Y. Liu, L. Zhang, Y. Yang, L. Zhou, L. Ren, F. Wang, R. Liu, Z. Pang, M.J. Deen, A novel cloud-based framework for the elderly healthcare services using digital twin, IEEE Access 7 (2019) 49088–49101.
[128] R.K. Phanden, P. Sharma, A. Dubey, A review on simulation in digital twin for aerospace, manufacturing and robotics, Mater. Today Proc. 38 (2021) 174–178.
[129] G. Garg, V. Kuts, G. Anbarjafari, Digital twin for fanuc robots: industrial robot programming and simulation using virtual reality, Sustainability 13 (18) (2021) 10336.
[130] L. Pérez, S. Rodríguez-Jiménez, N. Rodríguez, R. Usamentiaga, D.F. García, Digital twin and virtual reality based methodology for multi-robot manufacturing cell commissioning, Appl. Sci. 10 (10) (2020) 3633.
[131] K. Dröder, P. Bobka, T. Germann, F. Gabriel, F. Dietrich, A machine learning-enhanced digital twin approach for human-robot-collaboration, Proc. CIRP 76 (2018) 187–192.
[132] L. Wang, et al., Application and development prospect of digital twin technology in aerospace, IFAC-PapersOnLine 53 (5) (2020) 732–737.
[133] M. Xiong, H. Wang, Q. Fu, Y. Xu, Digital twin–driven aero-engine intelligent predictive maintenance, Int. J. Adv. Manuf. Technol. 114 (11) (2021) 3751–3761.
[134] E. Hinchy, C. Carcagno, N. O'Dowd, C. McCarthy, Using finite element analysis to develop a digital twin of a manufacturing bending operation, Proc. CIRP 93 (2020) 568–574.
[135] M. Furuya, Digital twin to digital triplet: Machine learning, additive manufacturing and computational fluid dynamics simulations, in: AIP Conference Proceedings, vol. 2659, AIP Publishing, 2022.
[136] N. Hansson, Modelling of production flow at siemens energy: Digital twin with plans toward statistical process control, 2021.
[137] S. Gupta, S. Modgil, A. Gunasekaran, Big data in lean six sigma: a review and further research directions, Int. J. Prod. Res. 58 (3) (2020) 947–969.
[138] H. Xu, J. Wu, Q. Pan, X. Guan, M. Guizani, A survey on digital twin for industrial Internet of things: applications, technologies and tools, IEEE Commun. Surv. Tutor. (2023).
[139] A. Rajendran, G. Asokan, Real time monitoring of machining process and data gathering for digital twin optimization, 2021.
[140] E. Baalbergen, J. de Marchi, A. Offringa, S. Hengeveld, B. Troost, K. He, R. Koppert, W. van den Eijnde, Applying digital twin technology in thermoplastic composites production: Supporting process monitoring, optimization and automation for real-time efficiency and smart quality control, 2020.
[141] J. Shi, Z. Pan, L. Jiang, X. Zhai, An ontology-based methodology to establish city information model of digital twin city by merging bim, gis and iot, Adv. Eng. Inform. 57 (2023) 102114.
[142] K. Valaskova, J. Oláh, J. Popp, G. Lăzăroiu, Virtual modeling and remote sensing technologies, spatial cognition and neural network algorithms, and visual analytics tools in urban geopolitics and digital twin cities, Geopolitics, History and International Relations 14 (2) (2022) 9–24.
[143] W. Jia, W. Wang, Z. Zhang, From simple digital twin to complex digital twin part I: a novel modeling method for multi-scale and multi-scenario digital twin, Adv. Eng. Inform. 53 (2022) 101706.
[144] S. Iizuka, Y. Xuan, C. Takatori, H. Nakaura, A. Hashizume, Environmental impact assessment of introducing compact city models by downscaling simulations, Sustain. Cities Soc. 63 (2020) 102424.
[145] R. Patel, R. Goyal, V. Ramasubramanian, M. Sudeep, et al., Markov chain based crop forecast modeling software, J. Indian Soc. Agric. Stat. 67 (3) (2013) 371–379.
[146] P. Mensik, M. Starỳ, D. Marton, Water management software for controlling the water supply function of many reservoirs in a watershed, Water Resour. 42 (2015) 133–145.
[147] A. Hartmann, N. Goldscheider, T. Wagener, J. Lange, M. Weiler, Karst water resources in a changing world: review of hydrological modeling approaches, Rev. Geophys. 52 (3) (2014) 218–242.
[148] M.A. Musen, B. Middleton, R.A. Greenes, Clinical decision-support systems, in: Biomedical Informatics: Computer Applications in Health Care and Biomedicine, Springer, 2021, pp. 795–840.
[149] K.M. Tsiouris, D. Gatsios, V. Tsakanikas, A.A. Pardalis, I. Kouris, T. Androutsou, M. Tarousi, N. Vujnovic Sedlar, I. Somarakis, F. Mostajeran, et al., Designing interoperable telehealth platforms: bridging iot devices with cloud infrastructures, Enterp. Inf. Syst. 14 (8) (2020) 1194–1218.
[150] Z. Wang, Y. OuYang, O. Kochan, Bidirectional Linkage Robot Digital Twin System Based on Ros, 2023 17th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), vol. 1, IEEE, 2023, pp. 1–5.
[151] R. Molinaro, J.-S. Singh, S. Catsoulis, C. Narayanan, D. Lakehal, Embedding data analytics and cfd into the digital twin concept, Comput. Fluids 214 (2021) 104759.
[152] M. Topaç, I. Bahar, Determinaton of the spring characteristic of a parabolic leaf spring used in a military vehicle by using non-linear finite element analysis, J. Polytech. 22 (1) (2019).
[153] A.T. Biggs, D.A. Hirsch, Using Monte Carlo simulations to translate military and law enforcement training results to operational metrics, J. Defense Model. Simul. 19 (3) (2022) 403–415.

[154] A. Antonakis, K. Giannakoglou, Optimisation of military aircraft engine maintenance subject to engine part shortages using asynchronous metamodel-assisted particle swarm optimisation and Monte-Carlo simulations, Int. J. Syst. Sci.: Oper. Logist. 5 (3) (2018) 239–252.
[155] F. Tao, M. Zhang, Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing, IEEE Access 5 (2017) 20418–20427.
[156] A. Das, M.P. Sarma, K.K. Sarma, N. Mastorakis, Design of an Iot Based Real Time Environment Monitoring System Using Legacy Sensors, MATEC Web of Conferences, vol. 210, EDP Sciences, 2018, p. 03008.
[157] Z. Ling, C. Gao, C. Sano, C. Toe, Z. Li, X. Fu, Stir: a smart and trustworthy iot system interconnecting legacy ir devices, IEEE Int. Things J. 7 (5) (2020) 3958–3967.
[158] J.A.d. Rosas, V. Brito, L. Brito Palma, J. Barata, Approach to adapt a legacy manufacturing system into the iot paradigm, Int. J. Interact. Mob. Technol. 11 (5) (2017).
[159] C. Ruah, O. Simeone, B. Al-Hashimi, A Bayesian framework for digital twin-based control, monitoring, and data collection in wireless systems, IEEE J. Sel. Areas Commun. (2023).
[160] J. Monteiro, J. Barata, M. Veloso, L. Veloso, J. Nunes, A scalable digital twin for vertical farming, J. Ambient Intell. Humaniz. Comput. 14 (10) (2023) 13981–13996.
[161] D. An, Y. Chen, A digital twin enabled Internet of living things (iolt) framework for soil carbon management, in: 2022 18th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications, IEEE, MESA, 2022, pp. 1–6.
[162] M.A. Scholl, K. Stine, J. Hash, P. Bowen, L.A. Johnson, C.D. Smith, D. Steinberg, An introductory resource guide for implementing the health insurance portability and accountability act (hipaa) security rule, 2008.
[163] S.N. Duda, N. Kennedy, D. Conway, A.C. Cheng, V. Nguyen, T. Zayas-Cabán, P.A. Harris, Hl7 fhir-based tools and initiatives to support clinical research: a scoping review, J. Am. Med. Inform. Assoc. 29 (9) (2022) 1642–1653.
[164] S. Rundel, R. De Amicis, Leveraging digital twin and game-engine for traffic simulations and visualizations, Front. Virtual Real. 4 (2023) 1048753.
[165] J.D. de Hoz Diego, A. Temperekidis, P. Katsaros, C. Konstantinou, An iot digital twin for cyber-security defence based on runtime verification, in: International Symposium on Leveraging Applications of Formal Methods, Springer, 2022, pp. 556–574.
[166] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, H. Wang, Blockchain challenges and opportunities: a survey, Int. J. Web Grid Serv. 14 (4) (2018) 352–375.
[167] T. Li, A.K. Sahu, A. Talwalkar, V. Smith, Federated learning: challenges, methods, and future directions, IEEE Signal Process. Mag. 37 (3) (2020) 50–60.
[168] R. Sahal, S.H. Alsamhi, K.N. Brown, D. O'shea, C. McCarthy, M. Guizani, Blockchain-empowered digital twins collaboration: smart transportation use case, Machines 9 (9) (2021) 193.
[169] J.M. Krotkiewicz, An in-Depth Look into Cryptographic Hashing Algorithms, Algoma University at Sault Ste. Marie, 2016.
[170] H. Gasmi, A. Belhi, A. Hammi, A. Bouras, B. Aouni, I. Khalil, Blockchain-based manufacturing supply chain management using hyperledger fabric, in: IFIP International Conference on Product Lifecycle Management, Springer, 2021, pp. 305–318.
[171] A. Belhi, H. Gasmi, A. Hammi, A. Bouras, B. Aouni, I. Khalil, A broker-based manufacturing supply chain integration with blockchain: managing odoo workflows using hyperledger fabric smart contracts, in: IFIP International Conference on Product Lifecycle Management, Springer, 2021, pp. 371–385.
[172] S.S. Kushwaha, S. Joshi, D. Singh, M. Kaur, H.-N. Lee, Ethereum smart contract analysis tools: a systematic review, IEEE Access 10 (2022) 57037–57062.
[173] Y. Yao, H. Liang, X. Li, J. Zhang, J. He, Sensing urban land-use patterns by integrating Google tensorflow and scene-classification models, preprint, arXiv:1708.01580, 2017.
[174] E.M. Agyemang-Duah, Investigating specialty crop farmers' preferences for contract design and attitudes towards blockchain-based smart contracts, 2023.
[175] N. Kshetri, Blockchain-based smart contracts to provide crop insurance for smallholder farmers in developing countries, IT Prof. 23 (6) (2021) 58–61.
[176] D. Mohanty, R3 Corda for Architects and Developers: With Case Studies in Finance, Insurance, Healthcare, Travel, Telecom, and Agriculture, Apress, 2019.
[177] A. Ziller, A. Trask, A. Lopardo, B. Szymkow, B. Wagner, E. Bluemke, J.-M. Nounahon, J. Passerat-Palmbach, K. Prakash, N. Rose, et al., Pysyft: a library for easy federated learning, in: Federated Learning Systems: Towards Next-Generation AI, 2021, pp. 111–139.
[178] R. Sahal, S.H. Alsamhi, K.N. Brown, D. O'Shea, B. Alouffi, et al., Blockchain-based digital twins collaboration for smart pandemic alerting: decentralized Covid-19 pandemic alerting use case, Comput. Intell. Neurosci. (2022) 2022.
[179] G.A. Oliva, A.E. Hassan, Z.M. Jiang, An exploratory study of smart contracts in the Ethereum blockchain platform, Empir. Softw. Eng. 25 (2020) 1864–1904.
[180] C. Zhang, S. Li, J. Xia, W. Wang, F. Yan, Y. Liu, {BatchCrypt}: efficient homomorphic encryption for {Cross-Silo} federated learning, in: 2020 USENIX Annual Technical Conference, in: USENIX ATC, vol. 20, 2020, pp. 493–506.
[181] S.H. Alsamhi, B. Lee, Blockchain-empowered multi-robot collaboration to fight Covid-19 and future pandemics, IEEE Access 9 (2020) 44173–44197.
[182] F. Kureshi, D. Makwana, U. Bodkhe, S. Tanwar, P. Chaturvedi, Blockchain based humans-agents interactions/human-robot interactions: a systematic literature review and research agenda, Robotic Process Automation (2023) 139–165.
[183] X. Shu, B. Ding, J. Luo, X. Fu, M. Xie, Z. Li, A hashgraph-based knowledge sharing approach for mobile robot swarm, in: Collaborative Computing: Networking, Applications and Worksharing: 17th EAI International Conference, CollaborateCom 2021, Virtual Event, October 16-18, 2021, Proceedings, Part II 17, Springer, 2021, pp. 158–172.
[184] M. Alsayegh, P. Vanegas, A.A.R. Newaz, L. Bobadilla, D.A. Shell, Privacy-preserving multi-robot task allocation via secure multi-party computation, in: 2022 European Control Conference (ECC), IEEE, 2022, pp. 1274–1281.
[185] R. Saha, G. Kumar, M. Conti, T. Devgun, T.-h. Kim, M. Alazab, R. Thomas, Dhacs: smart contract-based decentralized hybrid access control for industrial Internet-of-things, IEEE Trans. Ind. Inform. 18 (5) (2021) 3452–3461.